Charles University in Prague

Faculty of Science

Department of Genetics and Microbiology

Microbiology

Mgr. Tomáš Větrovský

Combination of biochemical and high-throughput-sequencing approaches to study the role of Actinobacteria and fungi in the decomposition of plant biomass

Studium úlohy Aktinobakterií a hub účastnících se degradace rostlinné biomasy kombinací biochemických a moderních sekvenačních metod

Ph.D. Thesis

Supervisor: **RNDr. Petr Baldrian, Ph.D.**

Institute of Microbiology

of the Czech Academy of Sciences

Prague, 2015

Declaration:

I declare that I wrote this work on my own and all sources and literature are properly cited. I also declare, that this work or its major part was not previously used for obtaining of the same or any other academic degree.

Prague, 23.11.2015 Tomáš Větrovský

Table of contents

Acknowledgements

First I would like to acknowledge my supervisor RNDr. Petr Baldrian, Ph.D. for his great help and friendly attitude over the period of my project and for the opportunity to obtain valuable skills in various areas of my research. Further, I would like to thank all my colleagues and friends from the Institute of Microbiology for their advice, assistance and collaboration. I have learned a lot from them and they were creating a pleasant working environment. My thanks also go to Doc. RNDr. Jiří Gabriel, DrSc., Mgr. Miroslav Kolařík Ph.D and Doc. Kari Steffen Ph.D. and his team at the University of Helsinky, where the experiments with radiolabelled substances were carried out. Finally I would also like to thank my wife for her patience and support and my family for the permanent encouragement.

This work was supported by the Czech Science Foundation (Projects No. 526/08/0751, 13-06763S), by the Czech Academy of Sciences (Project No. IAA603020901), and by the Ministry of Education, Youth and Sports of the Czech Republic (Project No. LD12048).

Abstract

Dead plant biomass is a key pool of carbon in terrestrial ecosystems. Its decomposition in soil environments is thus an essential process of the carbon cycle. Fungi are considered to be the primary decomposers in soil ecosystems because of their physiological adaptations and enzymatic apparatus composed from highly effective oxidative and hydrolytic enzymes. Many recent works show that in addition to fungi, bacteria may also play a significant role in lignocellulose decomposition and among bacteria, the members of the phylum Actinobacteria are often regarded to significantly contribute to cellulose and lignocellulose decomposition.

This thesis is focused on the evaluation of the role that fungi and Actinobacteria play in dead plant biomass degradation. First, it explored mechanisms involved in degradation, in particular the enzymatic breakdown of major lignocellulose components as cellulose, hemicelluloses and lignin. Enzymatic apparatus of the saprotrophic fungus *Fomes fomentarius* was explored both in vitro as well as in vivo. Several Actinobacteria were isolated from soil and comparative experiments, investigating production of hydrolytic enzymes, were carried out to track the transformation of polysaccharides and lignin by these strains.

To explain the roles of lignocellulose decomposers in complex environments like soil, the community composition of fungi and bacteria, with special focus on Actinobacteria, was investigated. The use of next generation sequencing (NGS) methods to cover this task required implementation and design of appropriate tools for data handling, not available at the time when the studies were conducted. This resulted in the development of the software pipeline SEED. With the help of this tool, active degraders in forest soil were identified by the comparison of RNA and DNA communities. Further, it was demonstrated that some microbial taxa show high RNA/DNA ratio or were detected only in RNA pool and thus they are underestimated or missing in studies based on DNA analysis. Results also confirmed the importance of Actinobacteria showing that they belong to the most active bacterial groups especially in soil organic horizon.

To provide an in-depth analysis of actinobacterial communities along the gradient of heavy metal contamination, where they were expected to represent the most metabolically active group, the method for a selective community composition analysis of the

Actinobacteria using 454 pyrosequencing was developed that allows community composition at a high resolution. The study showed that diversity of Actinobacteria was unaffected by heavy metal content, but the contamination changed the community composition significantly. Results also confirmed that Actinobacteria thrive better than other bacteria in contaminated soils and may thus serve as important degraders of lignocellulose.

Finally, to get the most accurate estimates of the relative abundance of fungal and bacterial taxa obtained by NGS, methods improving these estimates were proposed. This part of study investigated the impact of the variability of 16S copy numbers among the bacterial genomes on the diversity and abundances of different taxonomical groups and steps to refine the abundance estimates were suggested. In the case of fungal communities, an alternative marker derived from the single-copy gene *rpb2* was compared with the widely used multi-copy ITS with the aim to reduce the problems of community analysis due to the intragenome variability and multicopy nature of the latter marker. Broad taxonomic coverage, suitability for taxonomic assignments and sufficient variation for the use in phylogenetic analyses were confirmed for the newly proposed marker.

Abstrakt

Odumřelá rostlinná biomasa je klíčovým zdrojem uhlíku v pevninských ekosystémech, a proto je její rozklad pro koloběh tohoto prvku na zemi esenciální. Mezi nejvýznamnější rozkladače jsou řazeny především houby, a to zejména díky jejich schopnosti produkovat vysoce účinné oxidativní a hydrolytické enzymy. Mnoho současných prací dále ukazuje, že kromě hub mohou být dalšími důležitými rozkladači celulózy a lignocelulózy také bakterie, zejména zástupci kmene Actinobacteria.

Tato disertační práce je zaměřena na zhodnocení role hub a aktinobakterií při rozkladu odumřelé rostlinné biomasy. V rámci jejího řešení byly nejdříve zkoumány mechanismy tohoto rozkladu, týkající se především rozkladu hlavních komponent lignocelulózy, celulózy, hemicelulóz a ligninu pomocí extracelulárních enzymů. U saprotrofní houby *Fomes fomentarius* byl důkladně popsán enzymový aparát a to jak *in vitro*, tak *in vivo*. Schopnost produkce hydrolytických enzymů byla dále testována u aktinobakterií isolovaných z půdy a u vybraných kmenů byly provedeny experimenty sledující osud polysacharidů a ligninu v průběhu jejich dekompozice.

Pro objasnění úlohy rozkladačů lignocelulózy v komplexních prostředích, jako je půda, bylo zkoumáno složení společenstev hub a bakterií se speciálním zaměřením na aktinobakterie. K dosažení tohoto cíle bylo využito metod sekvenování nové generace (NGS), které vyžadovalo navržení a implementaci nástrojů pro práci se získanými daty, v dané době ještě nedostupných. Tato snaha vyústila ve vytvoření programu pro zpracování sekvenačních dat SEED. S pomocí tohoto programu byli na základě porovnání RNA a DNA komunit odhaleni aktivní mikrobiální rozkladači v lesní půdě. Z porovnání je patrné, že mnohé významné taxonomické skupiny mikroorganismů mají vysoký poměr RNA/DNA nebo jsou obsaženy pouze v RNA, a tak mohou být podceněny nebo úplně chybět ve studiích, kde byla analyzována pouze DNA. Výsledky také potvrdily význam aktinobakterií jako jedné z nejaktivnějších bakteriálních skupin, zejména v organickém horizontu půdy.

Pro získání detailního popisu společenstva aktinobakterií žijících v půdách kontaminovaných těžkými kovy, ve kterých mohou představovat metabolicky nejaktivnější skupinu mikroorganismů, byla vyvinuta metoda pro jejich selektivní analýzu pomocí 454 pyrosekvenace s vysokým rozlišením. Použití této metody ukázalo, že diversita aktinobakterií není ovlivněna obsahem těžkých kovů, avšak koncentrace těchto kovů významně ovlivňují

složení jejich společenstva. Výsledky také podporují předpoklad, že aktinobakterie v těchto extrémních podmínkách prospívají lépe než jiné bakterie, a mohou zde představovat důležité rozkladače lignocelulózy.

Práci uzavírají studie zaměřené na zlepšení odhadů relativních četností taxonů hub a bakterií zkoumaných metodami NGS. V případě bakterií byl analyzován vliv variability jednotlivých kopií 16S rRNA genů a jejich počtu v bakteriálních genomech na odhad diverzity a četnosti různých taxonomických skupin a byla navržena metoda umožňující zpřesnění těchto odhadů. V případě houbových komunit bylo provedeno srovnání analýz houbových společenstev získaných pomocí běžně používaného markeru ITS a alternativního markeru odvozeného od genu *rpb2*, vyskytujícího se převážně v jedné kopii. Cílem bylo zejména snížit problémy spojené s vnitrogenomovou variabilitou a výskytem mnoha kopií ITS markeru v houbových genomech. U nově navrženého markeru *rpb2* se podařilo prokázat široké taxonomické pokrytí, vhodnost pro identifikaci taxonomického zařazení a dostatečnou variabilitu pro použití ve fylogenetických analýzách.

List of Abbreviations

BG	1,4-β-glucosidase
bp	base pair
CBH	cellobiohydrolase
DGGE	denaturing gradient gel electrophoresis
DNA	deoxyribonucleic acid
EG	endoglucanase
GH	glycoside hydrolase
NGS	next generation sequencing
OTU	operational taxonomic unit
PCR	polymerase chain reaction
PMO	polysaccharide monooxygenase
RNA	ribonucleic acid
rRNA	ribosomal ribonucleic acid
SMRT	single molecule real time sequencing
T-RFLP	terminal-restriction fragment length polymorphism

1. Introduction

The thesis focuses on the characterization of main drivers involved in dead plant biomass degradation which can be represented by saprotrophic fungi and some bacteria, especially Actinobacteria.

During the years of my PhD study in the Laboratory of Environmental Microbiology the direction of my work has changed several times and branched to cover biochemistry, molecular biology, and bioinformatics. Even though some of the pieces of my work appeared to be quite distant from the original topic of my thesis, I am convinced that they all represent integral parts of my research efforts, shaped largely by the arrival of novel methods in molecular ecology. These methods changed the paradigm in molecular biology, especially the expansion of next generation sequencing (NGS).

The aims of my research were:

1. To compare enzyme production in common wood rotting and litter decomposing fungi and search for the "model" fungus presented in our forest environment.
2. To explore ecology of coarse wood colonized by white rot fungi *Fomes fomentarius*, measure local distribution of enzyme activity and evaluate the influence of fungal and bacterial community to enzyme production.
3. Optimize the method for monitoring of hydrolytic enzymes activities on various substrate surfaces at high resolution.
4. Screen the cellulose degrading enzymes produced by soil Actinobacteria and explore their ability to decompose lignin.
5. Develop methods for NGS data processing and analysis to identify the active microbial decomposers through the comparison of DNA and RNA communities.
6. Get in-depth insight into the actinobacterial community composition in heavy metal contaminated soils where they may potentially represent the major metabolically active group.
7. Improve the estimates of the relative abundance of fungal and bacterial taxa obtained by NGS.

In my initial works, in order to understand the processes involved in lignocellulose degradation, I focused on biochemical apparatus of degraders, specifically on oxidative and hydrolytic enzymes. This part of my work is represented by the comparison of extracellular enzyme production of saprotrophic basidiomycetes and isolation and characterization of a novel enzyme from the white-rot fungus *Fomes fomentarius* (Paper I), followed by a detailed study of the ecology of coarse wood decomposition by *Fomes fomentariu*s where the role of bacteria in degradation was also discussed (Paper II). I have also contributed to the development of the method for high resolution monitoring of enzyme activities on surfaces of natural substrates (e.g. decomposing wood with fungal fruiting bodies, decayed leaves etc.; Paper III).

Although fungi are often regarded as major lignocellulose decomposers, there are some groups of bacteria expected to significantly contribute to lignocellulose degradation, especially the Actinobacteria. To explore their ability to decompose poorly degradable compounds which compose dead plant biomass (polyphenols as lignin or his derivatives), analysis of co-metabolic transformation of radio labelled compounds by selected *Streptomyces* strains was investigated (Paper IV). Before those experiments, isolation and characterization of dozens of Actinobacteria was performed providing the information about their ability to produce polysaccharide hydrolases.

To understand the processes in the complex environment of soil, it is essential to know the microorganisms which live there. With the breakthrough development of massive parallel sequencing, generation of sequencing datasets containing up to millions of sequences within a few days, we had acquired the tool to get enough information to describe the microbial community composition. However, the data analysis of the early studies faced the problems how to process such huge datasets effectively, correctly and with proper interpretation. To resolve these problems, I have designed and developed the software tool called SEED and a workflow for data processing was proposed (Paper V). Still today, despite the presence of many well-established pipelines for amplicon data analyses, most of them are more difficult to comprehend and handle by biologists without the necessary command line skills and background in bioinformatics. SEED was created to provide an intuitive interface for fast bioinformatic analysis of PCR amplicons on desktop computers.

After the necessary tools were developed, the NGS methods were used to identify the active microbial decomposers through the comparison of DNA and RNA communities. To sample the community in the phase when decomposition of organic matter in soil prevails, the experiment was carried out in the beginning of winter, shortly after snowfall. To target an important decomposition process, apart from widely used markers derived from ribosome coding genes (16S for bacteria and ITS for fungi), fungal exocellulase gene *cbhI* was used for the comparison of the gene and transcript pools (Paper VI).

Because of the interest in Actinobacteria as potential degraders of lignocellulose I have developed the method for a detailed analysis of actinobacterial community composition in soil through massive parallel sequencing. The actinobacterial community in grassland soils along a gradient of heavy metal contamination was selected for the experiment, assuming that their mycelial growth and previously reported resistance to various heavy metals may result in their important role in decomposition in metal-contaminated locations (Paper VII).

Analyses of bacterial and fungal communities based on markers derived from ribosomal genes, especially the estimates of diversity and community composition, suffer from the fact that these marker genes can be present in multiple copies, often different within a single genome. We thus describe the variability of 16S rRNA genes in accessible fully sequenced bacterial genomes and propose a method for more accurate estimations (Paper VIII). While bacterial ribosomal genes are spread along the chromosome and the variability and number of their copies can be estimated from the known complete sequences of genomes, the known fungal ribosomal gene clusters are repetitions of up to hundreds of copies which can be hardly assembled and thus their exact numbers are unknown. To overcome the problem of complicated fungal abundance quantification, the community analysis based on alternative marker derived from single copy gene *rpb2* was proposed (Paper IX).

2. Scientific background

2.1. Dead plant biomass composition

Dead plant biomass is an organic material derived from plants as plant litter and wood. Dominant part of this material is represented by lignocellulose composed from cellulose, hemicelluloses and lignin (Fig. 1). Cellulose is a polysaccharide composed of linear chains of several hundreds to thousands of β(1->4) linked D-glucose units. Long chains of cellulose are stabilized by hydrogen bonds to make micro-fibrils which are organized into solid fibrous structures varying from highly organized, crystalline regions to amorphous, less organized ones (Osullivan, 1997). Hemicelluloses are hetero-polysaccharides made of linear and branched chains composed from several different monosaccharides. Those monosaccharides are often D-glucose, D-xylose, D-mannose, L-arabinose, D-galactose or D-glucuronic acid and proportion of individual saccharides varies among plants. Majority of hemicelluloses contain from two to six different sugar units linked by various types of glycosidic bonds. The length of the chains typically ranging between 50 to 200 units (Berg et al., 2003). In contrast to the polysaccharides, lignin is an amorphous branched polymer covalently linked to hemicellulose (Oinonen et al., 2015). Lignin is generated by oxidation and condensation of radicals of aromatic alcohols (monolignols). These aromatic alcohols are mainly coniferyl alcohol, sinapyl alcohol, *p*-coumaryl alcohol and their derivatives. Polymerization reactions produce a high variety of structures with different types of linkages most often represented by β-O-4 aryl ether bonds and lignin is thus a highly resistant, water-insoluble compound (Ralph et al., 2007). In addition, plant biomass also contains variable proportion of other compounds including proteins, pectins, leaf protecting compounds as cutins and tannins together with compounds derived from plant basal metabolism as amino acids, simple sugars and short chain fatty acids.

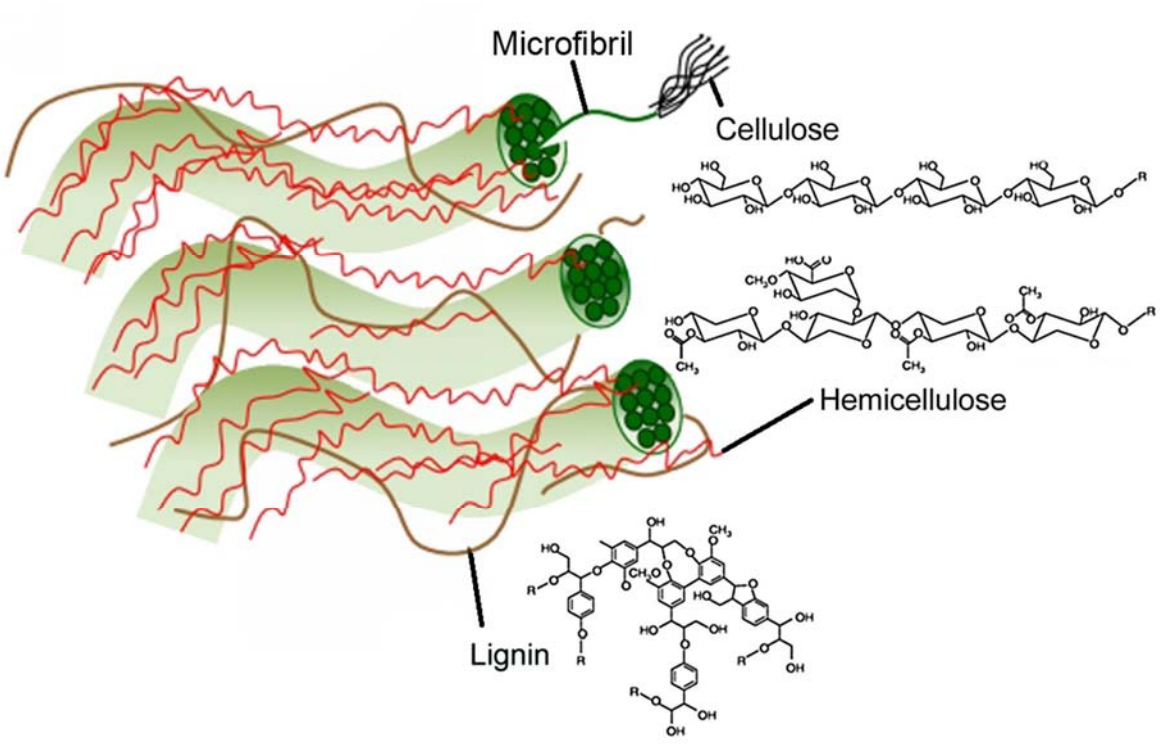


Figure 1. Structural organization of lignocellulose based on (Lee et al., 2014).

2.2. Mechanisms of lignocellulose degradation

The name “lignocellulose” highlights the two main components of dead plant biomass, lignin and cellulose. While the lignin represents highly resistant polyphenolic material degradable by a limited group of organisms, cellulose and other polysaccharides can be cleaved relatively easily and thus represent a good source of carbon and energy for a wide range of microorganisms and herbivores. Lignocellulose represents more than 60% of total biomass on Earth and thus the decomposition of lignocellulose in the soil environment is an essential process of the carbon cycle (Eriksson et al., 1990; Kuhad et al., 1997).

Fungi are often regarded to be the major lignocellulose decomposers (Baldrian et al., 2011) because of their highly effective extracellular enzymes and filamentous growth that make them well suited for the exploitation of bulky lignocellulose substrates (de Boer et al., 2005). The cellulolytic systems of fungi were traditionally thought to consist of endo- and exo-acting enzymes (Fig. 2).

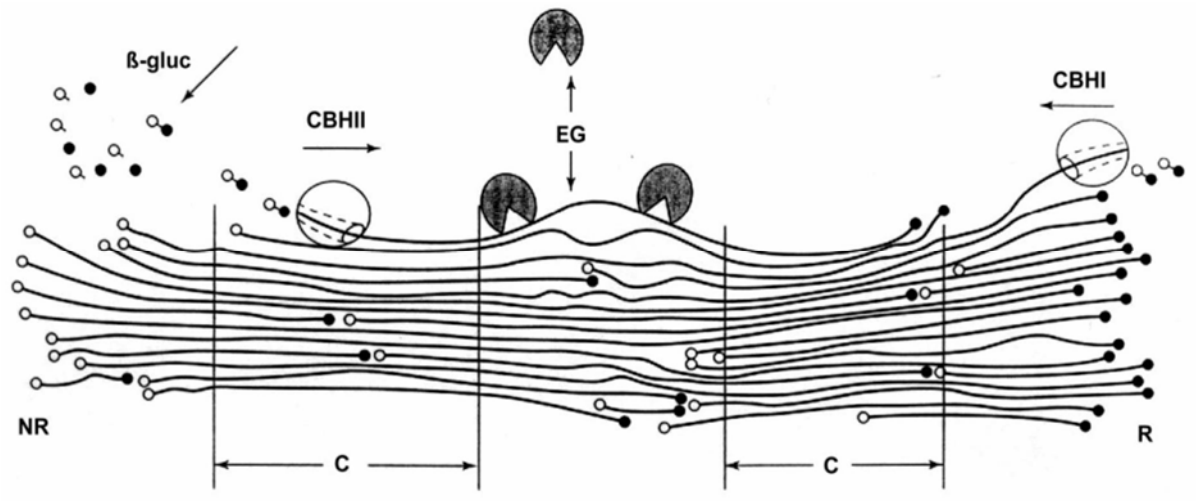


Figure 2. Cellulolytic system of fungi represented by endoglucanases (EG) introducing random cuts into the cellulose chains, cellobiohydrolases hydrolysing cellobiose units from the reducing (CBH I) or the nonreducing ends (CBH II) of cellulose microfibrils and 1,4-β-glucosidases (β-gluc) (Teeri, 1997)

The endoglucanases (EG) introduce random cuts into the polysaccharide chain and cellobiohydrolases (CBH; exocellulases) processively cleave the ends of the molecules. The resulting low molecular oligomers such as cellobiose are hydrolysed to glucose by enzymes like 1,4-β-glucosidase (BG) (Baldrian and Valášková, 2008). The idea that direct enzymatic cleavage is the sole mechanism of its decomposition seems, however, too simple. Nowadays we know that cellulose is often cleaved by oxidative reactions based on Fenton chemistry,

involvement of quinone reductases, cellobiose dehydrogenase or glycopeptides (Baldrian and Valášková, 2008) and by polysaccharide monooxygenases (PMO) (Žifčáková and Baldrian, 2012) (Fig. 3). The hemicellulolytic system is extensive, consisting of multiple glycosyl hydrolases that are specific for xylose-, mannose-, arabinose- and galactose-containing polysaccharides (Baldrian and Šnajdr, 2011).

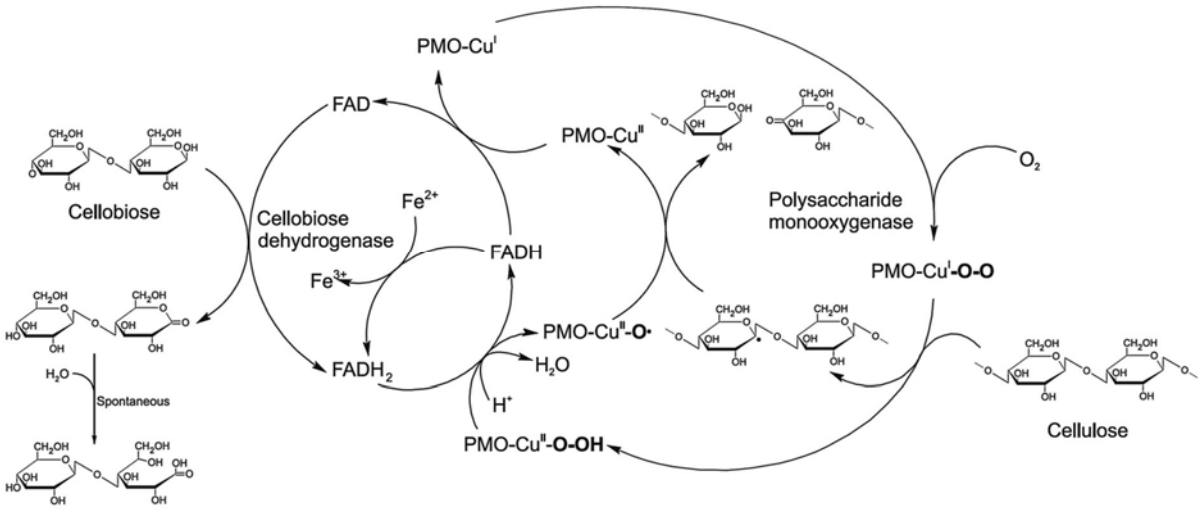


Figure 3. Reactions catalysed by polysaccharide monooxygenase (PMO) and the proposed cooperation of PMO with cellobiose dehydrogenase based on (Žifčáková and Baldrian, 2012).

Lignin degradation is mediated by oxidative enzymes, such as oxidases (laccases), peroxidases (lignin peroxidases, manganese peroxidases) and auxiliary enzymes (Fig. 4) (Martinez et al., 2005). Laccases are copper-containing oxidases which act on phenols and similar molecules, performing one-electron oxidations and peroxidases are oxidoreductases which oxidize phenolic compounds using hydrogen peroxide as a co-substrate (ten Have and Teunissen, 2001). Fungi vary in their production of extracellular enzymes and only certain groups, such as the saprotrophic white-rot fungi, are able to decompose all of the lignocellulose components (Eastwood et al., 2011).

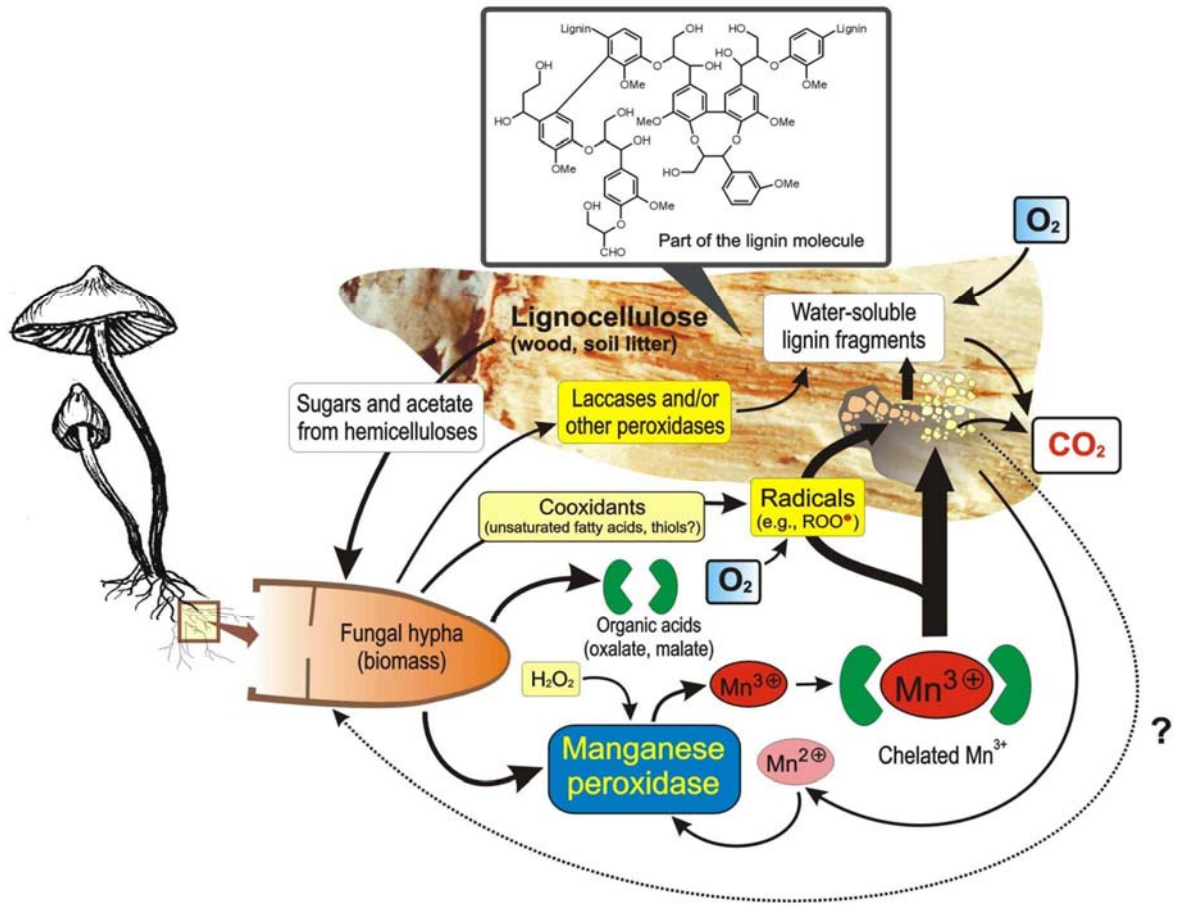


Figure 4. Lignin degradation by basidiomycetous fungi modified from (Hofrichter, 2002).

Theoretical potential of bacteria to degrade lignocellulose is studied for a long time and current advances in genome sequencing identified that genes involved in cellulose and lignocellulose decomposition are relatively widespread (Berlemont and Martiny, 2013, 2015) and bacteria may play a significant role in cellulose decomposition in the soil environments (Štursová et al., 2012). One of the most promising groups of bacteria is the group of Actinobacteria where ability of lignocellulose degradation was intensively studied (Anderson et al., 2012; Crawford, 1978; McCarthy, 1987) and where cellulolytic enzymes are often found in their sequenced genomes (Berlemont and Martiny, 2015).

2.3. Actionobacteria as lignocellulose degraders

Actinobacteria represent a phylum of Gram-positive bacteria with high GC content in their genomes. They are mainly known for their powerful secondary metabolism, especially production of various antibiotics, and typical filamentous mycelial growth (Ventura et al., 2007). Their filamentous growth may help them to access and utilize the polymeric substrates and make Actinobacteria good candidates for lignocellulose decomposition (Chater et al., 2010). It was proven that their relatively large genomes often harbor cellulolytic genes (Berlemont and Martiny, 2013) and many isolates have been found to degrade polysaccharides (Fig. 5) (Anderson et al., 2012; Enkhbaatar et al., 2012). Particularly the members of the genus *Streptomyces* were investigated for the ability to degrade lignocellulose and several enzymes involved in the degradation were isolated and characterized (Crawford, 1978; Chater et al., 2010).

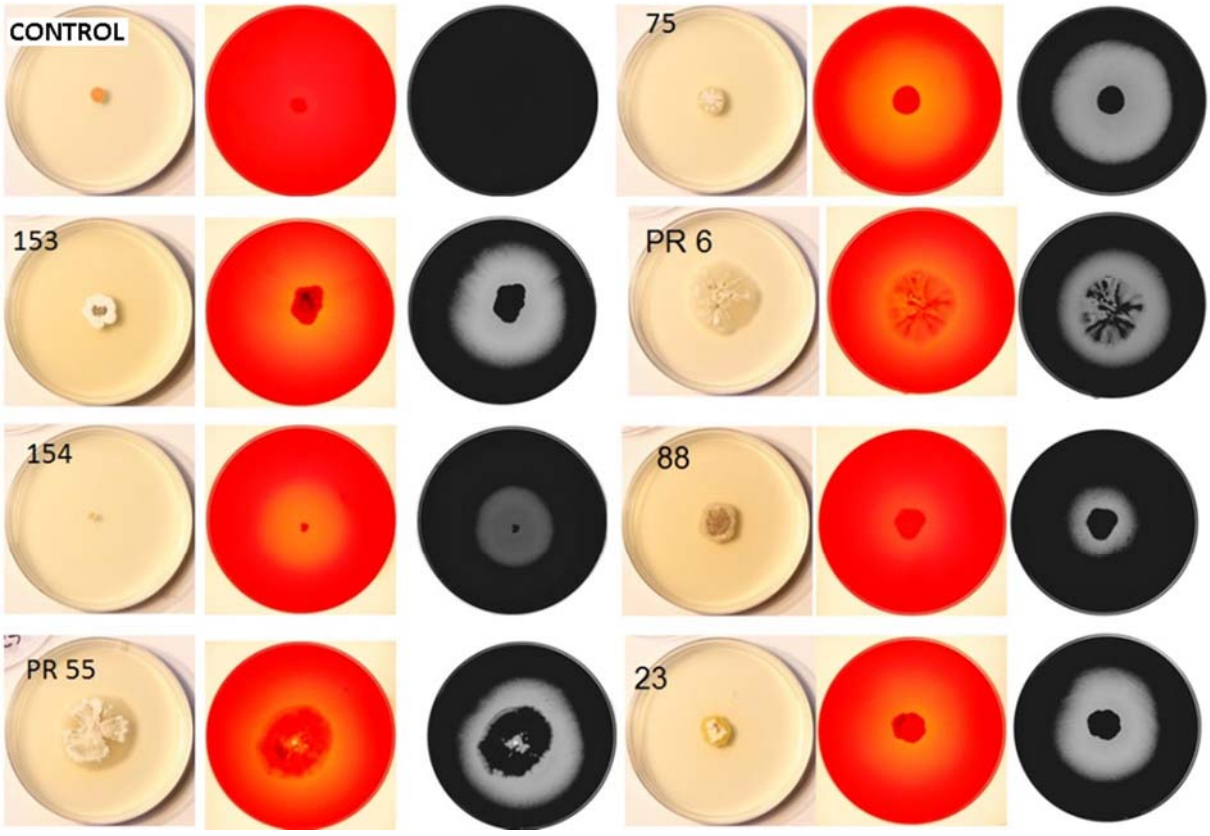


Figure 5. Carboxymethyl cellulose test shows the ability of endo-cellulase production of several actinobacterial strains isolated from soil (unpublished data).

Cellulolytic enzymes of Actinobacteria cover all steps necessary for complete cellulose decomposition. The production of endocellulases by the genera *Streptomyces*, *Cellulomonas* and *Acidothermus* were reported (Enkhbaatar et al., 2012; Sun et al., 2007; Yin et al., 2010), exocellulase activity was found in *Thermobifida*, *Cellulomonas* and *Cellulosimicrobium* (Song and Wei, 2010; Yoon and Choi, 2007; Zhang et al., 2012) and many actinobacterial β-glucosidases have been characterised in the above genera as well as in *Clavibacter*, *Terrabacter*, *Micrococcus*, *Microbacterium* and *Bifidobacterium* (An et al., 2010; Fan et al., 2011; Nakano et al., 1998; Nunoura et al., 1996; Quan et al., 2012). While information of hemicellulose-degrading systems is scarce, individual enzymes were reported in multiple genera, including *Streptomyces*, *Cellulomonas*, *Cellulosimicrobium* and *Kocuria* (Khanna and Gauri, 1993; Li et al., 2009; Oh et al., 2008; Petrosyan et al., 2002).

As mentioned, white-rot fungi are considered to be the major lignin degraders, nevertheless there are studies reporting the decomposition of natural and synthetic lignins by Actinobacteria. Several degradation experiments using isolated lignin, prepared ^{14}C synthetic lignins or model compounds, indicated that the genera *Arthrobacter*, *Nocardia* and *Streptomyces* may be capable of lignin utilization, although their efficiencies varied widely (Cartwright and Holdom, 1973; Crawford et al., 1982; Sutherland et al., 1979; Trojanowski et al., 1977).

Although Actinobacteria are less efficient degraders of lignocellulose than fungi, they have important role in the carbon cycle in soil. Their specific physiology in combination with common spore formation represent suitable traits for the survival of Actinobacteria under stress conditions such as heavy metal-contaminated soils; several resistance mechanisms developed by Actinobacteria were described (Bajkic et al., 2013; Ivshina et al., 2013; Schmidt et al., 2005). Due to their heavy metal resistance they may theoretically replace the more sensitive fungi as the main decomposers in heavy metal contaminated soils and may potentially represent an important metabolically active group in such soils (Harichova et al., 2012).

2.4. Fungal and bacterial communities in soils

Soil represents a complex environment with very high diversity of microorganisms where the structure and activity is dependent on the distribution of nutrients, the most abundant of which are the polysaccharides contained in the dead plant biomass (Prosser, 2002). These nutrients typically show uneven distribution along the soil profile and cause the vertical stratification of soil properties characteristic especially for forest soils. In forest soils, it is possible to observe clearly distinct horizons of plant litter on the top, followed by the organic (humic) horizon composed of recalcitrant residues resulting from enzymatic digestion of litter and mineral soil with low nutrient contents (Fig. 6) (Šnajdr et al., 2008). Moreover, in temperate forests, seasonal fluctuation of photosynthetic activity of trees generates unequal allocation of soil carbon during the year (Högberg et al., 2010; Kaiser et al., 2010) and represents another driver of spatial and temporal distribution of certain taxonomical groups of microbes such as the ectomycorrhizal fungi.

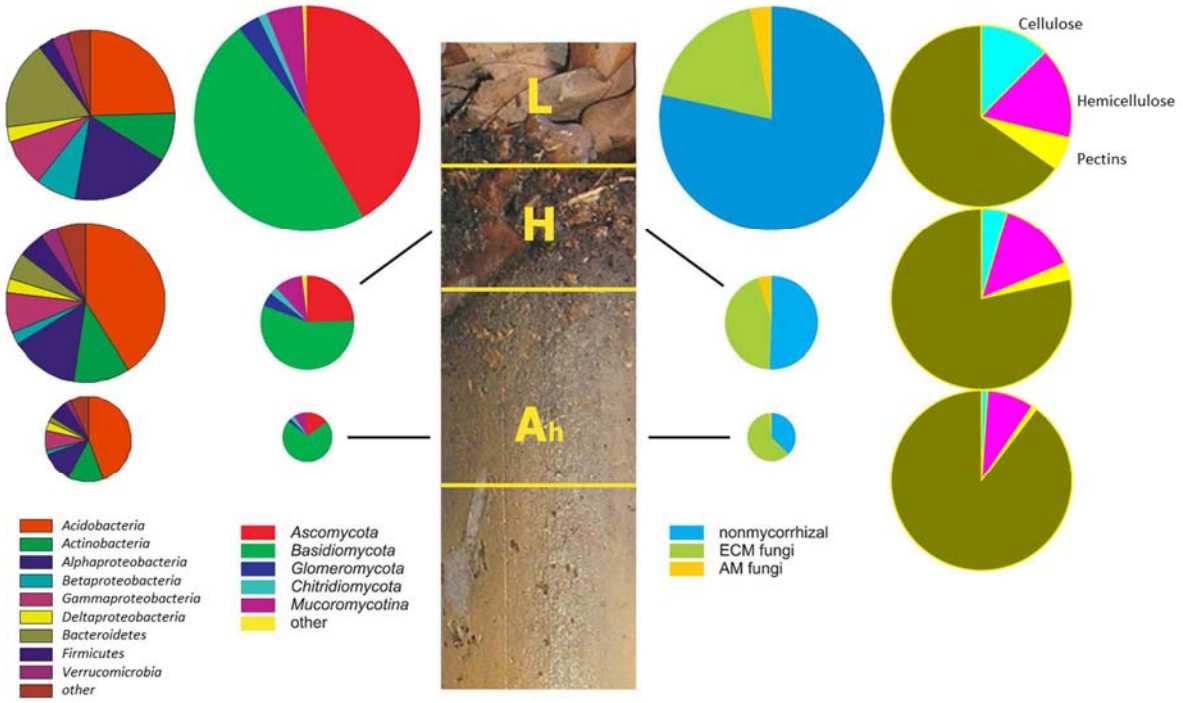


Figure 6. Vertical stratification of soil. The size of the circles reflects a relative abundance of microorganisms in certain soil horizon (Voříšková et al., 2014).

Fungal communities in boreal and temperate forests show vertical distribution with more or less complete spatial separation of saprotrophic and mycorrhizal fungi (Lindahl et al., 2007). Saprotrophic fungi primarily inhabit litter, where carbon can be accessed by their enzymatic apparatus, while mycorhizal fungi dominate in deeper layers where they mobilize nitrogen and provide it to plants (Lindahl et al., 2007; O'Brien et al., 2005). From previous culture-based studies, it is apparent that the main litter decomposers belong to the Ascomycota which preferentially remove cellulose (Koide et al., 2005a; Koide et al., 2005b; Osono et al., 2003) and to litter-decomposing Basidiomycota (Kubartova et al., 2009) dominant in later stages (Frankland, 1998; Osono, 2007; Voříšková and Baldrian, 2013). Endophyte contribution to early stages of litter decomposition is still uncertain but expected (Voříšková and Baldrian, 2013; Žifčáková et al., 2011). Mycorrhizal species are not supposed to participate substantially in lignocellulose degradation, but the recent genome studies proved the presence of cellulases in their genomes and there are some hypotheses that they can use decomposition abilities when mining for organic N (Bödeker et al., 2014; Köhler et al., 2015; Štursová et al., 2012).

While fungi dominate the decomposition of litter material, the importance of bacteria increases with soil depth (Bååth and Anderson, 2003). Comparing fungal and bacterial communities, bacteria seem to be more generally distributed and the composition of their communities is shaped mainly by environmental factors such as temperature, pH and water content (Cong et al., 2015). Especially pH is often considered as a significant factor which influences the bacterial community composition in forest soils. It was reported that bacterial abundance and diversity decrease with decreasing pH where Acidobacteria dominate in acidic soils (Lauber et al., 2009; Rousk et al., 2010), on the other hand litter horizon in certain acidic soils may exhibit higher phylogenetic diversity and reduced proportion of Acidobacteria in the community (Lopez-Mondejar et al., 2015).

Despite the continuous attention to the question how bacterial communities contribute to lignocellulose decomposition, and although many lignocellulose-decomposing bacteria were found particularly among Actinobacteria and Acidobacteria (Berlemont and Martiny, 2013), a clear answer is still missing.

2.5. The use of ribosomal genes for fungal and bacterial community analyses

The ribosomal ribonucleic acid (rRNA) genes are frequently used as a target for identification of isolated organisms or for microbial community analyses (Fig. 7) (Hugenholtz et al., 1998; Schoch et al., 2012). Novel sequencing techniques of the next-generation sequencing (NGS) allow targeted amplicon sequencing to produce at least thousands of sequence reads of amplicons of a certain target region. The sequence reads obtained by NGS differ both because of the community composition and sequencing errors. To reduce this high amount of observed diversity in the obtained datasets, the sequences are often grouped (clustered) on arbitrary defined similarity level forming an operational taxonomic unit (OTU) which is an operational definition of a taxonomical level (mostly species or group of species) often used when only DNA sequence data are available (Blaxter et al., 2005).

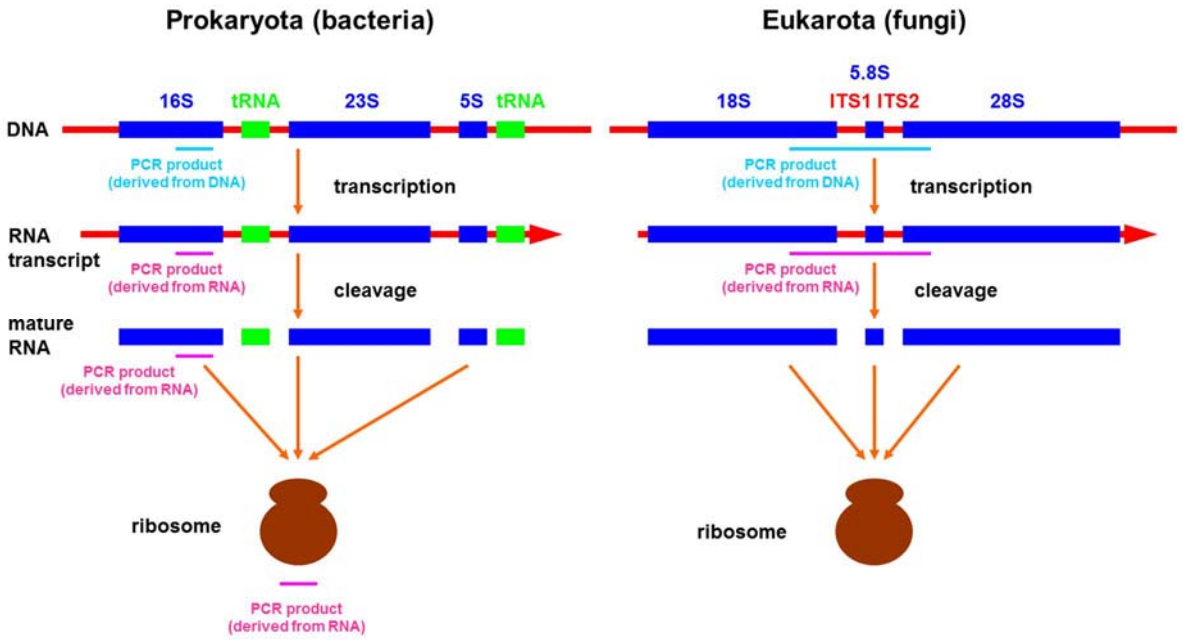


Figure 7. Most used target regions within the rRNA gene in prokaryotes and eukaryotes.

For bacteria, the most widespread and almost exclusive target for the description of their communities is 16S rRNA gene, containing several hypervariable regions and thus allowing to distinguishing between closely related species (Fig. 8). When the bacterial community composition is analyzed by amplicon sequencing of the 16S rRNA genes the obtained sequences are mostly clustered to OTUs using 97% sequence similarity threshold often regarded to represent bacterial species (Hagstrom et al., 2000). Number of OTUs thus

represents the diversity of analyzed community and the amount of sequences belonging to certain OTU is used as an approximation of the "species" abundance. This method of community composition description is often used without any customization, although it is known that copy numbers of 16S in bacterial genomes vary and intragenomic variability between individual copies is often observed (Bodilis et al., 2012; Klappenbach et al., 2001). This variability can cause overestimation or underestimation of abundance of certain bacterial taxa or inflation of diversity, but many authors did not take it into the account.

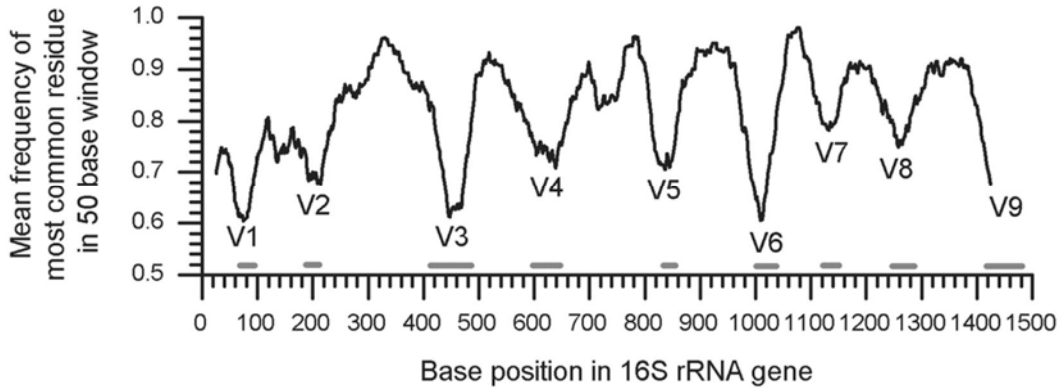


Figure 8. Variable regions within the 16S rRNA gene. The locations of the hypervariable regions are labeled, with gray bars on the x axis defining these regions as V1 to V9. Based on (Ashelford et al., 2005).

Up to now, more than 4.000 bacterial genomes were fully assembled (http://www.ncbi.nlm.nih.gov/genome/browse/). Through the analysis of fully sequenced genomes, it was observed that bacterial genomes may contain anything between 1 and 15 copies of the 16S and that the 16S copies in the same genomes may differ (Klappenbach et al., 2001). It is also suggested that no correlation exists between bacterial genome size and 16S rRNA copy number (Fogel et al., 1999). The knowledge of rRNA genes variation is thus essential for correct interpretation of sequencing data based on rRNA genes and for more accurate estimates of the relative abundances of bacteria.

For the description of fungal communities, the internal transcribed spacers (ITS) regions 1 and 2 of the rRNA are most often used instead of the rRNA molecules which are often insufficient for species discrimination (Kurtzman and Robnett, 1998; Scorzetti et al., 2002). Fungal genomes typically contain multiple copies of ribosomal genes which are estimated to range from dozens to hundreds and are typically grouped into repetitive

clusters on fungal chromosomes (Herrera et al., 2009; Raidl et al., 2005). Due to the repetitive nature, precise genome assembly and estimation of rDNA copy numbers is unfeasible. From the knowledge of the fungal genome sizes it is possible to estimate abundance of ITS in the genomes by using quantitative PCR and it seems that there is no correlation between phylogeny relation and ITS copy number, as the variation is high even within genera (Fig. 9). To make estimations of the relative abundances of fungal taxa more accurate, genes appearing in fungal genomes in a single copy that exhibit sufficient taxonomic discriminative power seem to represent a suitable alternative. β-Tubulin (*tub2*), translation elongation factor 1-α (*tef1α*), and the second largest subunit of RNA polymerase II (*rpb2*) are examples of such candidate genes for molecular taxonomy.

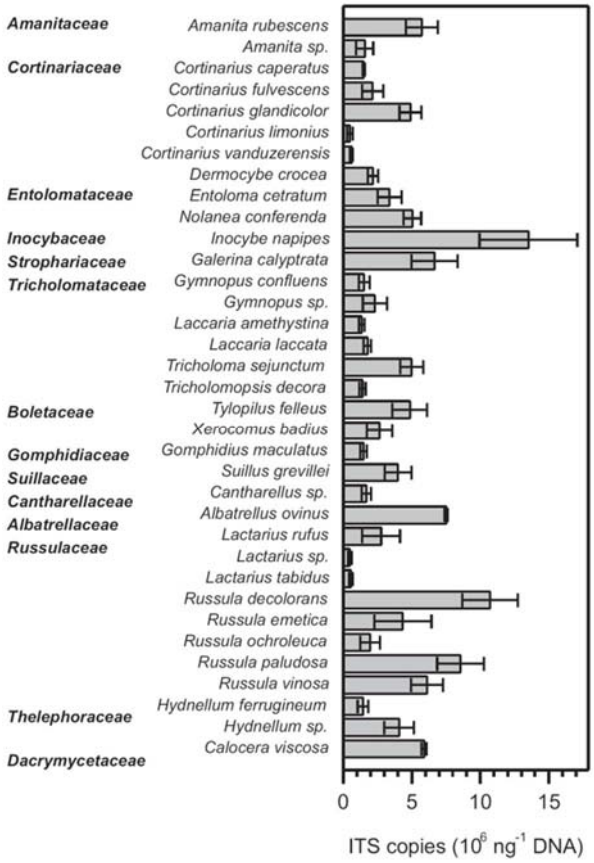


Figure 9. Content of ITS in the DNA of fungi from the *Picea abies* forest. The values represent means and standard deviations from three replicates (Baldrian et al., 2013).

2.6. Next generation sequencing data processing and analysis

Since next generation sequencing (NGS) became commercially available in 2005 (Metzker, 2005), it substantially increased the applicability of molecular biology in ecology research. While in past, labor intensive cloning followed by expensive Sanger sequencing and the use of community fingerprinting methods as denaturing gradient gel electrophoresis (DGGE) or terminal-restriction fragment length polymorphism (T-RFLP) suggested high diversity of soil bacteria (Brons and van Elsas, 2008; Hugenholtz et al., 1998), but without satisfactory amount of data to describe them, NGS provides from tens of thousands to hundreds of millions of nucleotide reads per run and cost of sequencing rapidly decreased as its techniques became popular all over the world. Molecular ecologists use NGS primarily to analyze community composition using ultra-deep sequencing of PCR products amplified from extracted DNA (or RNA) isolated from various environment.

At the beginning there were two major platforms developed by Roche and Illumina. While Roche machines are based on pyrosequencing, a bioluminescence method that measures the release of inorganic pyrophosphate by proportionally converting it into visible light using a series of enzymatic reactions, Illumina is based on sequencing-by-synthesis approach using fluorescently labeled reversible-terminator nucleotides (Kircher and Kelso, 2010). Originally the Roche 454 pyrosequencing was the most suitable platform for molecular ecologists because it provided longer amplicon reads (400-500 bp vs. 50-150 bp of Illumina) and thus allowed full sequencing of molecular barcodes longer than 250 bp. Nowadays, the Illumina platform provides up to 300 bp long paired-end reads which enable the assembly of 500 bp long targets and it is a generally preferred platform because of its very low price per sequence and accuracy highly exceeding 454 (Caporaso et al., 2012; Claesson et al., 2010). To complete, there are also alternative platforms as SOLiD, IonTorrent, Pacific Biosciences, Oxford Nanopore and etc. Some of these platforms are based on new revolutionary technologies. For example Pacific Biosciences has developed a process enabling single molecule real time sequencing (SMRT) using DNA polymerase molecules attached to the bottom of 50 nm-wide wells where addition of fluorescently labeled nucleotides emit fluorescence which is detected in real time (Eid et al., 2009) and thus provides very long (up to 18kb) sequencing reads, but the throughput is low, about 100 MB per run. On the other hand the Ion Torrent PGM is based on detecting the protons

released as nucleotides are incorporated during synthesis without necessity of chemically modified nucleotides or special optics for detection which rapidly reduce the machine and sequencing price (Rothberg et al., 2011). Even though these new platforms have several advantages as a cheap price per read or sequence length, the error rate is still so high that their use for amplicon sequencing is negligible (Quail et al., 2012). Comparison of current most used NGS platforms is summarized in Table 1.

Table 1. Comparison of next-generation sequencing platforms based on (Lee et al., 2013).

Company	Sequencing Principle	Detection	System platform	Read length (bp)	Number of Reads	Time/run	Throughput/run	Accuracy	Machine cost ($)	Advantage	Disadvantage
Illumina	Reversible terminator sequencing by synthesis	Fluorescence/ Optical	HiSeq 2500/1500	36/50/100	3 billion (SE)	2~11 days	600 GB	> 99%	740,000	Very high throughput; Cost-effectiveness; Steadily improving read lengths; Massive throughput	Long run time; Short read lengths; Expensive instrument; Lower error rate
			Genome Analyzer IIx	35/50/75/100	320 million (SE)	2~14 days	95 GB	> 99%	250,000	High throughput; The most widely used platform	Low multiplexing capability of samples
			MiSeq	25/36/100/ 150/250	17 million (SE)	4~27 hours	8.5 GB	> 99%	125,000	High throughput; Cost-effectiveness; Short run times; Appropriate throughput for microbial applications; Minimal hands-on time; High coverage	Short read lengths
Roche	Pyrosequencing	Optical	454 GS FLX+	700	1 million	23 hours	0.7 GB	99.997%	450,000	High throughput; Longer read lengths; Short run times; High coverage	Appreciable hands-on time; High reagent costs; Higher error rate in homopolymers regions
			454 GS Junior	400	1 million	10 hours	0.035 GB	> 99%	108,000	Longer read lengths; Short run times	
Helicos Biosciences	Single molecule sequencing	Fluorescence/ Optical	Heliscope	25~55 (average: 32)	600~800 million	8 days	37 GB	99.99%	999,000	Single-molecule nature of technology; Non-bias representation of templates for genome	Expensive instrument; Very short read lengths (increase cost and difficulty of assembly); Higher error rate
ABI Life Technologies	Ligation	Fluorescence/ Optical	5500 SOLiD	75+35	1.4 billion	7 days	90 GB	99.99%	350,000	High throughput; Lowest reagent cost	Long run times; Very short read lengths (increase cost and difficulty of assembly)
			5500xl SOLiD	75+35	2.8 billion	7 days	180 GB	99.99%	595,000	Very high throughput; Low error rate; Massive throughput	
	Proton detection	Change in pH detected by Ion-Sensitive Field Effect Transistors (ISFETs)	Ion Personal Genome Machine (PGM)	35/200/400	12 million	2 hours	2 GB	> 99%	80,000	Short run times; Low cost per sample; Appropriate throughput for microbial applications; Direct measurement of nucleobase incorporation events	Appreciable hands-on time; High reagent costs; Higher error rate in homopolymers (sequential washing steps)
			Ion Proton Chip I/II	Up to 200	60-80 million	2 hours	10 GB / 100 GB	> 99%	243,000	Short run times; Flexible chip reagents	Instrument not available at time of writing
Pacific Bioscience	Real-time, single molecule DNA sequencing	Fluorescence/ Optical	PacBio RS	Average: 3000	~50 K	2 hours	13 GB	84~85%	750,000	Short run times; Very long read lengths; Low reagent costs; Simple sample preparation	No paired reads; Highest error rates; Expensive instrument; Difficult installation
Oxford Nanopore	Nanopore exonuclease sequencing	Electrical Conductivity	gridION	Tens of Kb	4~10 million	According to experiment	Tens of GB	96%	According to experiment	Extremely long read lengths; Low cost of α-HL nanopore production; Customization; No fluorescent labeling; No optics	4% error rates; Cleaved nucleotide may be read in the wrong order; Difficult to fabricate a device with multiple parallel pores

When Illumina and 454 pyrosequencing are compared, one of the most problematic aspects of 454 pyrosequencing is the high quality decrease when homopolymeric regions are present in sequences (Fig. 10). This quality drop occurrs because the response of the emitted light intensity during the current nucleotide incorporation is not a linear function of homopolymeric region size and when the size of homopolymeric region is greater than specific size (above three nucleotides in a row) software is not able to estimate number of incorporated nucleotides with sufficient quality (Quince et al., 2011). This common type of error inflates the observed OTUs and then skews the estimates of diversity.

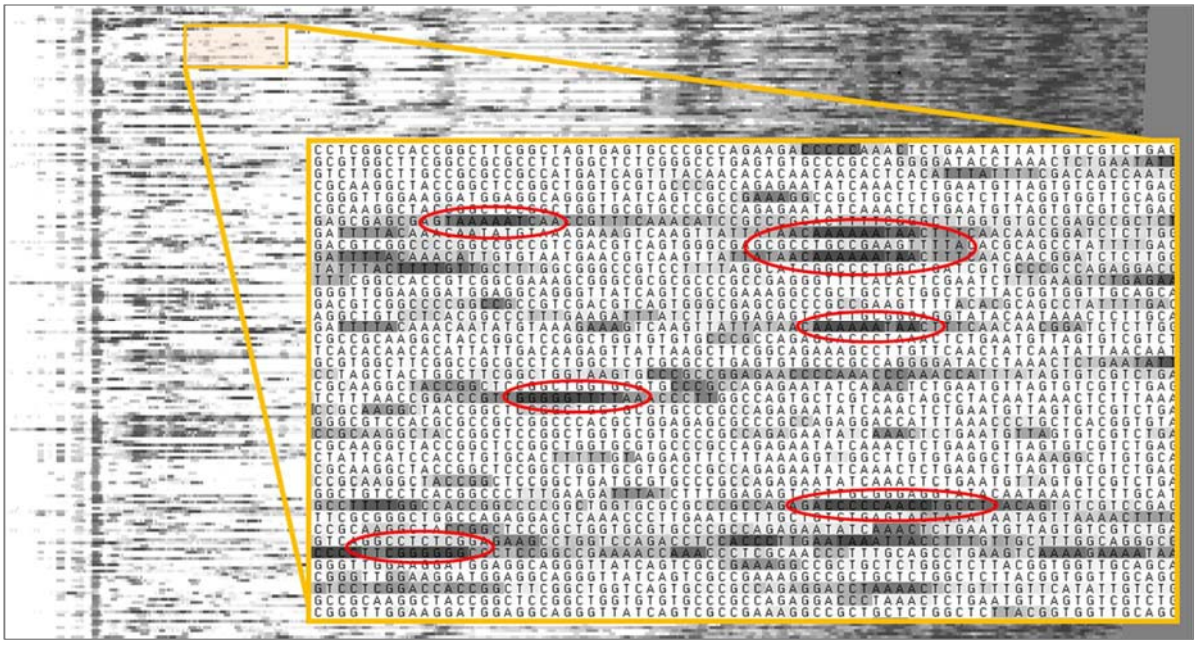


Figure 10. Typical sequence quality drop caused by homopolymeric regions reported in raw 454 pyrosequencing data.

To overcome this problem several approaches were developed from trimming the homopolymeric regions to defined maximal length (which makes the data artificial) to applying algorithms called denoisers (PyroNoise, AmpliconNoise, DeNoiser) to the data (Quince et al., 2011). Basic idea of denoising is based in grouping sequences based on their light intensity patterns (flowgrams) and estimation of the most probable variant for each group (Quince et al., 2009). Illumina does not suffer from such errors because light response after nucleotide incorporation is read after incorporation of each single nucleotide.

Another significant problem causing an inflation of observed OTU is the presence of chimeric sequences in the data. Chimeric sequences (chimeras) are hybrid products between multiple parent sequences often formed during the initial PCR, which can be falsely

interpreted as novel organisms (Fig. 11) (Haas et al., 2011). To remove chimeric sequences from the datasets two main approaches (or their combination) are often used. One approach is based on using curated reference database of non-chimeric sequences e.g. in program Chimera Slayer (Haas et al., 2011). This approach is hard to use when investigating unknown communities with very high diversity (e.g., the forest soil communities). The second approach identifies chimeras *de novo* and is based on identification of parent sequences directly from the dataset comparing the frequencies of sequence parts. Some software tools, such as UCHIME allowing to combine both of the chimera removal approaches (Edgar et al., 2011).

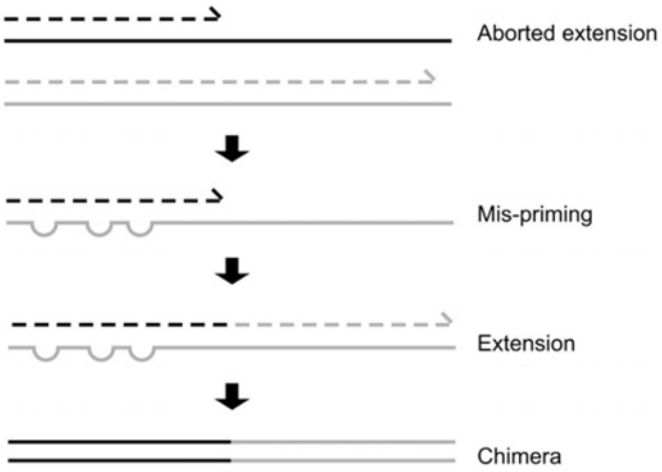


Figure 11. Formation of chimeric sequences during PCR. An aborted extension product from an earlier cycle of PCR can function as a primer in a subsequent PCR cycle. If this aborted extension product anneals to and primes DNA synthesis from an improper template, a chimeric molecule is formed.

Finally, the choice of the algorithm for clustering sequences into OTUs has significant influence on observed diversity. Large amount of amplicon reads force scientists to prefer fast heuristic algorithms over slow accurate hierarchical algorithms. Most used heuristic approach simplifies the problem of sequence comparison to a linear problem (Fig. 12) and was implemented in several programs, e.g. CD-HIT and USEARCH (Edgar, 2010; Li and Godzik, 2006). The latest improved version of fast heuristic program UPARSE has accuracy comparable to hierarchical algorithms (Edgar, 2013).

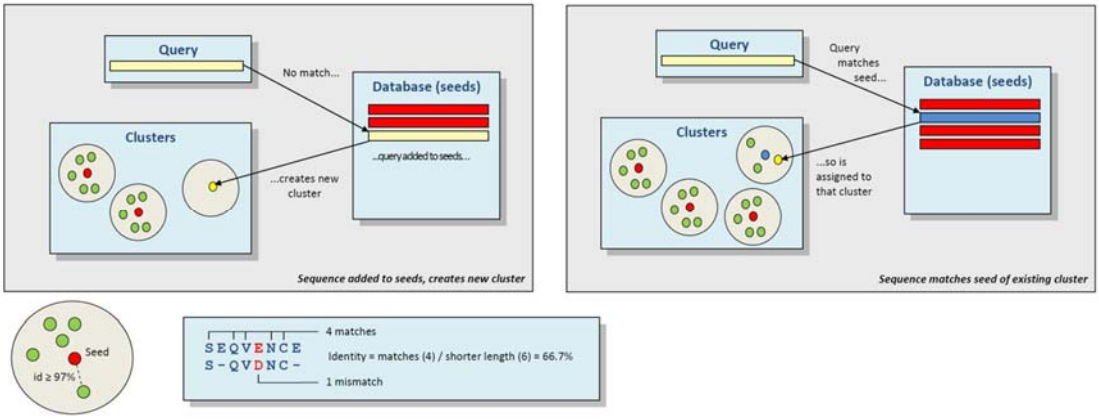


Figure 12. Heuristic approach of sequence clustering used in USEARCH. Sequences are sequentially compared with existing cluster "seeds" (first sequences in the cluster). If the sequence is similar enough with any existing "seed" from the database, it is added to the cluster and the next sequence is processed. When the sequence is dissimilar, new cluster is formed and sequence is used as its "seed". Adapted from (http://www.drive5.com/usearch/usearch.pdf)

Hierarchical algorithms compare every read with the others and are often based on alignment of processed sequences e.g.: MOTHUR, ESPRIT-tree (Cai and Sun, 2011; Schloss et al., 2009). Construction of such big alignments is often problematic and in cases of highly dissimilar sequences, as fungal ITS, it is not possible. Contribution of mentioned sequence processing steps to number of resulting OTUs is depicted in Figure 13.

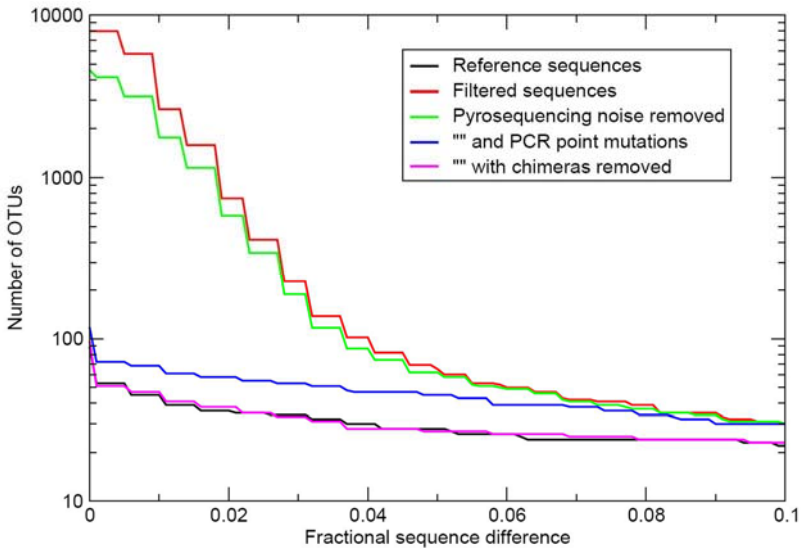


Figure 13. Contribution of several sequence processing steps to number of resulted OTUs in artificial community. Results are shown following filtering (red line), pyrosequencing noise removal by the PyroNoise (green line), further removal of PCR point mutations by the second SeqNoise stage of AmpliconNoise (blue line) and following removal of chimeric sequences (magenta line). For comparison the numbers of OTUs obtained by clustering the known reference sequences are shown in black. The y-axis is logarithmically scaled (Quince et al., 2011).

To simplify the processing of NGS data, programs covering all necessary steps were developed. These programs are called "pipelines" because they allow an appropriate custom workflow depending the user's needs. The first, often incomplete, pipelines for NGS data processing were MOTHUR (Schloss et al., 2009) and QIIME (Caporaso et al., 2010) and the software was originally developed for analyses of bacterial communities and only later modified for analysis of other targets, most importantly fungal communities. It has to be noted that in the time when the first datasets of 454-pyrosequencing were analyzed – which includes the first datasets obtained in our laboratory - these tools were either unavailable, incomplete or under development and did not represent a suitable alternative for data processing. Still up to now, the most widespread and well-established pipelines for amplicon data analysis are difficult to comprehend and handle by biologists who lack the necessary command line skills and background in bioinformatics. These reasons made it necessary to develop a suitable software application combining existing bioinformatic tools and an extended array of options for data handling with a user friendly interface.

3. List of publications

This thesis consists of the following papers:

I. **Větrovský, T.**, Baldrian, P., Gabriel, J. (2013) Extracellular Enzymes of the White-Rot Fungus *Fomes fomentarius* and Purification of 1,4-β-Glucosidase. Applied Biochemistry and Biotechnology 169: 100-109.

II. **Větrovský, T.**, Voříšková, J., Šnajdr, J., Gabriel, J., Baldrian, P. (2011) Ecology of coarse wood decomposition by the saprotrophic fungus Fomes fomentarius. Biodegradation 22: 709-718.

III. Baldrian, P., **Větrovský, T.** (2012) Scaling Down the Analysis of Environmental Processes: Monitoring Enzyme Activity in Natural Substrates at a Millimeter Resolution Scale. Applied and Environmental Microbiology 78: 3473-3475.

IV. **Větrovský, T.**, Steffen, K. T., Baldrian, P. (2014) Potential of cometabolic transformation of polysaccharides and lignin in lignocellulose by soil Actinobacteria. PLoS ONE 9, e89108.

V. **Větrovský, T.**, Baldrian, P. (2013) Analysis of soil fungal communities by amplicon pyrosequencing: current approaches to data analysis and the introduction of the pipeline SEED. Biology and Fertility of Soils 49: 1027-1037.

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

VII. **Větrovský, T.**, Baldrian P. (2015) An in-depth analysis of actinobacterial communities shows their high diversity in grassland soils along a gradient of mixed heavy metal contamination. Biology and Fertility of Soils 51: 827-837.

VIII. **Větrovský, T.**, Baldrian, P. (2013) The Variability of the 16S rRNA Gene in Bacterial Genomes and Its Consequences for Bacterial Community Analyses. PLoS ONE 8, e57923.

IX. **Větrovský, T.**, Kolařík, M., Žifčáková, L., Zelenka, T., Baldrian, P. (*in press*) The *rpb2* gene represents a viable alternative molecular marker for the analysis of environmental fungal communities. Molecular Ecology Resources, DOI: 10.1111/1755-0998.12456

Paper I

Větrovský, T., Baldrian, P., Gabriel, J. (2013) Extracellular Enzymes of the White-Rot Fungus *Fomes fomentarius* and Purification of 1,4-β-Glucosidase. Applied Biochemistry and Biotechnology 169: 100-109.

Production of the lignocellulose-degrading enzymes endo-1,4-β-glucanase, 1,4-β-glucosidase, cellobiohydrolase, endo-1,4-β-xylanase, 1,4-β-xylosidase, Mn peroxidase, and laccase was characterized in a common wood-rotting fungus *Fomes fomentarius*, a species able to efficiently decompose dead wood, and compared to the production in eight other fungal species. The main aim of this study was to characterize the 1,4-β-glucosidase produced by *F. fomentarius* that was produced in high quantities in liquid stationary culture (25.9 Uml^{-1}), at least threefold compared to other saprotrophic basidiomycetes, such as *Rhodocollybia butyracea*, *Hypholoma fasciculare*, *Irpex lacteus*, *Fomitopsis pinicola*, *Pleurotus ostreatus*, *Piptoporus betulinus*, and *Gymnopus* sp. (between 0.7 and 7.9 Uml^{-1}). The 1,4-β-glucosidase enzyme was purified to electrophoretic homogeneity by both anion-exchange and size-exclusion chromatography. A single 1,4-β-glucosidase was found to have an apparent molecular mass of 58 kDa and a pI of 6.7. The enzyme exhibited high thermotolerance with an optimum temperature of 60 °C. Maximal activity was found in the pH range of 4.5–5.0, and K_M and V_{max} values were 62 μM and 15.8 μmol min^{-1} l^{-1}, respectively, when *p*-nitrophenylglucoside was used as a substrate. The enzyme was competitively inhibited by glucose with a K_i of 3.37 mM. The enzyme also acted on *p*-nitrophenylxyloside, *p*-nitrophenylcellobioside, *p*-nitrophenylgalactoside, and *p*-nitrophenylmannoside with optimal pH values of 6.0, 3.5, 5.0, and 4.0–6.0, respectively. The combination of relatively low molecular mass and low K_M value make the 1,4-β-glucosidase a promising enzyme for biotechnological applications.

Paper II

Větrovský, T., Voříšková, J., Šnajdr, J., Gabriel, J., Baldrian, P. (2011) Ecology of coarse wood decomposition by the saprotrophic fungus *Fomes fomentarius*. Biodegradation 22: 709-718.

Saprotrophic wood-inhabiting basidiomycetes are the most important decomposers of lignin and cellulose in dead wood and as such they attracted considerable attention. The aims of this work were to quantify the activity and spatial distribution of extracellular enzymes in coarse wood colonised by the white-rot basidiomycete *Fomes fomentarius* and in adjacent fruitbodies of the fungus and to analyse the diversity of the fungal and bacterial community in a fungus-colonised wood and its potential effect on enzyme production by *F. fomentarius*. Fungus-colonised wood and fruitbodies were collected in low management intensity forests in the Czech Republic. There were significant differences in enzyme production by *F. fomentarius* between *Betula pendula* and *Fagus sylvatica* wood, the activity of cellulose and xylan-degrading enzymes was significantly higher in beech wood than in birch wood. Spatial analysis of a sample *B. pendula* log segment proved that *F. fomentarius* was the single fungal representative found in the log. There was a high level of spatial variability in the amount of fungal biomass detected, but no effects on enzyme activities were observed. Samples from the fruiting body showed high β-glucosidase and chitinase activities compared to wood samples. Significantly higher levels of xylanase and cellobiohydrolase were found in samples located near the fruitbody (proximal), and higher laccase and Mn-peroxidase activities were found in the distal ones. The microbial community in wood was dominated by the fungus (fungal to bacterial DNA ratio of 62-111). Bacterial abundance composition was lower in proximal than distal parts of wood by a factor of 24. These results show a significant level of spatial heterogeneity in coarse wood. One of the explanations may be the successive colonization of wood by the fungus: due to differential enzyme production, the rates of biodegradation of coarse wood are also spatially inhomogeneous.

Paper III

Baldrian, P., **Větrovský, T.** (2012) Scaling Down the Analysis of Environmental Processes: Monitoring Enzyme Activity in Natural Substrates at a Millimeter Resolution Scale. Applied and Environmental Microbiology 78: 3473-3475.

Natural environments often show high levels of spatial heterogeneity. With a methodology based on the immobilization of fluorescent substrates, the distribution of extracellular enzymes can be studied at a 2.3-mm resolution with a detection limit of 1.8 $nmol \cdot h^{-1} \cdot cm^{-2}$. The method is applicable to environmental samples such as wood, litter, soil, or fungal colonies.

Paper IV

Větrovský, T., Steffen, K. T., Baldrian, P. (2014) Potential of cometabolic transformation of polysaccharides and lignin in lignocellulose by soil Actinobacteria. PLoS ONE 9, e89108.

While it is known that several Actinobacteria produce enzymes that decompose polysaccharides or phenolic compounds in dead plant biomass, the occurrence of these traits in the environment remains largely unclear. The aim of this work was to screen isolated actinobacterial strains to explore their ability to produce extracellular enzymes that participate in the degradation of polysaccharides and their ability to cometabolically transform phenolic compounds of various complexities. Actinobacterial strains were isolated from meadow and forest soils and screened for their ability to grow on lignocellulose. The potential to transform ^{14}C-labelled phenolic substrates (dehydrogenation polymer (DHP), lignin and catechol) and to produce a range of extracellular, hydrolytic enzymes was investigated in three strains of *Streptomyces* spp. that possessed high lignocellulose degrading activity. Isolated strains showed high variation in their ability to produce cellulose- and hemicellulose-degrading enzymes and were able to mineralize up to 1.1% and to solubilize up to 4% of poplar lignin and to mineralize up to 11.4% and to solubilize up to 64% of catechol, while only minimal mineralization of DHP was observed. The results confirm the potential importance of Actinobacteria in lignocellulose degradation, although it is likely that the decomposition of biopolymers is limited to strains that represent only a minor portion of the entire community, while the range of simple, carbon-containing compounds that serve as sources for actinobacterial growth is relatively wide.

Paper V

Větrovský, T., Baldrian, P. (2013) Analysis of soil fungal communities by amplicon pyrosequencing: current approaches to data analysis and the introduction of the pipeline SEED. Biology and Fertility of Soils 49: 1027-1037.

Fungi are important in soils as both decomposers and plant symbionts, and an understanding of the composition of their complex communities is thus indispensable to answer a variety of ecological questions. 454 Pyrosequencing is currently the method of choice for the in-depth analysis of fungal communities. However, the interpretation of its results is complicated by differences in data analysis approaches that make inter-study comparisons difficult. The pyrosequencing studies published so far have also used variable molecular targets in fungal rDNA. Although the ITS region and, in particular, ITS1 appear to be the most frequent sequencing targets, the use of various primers with different coverages of fungal groups remains a serious problem. Sequence length limits also vary widely across studies, and in many studies, length differences may negatively affect sequence similarity clustering or identification. Unfortunately, many studies neglect the need to correct for method-dependent errors, such as pyrosequencing noise or chimeric sequences. Even when performed, error rates in sequences may be high, and consensus sequences created by sequence clustering therefore better represent operational taxonomic units. We recommend a data analysis workflow that includes sequence denoising, chimera removal, sequence trimming before clustering and random resampling before calculating diversity parameters. The newly developed free pipeline (SEED) introduced here can be used to perform all the required analytical steps. The improvement and unification of data analysis procedures should make future studies both more reliable and comparable and allow metastudies to be performed to provide more general views on fungal diversity, biogeography or ecology.

Paper VI

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

Soils of coniferous forest ecosystems are important for the global carbon cycle, and the identification of active microbial decomposers is essential for understanding organic matter transformation in these ecosystems. By the independent analysis of DNA and RNA, whole communities of bacteria and fungi and its active members were compared in topsoil of a *Picea abies* forest during a period of organic matter decomposition. Fungi quantitatively dominate the microbial community in the litter horizon, while the organic horizon shows comparable amount of fungal and bacterial biomasses. Active microbial populations obtained by RNA analysis exhibit similar diversity as DNA-derived populations, but significantly differ in the composition of microbial taxa. Several highly active taxa, especially fungal ones, show low abundance or even absence in the DNA pool. Bacteria and especially fungi are often distinctly associated with a particular soil horizon. Fungal communities are less even than bacterial ones and show higher relative abundances of dominant species. While dominant bacterial species are distributed across the studied ecosystem, distribution of dominant fungi is often spatially restricted as they are only recovered at some locations. The sequences of *cbhI* gene encoding for cellobiohydrolase (exocellulase), an essential enzyme for cellulose decomposition, were compared in soil metagenome and metatranscriptome and assigned to their producers. Litter horizon exhibits higher diversity and higher proportion of expressed sequences than organic horizon. Cellulose decomposition is mediated by highly diverse fungal populations largely distinct between soil horizons. The results indicate that low-abundance species make an important contribution to decomposition processes in soils.

Paper VII

Větrovský, T., Baldrian P. (2015) An in-depth analysis of actinobacterial communities shows their high diversity in grassland soils along a gradient of mixed heavy metal contamination. Biology and Fertility of Soils 51: 827-837.

Several previous studies indicated that Actinobacteria may be enriched in soils with elevated content of heavy metals. In this study, we have developed a method for the in-depth analysis of actinobacterial communities in soil through phylum-targeted high-throughput sequencing and used it to address this question and examine the community composition in grassland soils along a gradient of heavy metal contamination (Cu, Zn, Cd, Pb). The use of the 16Sact111r primer specific for Actinobacteria resulted in a dataset obtained by pyrosequencing where over 98 % of the sequences belonged to Actinobacteria. The diversity within the actinobacterial community was not affected by the heavy metals, but the contamination was the most important factor affecting community composition. The most significant changes in community composition were due to the content of Cu and Pb, while the effects of Zn and Cd were relatively minor. For the most abundant actinobacterial taxa, the abundance of taxa identified as members of the genera *Acidothermus*, *Streptomyces*, *Pseudonocardia*, *Janibacter* and *Microlunatus* increased with increasing metal content, while those belonging to *Jatrophihabitans* and *Actinoallomurus* decreased. The genus *Ilumatobacter* contained operational taxonomic units (OTUs) that responded to heavy metals both positively and negatively. This study also confirmed that Actinobacteria appear to be less affected by heavy metals than other bacteria. Because several Actinobacteria were also identified in playing a significant role in cellulose and lignocellulose decomposition in soil, they potentially represent important decomposers of organic matter in such environments.

Paper VIII

Větrovský, T., Baldrian, P. (2013) The Variability of the 16S rRNA Gene in Bacterial Genomes and Its Consequences for Bacterial Community Analyses. PLoS ONE 8, e57923.

16S ribosomal RNA currently represents the most important target of study in bacterial ecology. Its use for the description of bacterial diversity is, however, limited by the presence of variable copy numbers in bacterial genomes and sequence variation within closely related taxa or within a genome. Here we use the information from sequenced bacterial genomes to explore the variability of 16S rRNA sequences and copy numbers at various taxonomic levels and apply it to estimate bacterial genome and DNA abundances. In total, 7,081 16S rRNA sequences were *in silico* extracted from 1,690 available bacterial genomes (1–15 per genome). While there are several phyla containing low 16S rRNA copy numbers, in certain taxa, e.g., the *Firmicutes* and *Gammaproteobacteria*, the variation is large. Genome sizes are more conserved at all tested taxonomic levels than 16S rRNA copy numbers. Only a minority of bacterial genomes harbors identical 16S rRNA gene copies, and sequence diversity increases with increasing copy numbers. While certain taxa harbor dissimilar 16S rRNA genes, others contain sequences common to multiple species. Sequence identity clusters (often termed operational taxonomic units) thus provide an imperfect representation of bacterial taxa of a certain phylogenetic rank. We have demonstrated that the information on 16S rRNA copy numbers and genome sizes of genome-sequenced bacteria may be used as an estimate for the closest related taxon in an environmental dataset to calculate alternative estimates of the relative abundance of individual bacterial taxa in environmental samples. Using an example from forest soil, this procedure would increase the abundance estimates of *Acidobacteria* and decrease these of *Firmicutes*. Using the currently available information, alternative estimates of bacterial community composition may be obtained in this way if the variation of 16S rRNA copy numbers among bacteria is considered.

Paper IX

Větrovský, T., Kolařík, M., Žifčáková, L., Zelenka, T., Baldrian, P. The *rpb2* gene represents a viable alternative molecular marker for the analysis of environmental fungal communities. Molecular Ecology Resources *in press*, DOI: 10.1111/1755-0998.12456

Although the commonly used internal transcribed spacer region of rDNA (ITS) is well suited for taxonomic identification of fungi, the information on the relative abundance of taxa and diversity is negatively affected by the multi-copy nature of rDNA and the existence of ITS paralogues. Moreover, due to high variability, ITS sequences cannot be used for phylogenetic analyses of unrelated taxa. The part of single-copy gene encoding the second largest subunit of RNA polymerase II (*rpb2*) was thus compared with first spacer of ITS as an alternative marker for the analysis of fungal communities in spruce forest topsoil and their applicability was tested on a comprehensive mock community. In soil, *rpb2* exhibited broad taxonomic coverage of the entire fungal tree of life including basal fungal lineages. The gene exhibited sufficient variation for the use in phylogenetic analyses and taxonomic assignments, although it amplifies also paralogues. The fungal taxon spectra obtained with *rbp2* region and ITS1 corresponded, but sequence abundances differed widely, especially in the basal lineages. The proportions of OTU counts and read counts of major fungal groups were close to the reality when *rpb2* was used as a molecular marker while they were strongly biased towards the Basidiomycota when using the ITS primers ITS1/ITS4. Although the taxonomic placement of *rbp2* sequences is currently more difficult than of the ITS sequences, its discriminative power, quantitative representation of community composition and suitability for phylogenetic analyses represent significant advantages.

4. Methods

Bioinformatic analysis of sequencing data (paper V, VI, VII, VIII, IX)

Cultivation of fungi and bacteria (paper I, IV)

Diversity and statistical analysis (paper VI, VII, VIII, IX)

Enzyme assays (paper I, II, III, IV)

Enzyme purification and characterization (paper I)

Library preparation for tag-encoded amplicon pyrosequencing (paper VI, VII)

Measurement of substrate utilization using 14C-labelled substrates (paper IV)

Software development (paper V)

Soil sampling (paper VI, VII)

Taxonomical identification of fungal and bacterial strains (paper II, IV, VIII, IX)

Wood sampling (paper II)

5. Results and Discussion

Saprotrophic wood-inhabiting basidiomycetes play the key role in lignin and cellulose decomposition of dead plant biomass and thus the knowledge of decomposition mechanisms they use represents an important question. The *in vitro* characterization of major enzymes involved in lignocellulose decomposition produced by several saprotrophic basidiomycetes *Rhodocollybia butyracea*, *Hypholoma fasciculare*, *Irpex lacteus*, *Fomitopsis pinicola*, *Pleurotus ostreatus*, *Piptoporus betulinus*, *Gymnopus* sp. and *Fomes fomentarius* showed, as expected, that white-rot species produced high levels of lignin-degrading enzymes, brown-rot fungi produced only cellulolytic and hemicellulolytic enzymes, and litter-colonizing saprotrophic fungi produced both types of enzymes (Tab. 2).

Table 2. Mean production of extracellular enzymes by saprotrophic basidiomycetes (Paper I).

Fungus	EG	EX	CBH	bG	bX	MnP	Lac
Gymnopus erythropus	3.1	46.5	0.90	4.7	2.3	5.9	4.2
Rhodocollybia butyracea	15.8	18.8	0.62	6.2	0.38	74.4	18.1
Hypholoma fasciculare	14.7	19.3	0.77	7.9	0.27	74.8	8.4
Irpex lacteus	25.8	8.3	0.94	2.9	0.39	47.8	1.4
Fomitopsis pinicola	29.9	14.7	0.26	5.0	0.53	0.0	0.0
Pleurotus ostreatus	3.2	1.8	0.34	3.9	0.21	6.6	16.7
Fomes fomentarius	20.7	22.0	10.0	25.9	1.6	69.8	206.8
Piptoporus betulinus	5.6	2.3	0.74	0.65	0.03	0.0	0.0

Enzyme activities ($mUml^{-1}$) were measured after 7, 14, 21, 28, and 35 days of culture in liquid CLN medium; the data represent means of these five measurements (EG endoglucanase, EX endoxylanase, CBH cellobiohydrolase, bG β-glucosidase, bX β-xylosidase, MnP Mn peroxidase, Lac laccase)

Because cellulases are currently the third largest group of industrial enzymes used worldwide according to market size (Acharya and Chaudhary, 2012) and β-glucosidase has a broad range of uses in biotechnological applications (Duan and Feng, 2010) this enzyme produced by *F. fomentarius* in very high amount was purified and characterized (Paper I). The *F. fomentarius* 1,4-β-glucosidase is similar to the small extracellular enzymes produced by *Pleurotus* or *Phanerochaete* spp. (Baldrian and Valášková, 2008). The enzyme was observed to have a relatively high pI of 6.4, which is typical for an intracellular enzyme (Baldrian and Valášková, 2008). The optimal pH (6.0) and temperature (45–75 °C) fall within the range observed for other fungal β-glucosidases (Baldrian and Valášková, 2008), but the high enzyme activity at 60 °C may be beneficial for biotechnological applications that require high process rates. More importantly, the enzyme has a low K_M of 62 μM. The only lower K_M

value observed is that of the β-glucosidase of *Gloeophyllum trabeum*, which has limited commercial applications due to its high molecular mass (Herr et al., 1978). Except β-glucosidase *F. fomentarius* exhibited also higher production of laccase and cellobiohydrolase than any of the other tested fungi and thus was selected for further experiments.

To better understand how the degradative enzymes work under *in vivo* conditions, the production and spatial distribution of extracellular enzymes in coarse wood colonized by the *Fomes fomentarius* were explored (Paper II). Enzyme production varied significantly depending on host tree. The activity of cellulose and xylan-degrading enzymes was significantly higher in beech wood (*Fagus sylvatica*) compared to birch wood (*Betula pendula*). To get the detailed information about ecology of fungus colonized wood an in-depth analysis of a *B. pendula* log segment was done. To evaluate the influence of the present fungal and bacterial community on enzyme production of dominant fungus the microbial community in wood was analyzed (Fig. 14). Bacterial diversity was high which is in accordance with previous studies on bacteria associated with decaying wood (Folman et al., 2008; Valášková et al., 2009; Zhang et al., 2008a) and may indicate the fact that bacteria and fungi actually occupy different decomposition niches and do not directly compete. On the other hand, *F. fomentarius* was surprisingly the single fungal representative in the log according to the denaturing gradient gel electrophoresis, which is more characteristic for initial decay (Vainio and Hantula, 2000; Valášková et al., 2009), since in later phases a relatively rich fungal community usually develops (Lonsdale et al., 2008; Zhang et al., 2008b). Relatively high spatial variability in fungal biomass without any correlation to enzyme activities was observed, which showed high BG and chitinase activities in samples from the fruiting body compared to wood samples. Significantly higher levels of xylanase and cellobiohydrolase were found in samples located near the fruitbody, and higher laccase and Mn-peroxidase activities were found in the distal ones. This corresponds well with previous observations described by Hatakka (2005) who reports that preferential removal of lignin (selective delignification) is performed before extensive utilization of wood polysaccharides during lignocellulose transformation by white-rot fungi. Activities of N-acetylglucosaminidase and laccase, enzymes potentially involved in the interactions among microorganisms (Rast et al., 2003), exhibited positive correlation. Due to its participation in chitin degradation, N-acetylglucosaminidase may be also an indicator of active rearrangement of fungal mycelia (Lindahl and Finlay, 2006). The biomass of bacteria in wood

was very low when compared to fungal biomass content, which supports the view that bacteria are not supposed to be efficient decomposers of bulky lignocellulose like wood (de Boer et al., 2005). As a consequence, it is most likely that the measured enzyme activities reflected the degradative activity of the fungus. It seems that colonization of wood by fungus is often done in consecutive processes which were apparent from spatially inhomogeneous rates of coarse wood biodegradation and differential enzyme production.

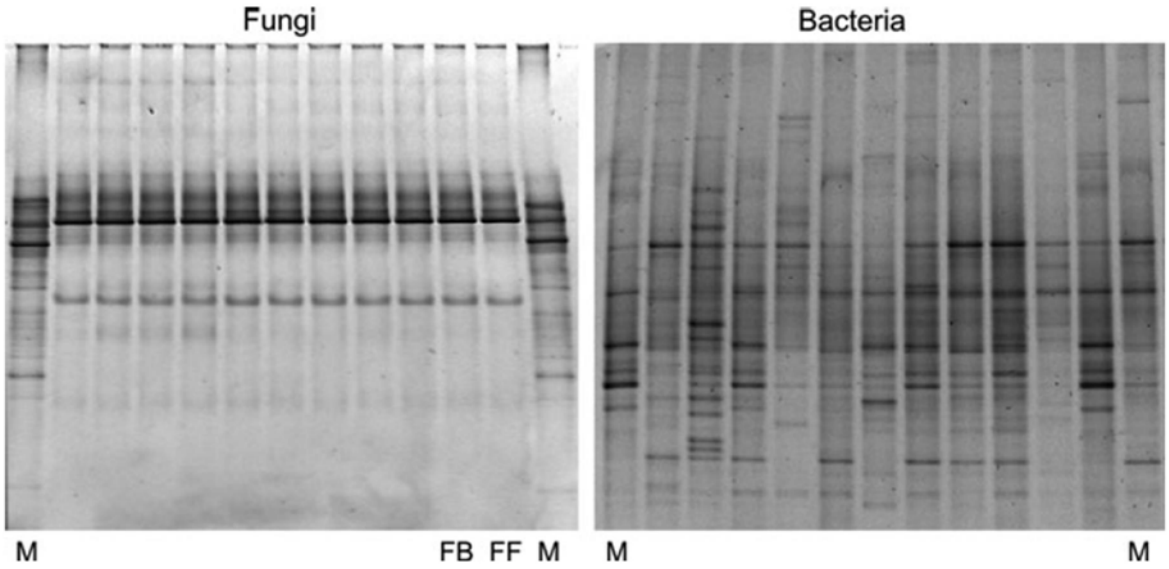


Figure 14. DGGE analysis of fungal and bacterial community in the *Betula pendula* log segment colonised by *Fomes fomentarius*. M - marker, FB - DNA isolated from *F. fomentarius* fruitbody, FF - DNA extracted from the pure culture of *Fomes fomentarius* CCBAS534, all other lanes DNA extracted from wood (Paper II).

Hydrolytic enzymes, including cellobiohydrolase,β-xylosidase, β-glucosidase, and N-acetylglucosaminidase which were measured in the case of colonized wood (paper II) and several others can be measured by enzyme assays using fluorogenic 4-methylumbelliferyl (MUB) or 4-amidomethylcoumaryl-labeled enzyme substrates with sufficient sensitivity based on the fluorescence of reaction products (Baldrian, 2009; Vepsäläinen et al., 2001). With a methodology based on the immobilization of fluorescent substrates, a method was developed for enzyme measurement at a 2.3-mm resolution with a detection limit of 1.8 $nmol \cdot h^{-1} \cdot cm^{-2}$ (Paper III) which should be used for detailed studies of spatial heterogeneity in various substrates like the fungus-colonized wood in Paper II (Fig. 15).

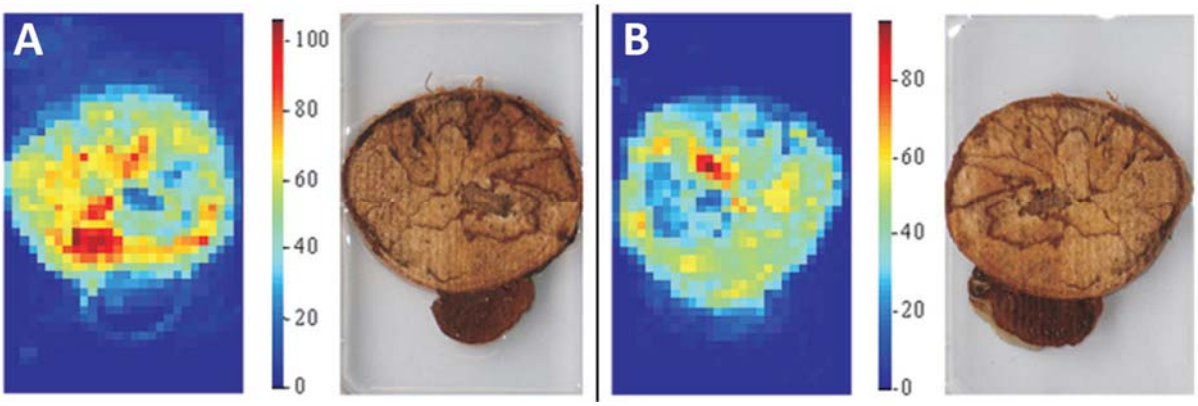


Figure 15. Distribution of hydrolytic enzymes over a cross-section of a *Betula pendula* branch colonized by wood-decomposing fungi, including *Fomes fomentarius* (fruit body). A) β-glucosidase B) cellobiohydrolase (Paper III).

It is well known that several Actinobacteria produce enzymes that decompose polysaccharides or phenolic compounds in dead plant biomass (McCarthy, 1987; Warren, 1996) and may thus fulfill the same ecological role as fungi. To get an insight into the actinobacterial lignocellulose decomposition, more than seventy isolated actinobacterial strains from meadow and forest soils were screened for the production of extracellular cellulolytic enzymes BG and CBH (Paper IV). Of these strains, 32% did not produce any of the tested enzymes, while 31% produced both of them. The active strains were represented mainly by members of the genus *Streptomyces*, *Amycolatopsis*, *Curtobacterium*, *Kribbella*, *Microbispora*, *Micromonospora* and *Nocardia* where genera *Amycolatopsis*, *Kribella*, *Micromonospora*, *Nocardia* and *Streptomyces* corresponded well with the presence of the corresponding genes in their genomes (Berlemont and Martiny, 2013). CBH often represents the rate-limiting enzyme in the decomposition of cellulose, the most abundant and rapidly decomposable polysaccharide in plant litter (Šnajdr et al., 2011; Štursová et al., 2012) thus the strains with high production of CBH were selected for further experiments. Also, from those most active strains, three *Streptomyces* spp. strains were selected to assess whether bacteria that utilize polysaccharides are also able to co-metabolize, transform or even mineralize the phenolic compounds within lignocellulose, and the transformation of ^{14}C-labelled phenolic compounds (dehydrogenation polymer (DHP), lignin and catechol) was studied. During 76 days of cultivation with ^{14}C-labelled phenolic substrates amount of radioactive CO_2 released was measured. Tested strains were able to mineralize up to 1.1% of

poplar lignin (Fig. 16) and to solubilize up to 4% of it. Catechol was mineralized up to 11.4% and solubilized up to 64%, while only minimal mineralization of DHP was observed.

The mineralization of lignin by *Streptomycetes* was first reported by Crawford (1978), who found that mineralization rates of ^{14}C-lignin labelled fir varied from 1.5 to 3% after nearly 42 days. Similar rates were also reported by Pasti et al. (1990). Occasionally, higher lignin mineralization rates were reported, such as 2.9% after 10 days by *Arthrobacter* sp. (Kerr et al., 1983) or up to 5% after 15 days by *Nocardia* sp. (Trojanowski et al., 1977). In comparison, the mineralization rates that were observed in our isolates seemed to be relatively low, but care must be taken in these comparisons because the modes of lignin labelling and preparation greatly affect mineralization rates. The mineralization rates of the recalcitrant lignin model compound ^{14}C-DHP that were reported in this study indicates that the ability of the studied Actinobacteria to catalyze the complete decomposition of lignin is low and that co-metabolic lignin degradation during the growth of active decomposers on lignocellulose is rather negligible.

Although solubilization is the first important step in lignin decomposition because it makes the polar lignin residues available to other microorganisms, only a few studies have reported the amount of solubilized lignin. Our results indicate only a limited solubilization of lignin, and these data are comparable to those published for several *Streptomycetes* (6–10%) (Pasti et al., 1990). Catechol is the first intermediate product of phenol degradation, and its cleavage is a critical step in the aerobic degradation of aromatic compounds in microorganisms (Krastanov et al., 2013). Several bacterial strains that are capable of catechol degradation belong to various phyla, such as Actinobacteria, *Alphaproteobacteria*, *Betaproteobacteria* or *Gammaproteobacteria*. Catechol degradation was investigated mainly in studies that focused on the treatment of phenolic wastes, and efficient degraders were found in the genera *Pseudomonas*, *Acinetobacter* and *Klebsiella* (El Azhari et al., 2010). All of the isolates in this study were able to efficiently degrade or solubilize catechol. Although the importance of phenolics as a growth substrate for bacteria is not clear, low molecular mass phenolic compounds are relatively common in litter (Osono and Takeda, 2005) and their transformation might be of importance.

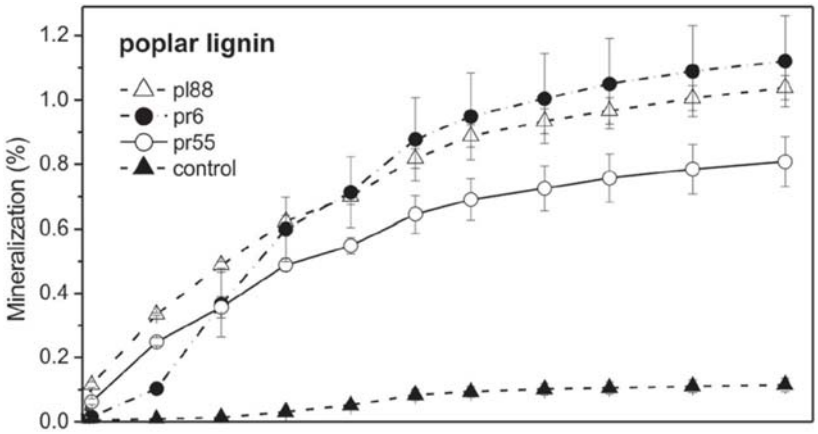


Figure 16. Mineralization of lignin by selected Actinobacteria. Time course of $^{14}CO_2$ production during the transformation of ^{14}C-poplar lignin in wheat straw microcosms by the selected *Streptomycetes*. Control treatments contained sterile straw. The data represent the means and standard errors (Paper IV).

The results showed that Actinobacteria could be significantly involved in lignocellulose degradation, although it is likely that the decomposition of biopolymers is limited to strains that represent only a minor portion of the entire community. Actinobacterial cometabolic degradation of lignin seems to be limited when compared to saprotrophic fungi, but they may significantly contribute to the solubilization of phenolics, especially low-molecular-mass compounds.

As shown in Paper II, in case of bacterial microbial community analysis (figure 14) microbial diversity can be very high. Investigation of microbial communities using comparative approaches as DGGE, T-RFLP or cloning and Sanger sequencing doesn't provide sufficient resolution (or enough information) to describe their diversity or phylogenetic classification. Majority of plant biomass decomposition is carried out in soil (in particular in litter layer) and the microbial diversity there could be many times higher than in fungus colonized wood (Buee et al., 2009; Roesch et al., 2007). To get insight into the fungal and bacterial diversity and community composition and to identify important groups potentially responsible for lignocellulose degradation in soils, amplicon pyrosequencing was the method of choice. Therefore, it was need to acquire appropriate bioinformatics tools and get the knowledge of necessary data processing steps. Bacterial community NGS data analyses were well established in contrast to fungal communities, therefore there was a need of a unified workflow for fungal-derived sequence data. Even though there were attempts to standardize the description and publication of NGS datasets of fungal communities (Nilsson et al., 2011), investigation of 40 publications concerning fungal community analyses from time period

between years 2009 and 2013 (Paper V) showed that many authors underestimated important steps in data treatment (e.g. denoising and chimera removal) (Fig. 17). This inconsistency of methodologies in the published studies makes use of the wealth of information derived by pyrosequencing for inter-study comparisons or meta-studies extremely difficult.

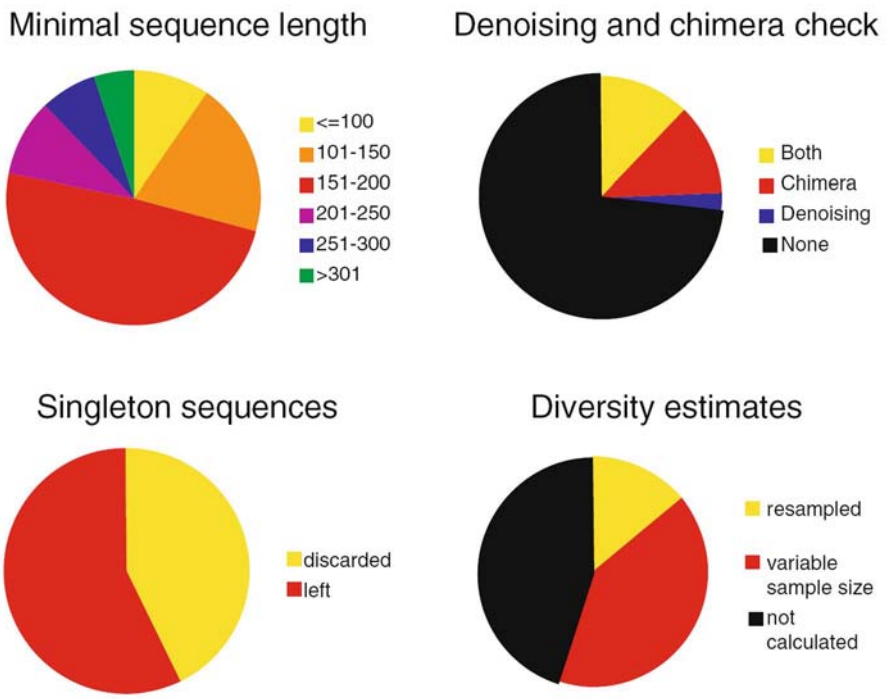


Figure 17. Overview of approaches used to analyse sequences derived by amplicon pyrosequencing of fungal rDNA based on 42 recently published studies (Paper V).

Paper V proposed a suitable workflow for fungal-amplicon pyrosequencing data and incorporated all important steps of this workflow into the developed software pipeline and sequence editor SEED. The developed software was also used to process bacteria-derived data in Papers VI and VII and data from amplicon pyrosequencing of functional genes *cbhI* and *rpb2* in Papers VI and IX. NGS data processing steps in the latest version of SEED are depicted in Figure 18.

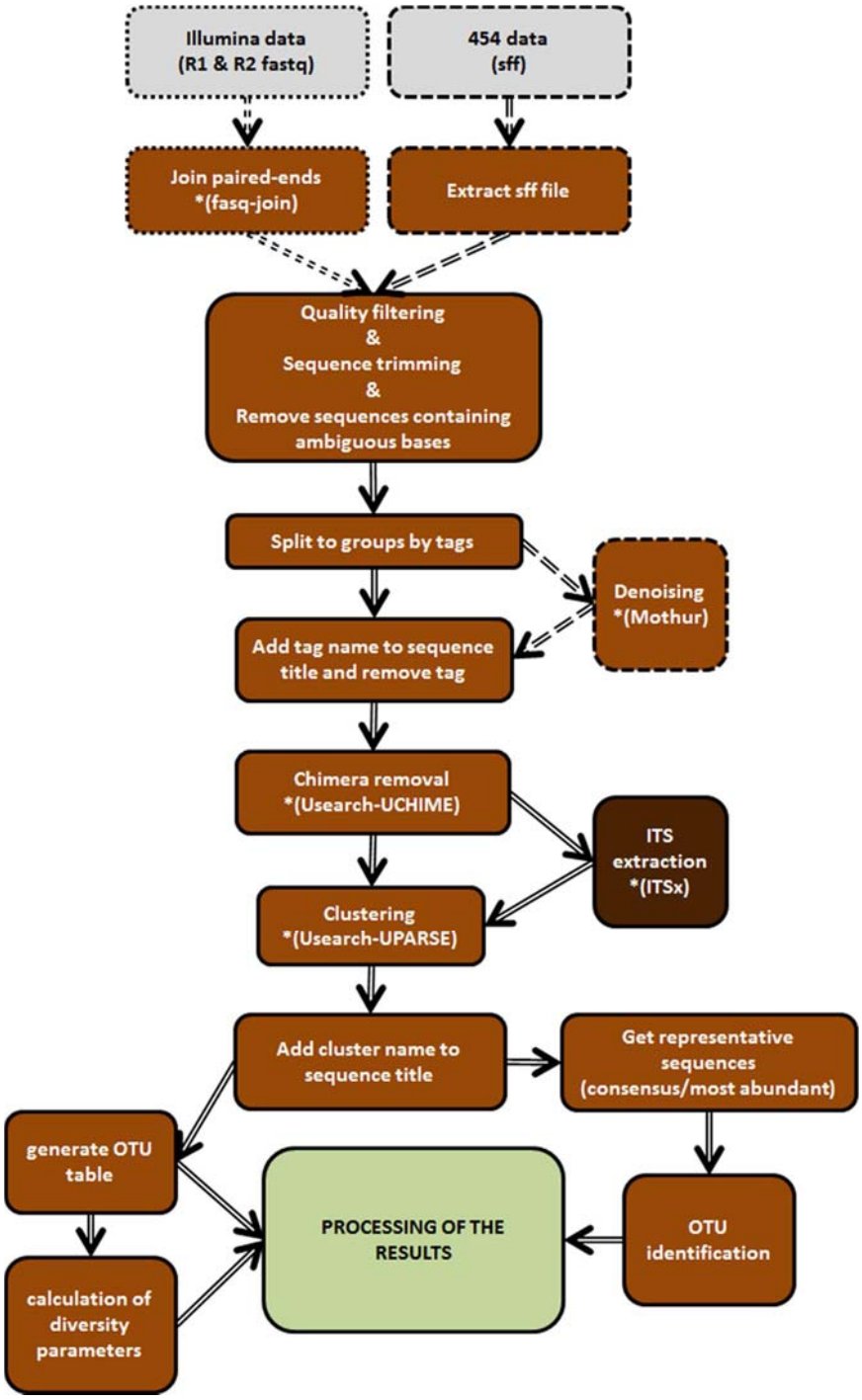


Figure 18. Optimal processing desingn for 454 and Illumina proposed for fungal ITS and bacterial 16S rDNA. All steps which can be done in SEED pipeline (version 1.2.3) are (light/dark) brown. Steps specific for 454 pyrosequencing are marked by dashed lines and for Illumina by dotted lines. Required external programs performing specific step are marked by asterisk.

While DNA-derived approaches to community analyses can reveal image of total community of microorganisms presented in sample and probably of some already extinct, although is assumed that degradation of DNA in the soil environment by DNases could be

quite fast (Blum et al., 1997). RNA-derived approaches may bring unique insights into the metabolically active part of the community due to rapid RNA turnover (Anderson and Parkin, 2007). Using the method of targeted amplicon library preparation for 454 pyrosequencing the structure of DNA as well as RNA-derived community of fungi and bacteria in spruce (*Picea abies*) forest soil was analyzed for the first time and to specifically target an important decomposition process, diversity estimation of a functional eukaryotic gene for cellobiohydrolase (exocellulase) *cbhI* was investigated (Paper VI). While information about all bacteria and fungi presented in the samples was obtained by DNA-derived 16S and ITS amplicons, RNA-derived sequences of bacteria represented the active part of the community that produce and contain ribosomes and fungal RNA-derived ITS region provide the information about species synthesizing ribosomes at given moment (Anderson, 2008). CBH represents the rate limiting enzyme in the decomposition of cellulose and the identification of the producers of the corresponding gene *cbhI* can reveal important fungal cellulose decomposers. Because the forest soil is highly stratified, litter and humic horizons (layers) were analyzed separately.

Analysis of fungal and bacterial communities revealed considerably higher diversity of bacteria than fungi and show that the litter and humic horizons differ significantly in both the total and relative amounts of bacterial and fungal biomass. While the diversity estimates for DNA- and RNA-derived communities of bacteria were similar and composition largely overlapped, the fungal communities showed that a more diverse community was present in RNA, 18% of fungal OTUs were found only in the RNA community, and 2% were found exclusively in the DNA community. Among *cbhI* sequences, representing cellulolytic members of the fungal community, 27% were found only in DNA and 15% only in RNA. Results shows that the DNA sequencing approaches miss a significant and functionally relevant part of microbial communities and the current knowledge largely based on this approach is incomplete.

Fungal community composition consisted mainly of Dikarya, Basidiomycota (53.5%) and Ascomycota (41.1%). Based on previous studies, large difference between litter and organic horizons community composition with higher abundance of saprotrophic fungi in litter and ectomycorrhizal species in deeper soil was expected (Edwards and Zak, 2010; Lindahl et al., 2007). Even though our results showed differences between the two horizons with 42% of abundant species exclusively recovered from either the L or the H horizon, the

ectomycorrhizal fungi were highly dominant in both horizons probably due to shallow rooting of *P. abies*. The most abundant genera of ectomycorrhizal fungi found in this study were *Piloderma* and *Tylospora* spp. also the most abundant in the boreal *P. abies* forests in Finland and Sweden (Korkama et al., 2006; Rosling et al., 2003; Wallander et al., 2010).

Despite several reports that bacterial abundance and diversity decrease with decreasing soil pH (Lauber et al., 2009; Rousk et al., 2010), a highly diverse bacterial community was found in our strongly acidic soil. *Proteobacteria*, *Acidobacteria* and Actinobacteria were the most dominant phyla comprising 80–90% of all DNA derived sequences in both horizons and this dominance was even stronger in the RNA pool, where especially Actinobacteria showed high presence in the active part of the community in H horizon and where they may contribute in recalcitrant phenolic compounds decomposition (Fig. 19).

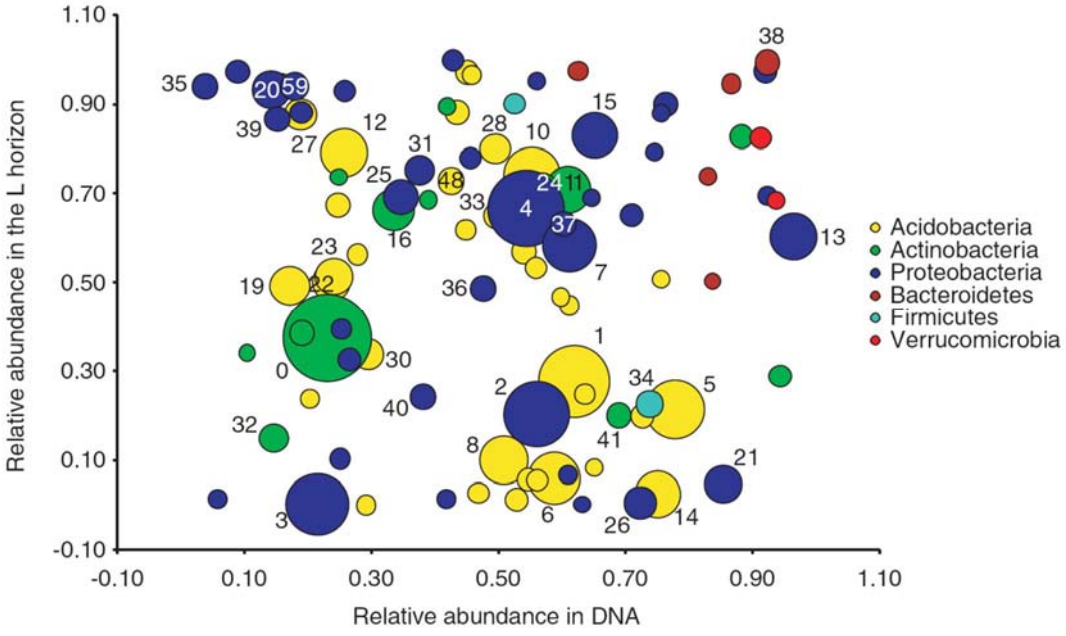


Figure 19. Distribution of major bacterial OTUs (represented by circles) from *Picea abies* forest topsoil between the L and H horizons and between DNA and RNA. The data represent mean values from four sampling sites. Symbol areas correspond to relative abundance in the combined set of DNA and RNA sequences from both horizons (Paper VI).

From the comparison of active and total communities in forest soils (Paper VI) it is apparent that Actinobacteria represent an important fraction of the active part of the microbial community in soil. Because their heavy metal resistance (Ivshina et al., 2013; Schmidt et al., 2005) they may potentially represent the major metabolically active group in

heavy metal contaminated soils where they could compete with more sensitive fungi (Harichova et al., 2012). In order to get deeper insight into actinobacterial community composition, a method for selective amplicon pyrosequencing of Actinobacteria was developed and tested on actinobacterial communities in grassland soils along the gradient of heavy metal contamination (Cd, Cu, Zn and Pb) (Paper VII). Using quantitative PCR, it was demonstrated that the amount of bacteria negatively correlated with heavy metal concentration, but Actinobacteria stayed unchanged which is in agreement with study of Berg et al. (2012) where the relative abundance of Actinobacteria-dominating soils across the Cu gradient remained stable with increasing bioavailable Cu and Gremion et al. (2003) have shown by a combined analysis of 16S rDNA and 16S rRNA that the actinomycete species dominate over all prokaryote diversity in heavy metal-contaminated soil, indicating that they might be a dominant part of the metabolically active bacteria in heavy metal-contaminated soils as prevoiusly suggested. The diversity of Actinobacteria in samples seemed to be unaffected by increasing heavy metal content, but heavy metals had a slight effect on community composition with Cu found to be the most significant factor (Fig. 20).

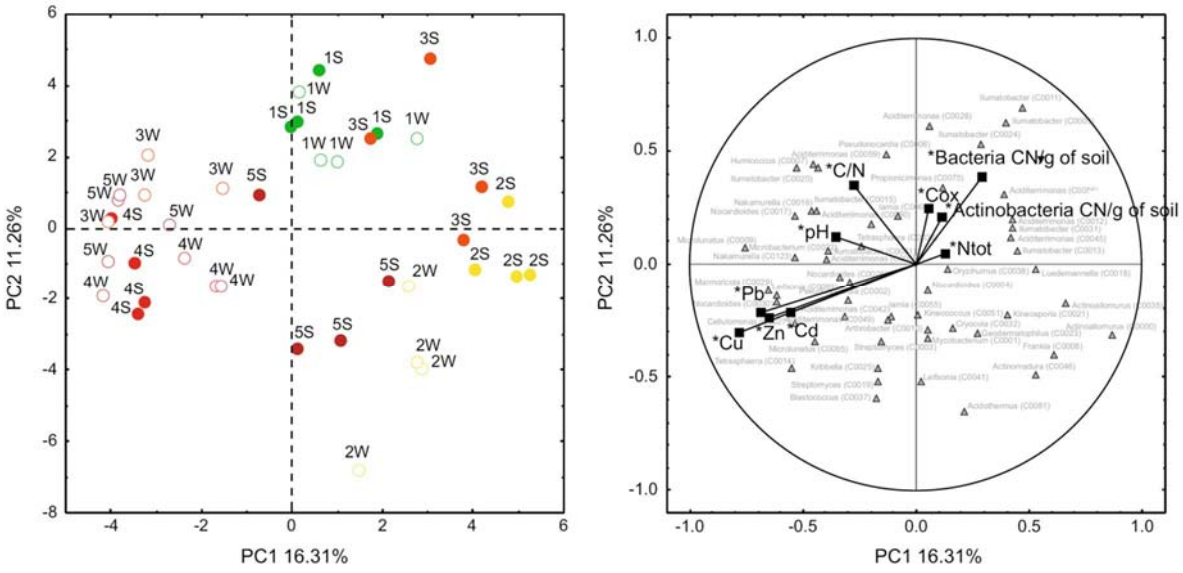


Figure 20. Principal component analysis of relative abundances of genomes of those actinobacterial OTUs representing at least 1% of total abundance within a sample for at least three samples. Heavy metal concentrations, environmental variables and relative bacterial and actinobacterial abundance are plotted as additional variables (Paper VII).

The abundance of *Acidothermus*, *Streptomyces*, *Pseudonocardia*, *Janibacter* and *Microlunatus* genera were positively correlated with Cu concentration. The Cu and Sb

resistance of members of the genera *Streptomyces* and *Janibacter* was previously described (Shi et al., 2013), and the heavy metal resistance of certain *Streptomyces* spp. is well known (Schmidt et al., 2005; So et al., 2001). There were also found several genera where can be suspected ability to degrade lignocellulose, especially mentioned *Streptomyces* (Paper IV), further genus *Acidothermus*, which also includes efficient degraders of cellulose, e.g. *Acidothermus cellulolyticus* (Blumer-Schuette et al., 2014), and *Microlunatus*, which are involved in the decomposition of plant biomass (Fan et al., 2014). In addition, the abundance of *Pseudonocardia* was high in our study area and positively correlated with the heavy metal content. The members of this genus are known as degraders of various organic compounds; *Pseudonocardia dioxanivorans* is able to use 1,4-dioxane as a sole carbon source and its growth is stimulated by the addition of Zn (Pornwongthong et al., 2014). In contrast, *Jatrophihabitans* and *Actinoallomurus* represented the genera with a significant negative response to heavy metal content. Members of both genera were described as endophytes of plants, which were isolated mainly from plant roots (Madhaiyan et al., 2013; Matsumoto et al., 2012), and thus the effects of metals on their occurrence might potentially be mediated by the effects on their plant hosts. This study demonstrated the potential of specific amplification combined with next-generation sequencing, here used to analyze the actinobacterial community composition in metal polluted soil. Actinobacteria seem to be less affected by heavy metals than other bacteria, and considering their metabolic potential, they can provide important soil functions, such as a contribution to organic matter decomposition.

In 454 pyrosequencing data sets, the abundance of reads belonging to a certain OTU is commonly interpreted as a measure of taxon abundance, useful for quantitative comparisons of community similarity. It was shown that incorrect data handling could alter read abundance and reduce the utility of quantitative metrics (Amend et al., 2010), but another important bias in relative abundances of taxa or their diversity which can be caused by the multi-copy nature of the target genes (as in case of widely used rRNA genes) was commonly overlooked in molecular studies. The most commonly used target in ecology studies of bacterial communities is 16S rRNA gene, because of its universal distribution and sufficient variability to determine phylogenetic relationships among taxa (Daubin et al., 2003; Head et al., 1998). Due to recent advances in bacterial genomics there are many genomes of bacteria species covering all common bacterial phyla allowing to obtain

sequences of 16S genes, estimate 16S rRNA genes copy numbers and investigate their variability within individual genomes as well as between closely related taxa. We investigated a total of 1,690 fully sequenced bacterial genomes belonging to 909 bacterial species and 454 genera where total of 7,081 16S rRNA genes were identified (Paper VIII). 16S rRNA gene copy numbers varied from 1 to 15 copies per genome with average of 4.2 and although some 15% of bacterial genomes contain only a single 16S rRNA copy, it seems that most bacterial phyla may contain bacteria with more than one copy and one half of the currently analyzed genomes harbor five or more copies (Fig. 21). Previous works showed that 16S rRNA copy numbers have shown both narrow and wide variation within the tested bacterial genera (Candela et al., 2004; Lee et al., 2008) and Pei et al. found 10% of genomes out of 568 analyzed bacterial species with >1% difference in 16S rRNA sequences (Pei et al., 2010).

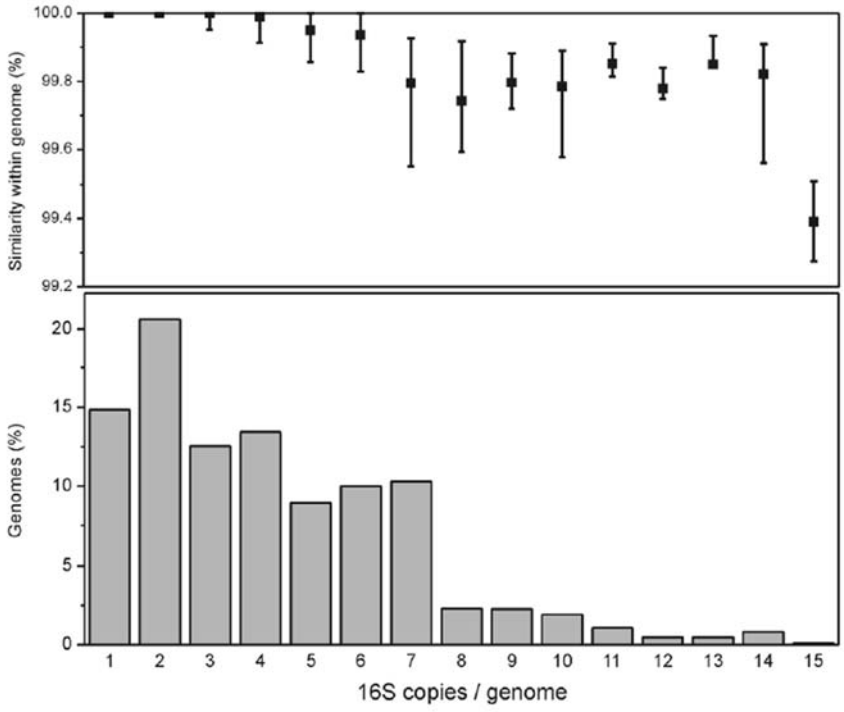


Figure 21. 16S rRNA within-genome similarity and copy numbers in bacterial genomes. Upper panel: the similarity of genomes with various copy numbers: the values indicated represent the first, the second and the third quartile. Lower panel: distribution of 16S rRNA copy numbers per genome in 1,690 sequenced bacterial genomes (Paper VIII).

Paper VIII shows that genome sizes appear to be more conserved than 16S rRNA copy numbers at all tested taxonomical levels and only a minor part of the bacterial genomes contain identical gene copies and so the genomes with more 16S rRNA copies tend to carry

more diverse variants of the gene. Although genomes with higher dissimilarity seem to be relatively rare (2.4% of genomes have 16S rRNA sequences with <99% mean similarity), highly dissimilar 16S rRNA sequences from the same genome were demonstrated to still carry a conserved secondary structure (Pei et al., 2010). The existence of highly dissimilar 16S rRNA sequences in certain genomes and the fact that 16S rRNA similarity within such genome can be lower than among certain bacterial genera may speak in favor of their evolution by horizontal gene transfer (Pei et al., 2010). Interspecific or intergeneric similarity can also significantly affect OTU construction: in certain genera, member species had exactly the same 16S rRNA sequence. If the standard level of 16S rRNA similarity of 97% is used, a few species, even genomes, can fall to different OTUs due to intragenomic or intraspecific differences. In contrast, species of several genera will not be separated at the same similarity level because in 41.7% of genera, the 16S rRNA differences were lower than 97%. This, however, depends on the part of the 16S rRNA gene used for OTU construction. On the example of forest soil bacteria described in Paper VI it could be demonstrated that the recently widely applied 16S rRNA-based abundance estimates provide an imperfect description of bacterial community composition. In general, abundance estimates based on the 16S rRNA sequence counts tend to underestimate the abundance of taxa with low 16S rRNA copy numbers such as the *Acidobacteria* and to overestimate taxa with high 16S rRNA copy numbers such as *Gammaproteobacteria* and *Firmicutes*. For example, the abundance of 16S rRNA sequences of *Acidobacteria* in certain organic-rich low-pH soils was found to be over 60% (Lauber et al., 2009). Based on our results, it is possible that their dominance in genome counts is even greater than this, and they may largely dominate in such environments. Based on this work, it was proposed that 16S rRNA copy numbers together with genome sizes of closest related taxon may be used for more precise estimates of the relative abundance of individual bacterial taxa in environmental samples. This approach should be a standard part of 16S-derived data processing and it was also applied in estimation of relative abundances of Actinobacteria in Paper VII.

Although there are limitations in the use of rRNA genes (16S rRNA) for bacteria community studies, particularly due to the mentioned multi-copy nature, the increasing amount of sequenced bacterial genomes could help to get sufficiently accurate predictions. In the case of fungi, the situation is less favorable. Instead of using fungal rRNA coding genes which don’t have sufficient discriminative power, internal transcribed spacer regions of

rDNA (ITS1 and ITS2) are well suited for their taxonomic identification (Schoch et al., 2012) and similarly to bacterial rRNA genes the information on the relative abundance of taxa and diversity is negatively affected by the multi-copy nature of rDNA and the possible existence of ITS paralogues. Compared to bacteria the rDNA clusters presented in fungal genomes can be significantly bigger ranging from 1 to 200 copies (Herrera et al., 2009; Raidl et al., 2005). This highly repetitive region is problematic to assemble and thus the number of copies cannot be derived from sequenced fungal genomes as in case of bacteria. Moreover, due to high sequential and length variability of ITS regions phylogenetic analyses of unrelated taxa are problematic and due to the variable length there is a strong PCR bias against species with longer amplicons which should largely affect the results of community studies (Ihrmark et al., 2012). To get an alternative view of fungal community composition we used the single-copy gene encoding the second largest subunit of RNA polymerase II (*rpb2*) for the analysis of fungal communities by amplicon pyrosequencing in spruce forest topsoil and we compared the obtained results with ITS1 derived data. The applicability of *rpb2* as an alternative marker was verified on comprehensive mock community composed of 130 species (Paper IX). Surprisingly, even though we used basidiomycete-specific *rpb2* primers (Matheny, 2005), this gene produced a more taxonomically diverse set of fungal sequences than the universal ITS primers and both markers produce proportionally similar number of basidiomycete reads (ITS: 51%, rpb2: 58% of all reads). The *rpb2* data set contained much more non-fungal sequences than the ITS data set (61% vs. 1% of reads), but on the other hand greatly exceeded the ITS in the diversity and abundance of sequences classified into various groups of the basal fungal lineages (41% of OTUs, 23% of reads). Basal fungal lineages are often zoosporic, difficult to cultivate, widespread in water ecosystems or soil, and their members live as saprobes, symbionts or parasites (Jones et al., 2011; Marano et al., 2012). Freeman et al. (2009) found that Chytridiomycota constituted over 70% of rDNA sequences in high-elevation soils without vegetation cover and their high abundance was also reported from periodically flooded alpine tundra soil under the snow cover. It is possible that zoosporic fungi were abundant in the studied ecosystem and yet were underestimated by the ITS marker due to the lack of complementarity of ITS primers or possibly due to the proportionally lower rDNA copy numbers per genome in these fungi.

The published fungal diversity estimates from environmental samples based on ITS analyses are notoriously high (Blackwell, 2011) often due to presence of chimeras, deeper

paralogues in multi-copy markers and pseudogenes. In our environmental samples and mock community, the numbers of singletons and chimeras were much higher in the ITS data set. The ITS data set may theoretically contain deeper intragenomic paralogues, which increased number of singletons and OTUs in the ITS data set. The fact that only *rpb2* sequences translatable into protein were used in our phylogenetic analysis partly eliminates the risk of chimeras or pseudogenes. Although there is the risk that unrecognized paralogues are present in the *rpb2* data set, such as, for example, in another protein-coding gene, *β*-tubulin (Hubka and Kolařík, 2012), genes paralogous to *rpb2* have not been reported thus far. The probability that the phylogenetically distant sequence clusters recovered using *rpb2* represent such paralogues is thus low, and it is possible that these sequences correspond to already known fungal groups, whose *rpb2* sequences are missing in public databases. The unassigned lineages may also represent novel ones that were not previously amplified using ITS primers and the advantages of degenerate primers in the discovery of new uncultured fungi should be further evaluated.

Similar levels of OTU inflation were observed for both markers in the soil community as in the defined mock community perhaps suggesting that the real diversity in the soil is closer to the estimate obtained using the *rpb2*. It should be also noted that the distribution of sequences among taxa of the mock community was slightly more even in the *rpb2* data set where 80% sequences were represented by 12 species compared to seven species in the ITS. Still, the reason for highly uneven distribution of read counts among species that was also observed in previous studies (Ihrmark et al., 2012), remains unclear: theoretically, it can be due to the combination of primer specificity, PCR preference, DNA quality or other factors. The present study shows that the single-copy, protein-encoding gene *rpb2* may be a viable option for fungal metabarcoding. Paper IX shows general agreement in the identity of the fungal genera and fungal classes recovered using *rpb2* marker and ITS, but no such agreement was found in the sequence abundance of the main genera or major fungal groups (Fig. 22). The single copy nature and the constant length of the *rpb2* amplicon represent an important advantage for proper estimations of diversity and relative abundance. The *rpb2* gene has taxonomic sensitivity superior to the ITS (Schoch et al., 2012), and our results reveal that this sequence is well suited for the study of basal fungal lineages. The use of a translatable protein coding gene also enables the identification of potential pseudogenes and the construction of robust phylogenetical trees. Although the precise taxonomic

placement of *rbp2* sequences is currently more difficult than it is for ITS due to the lower representation in GenBank, the phylogenetic discriminative power, better quantitative representation of the community composition and suitability for phylogenetic analyses may represent comparative advantages for *rbp2* over the use of ITS and make this molecular marker useful for studies in fungal ecology and diversity.

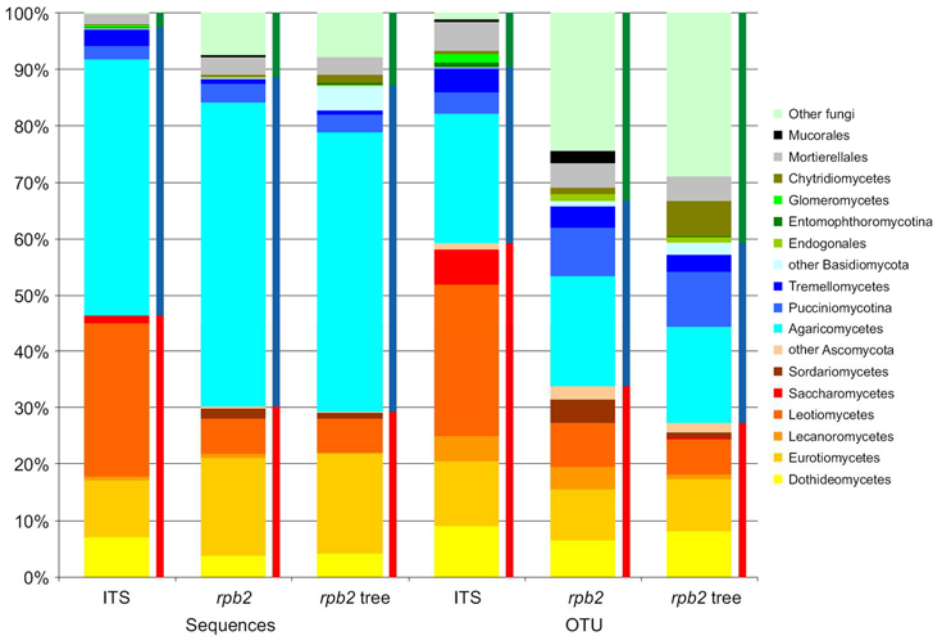


Figure 22. Abundance of sequences and OTUs at the level of fungal class identified from the ITS and *rpb2* data sets. Relative composition of fungal communities based on sequence and OTU counts. The thin bars use red colour to indicate the Ascomycota, blue for the Basidiomycota and green for other fungal lineages (Paper IX).

6. Conclusions

This work consists of 9 papers that contribute to the understanding of the role of fungi and bacteria play in dead plant biomass decomposition and to community composition analysis based on NGS. We analyzed both, the colonized wood as homogenous substrate where saprotrophic fungi are the dominating decomposers and soil environment with very high complexity and heterogeneity where hundreds of bacterial and fungal species live close to each other.

While enzymes involved in decomposition of saprotrophic white-rot basidiomycete *Fomes fometarius* were investigated and their spatial distribution was measured. Several soil Actinobacteria were screened for the extracellular enzyme production and their ability to degrade phenolic substrates was demonstrated.

To profit from advantages of NGS to describe the fungal and bacterial communities in soil bioinformatic software tool SEED was developed to facilitate NGS data processing. This tool was used in data processing and in-depth analyses of complex microbial communities from environmental samples. For the first time was done a comparison of fungal and bacterial communities in forest soil at a depth that allowed diversity estimation based on DNA and RNA analyses together with analysis of eukaryotic functional gene involved in decomposition. In addition, a detailed description of actinobacterial community composition along the gradient of heavy metal contamination was done showing their tolerance to heavy metal contamination and potential to contribute in decomposition processes.

Finally, several problematic issues of community analyses trough amplicon sequencing were investigated where major pitfalls are related with multi-copy nature of rRNA genes generally used for fungal and bacterial community composition analyses. Solutions by using information from known fully sequenced bacterial genomes and by using alternative single copy gene for fungal community investigation were proposed.

The papers included in this thesis reflect incremental significance of modern molecular methods especially NGS for microbial ecology and some of them strongly contribute to better knowledge of using those methods and better data handling. Soil ecosystems represent valuable sources of knowledge about important processes with global impact. Despite the many advances in their understanding they still remain highly unexplored. More effort is thus necessary to get better insight into soil ecosystem

functioning which could be accomplished through the progress in soil metagenomics and metatranscriptomic together with the new improvements in bioinformatics and computation technologies.

7. References

Acharya, S. and Chaudhary, A. (2012) Bioprospecting thermophiles for cellulase production: a review. Brazilian Journal of Microbiology 43, 844-856.

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

An, D.-S., Cui, C.-H., Lee, H.-G., Wang, L., Kim, S.C., Lee, S.-T., Jin, F., Yu, H., Chin, Y.-W., Lee, H.-K., Im, W.-T. and Kim, S.-G. (2010) Identification and characterization of a novel *Terrabacter ginsenosidimutans* sp nov beta-Glucosidase That Transforms Ginsenoside Rb1 into the Rare Gypenosides XVII and LXXV. Applied and Environmental Microbiology 76, 5827-5836.

Anderson, I., Abt, B., Lykidis, A., Klenk, H.-P., Kyrpides, N. and Ivanova, N. (2012) Genomics of aerobic cellulose utilization systems in actinobacteria. Plos One 7, e39331.

Anderson, I.C. and Parkin, P.I. (2007) Detection of active soil fungi by RT-PCR amplification of precursor rRNA molecules. Journal of microbiological methods 68, 248-253.

Ashelford, K.E., Chuzhanova, N.A., Fry, J.C., Jones, A.J. and Weightman, A.J. (2005) At least 1 in 20 16S rRNA sequence records currently held in public repositories is estimated to contain substantial anomalies. Applied and Environmental Microbiology 71, 7724-7736.

Baath, E. and Anderson, T.H. (2003) Comparison of soil fungal/bacterial ratios in a pH gradient using physiological and PLFA-based techniques. Soil Biology & Biochemistry 35, 955-963.

Bajkić, S., Narancić, T., Dokić, L., Dordević, D., Nikodinović-Runić, J., Morić, I. and Vasiljević, B. (2013) Microbial diversity and isolation of multiple metal-tolerant bacteria from surface and underground pits within the copper mining and smelting complex bor. Archives of Biological Sciences 65, 375-386.

Baldrian, P. (2009) Microbial enzyme-catalyzed processes in soils and their analysis. Plant Soil and Environment 55, 370-378.

Baldrian, P. and Šnajdr, J. (2011) Lignocellulose-degrading enzymes in soils, in: Shukla, G., Varma, A. (Eds.), Soil Enzymology 22, 167-186.

Baldrian, P. and Valášková, V. (2008) Degradation of cellulose by basidiomycetous fungi. FEMS Microbiology Reviews 32, 501-521.

Baldrian, P., Větrovský, T., Cajthaml, T., Dobiašová, P., Petránková, M., Šnajdr, J. and Eichlerová, I. (2013) Estimation of fungal biomass in forest litter and soil. Fungal Ecology 6, 1-11.

Baldrian, P., Voříšková, J., Dobiášová, P., Merhautová, V., Lisá, L. and Valášková, V. (2011) Production of extracellular enzymes and degradation of biopolymers by saprotrophic microfungi from the upper layers of forest soil. Plant and Soil 338, 111-125.

Berg, B., McClaugherty, C., Berg, B. and McClaugherty, C. (2003) Plant litter: Decomposition, humus formation, carbon sequestration. Springer, ISBN 978-3-642-38821-7.

Berg, J., Brandt, K.K., Al-Soud, W.A., Holm, P.E., Hansen, L.H., Sorensen, S.J. and Nybroe, O. (2012) Selection for Cu-tolerant bacterial communities with altered composition, but unaltered richness, via long-term Cu exposure. Applied and Environmental Microbiology 78, 7438-7446.

Berlemont, R. and Martiny, A.C. (2013) Phylogenetic Distribution of Potential Cellulases in Bacteria. Applied and Environmental Microbiology 79, 1545-1554.

Berlemont, R. and Martiny, A.C. (2015) Genomic potential for polysaccharide deconstruction in bacteria. Applied and Environmental Microbiology 81, 1513-1519.

Blackwell, M. (2011) The fungi: 1, 2, 3 ... 5.1 million species? American Journal of Botany 98, 426-438.

Blaxter, M., Mann, J., Chapman, T., Thomas, F., Whitton, C., Floyd, R. and Abebe, E. (2005) Defining operational taxonomic units using DNA barcode data. Philosophical Transactions of the Royal Society B: Biological Sciences 360, 1935-1943.

Blum, S.A.E., Lorenz, M.G. and Wackernagel, W. (1997) Mechanism of retarded DNA degradation and prokaryotic origin of DNases in nonsterile soils. Systematic and Applied Microbiology 20, 513-521.

Blumer-Schuette, S.E., Brown, S.D., Sander, K.B., Bayer, E.A., Kataeva, I., Zurawski, J.V., Conway, J.M., Adams, M.W.W. and Kelly, R.M. (2014) Thermophilic lignocellulose deconstruction. FEMS Microbiology Reviews 38, 393-448.

Bödeker, I.T.M., Clemmensen, K.E., de Boer, W., Martin, F., Olson, A. and Lindahl, B.D. (2014) Ectomycorrhizal *Cortinarius* species participate in enzymatic oxidation of humus in northern forest ecosystems. New Phytologist 203, 245-256.

Bodilis, J., Nsigue-Meilo, S., Besaury, L. and Quillet, L. (2012) Variable copy number, intra-genomic heterogeneities and lateral transfers of the 16S rRNA gene in *Pseudomonas*. PLoS ONE 7, e35647.

Brons, J.K. and van Elsas, J.D. (2008) Analysis of bacterial communities in soil by use of denaturing gradient gel electrophoresis and clone libraries, as influenced by different reverse primers. Applied and Environmental Microbiology 74, 2717-2727.

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

Cai, Y. and Sun, Y. (2011) ESPRIT-Tree: hierarchical clustering analysis of millions of 16S rRNA pyrosequences in quasilinear computational time. Nucleic Acids Research 39, e95.

Candela, M., Vitali, B., Matteuzzi, D. and Brigidi, P. (2004) Evaluation of the rrn operon copy number in *Bifidobacterium* using real-time PCR. Letters in Applied Microbiology 38, 229-232.

Caporaso, J.G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F.D., Costello, E.K., Fierer, N., Pena, A.G., Goodrich, J.K., Gordon, J.I. et al. (2010) QIIME allows analysis of high-throughput community sequencing data. Nature Methods 7, 335-336.

Caporaso, J.G., Lauber, C.L., Walters, W.A., Berg-Lyons, D., Huntley, J., Fierer, N., Owens, S.M., Betley, J., Fraser, L., Bauer, M. et al. (2012) Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. ISME Journal 6, 1621-1624.

Cartwright, N.J. and Holdom, K.S. (1973) Enzymic lignin, its release and utilization by bacteria. Microbios 8, 7-14.

Chater, K.F., Biro, S., Lee, K.J., Palmer, T. and Schrempf, H. (2010) The complex extracellular biology of Streptomyces. FEMS Microbiology Reviews 34, 171-198.

Claesson, M.J., Wang, Q., O'Sullivan, O., Greene-Diniz, R., Cole, J.R., Ross, R.P. and O'Toole, P.W. (2010) Comparison of two next-generation sequencing technologies for resolving highly complex microbiota composition using tandem variable 16S rRNA gene regions. Nucleic Acids Research 38, e200.

Cong, J., Yang, Y., Liu, X., Lu, H., Liu, X., Zhou, J., Li, D., Yin, H., Ding, J. and Zhang, Y. (2015) Analyses of soil microbial community compositions and functional genes reveal potential consequences of natural forest succession. Scientific Reports 5, 10007.

Crawford, D.L. (1978) Lignocellulose decomposition by selected streptomyces strains. Applied and Environmental Microbiology 35, 1041-1045.

Crawford, D.L., Barder, M.J., Pometto, A.L. and Crawford, R.L. (1982) Chemistry of softwood lignin degradation by *Streptomyces viridosporus*. Archives of Microbiology 131, 140-145.

Daubin, V., Moran, N.A. and Ochman, H. (2003) Phylogenetics and the cohesion of bacterial genomes. Science 301, 829-832.

de Boer, W., Folman, L.B., Summerbell, R.C. and Boddy, L. (2005) Living in a fungal world: impact of fungi on soil bacterial niche development. FEMS Microbiology Reviews 29, 795-811.

Duan, C.J. and Feng, J.X. (2010) Mining metagenomes for novel cellulase genes. Biotechnology Letters 32, 1765-1775.

Eastwood, D.C., Floudas, D., Binder, M., Majcherczyk, A., Schneider, P., Aerts, A., Asiegbu, F.O., Baker, S.E., Barry, K., Bendiksby, M. et al. (2011) The plant cell wall-decomposing machinery underlies the functional diversity of forest fungi. Science 333, 762-765.

Edgar, R.C. (2010) Search and clustering orders of magnitude faster than BLAST. Bioinformatics 26, 2460-2461.

Edgar, R.C. (2013) UPARSE: highly accurate OTU sequences from microbial amplicon reads. Nature Methods 10, 996-998.

Edgar, R.C., Haas, B.J., Clemente, J.C., Quince, C. and Knight, R. (2011) UCHIME improves sensitivity and speed of chimera detection. Bioinformatics 27, 2194-2200.

Edwards, I.P. and Zak, D.R. (2010) Phylogenetic similarity and structure of Agaricomycotina communities across a forested landscape. Molecular Ecology 19, 1469-1482.

Eid, J., Fehr, A., Gray, J., Luong, K., Lyle, J., Otto, G., Peluso, P., Rank, D., Baybayan, P., Bettman, B. et al. (2009) Real-time DNA sequencing from single polymerase molecules. Science 323, 133-138.

El Azhari, N., Devers-Lamrani, M., Chatagnier, G., Rouard, N. and Martin-Laurent, F. (2010) Molecular analysis of the catechol-degrading bacterial community in a coal wasteland heavily contaminated with PAHs. Journal of Hazardous Materials 177, 593-601.

Enkhbaatar, B., Temuujin, U., Lim, J.-H., Chi, W.-J., Chang, Y.-K. and Hong, S.-K. (2012) Identification and characterization of a xyloglucan-specific family 74 glycosyl hydrolase from *Streptomyces coelicolor* A3(2). Applied and Environmental Microbiology 78, 607-611.

Eriksson, K.-E.L., Blanchette, R. and Paul, A. (1990) Microbial and enzymatic degradation of wood and wood components. Springer, ISBN 978-3-642-46687-8.

Fan, F., Yin, C., Tang, Y., Li, Z., Song, A., Wakelin, S.A., Zou, J. and Liang, Y. (2014) Probing potential microbial coupling of carbon and nitrogen cycling during decomposition of maize residue by 13C-DNA-SIP. Soil Biology and Biochemistry 70, 12-21.

Fan, H.-X., Miao, L.-L., Liu, Y., Liu, H.-C. and Liu, Z.-P. (2011) Gene cloning and characterization of a cold-adapted beta-glucosidase belonging to glycosyl hydrolase family 1 from a psychrotolerant bacterium Micrococcus antarcticus. Enzyme and Microbial Technology 49, 94-99.

Fogel, G.B., Collins, C.R., Li, J. and Brunk, C.F. (1999) Prokaryotic genome size and SSU rDNA copy number: Estimation of microbial relative abundance from a mixed population. Microbial Ecology 38, 93-113.

Folman, L.B., Klein Gunnewiek, P.J., Boddy, L. and de Boer, W. (2008) Impact of white-rot fungi on numbers and community composition of bacteria colonizing beech wood from forest soil. FEMS Microbiology Ecology 63, 181-191.

Frankland, J.C. (1998) Fungal succession - unravelling the unpredictable. Mycological Research 102, 1-15.

Freeman, K.R., Martin, A.P., Karki, D., Lynch, R.C., Mitter, M.S., Meyer, A.F., Longcore, J.E., Simmons, D.R. and Schmidt, S.K. (2009) Evidence that chytrids dominate fungal communities in high-elevation soils. Proceedings of the National Academy of Sciences of the United States of America 106, 18315-18320.

Gremion, F., Chatzinotas, A. and Harms, H. (2003) Comparative 16S rDNA and 16S rRNA sequence analysis indicates that Actinobacteria might be a dominant part of the metabolically active bacteria in heavy metal-contaminated bulk and rhizosphere soil. Environmental Microbiology 5, 896-907.

Haas, B.J., Gevers, D., Earl, A.M., Feldgarden, M., Ward, D.V., Giannoukos, G., Ciulla, D., Tabbaa, D., Highlander, S.K., Sodergren, E. et al. (2011) Chimeric 16S rRNA sequence formation and detection in Sanger and 454-pyrosequenced PCR amplicons. Genome Research 21, 494-504.

Hagstrom, A., Pinhassi, J. and Zweifel, U.L. (2000) Biogeographical diversity among marine bacterioplankton. Aquatic Microbial Ecology 21, 231-244.

Harichová, J., Karelová, E., Pangallo, D. and Ferianc, P. (2012) Structure analysis of bacterial community and their heavy-metal resistance determinants in the heavy-metal-contaminated soil sample. Biologia 67, 1038-1048.

Hatakka, A. (2005) Biodegradation of lignin. Biopolymers Online, 3527600035.bpol1005.

Head, I.M., Saunders, J.R. and Pickup, R.W. (1998) Microbial evolution, diversity, and ecology: A decade of ribosomal RNA analysis of uncultivated microorganisms. Microbial Ecology 35, 1-21.

Herr, D., Baumer, F. and Dellweg, H. (1978) Purification and properties of an extracellular β-glucosidase from Lenzites trabea. European Journal of Applied Microbiology and Biotechnology 5, 29-36.

Herrera, M.L., Vallor, A.C., Gelfond, J.A., Patterson, T.F. and Wickes, B.L. (2009) Strain-Dependent Variation in 18S Ribosomal DNA Copy Numbers in *Aspergillus fumigatus*. Journal of Clinical Microbiology 47, 1325-1332.

Hofrichter, M. (2002) Review: lignin conversion by manganese peroxidase (MnP). Enzyme and Microbial Technology 30, 454-466.

Högberg, M.N., Briones, M.J.I., Keel, S.G., Metcalfe, D.B., Campbell, C., Midwood, A.J., Thornton, B., Hurry, V., Linder, S., Nasholm, T. and Hogberg, P. (2010) Quantification of effects of season and nitrogen supply on tree below-ground carbon transfer to ectomycorrhizal fungi and other soil organisms in a boreal pine forest. New Phytologist 187, 485-493.

Hubka, V. and Kolařík, M. (2012) Beta-tubulin paralogue *tubC* is frequently misidentified as the *benA* gene in *Aspergillus* section *Nigri* taxonomy: primer specificity testing and taxonomic consequences. Persoonia 29, 1-10.

Hugenholtz, P., Goebel, B.M. and Pace, N.R. (1998) Impact of culture-independent studies on the emerging phylogenetic view of bacterial diversity. Journal of Bacteriology 180, 4765-4774.

Ihrmark, K., Bodeker, I.T.M., Cruz-Martinez, K., Friberg, H., Kubartova, A., Schenck, J., Strid, Y., Stenlid, J., Brandstrom-Durling, M., Clemmensen, K.E. and Lindahl, B.D. (2012) New primers to amplify the fungal ITS2 region - evaluation by 454-sequencing of artificial and natural communities. FEMS Microbiology Ecology 82, 666-677.

Ivshina, I.B., Kuyukina, M.S. and Kostina, L.V. (2013) Adaptive mechanisms of nonspecific resistance to heavy metal ions in alkanotrophic actinobacteria. Russian Journal of Ecology 44, 123-130.

Jones, M.D., Forn, I., Gadelha, C., Egan, M.J., Bass, D., Massana, R. and Richards, T.A. (2011) Discovery of novel intermediate forms redefines the fungal tree of life. Nature 474, 200-203.

Kaiser, C., Koranda, M., Kitzler, B., Fuchslueger, L., Schnecker, J., Schweiger, P., Rasche, F., Zechmeister-Boltenstern, S., Sessitsch, A. and Richter, A. (2010) Belowground carbon allocation by trees drives seasonal patterns of extracellular enzyme activities by altering microbial community composition in a beech forest soil. New Phytologist 187, 843-858.

Kerr, T.J., Kerr, R.D. and Benner, R. (1983) Isolation of a bacterium capable of degrading peanut hull lignin. Applied and Environmental Microbiology 46, 1201-1206.

Khanna, S. and Gauri (1993) Regulation, purification, and properties of xylanase from cellulomonas-fimi. Enzyme and Microbial Technology 15, 990-995.

Kircher, M. and Kelso, J. (2010) High-throughput DNA sequencing - concepts and limitations. Bioessays 32, 524-536.

Klappenbach, J.A., Saxman, P.R., Cole, J.R. and Schmidt, T.M. (2001) rrndb: the Ribosomal RNA Operon Copy Number Database. Nucleic Acids Research 29, 181-184.

Köhler, A., Kuo, A., Nagy, L.G., Morin, E., Barry, K.W., Buscot, F., Canback, B., Choi, C., Cichocki, N., Clum, A. et al. (2015) Convergent losses of decay mechanisms and rapid turnover of symbiosis genes in mycorrhizal mutualists. Nature Genetics 47, 410-U176.

Koide, K., Osono, T. and Takeda, H. (2005a) Colonization and lignin decomposition of Camellia japonica leaf litter by endophytic fungi. Mycoscience 46, 280-286.

Koide, K., Osono, T. and Takeda, H. (2005b) Fungal succession and decomposition of Camellia japonica leaf litter. Ecological Research 20, 599-609.

Korkama, T., Pakkanen, A. and Pennanen, T. (2006) Ectomycorrhizal community structure varies among Norway spruce (*Picea abies*) clones. New Phytologist 171, 815-824.

Krastanov, A., Alexieva, Z. and Yemendzhiev, H. (2013) Microbial degradation of phenol and phenolic derivatives. Engineering in Life Sciences 13, 76-87.

Kubartová, A., Ranger, J., Berthelin, J. and Beguiristain, T. (2009) Diversity and decomposing ability of saprophytic fungi from temperate forest Litter. Microbial Ecology 58, 98-107.

Kuhad, R.C., Singh, A. and Eriksson, K.E. (1997) Microorganisms and enzymes involved in the degradation of plant fiber cell walls. Advances in Biochemical Engineering / Biotechnology 57, 45-125.

Kurtzman, C.P. and Robnett, C.J. (1998) Identification and phylogeny of ascomycetous yeasts from analysis of nuclear large subunit (26S) ribosomal DNA partial sequences. Antonie Van Leeuwenhoek International Journal of General and Molecular Microbiology 73, 331-371.

Lauber, C.L., Hamady, M., Knight, R. and Fierer, N. (2009) Pyrosequencing-based assessment of soil pH as a predictor of soil bacterial community structure at the continental scale. Applied and Environmental Microbiology 75, 5111-5120.

Lee, C.-Y., Chiu, Y.-C., Wang, L.-B., Kuo, Y.-L., Chuang, E.Y., Lai, L.-C. and Tsai, M.-H. (2013) Common applications of next-generation sequencing technologies in genomic research. Translational Cancer Research 2, 33-45.

Lee, C.M., Sieo, C.C., Abdullah, N. and Ho, Y.W. (2008) Estimation of 16S rRNA gene copy number in several probiotic *Lactobacillus* strains isolated from the gastrointestinal tract of chicken. FEMS Microbiology Letters 287, 136-141.

Lee, H.V., Hamid, S.B.A. and Zain, S.K. (2014) Conversion of lignocellulosic biomass to nanocellulose: Structure and chemical process. Scientific World Journal 2014, 631013.

Li, C., Hong, Y., Shao, Z., Lin, L., Huang, X., Liu, P., Wu, G., Meng, X. and Liu, Z. (2009) Novel alkali-stable, cellulase-free xylanase from deep-sea *Kocuria* sp Mn22. Journal of Microbiology and Biotechnology 19, 873-880.

Li, W. and Godzik, A. (2006) Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics 22, 1658-1659.

Lindahl, B.D. and Finlay, R.D. (2006) Activities of chitinolytic enzymes during primary and secondary colonization of wood by basidiomycetous fungi. The New phytologist 169, 389-397.

Lindahl, B.D., Ihrmark, K., Boberg, J., Trumbore, S.E., Hogberg, P., Stenlid, J. and Finlay, R.D. (2007) Spatial separation of litter decomposition and mycorrhizal nitrogen uptake in a boreal forest. New Phytologist 173, 611-620.

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

Lopez-Mondejar, R., Voříšková, J., Větrovský, T. and Baldrian, P. (2015) The bacterial community inhabiting temperate deciduous forests is vertically stratified and undergoes seasonal dynamics. Soil Biology & Biochemistry 87, 43-50.

Madhaiyan, M., Hu, C.J., Kim, S.J., Weon, H.Y., Kwon, S.W. and Ji, L. (2013) *Jatrophihabitans endophyticus* gen. nov., sp. nov., an endophytic actinobacterium isolated from a surface-sterilized stem of *Jatropha curcas* L. International Journal of Systematic and Evolutionary Microbiology 63, 1241-1248.

Marano, A.V., Gleason, F.H., Barlocher, F., Pires-Zottarelli, C.L., Lilje, O., Schmidt, S.K., Rasconi, S., Kagami, M., Barrera, M.D., Sime-Ngando, T. et al. (2012) Quantitative methods for the analysis of zoosporic fungi. Journal of Microbiological Methods 89, 22-32.

Martinez, A.T., Speranza, M., Ruiz-Duenas, F.J., Ferreira, P., Camarero, S., Guillen, F., Martinez, M.J., Gutierrez, A. and del Rio, J.C. (2005) Biodegradation of lignocellulosics: Microbial chemical, and enzymatic aspects of the fungal attack of lignin. International Microbiology 8, 195-204.

Matheny, P.B. (2005) Improving phylogenetic inference of mushrooms with RPB1 and RPB2 nucleotide sequences (*Inocybe*; Agaricales). Molecular Phylogenetics and Evolution 35, 1-20.

Matsumoto, A., Fukuda, A., Inahashi, Y., Omura, S. and Takahashi, Y. (2012) *Actinoallomurus radicium* sp. nov., isolated from the roots of two plant species. International Journal of Systematic and Evolutionary Microbiology 62, 295-298.

McCarthy, A.J. (1987) Lignocellulose-degrading Actinomycetes. FEMS Microbiology Letters 46, 145-163.

Metzker, M.L. (2005) Emerging technologies in DNA sequencing. Genome Research 15, 1767-1776.

Nakano, H., Okamoto, K., Yatake, T., Kiso, T. and Kitahata, S. (1998) Purification and characterization of a novel beta-glucosidase from *Clavibacter michiganense* that hydrolyzes glucosyl ester linkage in steviol glycosides. Journal of Fermentation and Bioengineering 85, 162-168.

Nilsson, H., R., Tedersoo, L., Lindahl, B.D., Kjøller, R., Carlsen, T., Quince, C., Abarenkov, K., Pennanen, T., Stenlid, J., Bruns, T. et al. (2011) Towards standardization of the description and publication of next-generation sequencing datasets of fungal communities. New Phytologist 191, 314-318.

Nunoura, N., Ohdan, K., Yano, T., Yamamoto, K. and Kumagai, H. (1996) Purification and characterization of beta-D-glucosidase (beta-D-fucosidase) from *Bifidobacterium breve* clb acclimated to cellobiose. Bioscience Biotechnology and Biochemistry 60, 188-193.

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

Oh, H.W., Heo, S.Y., Kim do, Y., Park, D.S., Bae, K.S. and Park, H.Y. (2008) Biochemical characterization and sequence analysis of a xylanase produced by an exo-symbiotic bacterium of *Gryllotalpa orientalis*, *Cellulosimicrobium* sp. HY-12. Antonie van Leeuwenhoek 93, 437-442.

Oinonen, P., Zhang, L., Lawoko, M. and Henriksson, G. (2015) On the formation of lignin polysaccharide networks in Norway spruce. Phytochemistry 111, 177-184.

Osono, T. (2007) Ecology of ligninolytic fungi associated with leaf litter decomposition. Ecological Research 22, 955-974.

Osono, T., Fukasawa, Y. and Takeda, H. (2003) Roles of diverse fungi in larch needle-litter decomposition. Mycologia 95, 820-826.

Osono, T. and Takeda, H. (2005) Decomposition of organic chemical components in relation to nitrogen dynamics in leaf litter of 14 tree species in a cool temperate forest. Ecological Research 20, 41-49.

Osullivan, A.C. (1997) Cellulose: The structure slowly unravels. Cellulose 4, 173-207.

Pasti, M.B., Pometto, A.L., 3rd, Nuti, M.P. and Crawford, D.L. (1990) Lignin-solubilizing ability of actinomycetes isolated from termite (Termitidae) gut. Applied and Environmental Microbiology 56, 2213-2218.

Pei, A.Y., Oberdorf, W.E., Nossa, C.W., Agarwal, A., Chokshi, P., Gerz, E.A., Jin, Z., Lee, P., Yang, L., Poles, M. et al. (2010) Diversity of 16S rRNA genes within individual prokaryotic genomes. Applied and Environmental Microbiology 76, 3886-3897.

Petrosyan, P., Luz-Madrigal, A., Huitron, C. and Flores, M.E. (2002) Characterization of a xylanolytic complex from *Streptomyces* sp. Biotechnology Letters 24, 1473-1476.

Pornwongthong, P., Mulchandani, A., Gedalanga, P.B. and Mahendra, S. (2014) Transition metals and organic ligands influence biodegradation of 1,4-dioxane. Applied biochemistry and biotechnology 173, 291-306.

Prosser, J.I. (2002) Molecular and functional diversity in soil micro-organisms. Plant and Soil 244, 9-17.

Quail, M., Smith, M., Coupland, P., Otto, T., Harris, S., Connor, T., Bertoni, A., Swerdlow, H. and Gu, Y. (2012) A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. BMC Genomics 13, 341.

Quan, L.-H., Min, J.-W., Jin, Y., Wang, C., Kim, Y.-J. and Yang, D.-C. (2012) Enzymatic biotransformation of ginsenoside Rb1 to compound K by recombinant beta-glucosidase from *Microbacterium esteraromaticum*. Journal of Agricultural and Food Chemistry 60, 3776-3781.

Quince, C., Lanzen, A., Curtis, T.P., Davenport, R.J., Hall, N., Head, I.M., Read, L.F. and Sloan, W.T. (2009) Accurate determination of microbial diversity from 454 pyrosequencing data. Nature Methods 6, 639-U627.

Quince, C., Lanzen, A., Davenport, R.J. and Turnbaugh, P.J. (2011) Removing noise from pyrosequenced amplicons. BMC Bioinformatics 12.

Raidl, S., Bonfigli, R. and Agerer, R. (2005) Calibration of quantitative real-time TaqMan PCR by correlation with hyphal biomass and ITS copies in mycelia of *Piloderma croceum*. Plant Biology 7, 713-717.

Ralph, J., Brunow, G. and Boerjan, W. (2007) Lignins. Encyclopedia of Life Sciences, 9780470015902.a0020104.

Rast, D.M., Baumgartner, D., Mayer, C. and Hollenstein, G.O. (2003) Cell wall-associated enzymes in fungi. Phytochemistry 64, 339-366.

Roesch, L.F.W., Fulthorpe, R.R., Riva, A., Casella, G., Hadwin, A.K.M., Kent, A.D., Daroub, S.H., Camargo, F.A.O., Farmerie, W.G. and Triplett, E.W. (2007) Pyrosequencing enumerates and contrasts soil microbial diversity. ISME Journal 1, 283-290.

Rosling, A., Landeweert, R., Lindahl, B.D., Larsson, K.H., Kuyper, T.W., Taylor, A.F.S. and Finlay, R.D. (2003) Vertical distribution of ectomycorrhizal fungal taxa in a podzol soil profile. New Phytologist 159, 775-783.

Rothberg, J.M., Hinz, W., Rearick, T.M., Schultz, J., Mileski, W., Davey, M., Leamon, J.H., Johnson, K., Milgrew, M.J., Edwards, M. et al. (2011) An integrated semiconductor device enabling non-optical genome sequencing. Nature 475, 348-352.

Rousk, J., Bååth, E., Brookes, P.C., Lauber, C.L., Lozupone, C., Caporaso, J.G., Knight, R. and Fierer, N. (2010) Soil bacterial and fungal communities across a pH gradient in an arable soil. ISME Journal 4, 1340-1351.

Scorzetti, G., Fell, J.W., Fonseca, A. and Statzell-Tallman, A. (2002) Systematics of basidiomycetous yeasts: a comparison of large subunit D1/D2 and internal transcribed spacer rDNA regions. FEMS Yeast Research 2, 495-517.

Shi, Z., Cao, Z., Qin, D., Zhu, W., Wang, Q., Li, M. and Wang, G. (2013) Correlation models between environmental factors and bacterial resistance to antimony and copper. PLoS One 8, e78533.

Schloss, P.D., Westcott, S.L., Ryabin, T., Hall, J.R., Hartmann, M., Hollister, E.B., Lesniewski, R.A., Oakley, B.B., Parks, D.H., Robinson, C.J. et al. (2009) Introducing mothur: Open-source, platform-independent, community-supported software for describing and comparing microbial communities. Applied and Environmental Microbiology 75, 7537-7541.

Schmidt, A., Haferburg, G., Sineriz, M., Merten, D., Buchel, G. and Kothe, E. (2005) Heavy metal resistance mechanisms in actinobacteria for survival in AMD contaminated soils. Chemie Der Erde - Geochemistry 65, 131-144.

Schoch, C.L., Seifert, K.A., Huhndorf, S., Robert, V., Spouge, J.L., Levesque, C.A., Chen, W., Bolchacova, E., Voigt, K., Crous, P.W. et al. (2012) Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for Fungi. Proceedings of the National Academy of Sciences of the United States of America 109, 6241-6246.

So, N.W., Rho, J.Y., Lee, S.Y., Hancock, I.C. and Kim, J.H. (2001) A lead-absorbing protein with superoxide dismutase activity from *Streptomyces subrutilus*. FEMS Microbiology Letters 194, 93-98.

Song, J.-M. and Wei, D.-Z. (2010) Production and characterization of cellulases and xylanases of *Cellulosimicrobium cellulans* grown in pretreated and extracted bagasse and minimal nutrient medium M9. Biomass & Bioenergy 34, 1930-1934.

Sun, Y., Cheng, J.J., Himmel, M.E., Skory, C.D., Adney, W.S., Thomas, S.R., Tisserat, B., Nishimura, Y. and Yamamoto, Y.T. (2007) Expression and characterization of *Acidothermus cellulolyticus* E1 endoglucanase in transgenic duckweed *Lemna minor* 8627. Bioresource Technology 98, 2866-2872.

Sutherland, J.B., Blanchette, R.A., Crawford, D.L. and Pometto, A.L. (1979) Breakdown of douglas-fir phloem by a lignocellulose-degrading *Streptomyces*. Current Microbiology 2, 123-126.

Šnajdr, J., Cajthaml, T., Valášková, V., Merhautová, V., Petranková, M., Spetz, P., Leppanen, K. and Baldrian, P. (2011) Transformation of *Quercus petraea* litter: Successive changes in litter chemistry are reflected in differential enzyme activity and changes in the microbial community composition. FEMS Microbiology Ecology 75, 291-303.

Šnajdr, J., Valášková, V., Merhautová, V., Herinková, J., Cajthaml, T. and Baldrian, P. (2008) Spatial variability of enzyme activities and microbial biomass in the upper layers of *Quercus petraea* forest soil. Soil Biology & Biochemistry 40, 2068-2075.

Štursová, M., Zifčáková, L., Leigh, M.B., Burgess, R. and Baldrian, P. (2012) Cellulose utilization in forest litter and soil: Identification of bacterial and fungal decomposers. FEMS Microbiology Ecology 80, 735-746.

Teeri, T.T. (1997) Crystalline cellulose degradation: new insight into the function of cellobiohydrolases. Trends in Biotechnology 15, 160-167.

ten Have, R. and Teunissen, P.J.M. (2001) Oxidative mechanisms involved in lignin degradation by white-rot fungi. Chemical Reviews 101, 3397-3413.

Trojanowski, J., Haider, K. and Sundman, V. (1977) Decomposition of C14-labeled lignin and phenols by a *Nocardia* sp. Archives of Microbiology 114, 149-153.

Vainio, E.J. and Hantula, J. (2000) Direct analysis of wood-inhabiting fungi using denaturing gradient gel electrophoresis of amplified ribosomal DNA. Mycological Research 104, 927-936.

Valášková, V., de Boer, W., Gunnewiek, P.J.A.K., Pospíšek, M. and Baldrian, P. (2009) Phylogenetic composition and properties of bacteria coexisting with the fungus *Hypholoma fasciculare* in decaying wood. ISME Journal 3, 1218-1221.

Ventura, M., Canchaya, C., Tauch, A., Chandra, G., Fitzgerald, G.F., Chater, K.F. and van Sinderen, D. (2007) Genomics of Actinobacteria: Tracing the evolutionary history of an ancient phylura. Microbiology and Molecular Biology Reviews 71, 495-548.

Vepsäläinen, M., Kukkonen, S., Vestberg, M., Sirvio, H. and Niemi, R.M. (2001) Application of soil enzyme activity test kit in a field experiment. Soil Biology & Biochemistry 33, 1665-1672.

Voříšková, J. and Baldrian, P. (2013) Fungal community on decomposing leaf litter undergoes rapid successional changes. ISME Journal 7, 477-486.

Voříšková, J., Brabcová, V., Cajthaml, T. and Baldrian, P. (2014) Seasonal dynamics of fungal communities in a temperate oak forest soil. New Phytologist 201, 269-278.

Wallander, H., Johansson, U., Sterkenburg, E., Brandström Durling, M. and Lindahl, B.D. (2010) Production of ectomycorrhizal mycelium peaks during canopy closure in Norway spruce forests. New Phytologist 187, 1124-1134.

Warren, R.A.J. (1996) Microbial hydrolysis of polysaccharides. Annual Review of Microbiology 50, 183-212.

Yin, L.-J., Huang, P.-S. and Lin, H.-H. (2010) Isolation of cellulase-producing bacteria and characterization of the cellulase from the isolated bacterium *Cellulomonas* sp YJ5. Journal of Agricultural and Food Chemistry 58, 9833-9837.

Yoon, M.-H. and Choi, W.-Y. (2007) Characterization and action patterns of two beta-1,4-glucanases purified from *Cellulomonas uda* CS1-1. Journal of Microbiology and Biotechnology 17, 1291-1299.

Zhang, F., Hu, S.-N., Chen, J.-J., Lin, L.-B., Wei, Y.-L., Tang, S.-K., Xu, L.-H. and Li, W.-J. (2012) Purification and partial characterisation of a thermostable xylanase from salt-tolerant *Thermobifida halotolerans* YIM 90462(T). Process Biochemistry 47, 225-228.

Zhang, H.-B., Yang, M.-X. and Tu, R. (2008a) Unexpectedly high bacterial diversity in decaying wood of a conifer as revealed by a molecular method. International Biodeterioration & Biodegradation 62, 471-474.

Zhang, H.-B., Yang, M.-X., Tu, R., Gao, L. and Zhao, Z.-W. (2008b) Fungal communities in decaying sapwood and heartwood of a conifer *Keteleeria evelyniana*. Current Microbiology 56, 358-362.

Žifčáková, L. and Baldrian, P. (2012) Fungal polysaccharide monooxygenases: New players in the decomposition of cellulose. Fungal Ecology 5, 481-489.

Žifčáková, L., Dobiášová, P., Kolářová, Z., Koukol, O. and Baldrian, P. (2011) Enzyme activities of fungi associated with *Picea abies* needles. Fungal Ecology 4, 427-436.

I

Appl Biochem Biotechnol (2013) 169:100–109
DOI 10.1007/s12010-012-9952-9

Extracellular Enzymes of the White-Rot Fungus *Fomes fomentarius* and Purification of 1,4-β-Glucosidase

Tomáš Větrovský · Petr Baldrian · Jiří Gabriel

Received: 11 July 2012 / Accepted: 31 October 2012 /
Published online: 14 November 2012

Abstract Production of the lignocellulose-degrading enzymes endo-1,4-β-glucanase, 1,4-β-glucosidase, cellobiohydrolase, endo-1,4-β-xylanase, 1,4-β-xylosidase, Mn peroxidase, and laccase was characterized in a common wood-rotting fungus *Fomes fomentarius*, a species able to efficiently decompose dead wood, and compared to the production in eight other fungal species. The main aim of this study was to characterize the 1,4-β-glucosidase produced by *F. fomentarius* that was produced in high quantities in liquid stationary culture (25.9 $U ml^{-1}$), at least threefold compared to other saprotrophic basidiomycetes, such as *Rhodocollybia butyracea*, *Hypholoma fasciculare*, *Irpex lacteus*, *Fomitopsis pinicola*, *Pleurotus ostreatus*, *Piptoporus betulinus,* and *Gymnopus* sp. (between 0.7 and 7.9 $U ml^{-1}$). The 1,4-β-glucosidase enzyme was purified to electrophoretic homogeneity by both anion-exchange and size-exclusion chromatography. A single 1,4-β-glucosidase was found to have an apparent molecular mass of 58 kDa and a p*I* of 6.7. The enzyme exhibited high thermotolerance with an optimum temperature of 60 °C. Maximal activity was found in the pH range of 4.5–5.0, and K_M and V_{max} values were 62 μM and 15.8 μmol $min^{-1} l^{-1}$, respectively, when *p*-nitrophenylglucoside was used as a substrate. The enzyme was competitively inhibited by glucose with a K_i of 3.37 mM. The enzyme also acted on *p*-nitrophenylxyloside, *p*-nitrophenylcellobioside, *p*-nitrophenylgalactoside, and *p*-nitrophenylmannoside with optimal pH values of 6.0, 3.5, 5.0, and 4.0–6.0, respectively. The combination of relatively low molecular mass and low K_M value make the 1,4-β-glucosidase a promising enzyme for biotechnological applications.

Keywords Cellulose · 1,4-β-glucosidase · Glycosyl hydrolase · Saprotrophic basidiomycetes · Wood-rotting fungi · *Fomes fomentarius*

T. Větrovský · P. Baldrian · J. Gabriel (✉)
Laboratory of Environmental Microbiology, Institute of Microbiology of the ASCR, v.v.i., Vídeňská 1083, 14220 Praha 4, Czech Republic
e-mail: gabriel@biomed.cas.cz

Introduction

The white-rot basidiomycete *Fomes fomentarius* is widely distributed and has been found in northern and southern Africa, Asia, eastern North America, and Europe [31]. The species most typically grows on hardwoods and is probably a long-term heart rot pathogen that results in significant economical damage. In northern regions, this species is most commonly found on birch trees, while in the south, it is more commonly found on beech trees [30]. This species has been used extensively by humans for many centuries [29]: The black felt-like substance contained in the pouch of the 5,000-year-old Neolithic-man, Ötzi, [26] was determined to be made of loosely interwoven hyphae from the fruiting bodies of *F. fomentarius*. The fungus is also known in Chinese traditional medicine where it is used for the treatment of various diseases including oral ulcers, gastroenteric disorders, hepatocirrhosis, inflammation, and various cancers [1, 12]. Water or methanol extracts from the fungus have been tested for pharmacological activity [17, 18, 25]. The fungus produces both intracellular polysaccharides [13] and exopolysaccharides that have been demonstrated to have antitumor effects [12]. Other studies on the fungus have described the autofluorescence of *F. fomentarius* [38], biosorption of several dyes from aqueous solutions [20], element sorption and distribution in fruit bodies [16], and the effects of heavy metals on the growth of the fungus under laboratory conditions [3]. *F. fomentarius* is recognized as a primary colonizer of wood and had been shown to effectively exclude other fungi from the resource [6]. It produces high levels of hydrolytic and oxidative enzymes. Its proficient ability to transforms wood into fungal biomass is indicative of its high-efficiency degradative enzymatic system [36].

The typical cellulolytic systems of fungi are composed of multiple enzymes. The endoglucanases (EG) hydrolyse cellulose chains internally, cellobiohydrolases (CBH) hydrolyse cellobiose units externally either from the reducing or the nonreducing end of cellulose microfibrils, and 1,4-β-glucosidases (BG) hydrolyse the resulting cellobiose to glucose [4]. Currently, cellulases and related enzymes are used in the food, brewery, wine, animal feed, textile, laundry, pulp and paper industries, and for agriculture, biofuel production, as well as for research purposes. The demand for these enzymes is growing more rapidly than ever before, and developing a more economical supply for this demand has become the driving force for research on cellulases and other related enzymes [5].

While the production of catalase, superoxide dismutase, and ligninolytic enzymes by *F. fomentarius* has been studied and described previously [14, 15, 22], there is little knowledge concerning the production of its cellulytic enzymes except for the detection of these enzymes in the fungus-colonized wood [36]. To the best of our knowledge, only one paper, where the effects of enzyme extracts from white-rot fungi on the chemical composition and in vitro digestibility of wheat straw where studied, noted the production of endoglucanase and cellobiohydrolase by *F. fomentarius* in laboratory cultures [28]. In light of the growing demand for the production of biotechnologically relevant enzymes for cellulose decomposition for the transformation of plant biomass for such applications as biofuels production, the focus of the current study was to characterize the cellulolytic system of the common wood-rotting fungus, *F. fomentarius*, and to compare its extracellular enzyme production under laboratory conditions with other saprotrophic fungi. The highly-produced1,4-β-glucosidase was purified and characterized with respect to possible future biotechnological applications.

Materials and Methods

Organisms and Growth Conditions for Enzyme Production

The fungal strains *Rhodocollybia butyracea* (CCBAS 286), *Hypholoma fasciculare* (CCBAS 281), *Irpex lacteus* (CCBAS 279), *Fomitopsis pinicola* (CCBAS 585), *Pleurotus ostreatus* (CCBAS 476), *F. fomentarius* (CCBAS 534), *Piptoporus betulinus* (CCBAS 537), and *Gymnopus* sp. (CCBAS 287) were obtained from the Culture Collection of Basidiomycetes (the ASCR Institute of Microbiology, v.v.i., Prague, Czech Republic) and were routinely maintained on solid MEA media (20 gl^{-1} malt extract and 15 gl^{-1} agar). For enzyme production, fungi were grown at 25 °C for 35 days in 250 ml Erlenmeyer flasks with 40 ml of liquid CLN medium, including modifications from a previously described assay [33]. Each culture was performed in quadruplicate. Samples for enzyme activity assays were collected every 7 days. Activity was measured immediately after sampling. *F. fomentarius* cultures used for isolation of the 1,4-β-glucosidase were grown for 25 days in fifty 500-ml Erlenmeyer flasks with 100 ml of synthetic CLN medium is an N-limited medium with cellulose as the source of carbon and a high C/N ratio that is typical for wood (5 gl^{-1} cellulose, 0.2 gl^{-1} ammonium tartarate, 1 gl^{-1} KH_2PO_4, 0.2 gl^{-1} NaH_2PO_4, 0.5 gl^{-1} $MgSO_4 \cdot 7H_2O$, 50 mgl^{-1} $CaCl_2$, 50 mgl^{-1} $FeSO_4 \cdot 7H_2O$, 5 mgl^{-1} $ZnSO_4 \cdot 7H_2O$, 10 mgl^{-1} $CuSO_4 \cdot 5H_2O$, and 5 mgl^{-1} $MnSO_4 \cdot 4H_2O$), pH of the medium was adjusted to 6.0 with 5 M NaOH or HCl before autoclaving.

The effects of N addition on the production of extracellular enzymes by *F. fomentarius* was also studied as previous studies have shown that the activity of hydrolytic enzymes from some saprotrophic basidiomycetes can be up regulated by nitrogen addition [33, 37]. The N content in the CLN medium (0.015 gl^{-1}) was increased to 1.5, 4.5, and 15 gl^{-1} by the addition of NH_4NO_3 to the liquid medium (four replicates per treatment). Enzyme activity was observed at weekly intervals over 35 days of incubation.

Extracellular Enzyme Assays and Protein Determination

Laccase (EC 1.10.3.2) activity was measured by monitoring the oxidation of ABTS (2,2′-azinobis-3-ethylbenzothiazoline-6-sulfonic acid) in citrate-phosphate buffer (100 mM citrate, 200 mM phosphate, pH 5.0) at 420 nm [7].

Manganese peroxidase (MnP, EC 1.11.1.13) activity was measured using a succinate-lactate buffer (100 mM, pH 4.5) [23]. 3-methyl-2-benzothiazolinone hydrazone and 3,3-dimethylaminobenzoic acid were oxidatively coupled by the enzymes. The resulting purple indamine dye was detected using a spectrophotometer at 595 nm. The results were normalized by the enzyme activity of the samples without manganese for MnP or by the substitution of an equimolar amount of EDTA for MnS.

Endo-1,4-β-glucanase (EC 3.2.1.4) and endo-1,4-β-xylanase (EC 3.2.1.8) activities were measured with azo-dyed carbohydrate substrates (carboxymethyl cellulose and birchwood xylan, respectively) following the manufacturer's guidelines (Megazyme, Ireland). Individual sample reaction mixture contained 0.2 ml of 2 % dye substrate in 200 mM sodium acetate buffer (pH 5.0) and 0.2 ml of fungal sample. Samples were incubated at 40 °C for 60 min, and the reaction was stopped by the addition of 1 ml of ethanol followed by 10 s of vortexing and 10 min centrifugation (10,000×*g*) [35]. Released dye amounts were measured at 595 nm. Enzyme activity was calculated according to standard curves that correlated the amount of dye release with the release of reducing sugars.

Cellobiohydrolase (EC 3.2.1.91) activity was measured in microplates using *p*-nitrophenyl-β-D-cellobioside (*p*NPC). The reaction mixture contained 0.16 ml of 1.2 mM *p*NPC in 50 mM sodium acetate buffer (pH 5.0) and 0.04 ml of sample. These reaction mixtures were incubated at 40 °C for 90–120 min. The reaction was stopped by the addition of 0.1 ml of 0.5 M sodium carbonate, and absorbance was read at 400 nm. 1,4-β-Glucosidase (EC 3.2.1.21), 1,4-β-xylosidase (EC 3.2.1.37), and 1,4-β-*N*-acetylglucosaminidase (chitinase; EC 3.2.1.52) were assayed using *p*-nitrophenyl-β-D-glucoside, *p*-nitrophenyl-β-D-xyloside and *p*-nitrophenyl-*N*-acetyl-β-D-glucosaminide, respectively, utilizing the same method [35].

All spectrophotometric measurements were performed in a microplate reader (Infinite, Tecan) or an ultraviolet–visible spectrophotometer (Lambda 11, Perkin-Elmer). One unit of enzyme activity was defined as the amount of enzyme required to produce 1 μmol of the reaction product per min.

Purification of 1,4-β-Glucosidase from *Fomes fomentarius*

Enzyme purification was performed using ÄKTA purifier (Pharmacia LKB, Sweden; Table 1) until electrophoretic homogeneity was reached. The cell-free culture filtrate (2.8 l final volume) was concentrated to 25 ml by ultrafiltration (10-kDa membrane; Amicon, Millipore, USA) and desalted via a Sephadex G-25 PD-10 column (GE Healthcare, USA). Fractions that exhibited enzyme activity between subsequent steps were also desalted. The desalted retentate was subjected to ion-exchange chromatography on a pre-equilibrated (50 mM acetate buffer, pH 4.5) diethylaminoethyl (DEAE)-sepharose CL B6 column (Amersham Pharmacia Biotech, volume 10 ml). The proteins were eluted using a linear gradient of 0–1 M NaCl (50 mM Na-acetate buffer, pH 4.5, final volume 50 ml). Fractions exhibiting 1,4-β-glucosidase activity were collected and subjected to ion-exchange chromatography on a Mono Q HR 5/50 column using the same linear-gradient mobile phase as noted previously (25 ml). Final purification was performed on a Superdex 75 HR 10/30 column eluted with 50 mM acetate buffer and 0.15 M NaCl (pH 4.5). The elution profile contained a single peak that was electrophoretically homogeneous [sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE)].

Characterization of the Purified 1,4-β-Glucosidase from *Fomes fomentarius*

The 1,4-β-glucosidase from *F. fomentarius* was characterized using polyacrylamide gel electrophoresis (SDS-PAGE, 10 % gel) and analytical isoelectric focusing (IEF, 7.5 % gel; Multiphor II electrophoresis system, Pharmacia LKB, Sweden). The IEF gel was prepared using ampholines of p*I* 2.5–5.0 and p*I* 3.5–10.0 (Pharmacia LKB, Sweden), and the isoelectric point was estimated using a low p*I* protein calibration kit, p*I* 2.5–6.5 (Pharmacia

Table 1 Purification of *F. fomentarius* 1,4-β-glucosidase from the liquid CLN medium

Purification step	Total protein (μg)	Total activity (mU)	Specific activity[a] (mU/mg)	Purification (fold)
Culture filtrate	21,000	1,175	56.1	1.0
DEAE Sepharose, pH 4.5	9,020	871	96.6	1.7
Mono Q HR 5/5, pH 4.5	187	25.5	136	2.4
Superdex 75 HR 10/30, pH 6.0	52	14.7	282	5.0

[a] Enzyme activity per milligram protein; 1 U of enzyme activity was defined as the amount of enzyme required to produce 1 μmol of the reaction product per minute

LKB, Sweden) and a surface electrode (Orion, USA). Both SDS-PAGE and IEF gels were stained with the Silver Stain Plus kit (Bio-Rad, USA). For protein identity, IEF gels were also activity-stained with *p*NPG. Protein concentrations were determined using the Bradford method (Bio-Rad Protein Assay kit) with bovine serum albumin used for standards. The molecular mass of the enzyme was estimated by gel filtration (Superdex 200 HR 10/30 column, Pharmacia LKB) using gel filtration standards (Pharmacia LKB) and by SDS-PAGE using Sigma CK molecular mass markers.

The optimum pH for enzymatic hydrolysis was determined in 0.2 M citrate-phosphate buffer, pH 2.5–8.0, in microtitration plates. The effect of temperature on enzyme activity and stability was determined for the temperature range of 10–80 °C with the pH value for optimal enzyme activity. Substrate specificity was tested in 50 mM citrate-phosphate buffer, pH 5.0, for six substrates (Table 2) at 40 °C. The K_M was determined for *p*NPG, *p*NPM and *p*NPX in 50 mM citrate-phosphate buffer at the respective optimal pH values and temperatures. The inhibitory effect on *p*NPG cleavage was tested for glucose in 50 mM citrate-phosphate buffer to determine the K_i value. The kinetic parameters K_M and K_i were obtained by a nonlinear, least-square fitting procedure using the ordinary Michaelis–Menten equation (Lineweaver–Burk plots) and using a nonlinear regression with Microcal Origin Professional 7.0 software.

Results

Enzyme Production

Extracellular ligninolytic and cellulolytic enzymes production varied between the brown-rot, white-rot and litter-decomposing fungi (Table 3). The brown-rot fungi *P. betulinus* and *F. pinicola* produced polysaccharide hydrolases but not the oxidative ligninolytic enzymes laccase and Mn peroxidase. The highest activities of both laccase and Mn peroxidase were detected in the white-rot fungus *F. fomentarius*, and relatively high levels of these enzymes were produced in both *R. butyracea* and *H. fasciculare*. High production of the 1,4-β-xylosidase and endoxylanase enzymes was found in *Gymnopus* sp., and elevated activity of the endoglucanase enzyme was recorded in *F. pinicola*. In addition to the high production of lignin-degrading enzymes, *F. fomentarius* also exhibited high activities of both celobiohydrolase and 1,4-β-glucosidase.

Nitrogen addition to the CLN medium resulted in overall downregulation of extracellular enzymes production by *F. fomentarius* (Fig. 1). While the activities of 1,4-β-xylosidase and celobiohydrolase were increased during the first week of incubation in 1.5 g N/l conditions, they decreased every subsequent week. Nitrogen addition almost completely inhibited the

Table 2 Substrate specificity of *F. fomentarius* 1,4-βglucosidase

Substrate	Relative activity (%)
p-nitrophenyl-β-D-glucopyranoside	100
p-nitrophenyl-β-D-xylopyranoside	6
p-nitrophenyl-β-D-galactopyranoside	15
p-nitrophenyl-β-D-mannopyranoside	9
p-nitrophenyl-β-D-cellobioside	9
p-nitrophenyl-β-D-lactopyranoside	0

Table 3 Mean production of extracellular enzymes by saprotrophic basidiomycetes

Fungus	EG	EX	CBH	bG	bX	MnP	Lac
Gymnopus erythropus	3.1	46.5	0.90	4.68	2.25	5.9	4.2
Rhodocollybia butyracea	15.8	18.8	0.62	6.19	0.38	74.4	18.1
Hypholoma fasciculare	14.7	19.3	0.77	7.87	0.27	74.8	8.4
Irpex lacteus	25.8	8.3	0.94	2.85	0.39	47.8	1.4
Fomitopsis pinicola	29.9	14.7	0.26	5.02	0.53	0.0	0.0
Pleurotus ostreatus	3.2	1.8	0.34	3.88	0.21	6.6	16.7
Fomes fomentarius	20.7	22.0	10.03	25.89	1.62	69.8	206.8
Piptoporus betulinus	5.6	2.3	0.74	0.65	0.03	0.0	0.0

Enzyme activities ($mU ml^{-1}$) were measured after 7, 14, 21, 28, and 35 days of culture in liquid CLN medium; the data represent means of these five measurements

EG endoglucanase, *EX* endoxylanase, *CBH* cellobiohydrolase, *bG* β-glucosidase, *bX* β-xylosidase, *MnP* Mn peroxidase, *Lac* laccase)

production of 1,4-β-glucosidase, *N*-acetylglucosaminidase and laccase. For the production of extracellular enzymes by *F. fomentarius*, including 1,4-β-glucosidase, nitrogen-limited growth condition was optimal.

Purification and Characterization of 1,4-β-glucosidase from *Fomes fomentarius*

The 1,4-β-glucosidase from *F. fomentarius* was purified using a combination of ion-exchange and gel permeation chromatographies until electrophoretic homogeneity was achieved. The molecular mass of the enzyme was 58 kDa by SDS-PAGE (Fig. 2). IEF analysis indicated a single band at pH6.7. This band was verified as the 1,4-β-glucosidase by activity staining using *p*NPG as a substrate.

The enzyme exhibited a high level of activity against *p*NPG over a relatively broad pH range (4.0–6.0) with the highest activity detected between pH4.5 and 5.0. The determined optimal pH for *p*NPGAL as the substrate was similar at a pH5.0, and the optimum pH for the enzyme with *p*NPX as the substrate was 6.0. When *p*NPC and *p*NPM were used as substrates the optimum pH was more acidic (3.5 and 4.0, respectively; Table 4). The enzyme was active only in a narrow range of temperatures and the optimum temperature when *p*NPG was a substrate was determined to be 60 °C. At temperatures between 10 and 20 °C, the enzyme's activity was <6 % of the activity observed at 60 °C (Fig. 3).

The enzyme was observed to have activity against multiple substrates that contained different monosaccharides and disaccharides; however, the enzyme's ability to cleave galactose, xylose and mannose was only 6–15 % of the enzyme activity observed when *p*NPG was the substrate. The K_M for *p*NPG was determined to be at 62 μM. Activity of 1,4-β-glucosidase was competitively inhibited by the addition of glucose with a K_i value of 3.37 mM.

Discussion

In this study, eight fungal species belonging to three ecological groups (white-rot, brown-rot, and litter-associated saprotrophic fungi) were tested for extracellular enzyme production. As expected, white-rot species produced high levels of lignin-degrading enzymes, brown-rot

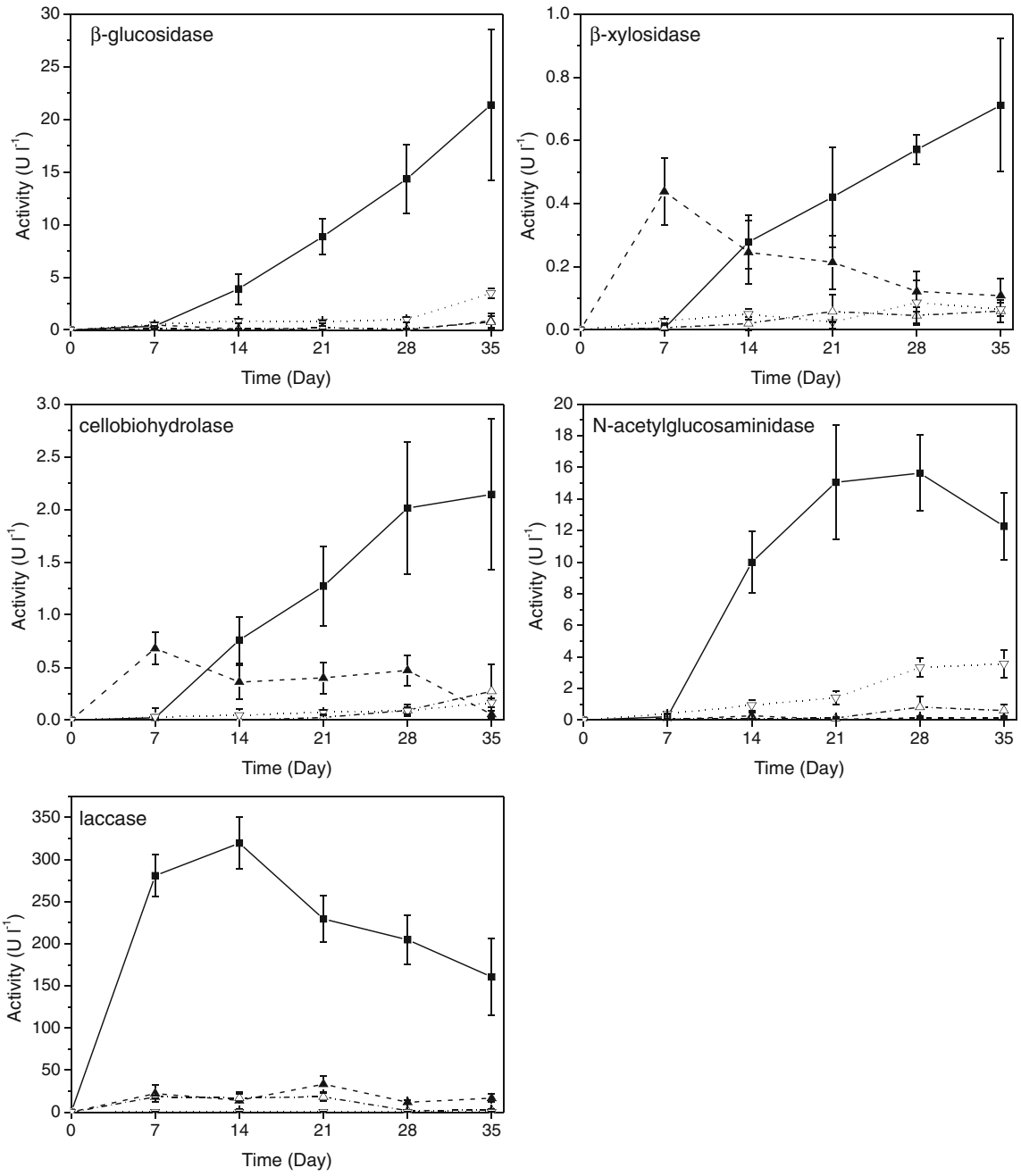


Fig. 1 Extracellular enzyme production by *F. fomentarius* in the liquid CLN medium without N addition (*squares*) and with the addition of 1.5 $gl^{-1}N$ (*full triangles*), 4.5 $gl^{-1}N$ (*open triangles*), and 15 $gl^{-1}N$ (*inverted triangles*) over time. These data represent the averages and standard errors of four replicates

fungi produced only cellulolytic and hemicellulolytic enzymes, and litter-colonizing saprotrophic fungi produced both types of enzymes (Table 3). The white-rot basidiomycete *F. fomentarius* exhibited higher production of laccase, celobiohydrolase, and 1,4-β-glucosidase, greater than any of the other tested fungi. Most importantly, *F. fomentarius* produced 1,4-β-glucosidase in amounts ten times higher than any other species tested. Thus, *F. fomentarius* was selected for the purification and characterization of 1,4-β-glucosidase. This enzyme has a broad range of uses in biotechnological applications [10].

While previous studies on enzyme production in white rote fungi, including *Ceriporiopsis aneirina*, *Ceriporiopsis resinascens* and *Dichomitus albidofuscus* [33] and *H. fasciculare* [37], have shown an increase in cellulolytic enzyme production with nitrogen addition, our study showed a decrease in enzyme activities in *F fomentarius* with nitrogen addition.

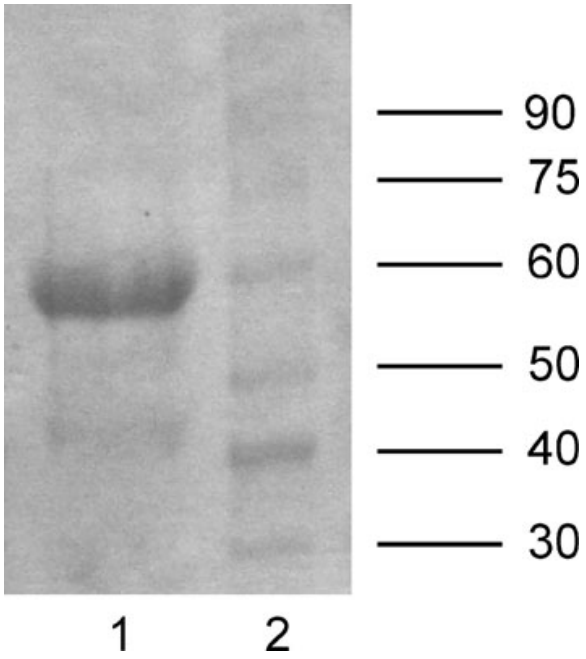


Fig. 2 SDS-PAGE of purified *Fomes fomentarius* 1,4-β-glucosidase. *Lane 1* purified enzyme, *lane 2* molecular mass markers. *Numbers* indicate molecular masses in kDa

Cellulases are currently the third largest group of industrial enzymes used worldwide according to market size, due to their use in cotton processing, in paper recycling, as detergent enzymes, in juice extraction, as animal feed additives, and as the enzymes used in bioethanol production [38]. Lignocellulosic biomass, a native substrate of fungal extracellular 1,4-β-glucosidases, is the most abundant and inexpensive source of bioethanol production. The capability of various filamentous fungi has been explored for the production of ethanol from this biomass, such as the genera *Aspergillus*, *Rhizopus*, *Mucor*, *Monilia*, *Neurospora*, *Fusarium*, *Trichoderma*, *Mucor*, and from the wood-decaying fungi *Trametes hirsuta* [24], *Phanerochaete chrysosporium* [2], and *Gloeophyllum trabeum* [27]. Cellulases are also increasingly being used in the textile industry. Their most successful application has been in the production of the stone-washed appearance of denim garments. Other processes that improve fabric appearance by removing fuzz fibers and pills or that have softening benefits have also been explored [8]. The increased use of cellulases is also a result of domestic fabric washing products where they are thought to aid detergency and to clean fiber surfaces, which improve appearance and color brightness [8]. Currently, these finishing and washing applications represent the largest market for cellulase enzymes worldwide.

Table 4 Effects of pH on the hydrolysis rates of various substrates by 1,4-β-glucosidase, purified from *F. fomentarius*

pH	*p*NPG	*p*NPX	*p*NPGal	*p*NPM	*p*NPC
2.5	0.0	3.0	0.9	1.4	0.2
3	9.0	1.8	0.9	1.5	0.3
3.5	36.0	5.2	7.4	5.1	9.0
4	70.0	6.0	12.2	9.0	1.2
4.5	100.0	5.3	7.4	8.7	1.3
5	100.0	4.2	15.0	7.1	2.0
5.5	76.0	3.7	8.4	6.8	0.0
6	47.0	2.7	6.0	4.0	0.0
7	9.0	1.6	1.5	1.2	0.0
8	8.0	1.4	0.3	1.8	0.0

The activity is expressed relative to the maximal activity of *p*NPG at optimal pH

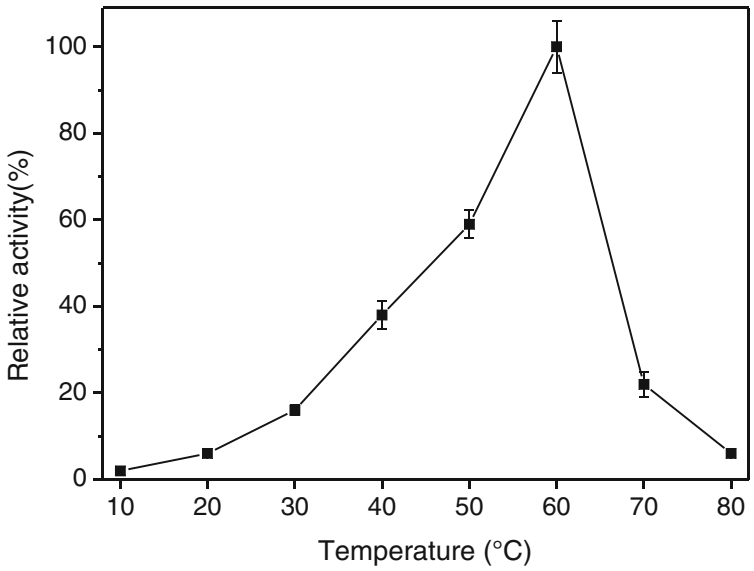


Fig. 3 Effect of temperature on the activity of *Fomes fomentarius* 1,4-β-glucosidase with *p*NPG serving as a substrate. These data represent the means and standard errors of four replicates

Commercial cellulases that are available for textiles are mainly produced from the fungi *Humicola insolens* (maximal activity at pH 7) and *Trichoderma reesei* (maximal activity at pH 5) [8]. The efficiency of the enzyme can be improved by its immobilization on various particles, but determining alternative enzymes that can be produced from various organisms would be desirable, considering the high levels of variability in the enzyme properties [4, 19].

The purified 1,4-β-glucosidase from *F. fomentarius* exhibited several interesting features that are relevant to its potential biotechnological application. While there are a wide range of molecular masses between fungal β-glucosidases that range from 35 in *P. ostreatus* to more than 400 in some *Phanerochaete chrysisporium* and *Trametes* spp. enzymes [6, 21, 32], the extracellular enzymes are usually smaller. The *F. fomentarius* enzyme is similar to the small extracellular enzymes produced by *Pleurotus* or *Phanerochaete* spp. [4]. The enzyme was observed to have a relatively high p*I* of 6.4, which is typical for an intracellular enzyme [4]. The optimal pH (6.0) and temperature (45–75 °C) fall within the range observed for other fungal β-glucosidases [4], but the high enzyme activity at 60 °C may be beneficial for biotechnological applications that require a high process rate. Similar to other β-glycosidases, the β-glucosidase of *F. fomentarius* has a relatively wide substrate range [9, 34], but the cleavage of 1,4-β-glycosidic bonds in cellobiose is the preferred reaction for the enzyme. More importantly, the enzyme has a low K_M of 62 μM. The only lower K_M value observed is that of the β-glucosidase of *Gloeophyllum trabeum*, which has limited commercial applications, due to its high molecular mass [11].

Compared to other fungal 1,4-β-glucosidases, the combination of a relatively low K_M with a high K_i for glucose indicates that the *F. fomentarius* enzyme is an efficient catalyst. Considering its low molecular mass and high activity at elevated temperatures, this enzyme is potentially suitable for biotechnological applications. This potential may be further explored by investigating the advantages of overexpression, heterologous expression or targeted mutagenesis of the enzyme. The ability of *F. fomentarius* to produce high titers of 1,4-β-glucosidases as well as laccase and Mn peroxidase suggests that this fungus may have potential in large-scale applications of lignocellulosic degradation (e.g., solid-state fermentation) .

Acknowledgments This work was supported by the Ministry of Education, Youth and Sports of the Czech Republic (OC08050, LA10001).

References

1. Abe, M., Takaoka, N., Idemoto, Y., Takagi, C., Imai, T., & Nakasaki, K. (2008). *International Journal of Food Microbiology, 124*, 199–203.
2. Bak, J. S., Ko, J. K., Choi, I. G., Park, Y. C., Seo, J. H., & Kim, K. H. (2009). *Biotechnology and Bioengineering, 104*, 471–482.
3. Baldrian, P., & Gabriel, J. (1997). *Folia Microbiologica, 42*, 521–523.
4. Baldrian, P., & Valášková, V. (2008). *FEMS Microbiology Reviews, 32*, 501–521.
5. Bhat, M. K. (2000). *Biotechnology Advances, 18*, 355–383.
6. Bhattacharjee, B., Roy, A., & Majumder, A. L. (1992). *Biochemistry International, 28*, 783–793.
7. Bourbonnais, R., & Paice, M. G. (1990). *FEBS Letters, 267*, 99–102.
8. Cavaco-Paulo, A. (1998). *Carbohydrate Polymers, 37*, 273–277.
9. Copa-Patino, J. L., & Broda, P. (1994). *Carbohydrate Research, 253*, 265–275.
10. Duan, C. J., & Feng, J. X. (2010). *Biotechnology Letters, 32*, 1765–1775.
11. Herr, D., Baumer, F., & Dellweg, H. (1978). *European Journal of Applied Microbiology and Biotechnology, 5*, 29–36.
12. Chen, W., Zhao, Z., Chen, S. F., & Li, Y. Q. (2008). *Bioresource Technology, 99*, 3187–3194.
13. Chen, W., Zhao, Z., & Li, Y. Q. (2011). *Carbohydrate Polymers, 85*, 369–375.
14. Jaszek, M., Zuchowski, J., Dajczak, E., Cimek, K., Graz, M., & Grzywnowicz, K. (2006). *International Biodeterioration & Biodegradation, 58*, 168–175.
15. Kapich, A. N., Prior, B. A., Lundell, T., & Hatakka, A. (2005). *Journal of Microbiological Methods, 61*, 261–271.
16. Kral, J., Voltr, J., Proska, J., Gabriel, J., Baldrian, P., Cerny, J., et al. (2005). *X-Ray Spectrometry, 34*, 341–344.
17. Lee, J. S. (2005). *Nutrition Research, 25*, 187–195.
18. Lee, J. S., Lim, M. O., Cho, K. Y., Cho, J. H., Chang, S. Y., & Nam, D. H. (2006). *Journal of Microbiology, 44*, 29–34.
19. Lynd, L. R., Weimer, P. J., van Zyl, W. H., & Pretorius, I. S. (2002). *Microbiology and Molecular Biology Reviews, 66*, 506.
20. Maurya, N. S., Mittal, A. K., Cornel, P., & Rother, E. (2006). *Bioresource Technology, 97*, 512–521.
21. Morais, H., Ramos, C., Matos, N., Forgacs, E., Cserhati, T., Almeida, V., et al. (2002). *Journal of Chromatography. B, Analytical Technologies in the Biomedical and Life Sciences, 770*, 111–119.
22. Neifar, M., Jaouani, A., Ellouze-Ghorbel, R., & Ellouze-Chaabouni, S. (2010). *Journal of Molecular Catalysis B: Enzymatic, 64*, 68–74.
23. Ngo, T. T., & Lenhoff, H. M. (1980). *Analytical Biochemistry, 105*, 389–397.
24. Okamoto, K., Sugita, Y., Nishikori, N., Nitta, Y., & Yanase, H. (2011). *Enzyme and Microbial Technology, 48*, 359–364.
25. Park, Y. M., Kim, I. T., Park, H. J., Choi, J. W., Park, K. Y., Lee, J. D., et al. (2004). *Biological & Pharmaceutical Bulletin, 27*, 1588–1593.
26. Peintner, U., Poder, R., & Pumpel, T. (1998). *Mycological Research, 102*, 1153–1162.
27. Rasmussen, M. L., Shrestha, P., Khanal, S. K., Pometto, A. L. R., & van Leeuwen, J. H. (2010). *Bioresource Technology, 101*, 3526–3533.
28. Rodrigues, M. A. M., Pinto, P., Bezerra, R. M. F., Dias, A. A., Guedes, C. V. M., Cardoso, V. M. G., et al. (2008). *Animal Feed Science and Technology, 141*, 326–338.
29. Roussel, B., Rapior, S., Charlot, C., Masson, C. L., & Bouti, P. (2002). *Revue d'Histoire de la Pharmacie, 50*, 599–614.
30. Schmidt, O. (2006) *Wood and tree fungi: Biology, damage, protection, and use.* Berlin: Springer
31. Schwarze, F. (1994). *Mycologist, 8*, 32–34.
32. Smith, M. H., & Gold, M. H. (1979). *Applied and Environmental Microbiology, 37*, 938–942.
33. Tomšovský, M., Popelářová, P., & Baldrian, P. (2009). *Folia Microbiologica, 54*, 74–80.
34. Valášková, V., & Baldrian, P. (2006). *Microbiology (UK), 152*, 613–3622.
35. Valášková, V., Šnajdr, J., Bittner, B., Cajthaml, T., Merhautová, V., Hoffichter, M., et al. (2007). *Soil Biology and Biochemistry, 39*, 2651–2660.
36. Větrovský, T., Voříšková, J., Šnajdr, J., Gabriel, J., & Baldrian, P. (2011). *Biodegradation, 22*, 709–718.
37. Voříšková, J., Dobiášová, P., Šnajdr, J., Vaněk, D., Cajthaml, T., Šantrůčková, H., et al. (2011). *Fungal Ecology, 4*, 417–426.
38. Žižka, Z., & Gabriel, J. (2006). *Folia Microbiologica, 51*, 109–113.

II

Biodegradation (2011) 22:709–718
DOI 10.1007/s10532-010-9390-8

ORIGINAL PAPER

Ecology of coarse wood decomposition by the saprotrophic fungus *Fomes fomentarius*

Tomáš Větrovský · Jana Voříšková · Jaroslav Šnajdr · Jiří Gabriel · Petr Baldrian

Received: 16 April 2010 / Accepted: 14 July 2010 / Published online: 29 July 2010

Abstract Saprotrophic wood-inhabiting basidiomycetes are the most important decomposers of lignin and cellulose in dead wood and as such they attracted considerable attention. The aims of this work were to quantify the activity and spatial distribution of extracellular enzymes in coarse wood colonised by the white-rot basidiomycete *Fomes fomentarius* and in adjacent fruitbodies of the fungus and to analyse the diversity of the fungal and bacterial community in a fungus-colonised wood and its potential effect on enzyme production by *F. fomentarius*. Fungus-colonised wood and fruitbodies were collected in low management intensity forests in the Czech Republic. There were significant differences in enzyme production by *F. fomentarius* between *Betula pendula* and *Fagus sylvatica* wood, the activity of cellulose and xylan-degrading enzymes was significantly higher in beech wood than in birch wood. Spatial analysis of a sample *B. pendula* log segment proved that *F. fomentarius* was the single fungal representative found in the log. There was a high level of spatial variability in the amount of fungal biomass detected, but no effects on enzyme activities were observed. Samples from the fruiting body showed high β-glucosidase and chitinase activities compared to wood samples. Significantly higher levels of xylanase and cellobiohydrolase were found in samples located near the fruitbody (proximal), and higher laccase and Mn-peroxidase activities were found in the distal ones. The microbial community in wood was dominated by the fungus (fungal to bacterial DNA ratio of 62-111). Bacterial abundance composition was lower in proximal than distal parts of wood by a factor of 24. These results show a significant level of spatial heterogeneity in coarse wood. One of the explanations may be the successive colonization of wood by the fungus: due to differential enzyme production, the rates of biodegradation of coarse wood are also spatially inhomogeneous.

Keywords Cellulose · Lignin · Microbial ecology · Saprotrophic basidiomycetes · Wood degradation

T. Větrovský · J. Voříšková · J. Šnajdr · J. Gabriel · P. Baldrian (✉)
Laboratory of Environmental Microbiology, Institute of Microbiology of the ASCR, v.v.i., Vídeňská 1083, 14220 Praha 4, Czech Republic
e-mail: baldrian@biomed.cas.cz

Introduction

Saprotrophic wood-inhabiting basidiomycetes are the most important decomposers of lignin and cellulose in dead wood and as such attracted considerable attention. These studies also included a thorough study of their wood decomposition (Ruel et al. 1994) and ultimately resulted in a detailed knowledge of the enzymology of their lignocellulose-decomposing extracellular enzymes that reflect either the brown-

rot mode of decay where wood polysaccharides are decomposed or the white-rot mode where lignin and polysaccharides are both utilized (Baldrian 2008; Baldrian and Valášková 2008; Arora and Sharma 2010; Lundell et al. 2010).

Although growth in wood and mycelial differentiation accompanied with the production of fungal fruitbodies are the most ecologically relevant modes of existence of wood-rotting fungi, relatively little attention has been paid to the production of enzymes under these conditions. This is striking since both the understanding of the decay process in fungus-colonised wood and the factors affecting fruitbody formation are essential for the understanding of the physiology of fungi causing biodeterioration of wood and may help both to develop preventive measures to reduce fungus-induced wood losses and to undertake measures for the conservation of endangered wood-associated fungi (Dahlberg et al. 2010). In previous papers, enzyme activity in wood was mostly studied under artificial laboratory conditions (e.g., Ruel et al. 1994; Lekounougou et al. 2008). Only recently, papers on enzyme activities in the environmental samples of wood colonised by brown-rot and white-rot fungi appeared, demonstrating high activities of several hydrolytic and oxidative enzymes in this environment (Valášková and Baldrian 2006; Valášková et al. 2009).

Fomes fomentarius is a common and economically important wood-rotting fungus in deciduous forests of Central Europe (Gabriel et al. 1997). It is a white-rot fungus causing heart rot of wood of several tree species including *Fagus sylvatica* and *Betula pendula* (Crockatt et al. 2008). The fungus often arrives at dead wood relatively early or colonizes already a living tree as a parasite but contrary to other early arriving fungi it exhibits better combative abilities and is not easily replaced (Niemela et al. 1995; Hiscox et al. 2010). The ecology of this fungus as early colonizer was confirmed by the study of Heilmann-Clausen and Boddy (2005) who found that wood pre-colonised by this species has a stimulatory effect on several secondary colonisers. As such, it can be presumed that large volumes of wood may be dominantly colonised by this fungus (although the cohabitation with other fungal species cannot be excluded) and enzyme activities measured in wood closely located to fungal fruitbodies are likely to reflect the activity of this single species. *F. fomentarius* is also ecologically relevant as a host of *Coleoptera* and other invertebrates and its antibiotic properties have found its use in medicine (Roussel et al. 2002; Teichert & Bondrup-Nielsen 2005).

The aim of this work was to quantify the activity and spatial distribution of extracellular enzymes in coarse wood colonised by the white-rot basidiomycete *F. fomentarius* and in adjacent fruitbodies of the fungus and to analyse the spatial diversity of enzymatic activities and the fungal and bacterial community in a fungus-colonised wood. We hypothesize that fungus-colonised wood which represents a nutritional base and fruitbodies as organs of reproduction would differ in the activity of extracellular enzymes, especially those acting on plant cell wall polymers. Since the fungus-colonised wood is likely not homogeneous in terms of the development of the fungal colony and the stage of fungal decay, we expected that the activity of extracellular enzymes would reflect this heterogeneity. Since the decay process changes the properties of wood significantly, we also analysed how this is reflected in the composition of microbial communities in wood.

Materials and methods

Sample collection

Fruitbodies of *Fomes fomentarius* and wood immediately adjacent to the fruitbodies were collected form beech (*Fagus sylvatica*) and birch (*Betula pendula*) trees in Jizerske hory Mts, Czech Republic, at 700–900 m a.s.l. in low management forests in a good health situation. Samples were used for the analysis of enzyme activities.

For the study of spatial variation of enzyme activity in wood colonised by *Fomes fomentarius*, a single birch log (8 cm diameter, collected from a standing tree) with a fruitbody was divided into 30 sections, approximately 16 cm^3 each (Fig. 1). Samples were homogenised and used for the analysis of enzyme activities, fungal biomass quantification and DNA extraction immediately after collection in August, 2008. Samples were described as "proximal" or "distal" based on their relative distance from the fruitbody (Fig. 1). For comparative molecular analyses, *Fomes fomentarius* CCBAS534 from the Culture Collection of Basidiomycetes of the Institute of

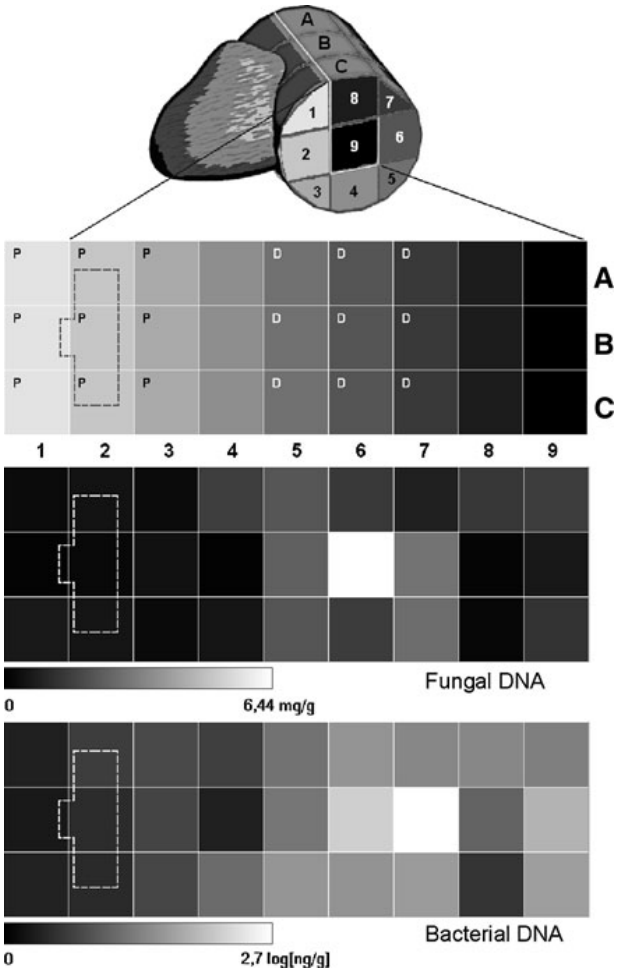


Fig. 1 Graphical representation of the sampling of birch log sample and the distribution of fungal and bacterial biomass in *Betula pendula* log segment colonised by *Fomes fomentarius*. *P* and *D* denote proximal and distal wood samples, fruitbody localisation on the map is indicated by a *dashed line*

Microbiology ASCR, Prague, was maintained on ME agar slants (20 g l^{-1} malt extract, 15 g l^{-1} agar).

Enzyme extraction and assays

Homogenized samples of colonized wood or fruiting bodies were extracted at 4°C for 2 h on an orbital shaker (100 rpm) with distilled water not to affect the enzyme stability during the extraction process, filtered through Whatman #5 filter paper and kept frozen at −18°C until enzyme activity analysis (Valášková and Baldrian 2006). Laccase (EC 1.10.3.2) activity was measured by monitoring the oxidation of ABTS (2,2′-azinobis-3-ethylbenzothiazoline-6-sulfonic acid) in citrate–phosphate (100 mM citrate, 200 mM phosphate) buffer (pH 5.0) at 420 nm (Bourbonnais and Paice 1990). Manganese peroxidase (MnP, EC 1.11.1.13) was assayed using succinate-lactate buffer (100 mM, pH 4.5). MBTH (3-methyl-2-benzothiazolinone hydrazone) and DMAB (3,3-dimethylaminobenzoic acid) were oxidatively coupled by the enzymes, and the resulting purple indamine dye was detected spectrophotometrically at 595 nm (Ngo and Lenhoff 1980). The results were corrected by the activities of the samples without manganese (for MnP)—the addition of manganese sulfate was substituted by an equimolar amount of ethylenediaminetetraacetate (EDTA).

Endo-1,4-β-glucanase (endocellulase; EC 3.2.1.4) and endo-1,4-β-xylanase (EC 3.2.1.8), were routinely measured with azo-dyed carbohydrate substrates (carboxymethyl cellulose and birchwood xylan, respectively) using the protocol of the supplier (Megazyme, Ireland). The reaction mixture contained 0.2 ml of 2% dyed substrate in 200 mM sodium acetate buffer (pH 5.0), and 0.2 ml sample. The reaction mixture was incubated at 40°C for 60 min and the reaction was stopped by adding 1 ml of ethanol followed by 10 s vortexing and 10 min centrifugation (10,000×g) (Valášková et al. 2007). The amount of released dye was measured at 595 nm and the enzyme activity was calculated according to standard curves correlating the dye release with the release of reducing sugars.

Cellobiohydrolase (exocellulase, EC 3.2.1.91) was assayed in microplates using *p*-nitrophenyl-β-D-cellobioside (PNPC).The reaction mixture contained 0.16 ml of 1.2 mM PNPC in 50 mM sodium acetate buffer (pH 5.0) and 0.04 ml sample. Reaction mixtures were incubated at 40°C for 90–120 min. The reaction was stopped by adding 0.1 ml of 0.5 M sodium carbonate, and absorbance was read at 400 nm. 1,4-β-glucosidase (EC 3.2.1.21), 1,4-β-xylosidase (EC 3.2.1.37) and 1,4-β-*N*-acetylglucosaminidase (EC 3.2.1.52) were assayed using *p*-nitrophenyl-β-D-glucoside, *p*-nitrophenyl-β-D-xyloside and *p*-nitrophenyl-*N*-acetyl-β-D-glucosaminide, respectively, using the same method (Valášková et al., 2007). All spectrophotometric measurements were done in a microplate reader (Infinite, Tecan) or a UV–VIS spectrophotometer (Lambda 11, Perkin-Elmer) and expressed per g dry mass of wood. One unit of enzyme activity was defined as the amount of enzyme forming 1 μmol of reaction product per min.

Quantification of fungal biomass

Total ergosterol was extracted and analyzed as described previously (Šnajdr et al. 2008). Samples (0.5 g) were sonicated (200 Hz) with 3 ml of 10% KOH in methanol at 70°C for 90 min. Distilled water (1 ml) was added and the samples were extracted three times with 2 ml of cyclohexane, evaporated under nitrogen, redissolved in methanol and analyzed isocratically using a Waters Alliance HPLC system (Waters, USA) with methanol as a mobile phase at a flow rate of 1 ml min^{-1}. Ergosterol was quantified by UV detection at 282 nm.

Statistics

Statistical tests were conducted using the software package Statistica 7 (StatSoft, USA). Differences between groups were tested by a one-way analysis of variance (ANOVA) followed by the Tukey post hoc test. Principal component analysis was used to identify the main sources of variability in the activities of extracellular enzymes in different parts of the birch log colonised by *Fomes fomentarius*. In all cases, differences at $P < 0.05$ were regarded as statistically significant.

Analysis of microbial communities

DNA from either birch log samples, adjacent fruitbody samples or the cultures of *Fomes fomentarius* CCBAS534 was isolated using modified Miller-SK method according to Sagova-Mareckova et al. (2008). Microbial community analysis was performed using denaturing gradient gel electrophoresis (DGGE) as described previously (Valášková and Baldrian 2009). Primers used for bacterial community analysis were 968-gc and 1378 targeting the V6-V8 hypervariable region of bacterial 16S rDNA (Table 1). For fungal community analysis, nested PCR was used for the amplification of the ITS1 region of fungal rDNA. A fragment comprising both ITS1 and ITS2 was amplified in the first PCR reaction using the primer pair ITS1f/ITS4. After the purification of the PCR product, the ITS1 region was specifically amplified in the second PCR reaction using the ITS1f-gc/ITS2 primers (Valášková and Baldrian 2009). The gel solution for bacterial community analysis consisted of 6% (w/v) acrylamide/bisacrylamide (37.5:1), in 0.5× TAE buffer, pH 8.3, containing 55 to 61% of the denaturant (100% denaturant corresponding to 7 M urea and 40% formamide). The gel solution for fungal community analysis consisted of 9% (w/v) of acrylamide/bisacrylamide (37.5:1) with a linear gradient of denaturant between 36 and 44%. The time of separation was around 17 h at 200 V in 0.5× TAE buffer (60°C). The gels were stained with ethidium bromide. In DGGE of fungal ITS amplicons, DNA isolated from *F. fomentarius* CCBAS534 was used for comparison. To confirm the identity of fungi in wood samples, dominant bands from the fungal community DGGE were cut off from lanes containing DNA extracted from wood and the lane of *F. fomentarius* CCBAS534. DNA was reamplified using the primers ITS1f and ITS2 (Table 1) in 50 µl reaction mixture. Each 50 µl reaction mixture contained 5 µl 10× buffer for DyNAzyme DNA Polymerase, 3 µl of bovine serum albumin (10 mg/ml), 2 µl of each

Table 1 PCR primers used in this work

Target	Primer	Sequence	Reference
Fungal ITS	ITS1	TCCGTAGGTGAACCTGCGG	White et al. (1990)
	ITS1f	CTTGGTCATTTAGAGGAAGTAA	Gardes and Bruns (1993)
	ITS1F-gc	CGCCCGCCGCGCGCGGCGGGCGGGGCGGG GGCACGGGGGGCTTGGTCATTTAGAGGAAGTAA	Valášková and Baldrian (2009)
	ITS2	GCTGCGTTCTTCATCGATGC	White et al. (1990)
	ITS2*	TTYGCTGYGTTCTTCATCG	Wakelin et al. (2007)
	ITS4	TCCTCCGCTTATTGATATGC	White et al. (1990)
Bacterial 16S rDNA	1108F	ATGGYTGTCGTCAGCTCGTG	Amann et al. (1995)
	1132R	GGGTTGCGCTCGTTGC	Wilmotte et al. (1993)
	1378	TCCTCCGCTTATTGATATGC	Heuer et al. (1997)

primer (0.01 mM), 1.6 µl of PCR Nucleotide Mix (10 mM each), 2 µl polymerase (2U μl^{-1}, DyNAZyme II DNA Polymerase) and 1 µl of the DNA template. PCR cycling conditions were 1× (95°C 3 min, 55°C 30 s, 72°C 1 min), 30× (95°C 30 s, 55°C 30 s, 72°C 1 min), 1× (95°C 30 s, 55°C 30 s, 72°C 10 min) (Valášková and Baldrian 2009). PCRs were run on TGradient thermocycler (Biometra, Germany), PCR products were sequenced as a single extension with primer ITS1f by Macrogen Inc. (Korea) using an ABI 3730 XL DNA Analyzer (Applied Biosystems). Sequences were manually edited and corrected prior to BLAST (blastn) search against the nucleotide database at NCBI (http://www.ncbi.nlm.nih.gov/blast).

Real-time PCR

Two sets of specific PCR primers were used to quantify the relative amounts of fungal and bacterial DNA, ITS1 and ITS2* for fungi (Table 1) and 1108F/1132R for bacteria, in real-time PCR assays based on the method of Leigh et al. (2007). Amplifications were performed on a StepOnePlus cycler (Applied Biosystems) using optical grade 96-well plates. Each 20 µl reaction mixture contained 10 µl SYBR Green Master Mix (Applied Biosystems), 0.9 µl BSA (10 mg/ml), 1.35 µl of each primer, 1.5 µl of template and 6.1 µl of water. The PCR cycling protocol was the same for fungal and bacterial DNA quantification: 56°C for 2 min; 95°C for 10 min; 95°C for 15 s and 60°C for 1 min (40 cycles). *Streptomyces lincolnensis* DNS 40335 and *Hypholoma fasciculare* CCBAS281 genomic DNA were used as standards, DNA content was expressed in nanogram of DNA per gram of wood dry mass. The ratio of fungal to bacterial DNA was calculated by dividing the content of fungal and bacterial DNA in individual samples.

Results

The comparison of extracellular enzyme distribution in wood and fruiting bodies showed that higher activity of the ligninolytic enzyme Mn-peroxidase was present in wood, while the concentrations in fruitbodies were low (Table 2). In *F. sylvatica*, also laccase activity was significantly higher in wood than in the fruitbody. *F. sylvatica* wood colonised by *F. fomentarius* exhibited significantly higher activities of Mn-peroxidase, endocellulase, cellobiohydrolase and especially of the xylanolytic enzymes endoxylanase and β-xylosidase where the differences between the two tree species were four to sevenfold (Table 2).

For the analysis of spatial variability of enzyme activity and microbial biomass content, *B. pendula* log in a medium stage of fungal decay was selected and a disc closely adjacent to a fruitbody of *F. fomentarius* was cut into approximately 2.5 × 2.5 cm samples (Fig. 1). While the fungal mycelium in the wood samples closely located to the fruitbody was apparent by the presence of fungal hyphal cords and mycelial sheaths, in the more distant samples, fungal biomass was less visually apparent. Significantly higher activities of both ligninolytic enzymes, laccase and Mn-peroxidase were found in distal samples, while the proximal samples exhibited higher activity of cellobiohydrolase (exocellulase) and endoxylanase. When enzyme activities were expressed in relative values,

Table 2 Enzyme activities in *Fagus sylvatica* and *Betula pendula* wood colonised by *Fomes fomentarius* and in its fruitbodies

	F. sylvatica wood	*F. sylvatica* fruitbodies	*B. pendula* wood	*B. pendula* fruitbodies
Laccase	0.39 ± 0.17a	0.09 ± 0.06b	0.21 ± 0.09ab	0.13 ± 0.09b
Mn-peroxidase	1.62 ± 0.45a	0.59 ± 0.18b	0.61 ± 0.26b	0.09 ± 0.05c
Endocellulase	9.46 ± 2.93a	7.39 ± 2.28a	2.28 ± 0.59b	1.11 ± 0.76b
Cellobiohydrolase	5.01 ± 1.01a	7.70 ± 1.91a	2.49 ± 0.80b	1.42 ± 0.52b
β-Glucosidase	12.9 ± 3.7b	38.7 ± 7.4a	19.9 ± 5.3ab	13.2 ± 1.6b
Endoxylanase	4.5 ± 2.0a	15.8 ± 7.0a	0.6 ± 0.2b	0.1 ± 0.1c
β-Xylosidase	18.1 ± 3.0a	19.0 ± 2.1a	4.1 ± 1.5b	1.5 ± 0.4b
N-Acetylglucosaminidase	62 ± 19b	140 ± 29a	86 ± 4b	14 ± 5c

Activities are expressed in U g^{-1} dry mass and represent averages and standard errors of means from 15 replicates per treatment
Different letters indicate statistically significant differences between sample types at $P < 0.05$

Table 3 Enzyme activities, ergosterol and DNA content in the *Betula pendula* log segment colonised by *Fomes fomentarius*

		Fruitbody (n = 3)	Proximal wood (n = 9)	Distal wood (n = 9)
Laccase*	(U g^{-1})	0.20 ± 0.20	0.03 ± 0.01	0.20 ± 0.05
Mn-peroxidase*	(U g^{-1})	0.00 ± 0.00	2.47 ± 1.00	5.63 ± 1.02
Endocellulase	(U g^{-1})	0.0 ± 0.0	13.1 ± 5.8	6.4 ± 3.8
Cellobiohydrolase*	(U g^{-1})	10.2 ± 4.2	5.8 ± 1.1	2.8 ± 0.9
β-Glucosidase	(U g^{-1})	119 ± 32	34 ± 6	22 ± 6
Endoxylanase*	(U g^{-1})	1.2 ± 1.2	9.7 ± 1.8	5.0 ± 1.4
β-Xylosidase	(U g^{-1})	4.99 ± 3.02	0.94 ± 0.56	3.14 ± 1.97
N-Acetylglucosaminidase	(U g^{-1})	369 ± 33	18 ± 5	22 ± 3
Ergosterol	(μg g^{-1})	n.d.	84 ± 13	53 ± 4
Fungal DNA*	(ng g^{-1})	n.d.	353 ± 58	2512 ± 539
Bacterial DNA*	(ng g^{-1})	n.d.	4 ± 1	95 ± 51
F/B ratio*		n.d.	111 ± 25	62 ± 15

Data represent averages and standard errors. Asterisks indicate statistically significant differences between samples from proximal and distal wood at $P < 0.05$

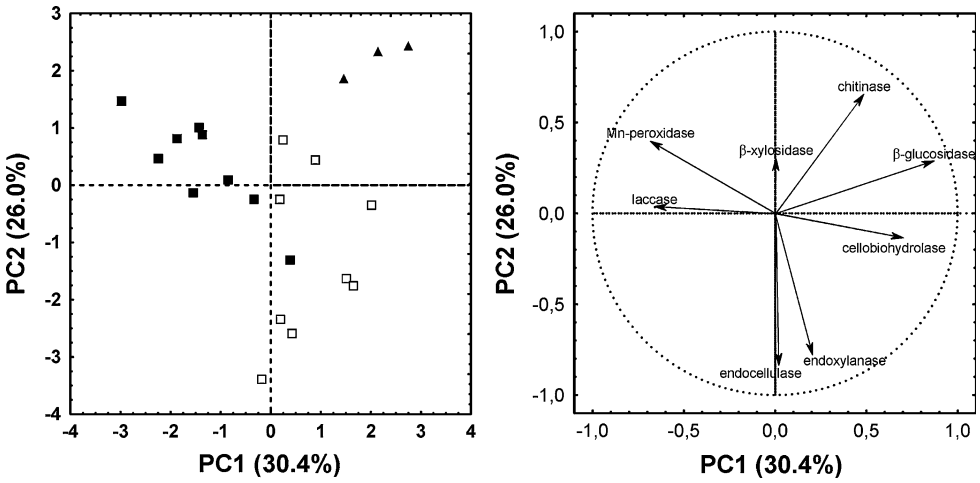


Fig. 2 Principal component analysis of the relative activities of extracellular enzymes in fruitbodies and *Betula pendula* log segment colonised by *Fomes fomentarius*. **a** PC loads of individual samples: fruitbodies (*triangles*), distal samples (*filled squares*), proximal samples (*open squares*); **b** PC loads of individual enzymes

proximal samples showed preferential occurrence of xylanase, β-glucosidase and cellobiohydrolase, while higher laccase and Mn-peroxidase were more important in the distal ones (Table 3). Samples from fruiting body showed high activity of β-glucosidase and chitinase. Principal component analysis of the relative activities of individual enzymes clearly separated proximal, distal and fruitbody-related samples (Fig. 2). Statistically significant differences between proximal and distal samples were found in the PC loads along the first ordinary axis. Fruitbody samples were separated from wood samples along the second PC axis and from the distal samples also along the first PC axis. The results show that the mycelial colony of the fungus in wood exhibits local differences in physiology. While the enzyme activities in wood localised close to fruitbody where the fungal colony is likely older and at a later decay stage produces preferentially polysaccharide hydrolases while in the more distant parts of wood, ligninolytic activity prevails. These observations are corroborated by the fact that endocellulase and endoxylanase activities correlated strongly positively ($P < 0.01, r = 0.746$) as well as cellobiohydrolase and β-glucosidase ($P < 0.01, r = 0.702$). Mn-peroxidase negatively correlated with both endocellulase and endoxylanase. *N*-acetyl-glucosaminidase and laccase correlated positively with each other ($P < 0.05$, r = 0.439).

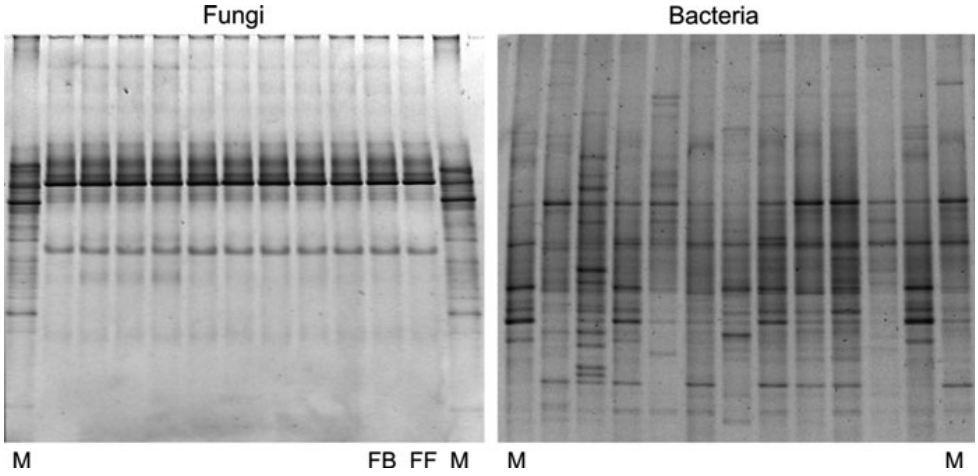


Fig. 3 DGGE analysis of fungal and bacterial community in the *Betula pendula* log segment colonised by *Fomes fomentarius*. DGGE analyses were performed with DNA of the fungal ITS 1 region (primers ITS1f/ITS4 nested with ITS1f-gc/ITS2) and of the bacterial V4-V6 region of 16S rDNA (primers 968-gc/1378). *M* marker, *FB* DNA isolated from *F. fomentarius* fruitbody, *FF* DNA extracted from the pure culture of *Fomes fomentarius* CCBAS534, *all other lanes* DNA extracted from wood

Ergosterol content in wood varied between 28 and 145 μg g^{-1} and did not significantly differ between proximal and distal samples of wood (84 ± 13 μg g^{-1} and 53 ± 4 μg g^{-1}, respectively; Table 3). Ergosterol content did not significantly correlate with any of the enzyme activities. When fungal biomass was analysed based on DNA content per gram of dry wood, its amount was significantly higher in the distal samples (Fig. 1). Also the amount of bacterial DNA in distal samples was significantly higher than in the proximal ones, despite high level of spatial variation. Fungal biomass was highly dominant in the substrate with fungal/bacterial DNA ratio of 111 in the proximal wood and 62 in the distal wood.

Analysis of DGGE gels showed that the banding pattern derived from *F. fomentarius* CCBAS 534 is similar to this obtained from the *F. fomentarius* fruitbody as well as those from the fungus-colonised *B. pendula* log (Fig. 3). Reamplification and sequencing of major DGGE bands derived from wood/fruitbody showed that these sequences share highest similarity (>99%) with the *Fomes fomentarius* strain sequence deposited in GenBank (gi 980706.7; data not shown). As a consequence, *F. fomentarius* represented the only dominant fungal species found in wood. In contrast, there was a rich bacterial community whose composition varied between samples (Fig. 3).

Discussion

The range of enzymes detected in *F. fomentarius*-colonised wood corresponded to the enzymes produced by other ligninolytic basidiomycetes in culture (Baldrian et al. 2010). Significant effect of wood type on the enzyme production was observed: the activity of cellulose and xylan-degrading enzymes was significantly higher in beech wood than in birch wood. Ligninolytic enzymes were found mainly in colonised wood while the activities in fruitbodies were low. In the brown-rot fungus *Piptoporus betulinus*, higher β-glucosidase activity was found in fungal fruitbodies while endocellulase was preferentially produced in wood colonised by the fungus (Valášková and Baldrian 2006). Similar differences were not observed in *F. fomentarius*.

During lignocellulose transformation by white-rot fungi (including *F. fomentarius*), preferential removal of lignin (selective delignification) is performed before extensive utilization of wood polysaccharides (Hatakka 2001). This corresponds well with our findings where wood adjacent to fungal fruitbodies (which is likely at the centre of the fungal colony and at a later stage of decomposition) exhibited higher activity of xylanase, β-glucosidase and cellobiohydrolase while more distant wood parts (in the earlier decomposition stages) showed higher activity of

laccase and Mn-peroxidase. Activities of *N*-acetylglucosaminidase and laccase, enzymes potentially involved in the interactions among microorganisms (Rast et al. 2003), exhibited positive correlation. Due to its participation in chitin degradation, *N*-acetylglucosaminidase may be also an indicator of active rearrangement of fungal mycelia (Lindahl and Finlay 2006).

In wood, primary wood-colonizing fungi spread to occupy the largest available volume of substrate. At places where different strains meet, coloured interaction zones develop. Within these domains, primary colonizers prevent the entry of other competing fungi until late decay (Boddy and Watkinson 1995). Interaction zones were not present in the analysed wood segment and the presence of a single dominant fungal species—*F. fomentarius*—was confirmed by the molecular methods. Due to a large volume of colonised wood, the fungal colony was able to recruit resources sufficient for fruitbody formation (Urcelay and Robledo 2009). The colonization of wood by a single fungal species is often found during initial decay (Vainio and Hantula 2000; Valášková et al. 2009), but in later phases a relatively rich fungal community may develop (Lonsdale et al. 2008; Zhang et al. 2008b).

Fungal biomass content in wood was relatively high, approximately twice as high as in spruce wood inoculated with *Phlebia radiata* or *Phanerochaete chrysosporium* (Niemenmaa et al. 2008). If we consider that 1 g of fungal biomass contains between 0.7 and 2.1 mg ergosterol, this would correspond to 10–150 mg of fungal biomass per g wood dry mass, which roughly corresponds to values in plant litter colonised by a single fungal species (unpublished data) and shows high efficiency of wood assimilation by *F. fomentarius*. Ergosterol content did not differ between proximal and distal wood while the qPCR showed higher amount of fungal DNA in distal wood. This bias may indicate the fact that distal wood contains more actively proliferating mycelium where the relative DNA content is higher than in older parts of mycelia.

Compared to fungal biomass, the biomass of bacteria in wood was very low. Indeed, bacteria are not supposed to be efficient decomposers of bulky lignocellulose like wood (de Boer et al. 2005) except water-saturated wood where fungi are limited by oxygen tension (Clausen 1996; Przybyl 2001). As a consequence, it is most likely that the measured enzyme activities reflected the degradative activity of the fungus. Previous studies on wood colonised by another white-rot basidiomycete, *Hypholoma fasciculare*, showed that bacterial counts were low in wood during initial colonization by the fungus while the abundance and diversity of bacteria was high in wood during later stages of decay (Folman et al. 2008; Valášková et al. 2009). This seems to reflect the fact that the originally high C/N and C/P content of wood decreases during decomposition which makes this substrate more suitable for bacteria (Boddy and Watkinson 1995). Our finding that polysaccharide hydrolysis is higher in proximal samples would suggest that there is more available substrate in the form of oligosaccharides available to bacteria. However, bacterial biomass and the F/B ratio in the studied sample were higher in distal than in the proximal samples.

The bacterial community in *F. fomentarius*-colonised wood was relatively rich and diverse among samples and contrasting with the extremely low diversity of fungi. This is in accordance to previous studies on bacteria associated with decaying wood (Folman et al. 2008; Zhang et al. 2008a; Valášková et al. 2009) and may indicate the fact that bacteria and fungi actually occupy different decomposition niches and do not directly compete while the establishment of other fungi in wood is prevented by the primary coloniser occupying the substrate.

Our results show that in addition to the substrate effect, a significant level of spatial heterogeneity in terms of enzyme activity, fungal and bacterial biomass is present in decaying wood homogeneously colonised by a single fungal species. Based on the differences between proximal and distal samples in a wood log, the wood decay stage may be one of the important driving factors of the variation in wood properties.

Acknowledgments This work was supported by the Czech Science Foundation (526/07/0620) by the Ministry of Education, Youth and Sports of the Czech Republic (OC155, ME10152 and LA10001), and by the Institutional Research Concept of the Institute of Microbiology of the ASCR, v.v.i. (AV0Z50200510).

References

Amann RI, Ludwig W, Schleifer KH (1995) Phylogenetic identification and in situ detection of individual microbial cells without cultivation. Microbiol Rev 59:143–169

Arora DS, Sharma RK (2010) Ligninolytic fungal laccases and their biotechnological applications. Appl Biochem Biotechnol 160:1760–1788

Baldrian P (2008) Enzymes of saprotrophic basidiomycetes. In: Boddy L, Frankland JC, van West P (eds) Ecology of saprotrophic basidiomycetes. Academic Press, Amsterdam, pp 19–41

Baldrian P, Valášková V (2008) Degradation of cellulose by basidiomycetous fungi. FEMS Microbiol Rev 32:501–521

Baldrian P, Voříšková J, Dobiášová P, Merhautová V, Lisá L, Valášková V (2010) Production of extracellular enzymes and degradation of biopolymers by saprotrophic microfungi from the upper layers of forest soil. Plant Soil. doi: 10.1007/s11104-010-0324-3

Boddy L, Watkinson SC (1995) Wood decomposition, higher fungi, and their role in nutrient redistribution. Can J Bot 73:S1377–S1383

Bourbonnais R, Paice MG (1990) Oxidation of non-phenolic substrates. An expanded role for laccase in lignin biodegradation. FEBS Lett 267:99–102

Clausen CA (1996) Bacterial associations with decaying wood: a review. Int Biodeter Biodegrad 37:101–107

Crockatt ME, Pierce GI, Camden RA, Newell PM, Boddy L (2008) Homokaryons are more combative than heterokaryons of *Hericium coralloides*. Fungal Ecol 1:40–48

Dahlberg A, Genney DR, Heilmann-Clausen J (2010) Developing a comprehensive strategy for fungal conservation in Europe: current status and future needs. Fungal Ecol 3:50–64

de Boer W, Folman LB, Summerbell RC, Boddy L (2005) Living in a fungal world: impact of fungi on soil bacterial niche development. FEMS Microbiol Rev 29:795–811

Folman LB, Gunnewiek P, Boddy L, de Boer W (2008) Impact of white-rot fungi on numbers and community composition of bacteria colonizing beech wood from forest soil. FEMS Microbiol Ecol 63:181–191

Gabriel J, Baldrian P, Rychlovsky P, Krenzelok M (1997) Heavy metal content in wood-decaying fungi collected in Prague and in the National Park Sumava in the Czech Republic. Bull Environ Contam Toxicol 59:595–602

Gardes M, Bruns TD (1993) ITS primers with enhanced specificity for basidiomycetes—application to the identification of mycorrhizae and rusts. Mol Ecol 2:113–118

Hatakka A (2001) Biodegradation of lignin. In: Steinbüchel A, Hofrichter M (eds) Biopolymers 1: lignin, humic substances and coal. Wiley-VCH, Weinheim, pp 129–180

Heilmann-Clausen J, Boddy L (2005) Inhibition and stimulation effects in communities of wood decay fungi: exudates from colonized wood influence growth by other species. Microb Ecol 49:399–406

Heuer H, Krsek M, Baker P, Smalla K, Wellington EMH (1997) Analysis of actinomycete communities by specific amplification of genes encoding 16S rRNA and gel-electrophoretic separation in denaturing gradients. Appl Environ Microbiol 63:3233–3241

Hiscox J, Baldrian P, Rogers HJ, Boddy L (2010) Changes in oxidative enzyme activity during interspecific mycelial interactions involving the white-rot fungus *Trametes versicolor*. Fungal Gene Biol. doi:10.1016/j.fgb.2010.03.007

Leigh MB, Pellizari VH, Uhlik O, Sutka R, Rodrigues J, Ostrom NE, Zhou JH, Tiedje JM (2007) Biphenyl-utilizing bacteria and their functional genes in a pine root zone contaminated with polychlorinated biphenyls (PCBs). ISME J 1:134–148

Lekounougou S, Mounguengui S, Dumarçay S, Rose C, Courty PE, Garbaye J, Gérardin P, Jacquot JP, Gelhaye E (2008) Initial stages of *Fagus sylvatica* wood colonization by the white-rot basidiomycete *Trametes versicolor*: enzymatic characterization. Int Biodeterior Biodegrad 61:287–293

Lindahl BD, Finlay RD (2006) Activities of chitinolytic enzymes during primary and secondary colonization of wood by basidiomycetous fungi. New Phytol 169: 389–397

Lonsdale D, Pautasso M, Holdenrieder O (2008) Wood-decaying fungi in the forest: conservation needs and management options. Eur J For Res 127:1–22

Lundell TK, Mäkelä MR, Hildén K (2010) Lignin-modifying enzymes in filamentous basidiomycetes—ecological, functional and phylogenetic review. J Basic Microbiol 50:5–20

Ngo TT, Lenhoff HM (1980) A sensitive and versatile chromogenic assay for peroxidase and peroxidase-coupled reactions. Anal Biochem 105:389–397

Niemela T, Renvall P, Pentilla R (1995) Interactions of fungi at late stages of wood decomposition. Ann Bot Fennici 32:141–152

Niemenmaa O, Galkin S, Hatakka A (2008) Ergosterol contents of some wood-rotting basidiomycete fungi grown in liquid and solid culture conditions. Int Biodeterior Biodegrad 62:125–134

Przybyl K (2001) Fungi and bacteria associated with the wet and brown wood in trunk of *Betula pendula* trees. Acta Soc Bot Polon 70:113–117

Rast DM, Baumgartner D, Mayer C, Hollenstein GO (2003) Cell wall-associated enzymes in fungi. Phytochemistry 64:339–366

Roussel B, Rapior S, Masson CL, Boutie P (2002) *Fomes fomentarius* (L.: Fr.) Fr.: a mushroom with varied uses. Cryptogamie Mycologie 23:349–366

Ruel K, Ambert K, Joseleau J-P (1994) Influence of the enzyme equipment of white-rot fungi on the patterns of wood degradation. FEMS Microbiol Rev 13:241–254

Sagova-Mareckova M, Cermak L, Novotna J, Plhackova K, Forstova J, Kopecky J (2008) Innovative methods for soil DNA purification tested in soils with widely differing characteristics. Appl Environ Microbiol 74:2902–2907

Šnajdr J, Valášková V, Merhautová V, Herinková J, Cajthaml T, Baldrian P (2008) Spatial variability of enzyme activities and microbial biomass in the upper layers of *Quercus petraea* forest soil. Soil Biol Biochem 40:2068–2075

Teichert S, Bondrup-Nielsen S (2005) Effect of spatial scale on habitat use of *Bolitotherus cornutus* (*Coleoptera* : *Tenebrionidae*). Can Entomol 137:192–201

Urcelay C, Robledo G (2009) Positive relationship between wood size and basidiocarp, production of polypore fungi in *Alnus acuminata* forest. Fungal Ecol 2:135–139

Vainio EJ, Hantula J (2000) Direct analysis of wood-inhabiting fungi using denaturing gradient gel electrophoresis of amplified ribosomal DNA. Mycol Res 104:927–936

Valášková V, Baldrian P (2006) Degradation of cellulose and hemicelluloses by the brown rot fungus *Piptoporus betulinus*—production of extracellular enzymes and characterization of the major cellulases. Microbiol SGM 152:3613–3622

Valášková V, Baldrian P (2009) Denaturing gradient gel electrophoresis as a fingerprinting method for the analysis of soil microbial communities. Plant Soil Environ 55:413–423

Valášková V, Šnajdr J, Bittner B, Cajthaml T, Merhautová V, Hoffichter M, Baldrian P (2007) Production of lignocellulose-degrading enzymes and degradation of leaf litter by saprotrophic basidiomycetes isolated from a *Quercus petraea* forest. Soil Biol Biochem 39:2651–2660

Valášková V, de Boer W, Gunnewiek P, Pospíšek M, Baldrian P (2009) Phylogenetic composition and properties of bacteria coexisting with the fungus *Hypholoma fasciculare* in decaying wood. ISME J 3:1218–1221

Wakelin SA, Colloff MJ, Harvey PR, Marschner P, Gregg AL, Rogers SL (2007) The effects of stubble retention and nitrogen application on soil microbial community structure and functional gene abundance under irrigated maize. FEMS Microbiol Ecol 59:661–670

White TJ, Bruns T, Lee S, Taylor J (1990) Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. In: Innins MA, Gelfand DH, Sninsky JJ, White TJ (eds) PCR protocols. Academic Press, San Diego, CA, pp 315–322

Wilmotte A, Van der Auwera G, De Wachter R (1993) Structure of the 16 S ribosomal RNA of the thermophilic cyanobacterium chlorogloeopsis HTF (‘mastigocladus laminosus HTF’) strain PCC7518, and phylogenetic analysis. FEBS Lett 317:96–100

Zhang HB, Yang MX, Tu R (2008a) Unexpectedly high bacterial diversity in decaying wood of a conifer as revealed by a molecular method. Int Biodeterior Biodegrad 62:471–474

Zhang HB, Yang MX, Tu R, Gao L, Zhao ZW (2008b) Fungal communities in decaying sapwood and heartwood of a conifer *Keteleeria evelyniana*. Curr Microbiol 56:358–362

III

AEM
Journals.ASM.org

Scaling Down the Analysis of Environmental Processes: Monitoring Enzyme Activity in Natural Substrates on a Millimeter Resolution Scale

Petr Baldrian and Tomáš Větrovský

Laboratory of Environmental Microbiology, Institute of Microbiology of the Academy of Sciences of the Czech Republic, v.v.i., Prague, Czech Republic

Natural environments often show high levels of spatial heterogeneity. With a methodology based on the immobilization of fluorescent substrates, the distribution of extracellular enzymes can be studied at a 2.3-mm resolution with a detection limit of 1.8 nmol · h^{-1} · cm^{-2}. The method is applicable to environmental samples such as wood, litter, soil, or fungal colonies.

The determination of enzymatic activities is a simple approach to the study of environmental processes mediated by fungi and bacteria. Thus, enzyme activities have been interpreted as indirect measures of microbial biomass, rhizosphere effects, soil productivity, or mineralization potential of naturally occurring substrates or xenobiotics (2). Fluorogenic 4-methylumbelliferyl (MUB)- or 4-amidomethylcoumaryl-labeled enzyme substrates have been introduced for measuring the activity of a wide variety of hydrolytic enzymes, including exoglucanases, phosphatases, peptidases, and several others, with sufficient sensitivity based on the fluorescence of reaction products (2, 21).

The spatial dependence of environmental variables, such as enzyme activities, has been previously studied at various levels of resolution, for example, in the soil ranging from centimeters to kilometers (7). While the picture of the spatial distribution of environmental variables provides meaningful results at any scale, depending on the question asked and sample size chosen, both the decreasing sampling distance and sample size result in a decrease of spatial autocorrelation range due to the presence of spatial heterogeneity at different scales (9). In several environments, the extent of such spatial heterogeneity is high, even at scales smaller than a few centimeters. This was documented in the case of spatial distribution of phyllosphere or litter-associated fungi (15), decomposing wood colonized by saprotrophic basidiomycetes or ascomycetes (4), enzyme activity variation within the litter and soil horizons (1, 6, 20), or the distribution of soil bacteria. Bacterial diversity in soils on a millimeter scale (11) and communities of *Archaea* recovered as spatially independent 0.1-g subsamples from within a single 20-mm-diameter soil core varied in their composition (14). The microbial biomass content and community composition, as well as the rates of microbe-catalyzed processes, have been demonstrated to vary considerably over a scale of several millimeters at the soil-litter interface (16–18). Additionally, lichen soil crusts show spatial heterogeneity at a comparable resolution (22). The understanding of the variability of microbe-catalyzed processes at such scales has been hindered by the limitations of sample size requirements for the analysis. Previous efforts of dense sampling of enzyme activity were obtained at a resolution of centimeters (19) or limited to measurements along linear transects (12). Here, we show a fluorimetric MUB-based enzyme assay suitable for the study of small-scale distribution of extracellular hydrolytic enzymes of fungi and bacteria over surfaces of various substrates.

Samples (fungal colonies growing on agar, thin slices of colonized wood, soil, or decaying leaf litter) were fixed into plastic plates and overlaid with a 1% low-melting-point agarose in a 50 mM Na-acetate buffer supplemented with appropriate MUB substrates immediately before application (45°C). After brief chilling at 4°C to solidify the agarose overlay, fluorescence was read at 40°C using a multimode microplate reader, Infinite M200 (TECAN, Austria), by scanning the surface of the gel at a rectangular 2.3- by 2.3-mm grid for 5- to 10-min intervals over a period of 30 to 120 min. The data were visualized in Origin 8 (Originlab, MA), and the geostatistical analysis (variogram construction and map construction by kriging) was performed in Surfer 8 (Golden Software, Inc., CO). Linear fitting was used to determine the relationships between fluorescence and MUB concentration and between fluorescence increase and the activity of purified β-glucosidase applied at various concentrations on the surface of an agarose gel and dried at room temperature under vacuum (see the text in the supplemental material).

The amount of MUB linearly correlated with detected fluorescence ($P < 0.0001$), and the recorded increase of fluorescence corresponded well with the activity of purified β-glucosidase applied to the gel surface ($P < 0.0001$; Fig. 1). The detection limit of the method was determined to be 1.8 nmol · h^{-1} · cm^{-2} as 3× the maximal background fluorescence change.

Visualization of enzyme activity was carried out on (i) colonies of saprotrophic basidiomycetous fungi on malt agar (Fig. 2a and b); (ii) thin sections of a dead *Betula pendula* branch colonized by fungi with a fruit body of *Fomes fomentarius* (Fig. 2c and d); (iii) *Quercus petraea* leaves decaying *in situ* (Fig. 2e and f); and (iv) profiles of *Quercus* sp. forest topsoil collected with a soil slicer (10) (Fig. 1). The results show that even at the scale of a few square centimeters, enzyme activity varied considerably; the coefficients of variation (CV = SD/mean) of enzyme activities in fungus-colonized wood were 0.31 ± 0.10 for cellobiohydrolase, 0.40 ± 0.08

Received 21 December 2011 Accepted 23 February 2012
Published ahead of print 2 March 2012
Address correspondence to Petr Baldrian, baldrian@biomed.cas.cz.
Supplemental material for this article may be found at http://aem.asm.org/.

doi:10.1128/AEM.07953-11

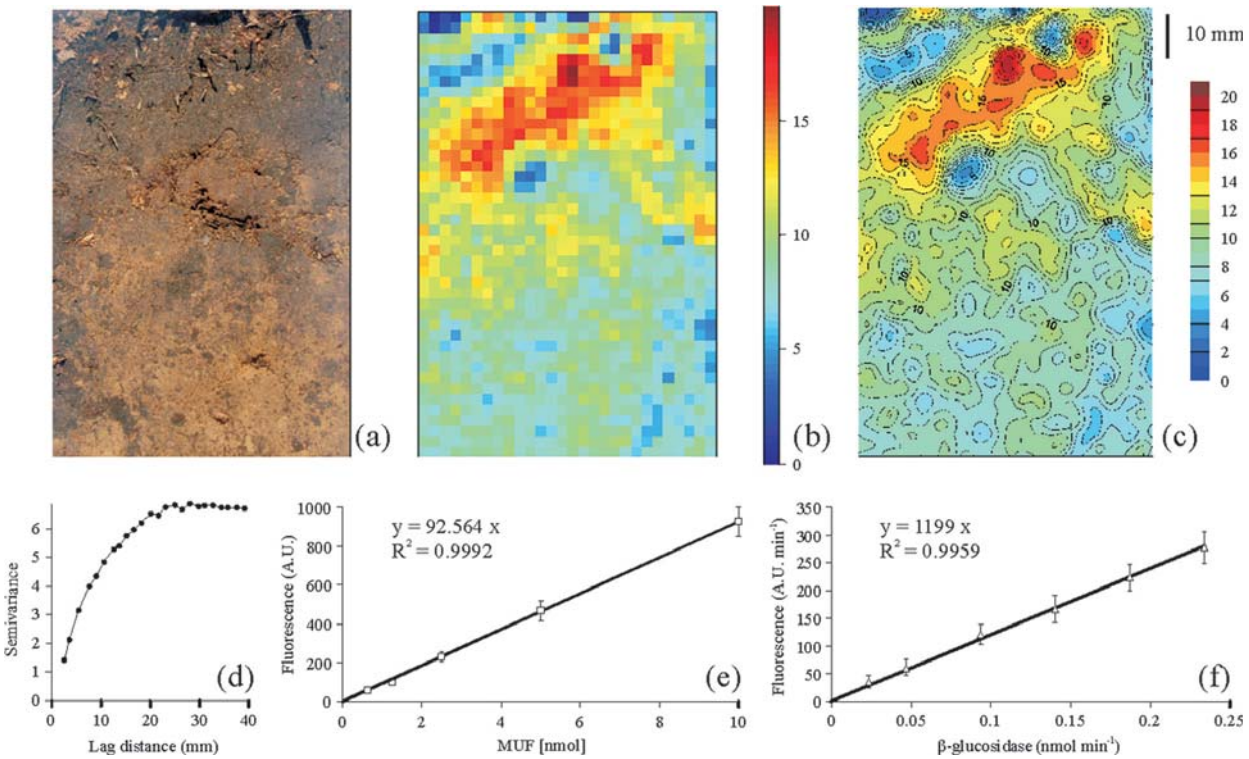


FIG 1 Distribution of β-glucosidase in the *Quercus* forest soil profile. (a) The soil profile of *Quercus* forest soil after litter removal (80-mm width by 120-mm depth); (b) measured values of β-glucosidase activity in nmol · h^{-1} · cm^{-2}; (c) krigged map of the distribution of β-glucosidase; (d) the plot of semivariance against distance; (e) linear fit of the observed fluorescence against the amount of MUF ($n = 6$ samples); (f) linear fit of the observed fluorescence change against the activity of purified β-glucosidase from almonds (Sigma, MO; $n = 6$ samples).

for β-glucosidase, 0.52 ± 0.21 for *N*-acetylglucosaminidase, and 0.56 ± 0.23 for β-xylosidase. In the decomposing leaves, the CV were 0.39 ± 0.08 for *N*-acetylglucosaminidase, 0.45 ± 0.12 for β-xylosidase, and 0.41 ± 0.22 for cellobiohydrolase. Even within single fungal colonies, the enzyme activities are unevenly distributed (Fig. 2).

In decaying wood, the enzyme activities spatially autocorrelated in a range of <30 mm (12 to 32 mm for cellobiohydrolase,

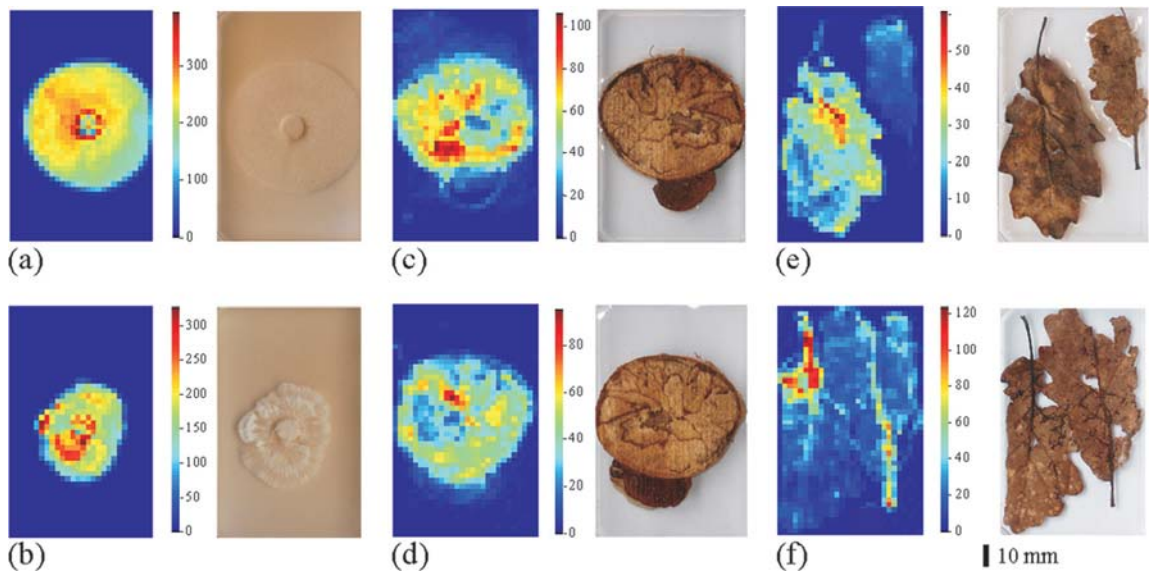


FIG 2 Distribution of hydrolytic enzymes across fungal colonies, fungus-colonized wood sections, and decaying leaves. Distribution of β-glucosidase (a) and phosphomonoesterase (b) across the colonies of the saprotrophic basidiomycete *Hypholoma fasciculare* on malt extract agar, β-glucosidase (c) and cellobiohydrolase (d) over a cross-section of a *Betula pendula* branch colonized by wood-decomposing fungi, including *Fomes fomentarius* (fruit body), *N*-acetylglucosaminidase (e), and cellobiohydrolase (exocellulase) (f) over the surface of *Quercus petraea* leaves decomposing for 10 months on the forest floor surface. Enzyme activities are expressed in nmol · h^{-1} · cm^{-2}.

β-xylosidase, β-glucosidase, and *N*-acetylglucosaminidase). The cross-section of the log showed borderlines between different species or genets of fungal colonizers (4), and high or low enzymatic activity was often associated with a particular patch colonized by a specific fungus (Fig. 2c and d), the spatial autocorrelation of the enzyme activity being thus largely affected by the size of these patches. The spatial autocorrelation in leaves ranged from 29 ± 21 mm for β-xylosidase, over 43 ± 19 mm for cellobiohydrolase, and up to 62 ± 16 mm for *N*-acetylglucosaminidase. This roughly corresponds to the sizes of areas in living plant leaves colonized by a single dominant fungal species (13), and we can hypothesize that the distribution of certain enzymes reflects the presence of their producers at this spatial scale.

Soil is a highly heterogeneous environment, and a link between the heterogeneity of soil physicochemical properties and microbial abundance has been demonstrated on a centimeter scale (8). When sampled at a coarse scale with 16-cm^2 samples analyzed over a 144-m^2 site, the CV of β-glucosidase activity in *Quercus* sp. forest mineral soil was 0.29 (20). In the same soil, where 0.053-cm^2 samples were collected over an area of only 0.0048 m^2, the CV was very similar (0.24 ± 0.06). This shows that enzyme activity is highly variable, even within a few square centimeters. The soil properties, microbial biomass, and enzyme activities of the soil from this study showed spatial autocorrelation in a range of tens of centimeters to more than 1 m if 16-cm^2 samples were analyzed (3); the autocorrelation range of β-glucosidase activity was 92 cm. Here, we show that the autocorrelation range drops to 21 ± 4 mm with a sample size reduction to 0.053 cm^2. To determine whether the current sample size and sampling frequency were appropriate for the estimation of the range, the data sets were resampled (i) by reducing the sampling depth to 1/4 by the analysis on 0.053-cm^2 samples taken at a double distance and (ii) by increasing the sample size to 0.212 cm^2 using the mean activities of four adjacent squares. These analyses delivered range estimates of 22 ± 4 mm and 20 ± 3 mm, respectively, showing that the sampling distance and sample size did not represent limitations for the reliable estimation of this important parameter.

There is a recent report on a method of enzyme measurement at a submillimeter scale in marine sediments (5). Unfortunately, due to the requirement of a foil adhesion onto the sample surface, this method is not applicable for dry samples with uneven surfaces. In terrestrial environments, enzyme activity distribution was previously studied using a Plexiglas window in soil with filter paper containing enzyme substrates that enabled the visualization of the presence or absence of enzymes but not the quantification of activity (6). At present, the lowest sample size used for the analysis of enzyme activity distribution was used for leaf discs of immersed litter, where 0.20-cm^2 samples were analyzed individually on a 2-cm grid (19). The method described here has an improved resolution and is technically more feasible, because whole surfaces are analyzed at once. The obvious limitation compared to classical methods of sampling is the 2-dimensional sampling that quantifies enzyme activity associated with surfaces rather than a volume of the sample. Despite this, the approach seems to be useful for the study of spatial distribution of hydrolytic enzymes in environmental samples with sufficient sensitivity. In addition, this work opens the possibility for future targeted sampling of microbial community composition under enzyme activity hot spots, an approach that can potentially increase our knowledge of microbial community structure-function relationships.

ACKNOWLEDGMENTS

This work was supported by the Ministry of Education, Youth, and Sports of the Czech Republic (grant no. LA10001 and ME10152).

REFERENCES

1. **Alarcon-Gutierrez E, et al.** 2009. Spatial variations of chemical composition, microbial functional diversity, and enzyme activities in a Mediterranean litter (Quercus ilex L.) profile. Pedobiologia **52**:387–399.
2. **Baldrian P.** 2009. Microbial enzyme-catalysed processes in soils and their analysis. Plant Soil Environ. **55**:370–378.
3. **Baldrian P, Merhautová V, Cajthaml T, Petránková M, Šnajdr J.** 2010. Small-scale distribution of extracellular enzymes, fungal, and bacterial biomass in *Quercus petraea* forest topsoil. Biol. Fertil. Soils **46**:717–726.
4. **Boddy L.** 2000. Interspecific combative interactions between wood-decaying basidiomycetes. FEMS Microbiol. Ecol. **31**:185–194.
5. **Cao Z, Zhu Q, Aller RC, Aller JY.** 2011. A fluorosensor for two-dimensional measurements of extracellular enzyme activity in marine sediments. Mar. Chem. **123**:23–31.
6. **Dong S, Brooks D, Jones MD, Grayston SJ.** 2007. A method for linking in situ activities of hydrolytic enzymes to associated organisms in forest soils. Soil Biol. Biochem. **39**:2414–2419.
7. **Ettema CH, Wardle DA.** 2002. Spatial soil ecology. Trends Ecol. Evol. **17**:177–183.
8. **Franklin RB, Mills AL.** 2009. Importance of spatially structured environmental heterogeneity in controlling microbial community composition at small spatial scales in an agricultural field. Soil Biol. Biochem. **41**:1833–1840.
9. **Franklin RB, Mills AL.** 2003. Multi-scale variation in spatial heterogeneity for microbial community structure in an eastern Virginia agricultural field. FEMS Microbiol. Ecol. **44**:335–346.
10. **Genney DR, Anderson IC, Alexander IJ.** 2006. Fine-scale distribution of pine ectomycorrhizas and their extramatrical mycelium. New Phytol. **170**:381–390.
11. **Grundmann GL.** 2004. Spatial scales of soil bacterial diversity—the size of a clone. FEMS Microbiol. Ecol. **48**:119–127.
12. **Lindahl BD, Finlay RD.** 2006. Activities of chitinolytic enzymes during primary and secondary colonization of wood by basidiomycetous fungi. New Phytol. **169**:389–397.
13. **Lodge DJ, Fisher PJ, Sutton BC.** 1996. Endophytic fungi of *Manilkara bidentata* leaves in Puerto Rico. Mycologia **88**:733–738.
14. **Nicol GW, Glover LA, Prosser JI.** 2003. Spatial analysis of archaeal community structure in grassland soil. Appl. Environ. Microbiol. **69**:7420–7429.
15. **Osono T.** 2006. Role of phyllosphere fungi of forest trees in the development of decomposer fungal communities and decomposition processes of leaf litter. Can. J. Microbiol. **52**:701–716.
16. **Poll C, Brune T, Begerow D, Kandeler E.** 2010. Small-scale diversity and succession of fungi in the detritusphere of rye residues. Microb. Ecol. **59**:130–140.
17. **Poll C, Ingwersen J, Stemmer M, Gerzabek MH, Kandeler E.** 2006. Mechanisms of solute transport affect small-scale abundance and function of soil microorganisms in the detritusphere. Eur. J. Soil Sci. **57**:583–595.
18. **Poll C, Marhan S, Ingwersen J, Kandeler E.** 2008. Dynamics of litter carbon turnover and microbial abundance in a rye detritusphere. Soil Biol. Biochem. **40**:1306–1321.
19. **Smart KA, Jackson CR.** 2009. Fine scale patterns in microbial extracellular enzyme activity during leaf litter decomposition in a stream and its floodplain. Microb. Ecol. **58**:591–598.
20. **Šnajdr J, et al.** 2008. Spatial variability of enzyme activities and microbial biomass in the upper layers of *Quercus petraea* forest soil. Soil Biol. Biochem. **40**:2068–2075.
21. **Vepsäläinen M, Kukkonen S, Vestberg M, Sirvio H, Niemi RM.** 2001. Application of soil enzyme activity test kit in a field experiment. Soil Biol. Biochem. **33**:1665–1672.
22. **Wu L, Lan SB, Zhang DL, Hu CX.** 2011. Small-scale vertical distribution of algae and structure of lichen soil crusts. Microb. Ecol. **62**:715–724.

IV

OPEN ACCESS Freely available online

PLOS ONE

Potential of Cometabolic Transformation of Polysaccharides and Lignin in Lignocellulose by Soil *Actinobacteria*

Tomáš Větrovský[1], Kari Timo Steffen[2], Petr Baldrian[1]*

1 Laboratory of Environmental Microbiology, Institute of Microbiology of the ASCR, v.v.i., Praha, Czech Republic, **2** Department of Applied Chemistry and Microbiology, University of Helsinki, Helsinki, Finland

Abstract

While it is known that several *Actinobacteria* produce enzymes that decompose polysaccharides or phenolic compounds in dead plant biomass, the occurrence of these traits in the environment remains largely unclear. The aim of this work was to screen isolated actinobacterial strains to explore their ability to produce extracellular enzymes that participate in the degradation of polysaccharides and their ability to cometabolically transform phenolic compounds of various complexities. Actinobacterial strains were isolated from meadow and forest soils and screened for their ability to grow on lignocellulose. The potential to transform ^{14}C-labelled phenolic substrates (dehydrogenation polymer (DHP), lignin and catechol) and to produce a range of extracellular, hydrolytic enzymes was investigated in three strains of *Streptomyces* spp. that possessed high lignocellulose degrading activity. Isolated strains showed high variation in their ability to produce cellulose- and hemicellulose-degrading enzymes and were able to mineralise up to 1.1% and to solubilise up to 4% of poplar lignin and to mineralise up to 11.4% and to solubilise up to 64% of catechol, while only minimal mineralisation of DHP was observed. The results confirm the potential importance of *Actinobacteria* in lignocellulose degradation, although it is likely that the decomposition of biopolymers is limited to strains that represent only a minor portion of the entire community, while the range of simple, carbon-containing compounds that serve as sources for actinobacterial growth is relatively wide.

Citation: Větrovský T, Steffen KT, Baldrian P (2014) Potential of Cometabolic Transformation of Polysaccharides and Lignin in Lignocellulose by Soil *Actinobacteria*. PLoS ONE 9(2): e89108. doi:10.1371/journal.pone.0089108

Editor: Michael Freitag, Oregon State University, United States of America

Received November 22, 2013; **Accepted** January 18, 2014; **Published** February 13, 2014

Funding: This work was supported by the Academy of Sciences of the Czech Republic (IAA603020901) and by the research concept of the Institute of Microbiology of the ASCR, v.v.i. (RVO61388971). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: baldrian@biomed.cas.cz

Introduction

Lignocellulose represents the dominant portion of plant biomass and is thus a key pool of carbon in terrestrial ecosystems. The decomposition of lignocellulose in soil environments, where it originates as aboveground or belowground litter, is thus an essential process of the carbon cycle. Microorganisms represent the key decomposers of lignocellulose in soils and especially fungi are often regarded as major lignocellulose decomposers [1], most likely because their larger, multicellular and often filamentous bodies are better suited for the exploitation of bulky substrates [2]. This potential has led to the evolution of efficient enzymatic systems responsible for the decomposition of biopolymers in several fungi [1,3–5].

The process of lignocellulose decomposition is mediated by extracellular enzymes that target its main components: the polysaccharides cellulose and hemicelluloses and polyphenolic lignin [6]. A wide array of enzymes is necessary for the complete decomposition of lignocellulose. The system for cellulose decomposition typically consists of endocellulases, cellobiohydrolases (exocellulases) and β-glucosidases. The hemicellulolytic system is composed of multiple glycosyl hydrolases that are specific for xylose-, mannose-, arabinose- and galactose-containing polysaccharides; and lignin degradation is mediated by oxidative enzymes, such as oxidases (laccases), peroxidases and auxiliary enzymes, that produce hydrogen peroxide [1,7,8].

Although fungi vary largely in their production of extracellular enzymes, several groups, including saprotrophic wood decomposers and cord-forming fungi, that inhabit litter and soil were shown to produce complete arrays of extracellular enzymes that decompose all of the components of lignocellulose [4,9]. Current advances in genome sequencing indicate that the theoretical potential of bacteria to degrade certain components of lignocellulose, e.g., cellulose, is relatively widespread [10]; and, for certain taxa, enzymes involved in decomposition were characterised [11]. Moreover, recent reports also show that bacteria may play a significant role in cellulose decomposition in soil environments [12]. However, the composition of bacterial enzymatic systems has not been systematically addressed, and it is difficult to estimate their potential to transform individual lignocellulose components.

Actinobacteria seem to be good candidates for efficient lignocellulose decomposition, and their filamentous growth may help them access and utilise polymeric substrates [13]. Therefore, the involvement of certain *Actinobacteria* in the degradation of polysaccharides or phenolic compounds in dead plant biomass is generally accepted [14,15]. This is based on previous reports that suggest the presence of decomposer traits in several actinobacterial taxa. In the case of cellulose, the production of endocellulase by

the genera *Streptomyces*, *Cellulomonas* and *Acidothermus* was reported [16–18], while efficient exocellulases often combined with xylanase activity were found in *Thermobifida*, *Cellulomonas* and *Cellulosimicrobium* [19–21]. β-Glucosidases have been characterised in the above genera as well as in *Clavibacter*, *Terrabacter*, *Micrococcus*, *Microbacterium* and *Bifidobacterium* [22–26]. Recently, many putative cellulose-degrading enzymes were found in the sequenced genomes of several *Actinobacteria* [27], and this phylum showed the highest percentage of genomes that harboured putative cellulolytic enzymes. Approximately 1/3 of the 514 characterised genomes harboured at least one putative cellulase [10]. Although information on hemicellulose-degrading enzymes is scarce, individual enzymes were reported in multiple genera, including *Streptomyces*, *Cellulomonas*, *Cellulosimicrobium* and *Kocuria* [28–31].

Although the major lignin degraders are white-rot fungi, there are also many reports about bacterial strains that are able to degrade lignin. In addition to the *Proteobacteria* and *Firmicutes* [32–34], these reports also mention actinobacterial taxa. Studies on the decomposition of natural and synthetic lignins, for example, isolated lignin, prepared ^{14}C-synthetic lignins or model compounds, indicated that the genera *Arthrobacter*, *Nocardia* and *Streptomyces* were capable of lignin utilisation, although their efficiencies varied widely and did not reach the level that was observed in ligninolytic fungi [35–38].

Despite the relative abundance of reports on the decomposer abilities of *Actinobacteria*, the current information remains rather fragmented. Different strains have been studied for the production of individual enzymes, and the abilities to use plant lignocellulose as a growth substrate were not studied in much detail. The aim of this study was to explore the ability of soil *Actinobacteria* to act as decomposers of dead plant biomass. To achieve this goal, a set of natural isolates was screened for their production of cellulolytic enzymes, and active strains were tested for their ability to use lignocellulose in the form of wheat straw as a growth source. The potential decomposers were further screened for the production of multiple lignocellulose-degrading enzymes and for their ability to cometabolically transform ^{14}C-labelled phenolic substances (lignin, DHP and catechol) during their growth on wheat straw. The *Actinobacteria* in this study were isolated from soil with mixed-metal pollution. Due to their heavy metal resistance, *Actinobacteria* are frequently found in such soils, and this group may replace the more sensitive fungi as the main decomposers in such soils [39,40]. We thus expected that the decomposition of lignocellulose would be a common trait of strains isolated from this environment.

Materials and Methods

Isolation of Bacterial Strains

The study was carried out on private land. No specific permission was required for the activities covered by this study. However, for more intensive research activities, individual private owners have to grant the permit to conduct such research. The field studies did not involve endangered or protected species. Bacterial strains were isolated from the organic horizons of the grassland and forest soils near Příbram, Czech Republic (49°42′22.207″N, 13°58′27.296″E). The soil is a cambisol with pH of approximately 5.5 and has a clay/silt/sand ratio of approx. 40:30:30% and an elevated heavy metal content (namely, Cu, Zn, Cd, Zn, Pb and As) due to its location near a polymetallic smelter [41]. The study was performed on private land; no specific permit was required for the activities performed as part of this study.

Physical and chemical treatments were used for the selective isolation of soil *Actinobacteria*. Soil samples were pre-treated by dry heating (120°C) and phenol treatment (1.5%), and water extracts

Table 1. Screening of actinobacterial strains for their ability to produce cellobiohydrolase and 1,4-β-glucosidase.

Strain	Cellobiohydrolase	1,4-β-Glucosidase
pl18	++	+++
pl21	−	+
pl23	+	++
pl28	+	++
pl36	+++	+++
pl41	+++	+++
pl67	++	−
pl70	−	++
pl73	−	+
pl75	−	+
pl77	−	+++
pl80	−	+
pl81	−	+
pl84	+	++
pl86	+	−
pl88	+++	+++
pl95	+++	+++
pl98	−	+
pl100	−	+
pl107	−	++
pl112	−	+
pl116	+++	+++
pl118	+	+++
pl123	+++	+++
pl124	++	+
pl129	+	+
pl131	−	++
pl134	−	+++
pl136	−	++
pl138	−	+++
pl149	+	++
pl150	−	+++
pl153	+++	+++
pl154	++	+++
pr10	+	−
pr22	−	+++
pr24	+	−
pr3	+	+
pr30	+	+++
pr4	−	+
pr40	++	−
pr41	+++	++
pr45	−	+++
pr48	−	++
pr49	+	+
pr52	++	−
pr55	+++	+
pr57	+	++

Table 1. Cont.

Strain	Cellobiohydrolase	1,4-β-Glucosidase
pr6	+++	++
pr7	+	−
pr9	++	−

Twenty-five isolates produced neither of the enzymes. Legend: "+++" high production of enzyme (activity >100 mU/mL), "++" average production (100> activity >10), small production (activity <10) and "–" no enzyme production.
doi:10.1371/journal.pone.0089108.t001

were used to inoculate plates containing selective media, either humic acid-vitamin agar (1 g L^{-1} humic acid, 0.5 g L^{-1} Na_2HPO_4, 7.7 g L^{-1} KCl, 0.05 g L^{-1} $MgSO_4.7H_2O$, 0.01 g L^{-1} $FeSO_4.7H_2O$, 0.02 g L^{-1} $CaCO_3$, B-vitamins: 0.5 mg L^{-1} thiamine-HCl, riboflavin, niacin, pyridoxine, inositol, Ca-pantothenate, *p*-aminobenzoic acid and 0.25 mg L^{-1} biotin, 18 g L^{-1} agar, pH 7.2) or lignin-soy bean flour-vitamin agar containing soil extract (1 g L^{-1} lignin, 0.2 g L^{-1} soy bean flour, 0.5 g L^{-1} Na_2HPO_4, 7.7 g L^{-1} KCl, 0.05 g L^{-1}, $MgSO_4.7H_2O$, 0.01 g L^{-1} $FeSO_4.7H_2O$, 0.02 g L^{-1} $CaCO_3$, B-vitamins (see above), 100 mL L^{-1} soil extract, 18 g L^{-1} agar, pH 7.5) that was supplemented with the antibiotics kanamycin (20 mg L^{-1}) and nalidic acid (10 mg L^{-1}) [42,43]. Pure cultures of bacteria were obtained from agar plates, and those strains that were identified as *Actinobacteria* were retained. Strains were stored in a sterile, 50% glycerol solution in 25 mM Tris at −20°C and subcultured on GYM agar (4 g L^{-1} glucose, 4 g L^{-1} yeast extract, 10 g L^{-1} malt extract, 2 g L^{-1} $CaCO_3$, 12 g L^{-1} agar, pH 7.2) at 25°C.

Screening for Lignocellulose-degrading Strains

Efficient decomposition of cellulose, which is the major component of dead plant biomass, depends on the production of cellobiohydrolase and β-glucosidase. To screen for the production of these two enzymes, isolated actinobacterial strains were cultivated in liquid GYM medium for 14 days at 25°C without agitation (three replicates). The cultivation liquid was collected, and the activities of the extracellular enzymes were measured spectrophotometrically as described previously [44]. Cellobiohydrolase (exocellulase, EC 3.2.1.91) and 1,4-β-glucosidase (EC 3.2.1.21) were assayed using 4-methylumbelliferyl-β-D-cellobioside and 4-methylumbelliferyl-β-D-glucopyranoside, respectively, in 50 mM sodium acetate buffer (pH 5.0). The reaction mixtures were incubated at 40°C for 120 min and terminated by sodium carbonate addition [44].

The activity of each strain was ranked on a scale of negative, low, medium or high. Cellobiohydrolase: negative = 0−0.5 μmol min^{-1} mL^{-1}, low = from >1 to 15 μmol min^{-1} mL^{-1}, medium from >15 to 50 μmol min^{-1} mL^{-1}, and high >50 μmol min^{-1} mL^{-1}. 1,4-β-glucosidase: negative = 0–1 μmol min^{-1} mL^{-1}, low = from >1 to 50, medium from >50 to 200 μmol min^{-1} mL^{-1}, and high >200 μmol min^{-1} mL^{-1}. Strains that produced cellobiohydrolase and 1,4-β-glucosidase and exhibited high activity of at least one of those enzymes were identified, and their ability to grow on lignocellulose as a carbon source was examined. One gram of air-dried, milled wheat straw was added into 100-mL, thick-walled flasks to form a uniform layer. Each flask was supplemented with 5 mL of distilled water and sterilised by autoclaving (2×30 min at 121°C with cooling to room temperature between the two cycles). The flasks were inoculated with 1 mL of cell suspension that had been pre-grown for three days on liquid GYM media. Triplicate flasks for each strain were incubated for 21 days at 25°C. After incubation, the enzymes were extracted in 15 mL of distilled water, and the extracts were filtered and used for enzyme activity measurements.

The activities of 1,4-β-glucosidase, cellobiohydrolase and 1,4-β-xylosidase in the extracts were assessed using 4-methylumbelliferyl-β-D-glucopyranoside, MUF-β-D-cellobioside and MUF-β-D-xylopyranoside, respectively, in 50 mM sodium acetate buffer, pH 5.0, as previously described [44]. Substrates (100 μL in DMSO) at a final concentration of 500 μM were combined with the three technical replicates of the 100-μL extracts in a 96-well multiwell plate. For the background fluorescence measurement, 100 μL of sodium acetate buffer was combined with 100 μL of the 4-methylumbelliferol standards to correct for fluorescence quenching. The multiwell plates were incubated at 40°C, and fluorescence was recorded from 5 min to 125 min using the Infinite microplate reader (TECAN, Austria) at an excitation wavelength of 355 nm and an emission wavelength of 460 nm. The quantitative enzymatic activities after blank subtraction were calculated based on standard curves of 4-methylumbelliferone, and enzyme activity was expressed per g of straw dry mass.

Of the 14 strains, only six exhibited visual growth and produced extracellular enzymes in straw. Among these, three strains, pl88, pr6 and pr55, that highly produced cellobiohydrolase were selected for the detailed characterisation of glycosyl hydrolase production and the decomposition of phenolic compounds.

To analyse the spectra of the extracellular enzymes that were produced by the bacterial strains, pl88, pr6 and pr55 were cultivated on diluted GYM media with either cellulose as a specific inducer or with finely milled wheat straw (mesh size 0.2 mm) as a complex inducer. In 50-mL flasks, 10 mL of 10× diluted GYM media was combined with 50 mg of cellulose or wheat straw and sterilised by autoclaving. The flasks were inoculated with 100 μL of cell suspension that had been pre-grown for three days in liquid GYM media. Triplicate flasks for each strain were incubated for 21 days at 25°C. After incubation, the 1,4-β-glucosidase, cellobiohydrolase and 1,4-β-xylosidase activities were measured, as described above. The activities of 1,4-α-glucosidase, 1,4-α-arabinosidase, 1,4-β-galactosidase, 1,4-β-mannosidase and 1,4-β-glucuronidase in the extracts were assessed using 4-methylumbelliferyl-α-D-glucopyranoside, 4-methylumbelliferyl-α-L-arabinopyranoside, 4-methylumbelliferyl-β-D-galactopyranoside, 4-methylumbelliferyl-β-D-mannopyranoside and 4-methylumbelliferyl-β-D-glucuronide, respectively, and the same method. The activities of endo-1,4-β-glucanase (endocellulase) and endo-1,4-β-xylanase (endoxylanase) were assayed using azo-dyed carboxymethyl cellulose and birchwood xylan, respectively, according to the manufacturer's instructions (Megazyme, Ireland). Reaction mixture containing 0.2 mL of a 2% dyed substrate in 200 mM sodium acetate buffer, pH 5.0, and 0.2 mL of sample was incubated at 40°C for 60 min, and the reaction was ended by adding 1 mL of ethanol followed by 10 s of vortexing and 10 min of centrifugation (10,000×g) [45]. The amount of released dye was measured at 595 nm, and the enzyme activity was calculated according to standard curves that correlated dye release with the release of reducing sugars.

Transformation of ^{14}C-labelled Phenolic Compounds

The transformation of phenolic compounds of various complexities was studied using ^{14}C-labelled compounds. $^{14}C_\beta$-labelled dehydrogenation polymer (^{14}C-DHP) was synthesised according to Brunow [46] and dissolved in a N,N-dimethylformamide-water suspension (1:20 v/v) [47]. $^{14}C_\beta$-labelled lignin was extracted from labelled poplar trees that were prepared according to Odier [48] and used as a solid material. $^{14}C_\beta$-labelled catechol in an ethanol

Table 2. Identification of selected actinobacterial strains and the accession numbers of their partial 16S rRNA gene sequences.

Strain	Accession No	Closest hit	Accession No	Similarity (%)	Coverage (%)
pl18	KC789721	*Micromonospora saelicesensis* strain L6	JN862845	99.6	99.3
pl36	KC789723	*Streptomyces ciscaucasicus* strain HBUM83169	EU841585	99.9	99.8
pl41	KC789724	*Curtobacterium flaccumfaciens* strain LMG 3645	NR025467	100.0	99.3
pl88	KC789730	*Streptomyces atratus* strain HBUM173340	FJ486302	100.0	94.2
pl95	KC789731	*Streptomyces aureus* strain HBUM174596	EU841581	99.4	100.0
pl116	KC789734	*Kribbella antibiotica* strain YIM 31530	NR029048	98.8	99.8
pl118	KC789719	*Streptomyces sanglieri* strain IHB B 6004	KF475877	99.6	99.9
pl123	KC789735	*Nocardia exalbida* W9709	GQ376167	99.2	100.0
pl153	KC789738	*Streptomyces hygroscopicus* subsp. *geldanus* strain NBRC14620	AB184606	99.8	99.9
pl154	KC789739	*Microbispora rosea* subsp. *rosea* strain A011	AB369120	99.8	95.6
pr6	KC789740	*Streptomyces mauvecolor* strain 7534	JN180187	99.2	99.3
pr30	KC789741	*Amycolatopsis saalfeldensis* strain HKI 0474	DQ792502	99.2	100.0
pr41	KC789742	*Streptomyces setonensis* strain 17-1	EU367980	99.4	99.1
pr55	KC789744	*Streptomyces sannanensis* strain 126195	JN180213	98.9	99.4

doi:10.1371/journal.pone.0089108.t002

solution (Sigma) that was mixed with water (1:18.75 v/v) was used directly.

The cometabolic transformation of phenolic compounds was studied in 100-mL, thick-walled flasks containing 1 g of air-dried and milled wheat straw, which was added to form a uniform layer. Each flask was supplemented with 5 mL of distilled water and sterilised by autoclaving (2×30 min at 121°C with cooling to room temperature between the two cycles). The flasks were inoculated with 1 mL of cell suspension that had been pre-grown for three days in liquid GYM media (five flasks per strain). Control flasks were left uninoculated. The following day, ^{14}C-labelled DHP, catechol and poplar lignin were added. In the DHP flasks, 750 µL of ^{14}C-DHP in a N,N-dimethylformamide-water suspension was added drop-wise onto the surface of the straw layer, which resulted in a final radioactivity of 127,500 dpm per flask. In the catechol flasks, 750 µL of ^{14}C-catechol in an ethanol-water solution was added, which resulted in a final radioactivity of 550,500 dpm per flask. In the lignin flasks, 10 mg of fine, ^{14}C-poplar lignin powder was added onto the surface of the straw layer, which resulted in a final radioactivity of 230,000 dpm per flask, and then, 750 µL of sterile water was added. The flasks were sealed with rubber septa and aluminium caps.

Incubation proceeded for 76 days at 24°C in the dark. Volatile compounds were flushed out of the flasks every week using sterile air, and CO_2 was trapped by bubbling the released air through two sequential flasks containing Opti-Fluor and Carbosorb/Opti-Fluor (Packard Instruments) every week. A liquid scintillation counter (Wallac 1411, WallacOy, Finland) was used to quantify the trapped $^{14}CO_2$ during the experiment. After incubation, the flasks supplemented with the ^{14}C-labelled substances were stored at −18°C until mass-balance extraction.

The flasks containing the residual ^{14}C material were extracted twice with 6 mL of distilled water. After the addition of water, the incubation flasks were shaken for 2 h (280 rpm) at room temperature on a table rotary shaker. The suspension was then poured into a 60-mL syringe containing a pre-weighed cotton plug, and the aqueous extract was pushed through the syringe. For measurements, 1 mL of water extract was diluted with 7 mL of distilled water, and the dilutions were mixed with 10 mL of LUMAGEL. The radioactivity of each extract was measured using a liquid scintillation counter (Model 1411, WallacOy, Finland). The cotton plugs that were used in the filtrations and the residual solid material were air dried and then burned in a combustion chamber (Junitek, Finland). The radioactivity was then counted using a liquid scintillation counter as reported previously [49]. The efficiency of combustion was verified using ^{14}C-labelled standards.

A one-way analysis of variance with the Fisher's least significant difference *post hoc* test was used to analyse the statistical significance of differences among treatments. Differences with a $P<0.05$ were regarded as statistically significant.

Identification of Actinobacterial Strains

DNA was isolated from the actinobacterial biomass that was obtained by cultivation in liquid GYM medium using the modified Miller-SK method [50]. Isolated genomic DNA was used as a template in PCR reactions using universal primers for the bacterial

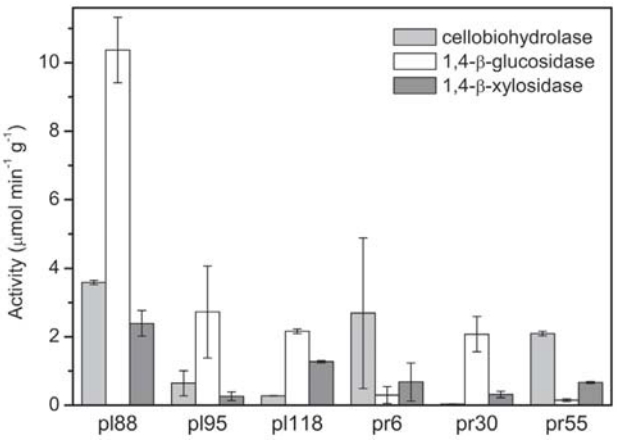


Figure 1. Production of cellobiohydrolase, β-glucosidase and β-xylosidase by *Actinobacteria*. Activity of cellobiohydrolase, 1, 4-β-glucosidase and 1,4-β-xylosidase after a 21-day cultivation of the selected actinobacterial strains on wheat straw. The data represent the means and standard errors.
doi:10.1371/journal.pone.0089108.g001

and archaeal 16 S rRNA gene, pH-T7 (5'-TAATACGACTCAC-TATAGAGAGTTTGATCCTGGCTCAG-3') and pA (5'-AAG-GAGGTGATCCAGCCGCA-3'). Each 50-µl reaction mixture contained 5 µl of 10× buffer for DyNAzyme DNA Polymerase (Finnzymes), 3 µl of purified BSA (10 mg mL^{-1}), 2 µl of each primer (0.01 mM), 1 µl of PCR Nucleotide Mix (10 mM each), 1.5 µl of DyNAZyme II DNA Polymerase (2 U μl^{-1}, Finnzymes) and 1 µl of isolated genomic DNA. The cycling conditions were as follows: 1× 94°C 5 min, 35× (94°C 1 min, 57°C for 45 s min and 72°C for 90 s) followed by 72°C for 10 min. The PCR products were directly sequenced by Macrogen (Seoul, Korea), and the sequences were manually edited using the BioEdit program (http://www.mbio.ncsu.edu/bioedit/bioedit.html) and corrected prior to a BLASTn search against the nucleotide database at the NCBI (http://www.ncbi.nlm.nih.gov/blast).

Results and Discussion

Seventy-six strains of soil *Actinobacteria* were isolated from the soils of the study area and screened for their ability to produce enzymes involved in cellulose decomposition, including cellobiohydrolase and 1,4-β-glucosidase. Of these strains, 32% did not produce any of the tested enzymes, while 31% produced both of them. The production of 1,4-β-glucosidase was more common (57% of strains) than that of cellobiohydrolase (41% of strains; Table 1). The percentage of strains that produced cellobiohydrolase roughly corresponded to the percentage of actinobacterial strains that harboured a gene for endocellulase or cellobiohydrolase (i.e., the glycosyl hydrolase family GH5, 6, 8, 9, 12, 44, 45 or 48), which was one-third of all of the sequenced actinobacterial genomes that were analysed in a recent study [10]. The percentage of strains that did not produce detectable amounts of any enzyme (32%) was higher than what was inferred from the analysis of the genomes, which was less than 20% of the genomes [10]. It is thus possible that some of the strains that harbour 1,4-β-glucosidase do not express the gene or show only low expression levels.

Fourteen strains that produced both enzymes, and highly produced at least one, were selected for further studies and were identified by 16S rRNA sequencing. Of these, eight strains showed highest similarity with members of the genus *Streptomyces,* while the best hits for the others were from the genera *Amycolatopsis*, *Curtobacterium*, *Kribbella*, *Microbispora*, *Micromonospora* and *Nocardia* (Table 2). The activity of cellobiohydrolase and 1,4-β-glucosidase in the genera *Amycolatopsis*, *Kribella*, *Micromonospora*, *Nocardia* and *Streptomyces* corresponded well with the presence of the corresponding genes in their genomes [10]. The currently analysed genomes of *Nocardia* did not contain a cellobiohydrolase gene, and the genomes of *Curtobacterium* and *Microbispora* have not been analysed. Despite their high cellulolytic activity, only six of the fourteen analysed isolates (pl88, pl95, pl118, pr6, pr30 and pr55) showed visually detectable growth on milled wheat straw after 21 days of culturing. All of these strains produced extracellular glycosyl hydrolases: cellobiohydrolase, 1,4-β-glucosidase and 1,4-

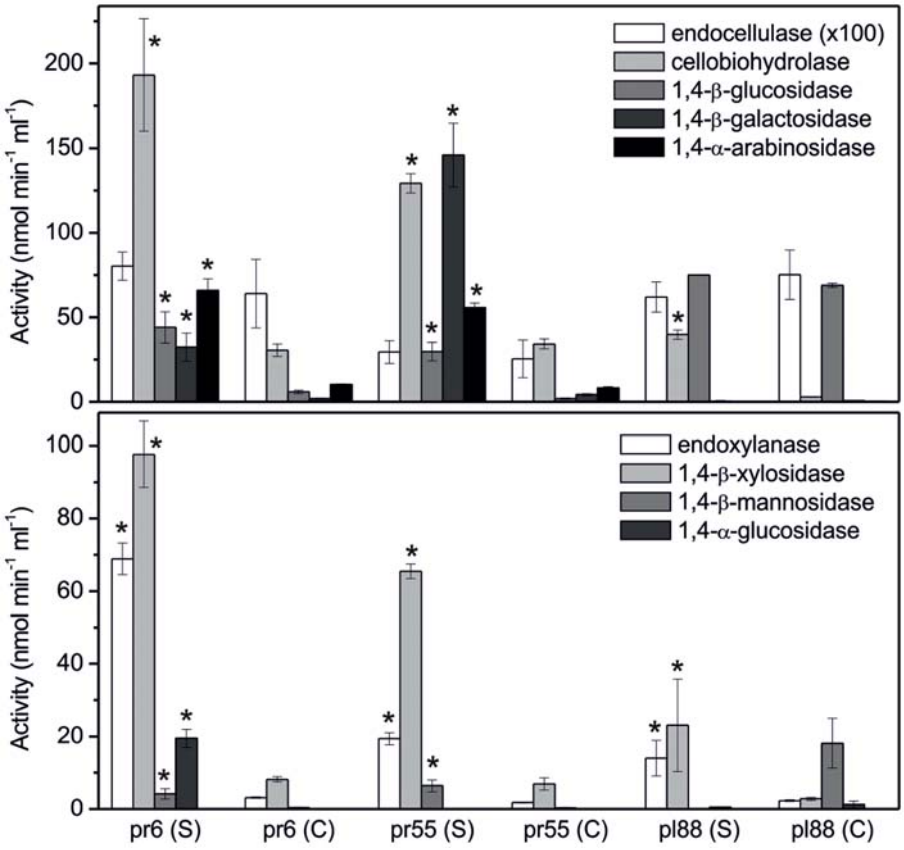


Figure 2. Production of hydrolytic enzymes by selected *Actinobacteria*. Activity of glycosyl hydrolases after a 21-day cultivation of the selected actinobacterial strains on wheat straw (S) and cellulose (C). The data represent the means and standard errors. The activity of endocellulase was multiplied 100× to fit the same scale. Asterisks indicate significant difference ($P<0.05$) in enzyme activity among treatments.
doi:10.1371/journal.pone.0089108.g002

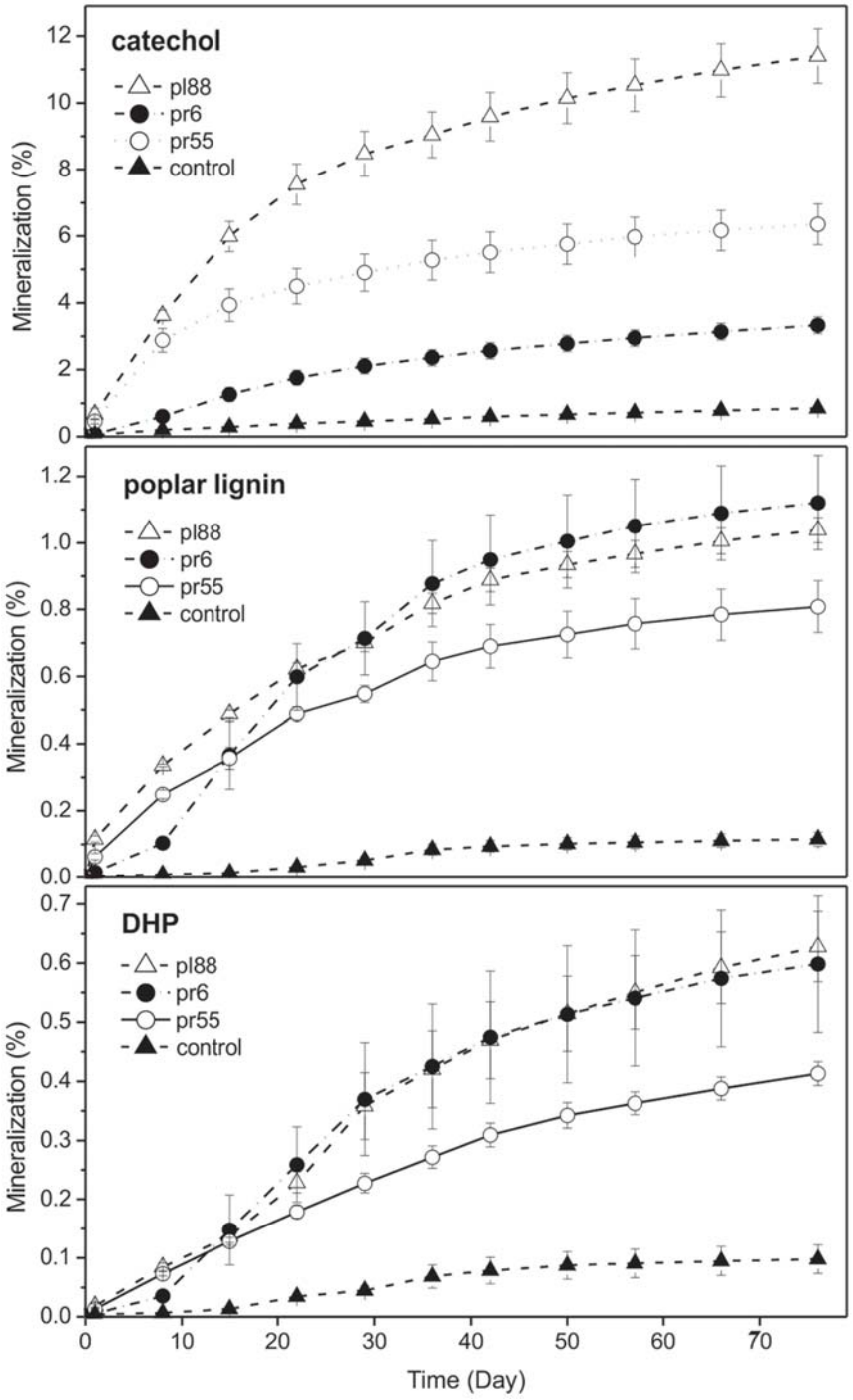


Figure 3. Mineralization of phenolic compounds by selected *Actinobacteria*. Time course of $^{14}CO_2$ production during the transformation of ^{14}C-catechol, ^{14}C-poplar lignin and ^{14}C-DHP in wheat straw microcosms by the selected actinobacterial strains. Control treatments contained sterile straw. The data represent the means and standard errors.
doi:10.1371/journal.pone.0089108.g003

β-xylosidase, although their activities differed (Figure 1). The strains that highly produced cellobiohydrolase (>2 μmol min^{-1} g^{-1} straw dry mass) were selected for further experiments. Cellobiohydrolase represents the rate-limiting enzyme in the decomposition of cellulose, the most abundant and rapidly decomposable polysaccharide in plant litter [12,51].

All studied strains produced a complete set of cellulolytic enzymes: endocellulase, cellobiohydrolase and 1,4-β-glucosidase, and all strains produced the xylanolytic enzymes endoxylanase and 1,4-β-xylosidase as well as the hemicellulases 1,4-β-galactosidase and 1,4-β-mannosidase. 1,4-α-Arabinosidase was only produced by pr6 and pr55, while amylase was produced by pr6

Table 3. Mass balance of ^{14}C-catechol, ^{14}C-poplar lignin and ^{14}C-DHP lignin after a 76-day incubation in wheat straw microcosms with the selected actinobacterial strains.

Substrate	Strain	Respired		Soluble		Bound	
catechol	pl88	11.64±0.85	a	65.2±0.7	c	23.2±0.6	c
	pr6	3.33±0.29	b	70.6±1.1	b	26.1±0.8	b
	pr55	6.36±0.60	c	74.4±0.5	a	19.2±1.1	d
	control	0.94±0.02	d	52.2±0.1	d	46.8±0.1	a
poplar lignin	pl88	1.04±0.03	a	2.7±0.1	c	96.2±0.1	b
	pr6	1.12±0.13	a	4.0±0.2	a	94.9±0.2	c
	pr55	0.81±0.07	b	3.4±0.2	b	95.8±0.2	b
	control	0.09±0.04	c	2.0±0.2	d	97.9±0.3	a
DHP	pl88	0.67±0.06	a	21.7±0.4	a	77.7±0.4	b
	pr6	0.64±0.11	a	22.4±0.5	a	76.9±0.6	b
	pr55	0.45±0.02	b	23.1±0.5	a	76.4±0.6	b
	control	0.11±0.03	c	18.9±1.1	b	81.0±1.1	a

The data (% of the total) represent the means and standard errors. Different letters indicate statistically significant differences at $P<0.05$.
doi:10.1371/journal.pone.0089108.t003

and pl88. The bacteria produced the same enzymes, regardless of whether cellulose or milled wheat straw was used as the carbon source (Figure 2). However, wheat straw, which is a complex substrate that contains various polysaccharides, increased the production of most hemicellulases and cellobiohydrolase and, in the case of pr6 and pr55, the production of 1,4-β-glucosidase. The titres of exocellulase were similar in both treatments. Remarkably, the production of endocellulase was 10–100× lower than that of endoxylanase, and the activities of 1,4-β-xylosidase on straw were higher than that of 1,4-β-glucosidase. Unfortunately, the production of xylanase was not studied in *Actinobacteria* in sufficient detail; however, studies on environmental isolates indicated that this ability might be relatively common [52]. When the genes of the most common endoxylanase, GH10, were screened across a range of different soils, actinobacterial genes were recovered with a relatively high frequency [53]. The activity of 1,4-β-arabinosidase was higher than that of 1,4-β-glucosidase in the pr6 and pr55 strains. It is thus possible that hemicellulose, and particularly xylan, represents a preferred source for lignocellulose-degrading *Actinobacteria*. Although cellulose decomposition was found to be more rapid in decomposing plant litter than that of hemicelluloses, the utilisation of xylose and arabinose-containing hemicelluloses was relatively fast [51]. Although the production of 1,4-β-mannosidase, 1,4-β-galactosidase and 1,4-α-arabinosidase by individual taxa of *Actinobacteria* was reported previously [54–56], we show that they belong to a set of enzymes that are simultaneously produced by saprotrophic taxa during their growth on lignocellulose. The enzyme 1,4-β-glucuronidase, which is involved in the degradation of pectins, was not produced by the tested strains.

To assess whether bacteria that utilise polysaccharides are also able to cometabolise, that is, transform or even mineralise the phenolic compounds within lignocellulose, their transformation of ^{14}C-labelled phenolic compounds was studied. Transformation was studied using compounds of various complexities that included the monomeric catechol (the precursor of lignin), the nonspecifically labelled poplar lignin that was isolated from plant tissues and $^{14}C_{\beta}$-labelled DHP, which is a specifically labelled lignin model compound. After 76 days of cultivation of the bacteria on wheat straw supplemented with the ^{14}C-labelled compounds, considerable mineralisation was only found for catechol: 11.4% for strain pl88, 6.4% for pr55 and 3.33% for pr6 (Figure 3). Mineralisation of poplar lignin was higher in pr6 and pl88 (>1%) than in pr55 (0.81%), and the same result was found for DHP, where mineralisation by pr55 was also lower than that of the other two strains. In all cases, inoculation with *Actinobacteria* resulted in significantly higher mineralisation than that of control flasks. The mass balance extraction of wheat straw showed that, in addition to the mineralisation of phenolic compounds, *Actinobacteria* increased the relative share of water-soluble phenolics, and this was most apparent for catechol, where insoluble ^{14}C compounds represented 47% of the total in the control, while they were 19–26% of the control after incubation with bacteria. The amounts of soluble ^{14}C in the DHP- and lignin-containing flasks also increased after bacterial treatment, but only slightly, by 2–4% and 0.7–2.0%, respectively (Table 3).

The mineralisation of lignin by *Streptomycetes* was first reported by Crawford, who found that mineralisation rates of ^{14}C-lignin-labelled fir varied from 1.5 to 3% after nearly 42 days [57]. Similar rates were also reported by Pasti et al. [58]. Occasionally, higher lignin mineralisation rates were reported, such as 2.9% after 10 days by *Arthrobacter* sp. [59] or up to 5% after 15 days by *Nocardia* sp. [38]. In comparison, the mineralisation rates that were observed in our isolates seemed to be relatively low, but care must be taken in these comparisons because the modes of lignin labelling and preparation greatly affect mineralisation rates. The mineralisation rates of the recalcitrant lignin model compound ^{14}C-DHP that were reported in this study indicates that the ability of the studied *Actinobacteria* to catalyse the complete decomposition of lignin is low and that cometabolic lignin degradation during the growth of active decomposers on lignocellulose is rather negligible.

Although solubilisation is the first important step in lignin decomposition because it makes the polar lignin residues available to other microorganisms, only a few studies have reported the amount of solubilised lignin. Our results indicate only a limited solubilisation of lignin, and these data are comparable to those published for several *Streptomycetes* (6–10%) [58].

Catechol is the first intermediate product of phenol degradation, and its cleavage is a critical step in the aerobic degradation of aromatic compounds in microorganisms [60]. Several bacterial strains that are capable of catechol degradation belong to various phyla, such as *Actinobacteria*, *Alphaproteobacteria*, *Betaproteobacteria* or *Gammaproteobacteria*. Catechol degradation was investigated mainly in studies that focused on the treatment of phenolic wastes, and efficient degraders were found in the genera *Pseudomonas*, *Acinetobacter* and *Klebsiella* [61]. All of the isolates in our study were able to efficiently degrade or solubilise catechol. Although the importance of phenolics as a growth substrate for bacteria is not clear, low molecular mass phenolic compounds are relatively common in litter [62] and their transformation might be of importance.

The results confirm the potential importance of *Actinobacteria* in lignocellulose degradation, although it is likely that the decomposition of biopolymers is limited to strains that represent only a minor portion of the entire community. Our results indicate that strains that are capable of growth on a complex lignocellulose substrate exist; these taxa are able to decompose cellulose and hemicelluloses, and at least some of them may prefer the latter. The importance of cometabolic degradation of lignin seems to be limited when compared with the abilities of saprotrophic fungi, but *Actinobacteria* may still contribute to the solubilisation of phenolics, especially low-molecular-mass compounds. Future studies using genomic, transcriptomic or proteomic approaches are needed to

explore the link between the genetic potential of *Actinobacteria* and their actual activities as decomposers of organic matter.

Acknowledgments

Markéta Marečková and Jan Kopecký are acknowledged for their help with the isolation of *Actinobacteria*.

Author Contributions

Conceived and designed the experiments: TV PB. Performed the experiments: TV KS. Analyzed the data: TV KS PB. Wrote the paper: TV PB.

References

1. Baldrian P, Šnajdr J (2011) Lignocellulose-Degrading Enzymes in Soils. In: Shukla G, Varma A, editors. Soil Enzymology. Berlin: Springer-Verlag Berlin. p.167–186.
2. de Boer W, Folman LB, Summerbell RC, Boddy L (2005) Living in a fungal world: impact of fungi on soil bacterial niche development. FEMS Microbiol Rev 29: 795–811.
3. Baldrian P (2008) Enzymes of Saprotrophic Basidiomycetes. In: Boddy L, Frankland JC, VanWest P, editors. Ecology of Saprotrophic Basidiomycetes. San Diego: Elsevier Academic Press Inc. p.19–41.
4. Eastwood DC, Floudas D, Binder M, Majcherczyk A, Schneider P, et al. (2011) The Plant Cell Wall–Decomposing Machinery Underlies the Functional Diversity of Forest Fungi. Science 333: 762–765.
5. Floudas D, Binder M, Riley R, Barry K, Blanchette RA, et al. (2012) The paleozoic origin of enzymatic lignin decomposition reconstructed from 31 fungal genomes. Science 336: 1715–1719.
6. Sjöström E (1993) Wood chemistry: Fundamentals and applications. 2nd edition ed: Academic Press (San Diego).
7. Martinez AT, Speranza M, Ruiz-Duenas FJ, Ferreira P, Camarero S, et al. (2005) Biodegradation of lignocellulosics: microbial chemical, and enzymatic aspects of the fungal attack of lignin. Int Microbiol 8: 195–204.
8. Theuerl S, Buscot F (2010) Laccases: toward disentangling their diversity and functions in relation to soil organic matter cycling. Biol Fertil Soils 46: 215–225.
9. Baldrian P, Voříšková J, Dobiášová P, Merhautová V, Lisá L, et al. (2011) Production of extracellular enzymes and degradation of biopolymers by saprotrophic microfungi from the upper layers of forest soil. Plant Soil 338: 111–125.
10. Berlemont R, Martiny AC (2013) Phylogenetic Distribution of Potential Cellulases in Bacteria. Appl Environ Microbiol 79: 1545–1554.
11. Lynd LR, Weimer PJ, van Zyl WH, Pretorius IS (2002) Microbial cellulose utilization: Fundamentals and biotechnology. Microbiol Mol Biol Rev 66: 506–577.
12. Štursová M, Žifčáková L, Leigh MB, Burgess R, Baldrian P (2012) Cellulose utilization in forest litter and soil: identification of bacterial and fungal decomposers. FEMS Microbiol Ecol 80: 735–746.
13. Chater KF, Biro S, Lee KJ, Palmer T, Schrempf H (2010) The complex extracellular biology of *Streptomyces*. FEMS Microbiol Rev 34: 171–198.
14. Warren RAJ (1996) Microbial hydrolysis of polysaccharides. Annu Rev Microbiol 50: 183–212.
15. McCarthy AJ (1987) Lignocellulose-degrading actinomycetes. FEMS Microbiol Rev 46: 145–163.
16. Enkhbaatar B, Temuujin U, Lim JH, Chi WJ, Chang YK, et al. (2012) Identification and Characterization of a Xyloglucan-Specific Family 74 Glycosyl Hydrolase from *Streptomyces coelicolor* A3(2). Appl Environ Microbiol 78: 607–611.
17. Sun Y, Cheng JJ, Himmel ME, Skory CD, Adney WS, et al. (2007) Expression and characterization of *Acidothermus cellulolyticus* E1 endoglucanase in transgenic duckweed *Lemna minor* 8627. Biores Technol 98: 2866–2872.
18. Yin LJ, Huang PS, Lin HH (2010) Isolation of Cellulase-Producing Bacteria and Characterization of the Cellulase from the Isolated Bacterium *Cellulomonas* Sp YJ5. J Agric Food Chem 58: 9833–9837.
19. Song JM, Wei DZ (2010) Production and characterization of cellulases and xylanases of *Cellulosimicrobium cellulans* grown in pretreated and extracted bagasse and minimal nutrient medium M9. Biomass Bioenergy 34: 1930–1934.
20. Yoon MH, Choi WY (2007) Characterization and action patterns of two beta-1,4-glucanases purified from *Cellulomonas uda* CS1–1. J Microbiol Biotechnol 17: 1291–1299.
21. Zhang F, Hu SN, Chen JJ, Lin LB, Wei YL, et al. (2012) Purification and partial characterisation of a thermostable xylanase from salt-tolerant *Thermobifida halotolerans* YIM 90462(T). Process Biochem 47: 225–228.
22. An DS, Cui CH, Lee HG, Wang L, Kim SC, et al. (2010) Identification and Characterization of a Novel *Terrabacter ginsenosidimutans* sp nov beta-Glucosidase That Transforms Ginsenoside Rb1 into the Rare Gypenosides XVII and LXXV. Appl Environ Microbiol 76: 5827–5836.
23. Fan HX, Miao LL, Liu Y, Liu HC, Liu ZP (2011) Gene cloning and characterization of a cold-adapted beta-glucosidase belonging to glycosyl hydrolase family 1 from a psychrotolerant bacterium *Micrococcus antarcticus*. Enzyme Microb Technol 49: 94–99.
24. Nakano H, Okamoto K, Yatake T, Kiso T, Kitahata S (1998) Purification and characterization of a novel beta-glucosidase from *Clavibacter michiganense* that hydrolyzes glucosyl ester linkage in steviol glycosides. J Ferment Bioeng 85: 162–168.
25. Nunoura N, Ohdan K, Yano T, Yamamoto K, Kumagai H (1996) Purification and characterization of beta-D-glucosidase (beta-D-fucosidase) from *Bifidobacterium breve* clb acclimated to cellobiose. Biosci Biotechnol Biochem 60: 188–193.
26. Quan LH, Min JW, Jin Y, Wang C, Kim YJ, et al. (2012) Enzymatic Biotransformation of Ginsenoside Rb1 to Compound K by Recombinant beta-Glucosidase from *Microbacterium esteraromaticum*. J Agric Food Chem 60: 3776–3781.
27. Anderson I, Abt B, Lykidis A, Klenk HP, Kyrpides N, et al. (2012) Genomics of Aerobic Cellulose Utilization Systems in *Actinobacteria*. PLOS One 7: e39331.
28. Khanna S, Gauri (1993) Regulation, purification, and properties of xylanase from *Cellulomonas fimi*. Enzyme Microb Technol 15: 990–995.
29. Li CJ, Hong YZ, Shao ZZ, Lin L, Huang XL, et al. (2009) Novel Alkali-Stable, Cellulase-Free Xylanase from Deep-Sea *Kocuria* sp Mn22. J Microbiol Biotechnol 19: 873–880.
30. Oh HW, Heo SY, Kim DY, Park DS, Bae KS, et al. (2008) Biochemical characterization and sequence analysis of a xylanase produced by an exo-symbiotic bacterium of *Gryllotalpa orientalis*, *Cellulosimicrobium* sp HY-12. Ant Leeuw Int J Gen Mol Microbiol 93: 437–442.
31. Petrosyan P, Luz-Madrigal A, Huitrón C, Flores M (2002) Characterization of a xylanolytic complex from *Streptomyces* sp. Biotechnol Lett 24: 1473–1476.
32. Robinson LE, Crawford RL (1978) Degradation of ^{14}C-labelled lignins by *Bacillus megaterium*. FEMS Microbiol Lett 4: 301–302.
33. Vicuna R (1988) Bacterial degradation of lignin. Enzyme Microb Technol 10: 646–655.
34. Zimmermann W (1990) Degradation of lignin by bacteria. J Biotechnol 13: 119–130.
35. Cartwrig N, Holdom KS (1973) Enzymic lignin, its release and utilization by bacteria. Microbios 8: 7–14.
36. Crawford DL, Barder MJ, Pometto AL, Crawford RL (1982) Chemistry of softwood lignin degradation by *Streptomyces viridosporus*. Arch Microbiol 131: 140–145.
37. Sutherland JB, Blanchette RA, Crawford DL, Pometto AL (1979) Breakdown of Douglas fir phloem by a lignocellulose-degrading *Streptomyces*. Curr Microbiol 2: 123–126.
38. Trojanowski J, Haider K, Sundman V (1977) Decomposition of ^{14}C-labelled lignin and phenols by a *Nocardia* sp. Arch Microbiol 114: 149–153.
39. Caliz J, Montserrat G, Marti E, Sierra J, Chung AP, et al. (2013) Emerging resistant microbiota from an acidic soil exposed to toxicity of Cr, Cd and Pb is mainly influenced by the bioavailability of these metals. J Soils Sediments 13: 413–428.
40. Harichova J, Karelova E, Pangallo D, Ferianc P (2012) Structure analysis of bacterial community and their heavy-metal resistance determinants in the heavy-metal-contaminated soil sample. Biologia 67: 1038–1048.
41. Mühlbachová G (2011) Soil microbial activities and heavy metal mobility in long-term contaminated soils after addition of EDTA and EDDS. Ecol Eng 37: 1064–1071.
42. Hayakawa M, Nonomura H (1987) Humic acid-vitamin agar, a new medium for the selective isolation of soil actinomycetes. J Ferment Technol 65: 501–509.
43. Hayakawa M, Momose Y, Yamazaki T, Nonomura H (1996) A method for the selective isolation of *Microtetraspora glauca* and related four-spored actinomycetes from soil. J Appl Bacteriol 80: 375–386.
44. Baldrian P (2009) Microbial enzyme-catalyzed processes in soils and their analysis. Plant Soil Environ 55: 370–378.
45. Valášková V, Šnajdr J, Bittner B, Cajthaml T, Merhautová V, et al. (2007) Production of lignocellulose-degrading enzymes and degradation of leaf litter by saprotrophic basidiomycetes isolated from a *Quercus petraea* forest. Soil Biol Biochem 39: 2651–2660.
46. Brunow G, Ammalahti E, Niemi T, Sipila J, Simola LK, et al. (1998) Labelling of a lignin from suspension cultures of *Picea abies*. Phytochem 47: 1495–1500.
47. Tuomela M, Oivanen P, Hatakka A (2002) Degradation of synthetic ^{14}C-lignin by various white-rot fungi in soil. Soil Biol Biochem 34: 1613–1620.
48. Odier E, Janin G, Monties B (1981) Poplar Lignin Decomposition by Gram-Negative Aerobic Bacteria. Appl Environ Microbiol 41: 337–341.
49. Šnajdr J, Steffen KT, Hofrichter M, Baldrian P (2010) Transformation of ^{14}C-labelled lignin and humic substances in forest soil by the saprobic basidiomycetes *Gymnopus erythropus* and *Hypholoma fasciculare*. Soil Biol Biochem 42: 1541–1548.
50. Sagova-Mareckova M, Cermak L, Novotna J, Plhackova K, Forstova J, et al. (2008) Innovative methods for soil DNA purification tested in soils with widely differing characteristics. Appl Environ Microbiol 74: 2902–2907.
51. Šnajdr J, Cajthaml T, Valášková V, Merhautová V, Petránková M, et al. (2011) Transformation of *Quercus petraea* litter: successive changes in litter chemistry are

reflected in differential enzyme activity and changes in the microbial community composition. FEMS Microbiol Ecol 75: 291–303.
52. She YL, Li XT, Sun BG, Lv YG, Song HX (2012) Screening of Actinomycetes with High Producing Xylanase. Adv Mat Res 365: 332–337.
53. Wang G, Meng K, Luo H, Wang Y, Huang H, et al. (2012) Phylogenetic Diversity and Environment-Specific Distributions of Glycosyl Hydrolase Family 10 Xylanases in Geographically Distant Soils. PLOS One 7: e43480.
54. Post DA, Luebke VE (2005) Purification, cloning, and properties of beta-galactosidase from *Saccharopolyspora erythraea* and its use as a reporter system. Appl Microbiol Biotechnol 67: 91–96.
55. Shi P, Yao G, Cao Y, Yang P, Yuan T, et al. (2011) Cloning and characterization of a new β-mannosidase from *Streptomyces* sp. S27. Enzyme Microb Technol 49: 277–283.
56. Tajana E, Fiechter A, Zimmermann W (1992) Purification and characerization of two alpha-L-arabinofuranosidases from *Streptomyces diastaticus*. Appl Environ Microbiol 58: 1447–1450.
57. Crawford DL (1978) Lignocellulose decomposition by selected streptomyces strains. Appl Environ Microbiol 35: 1041–1045.
58. Pasti MB, Pometto AL, Nuti MP, Crawford DL. Lignin-solubilizing ability of actinomycetes isolated from termite (*Termitidae*) gut. Appl Environ Microbiol 56: 2213–2218.
59. Kerr TJ, Kerr RD, Benner R (1983) Isolation of a Bacterium Capable of Degrading Peanut Hull Lignin. Appl Environ Microbiol 46: 1201–1206.
60. Krastanov A, Alexieva Z, Yemendzhiev H (2013) Microbial degradation of phenol and phenolic derivatives. Eng Life Sci 13: 76–87.
61. El Azhari N, Devers-Lamrani M, Chatagnier G, Rouard N, Martin-Laurent F (2010) Molecular analysis of the catechol-degrading bacterial community in a coal wasteland heavily contaminated with PAHs. J Hazard Mater 177: 593–601.
62. Osono T, Takeda H (2005) Decomposition of organic chemical components in relation to nitrogen dynamics in leaf litter of 14 tree species in a cool temperate forest. Ecol Res 20: 41–49.

Biol Fertil Soils (2013) 49:1027–1037
DOI 10.1007/s00374-013-0801-y

ORIGINAL PAPER

Analysis of soil fungal communities by amplicon pyrosequencing: current approaches to data analysis and the introduction of the pipeline SEED

Tomáš Větrovský · Petr Baldrian

Received: 1 December 2012 / Revised: 11 March 2013 / Accepted: 25 March 2013 / Published online: 16 April 2013

Abstract Fungi are important in soils as both decomposers and plant symbionts, and an understanding of the composition of their complex communities is thus indispensable to answer a variety of ecological questions. 454 Pyrosequencing is currently the method of choice for the in-depth analysis of fungal communities. However, the interpretation of its results is complicated by differences in data analysis approaches that make inter-study comparisons difficult. The pyrosequencing studies published so far have also used variable molecular targets in fungal rDNA. Although the ITS region and, in particular, ITS1 appear to be the most frequent sequencing targets, the use of various primers with different coverages of fungal groups remains a serious problem. Sequence length limits also vary widely across studies, and in many studies, length differences may negatively affect sequence similarity clustering or identification. Unfortunately, many studies neglect the need to correct for method-dependent errors, such as pyrosequencing noise or chimeric sequences. Even when performed, error rates in sequences may be high, and consensus sequences created by sequence clustering therefore better represent operational taxonomic units. We recommend a data analysis workflow that includes sequence denoising, chimera removal, sequence trimming before clustering and random resampling before calculating diversity parameters. The newly developed free pipeline (SEED) introduced here can be used to perform all the required analytical steps. The improvement and unification of data analysis procedures should make future studies both more reliable and comparable and allow meta-studies to be performed to provide more general views on fungal diversity, biogeography or ecology.

T. Větrovský · P. Baldrian (✉)
Laboratory of Environmental Microbiology,
Institute of Microbiology of the ASCR, v.v.i., Vídeňská 1083,
14220, Prague 4, Czech Republic
e-mail: baldrian@biomed.cas.cz

Keywords Fungal community · Internal transcribed spacer · Pyrosequencing pipeline · Ribosomal DNA · Sequencing errors · Soil microbial ecology

Introduction

Fungi are important in soils as both decomposers and plant symbionts. Traditional surveys based on macroscopic or microscopic features, such as fruit body surveys, microscopy of plant roots or isolation techniques, despite considerable progress, have been insufficient to describe fungal communities inhabiting soil environments. Molecular methods have recently greatly overcome these limitations to allow detection of unculturable community members. Since its first applications in 2009 (Buéé et al. 2009; Jumpponen and Jones 2009), amplicon pyrosequencing studies have focused on the diversity of fungal communities (e.g. Buéé et al. 2009; Jumpponen and Jones 2009; Öpik et al. 2009), the activity of fungal communities (Baldrian et al. 2012; Štursová et al. 2012) or functional genes (e.g. Baldrian et al. 2013; Voříšková and Baldrian 2013) of both total fungi and specific groups like the Glomeromycota (e.g. Dumbrell et al. 2011; Lekberg et al. 2012; Öpik et al. 2009). Pyrosequencing has become the method of choice for the in-depth analysis of fungal community composition.

Data accumulate with increasing numbers of studies, but the experimental approaches for data collection and analysis widely differ. This unfortunately greatly limits our ability to compare among studies and draw general conclusions regarding important questions such as estimating community diversity, evenness and composition or identifying important taxa. Data analysis appears to be an important area where further improvements and unification of experimental procedures are necessary. Past experience derived from published studies indicates which steps are

important and should be considered when designing data analysis workflows.

Most importantly, the complexity of the data and the specifics of the methods may cause several biases that affect the quality of the resulting sequence dataset and any subsequent statistical analyses or ecological considerations. These include pyrosequencing-specific errors, sometimes termed "sequencing noise" (Quince et al. 2009, 2011), unclear quality of low abundance sequences, the presence of chimeric sequences (Taylor and Houston 2011; Tedersoo et al. 2010) and PCR target-associated biases (Bellemain et al. 2010; Krüger et al. 2012).

Despite the development of alternative sequencing platforms (Shokralla et al. 2012), pyrosequencing will likely remain widely used in the near future. For this reason, we believe that the standardisation of methods for fungal community analysis is highly desirable because it will soon allow us to exploit the wealth of individual studies to deliver general statements regarding fungal diversity, biogeography or ecology.

Standards of data reporting that include information regarding the sampling site and its corresponding metadata, laboratory processing steps and data analysis were previously suggested (Nilsson et al. 2011). The aim of this paper was to describe the data analysis procedures previously used, indicate the limiting steps and suggest a simple data analysis workflow that can avoid potential problems. Because the processing of large-scale pyrosequencing-derived data may represent a methodological limitation, a newly developed software pipeline, SEED, is introduced in this paper that allows researchers to perform the required data analysis steps with a single, easy-to-use user interface.

Materials and methods

Meta-analysis of studies using amplicon pyrosequencing to explore fungal communities

Scientific publications using amplicon pyrosequencing to analyse fungal communities were retrieved. The source of sample material, molecular target (gene and primer pair) and number of sequences that were used for community analysis were recorded. With respect to the experimental methodology used for sequence processing, the minimum sequence length and the presence or absence of sequence processing steps (removal of pyrosequencing noise, removal of chimeric sequences, creation of similarity clusters, diversity analysis and sequence annotation) were recorded. The bioinformatic tools used for data cleanup, sequence clustering and annotation were also recorded (Table 1). The data retrieved from publications were used to analyse the approaches used in fungal community amplicon pyrosequencing.

Development of pipeline to analyse sequences obtained by amplicon pyrosequencing

Based on the previously applied approaches to amplicon pyrosequencing data analysis, the necessary steps were identified and the analysis workflow was proposed. The development of the optimized workflow was based on both the available knowledge from previous papers about the effects of certain data analysis steps on the resulting dataset quality (Schloss et al. 2009, Edgar et al. 2011) and on our own analysis of a sample dataset. For this purpose, the publicly available dataset deposited in MG Rast 4497081.3 that contains sequences of fungal internal transcribed spacer (ITS) region from oak leaves at different stages of decomposition (Voříšková and Baldrian 2013) was used. The aim was to analyse the effects of certain data analysis steps on the fungal diversity estimates and identification of operational taxonomic units (OTUs; defined as sequences clustered at a 97 % probability level). Specifically, we analysed (1) for each sample (n=21, 1129 sequences per sample) the effects of clustering sequences of original length (380–560 bases) versus sequences truncated to the same length on OTU richness, the Chao estimate and the number of singletons; (2) for each sample the effects of chimera removal on OTU richness, Chao and singletons; and (3) for the 150 most abundant OTUs, the quality of OTU identification was compared with OTUs represented either by random sequences or consensus sequence. The quality of identification was defined as the similarity of the query sequence and the most similar Sanger sequencing-derived sequence deposited in GenBank (http://www.ncbi.nlm.nih.gov/genbank/). This is based on the assumption that sequences containing errors are less similar to real sequences and that consensus construction should correct random errors in sequences. Sequence clustering and chimera removal were performed using default Usearch and Uchime settings (Edgar 2010, Edgar et al. 2011), and nucleotide BLAST (Altschul et al. 1990) was used to retrieve the closest hits from GenBank. Wilcoxon pair test was used to analyse the differences among dataset pairs. Differences at $P<0.01$ were regarded as statistically significant. The optimized data analysis workflow was used in the course of the development of a user-friendly data analysis pipeline.

The pipeline, SEED (http://www.biomed.cas.cz/mbu/lbwrf/seed/main.html), was created that enables users to perform the entire bioinformatic analysis of PCR amplicons according to the suggested workflow. The same pipeline was also used to perform all workflow testing steps outlined above. The functionality of the pipeline was tested with datasets from previous pyrosequencing projects with amplicon sequences for the fungal ITS region, bacterial 16S rDNA and the fungal *cbhI* exocellulase gene (Baldrian et al. 2012; Štursová et al. 2012; Větrovský and Baldrian 2013;

Table 1 Overview of studies using amplicon-based 454 pyrosequencing to analyse fungal communities along with important parameters of data processing

Sample[a]	Target	Primer pair(s)	Length[b]	Denoising	Chimera removal	Data cleanup[c]	Clustering[d]	Percentage	Rep. sequence[e]	Singleton removal	Diversity[f]	Annotation[g]	Size[h]	Reference
Soil	ITS	ITS1f/ITS2	100			EMBOSS	BLASTCLUST	97	R		YES	MEGAN	OOO	Buéé et al. (2009)
Plant	ITS	ITS1f/ITS2	200				CAP3	95	R	YES		BLASTN	OO	Jumpponen and Jones (2009)
Roots (G)	18S	NS31/AM1	160								YES	BLASTN	OOO	Öpik et al. (2009)
Sediment	ITS, 28S	ITS1f/ITS4, LRORf/LR5f	300			RDP	CD-HIT	97				MEGAN	OOO	Amend et al. (2010b)
Soil	ITS	ITS1f/ITS4	300			CLC Gen. W.	CD-HIT	97			YES (R)	MEGAN	O	Amend et al. (2010a)
Roots	ITS	ITS1f/ITS2	200				CAP3	95			YES	BLASTN	OO	Jumpponen et al. (2010b)
Plant	ITS	ITS1f/ITS4	200				CAP3	95		YES		BLASTN	OO	Jumpponen and Jones (2010)
Soil	ITS	ITS5/ITS4	200				CAP3	95		YES		BLASTN	OO	Jumpponen et al. (2010a)
Soil (G)	18S	AMV4.5NF /AMDGR, NS31/ Ammix	230					97					O	Lumini et al. (2010)
Plant	ITS	ITS1f/ITS2	150									BLASTN	OO	Ovaskainen et al. (2010)
Roots	ITS	ITS5/ITS2	140		YES	NEWBLER	Other	97	L	YES		BLASTN	OO	Tedersoo et al. (2010)
Soil	ITS	ITS1f/ITS4	200			SCATA	SCATA	98.5	C			SCATA	OO	Wallander et al. (2010)
Roots (G)	18S	WANDA/AM1	100									BLASTN	OOO	Dumbrell et al. (2011)
Soil, roots	28S	LROR/LR3	150			MOTHUR	Other	95		YES	YES	BLASTN	OOO	Gottel et al. (2011)
Soil	ITS	ITS1f/ITS2	200				CAP3	95	R	YES	YES	BLASTN	O	Hui et al. (2011)
Soil	ITS	ITS5/ITS2	196				Other	98			YES		OOO	Lentendu et al. (2011)
Soil	ITS	ITS1f/ITS2, ITS3/ITS4	60			Sequencher	Other	97			YES	BLASTN	OO	Mello et al. (2011)
Roots (G)	18S	NS31/AM1	160								YES	BLASTN	OOO	Moora et al. (2011)
Soil	ITS	ITS1/58A2R	200			CLOTU	BLASTCLUST	97		YES	YES	BLASTN	O	Xu et al. (2011)
Coral reef	28S	LROR/LR5	250	YES	YES	MOTHUR	Other	97		YES		MEGAN	OO	Amend et al. (2012)
Plant	ITS, 18S	ITS1f/ITS4, SSU817f/ SSU1536r	200		YES	PANGEA	CD-HIT	98	L	YES	YES	MEGAN	OOO	Arfi et al. (2012a)
Soil	ITS	ITS1f/ITS4	200		YES		CD-HIT	97–98	L	YES	YES		OO	Arfi et al. (2012b)
Soil	ITS	ITS1/ITS4	380	YES	YES	MOTHUR	CD-HIT (S)	97	C		YES (R)	PlutoF	OO	Baldrian et al. (2012)
Soil	18S	SSU817f/ SSU1536r	195				CAP3	97	R	YES	YES	BLASTN	?	Becklin et al. (2012)
Sediment	18S	SSU_F04/ SSU_R22, NF1/18Sr2b	200	YES	YES	QIIME	UCLUST	95–99	C			BLASTN	OOO	Bik et al. (2012)
Roots	ITS	ITS5/ITS2	150			CLOTU	BLASTCLUST	98.5		YES		CLOTU	OOO	Blaalid et al. (2012)
Plant	ITS	ITS1f/ITS2	100			RDP	UCLUST	97	A			BLASTN	OO	Cordier et al. (2012)
Plant, soil	ITS	ITS3/ITS4	200	YES		QIIME	UCLUST	97	A	YES		BLASTN	OOO	Davey et al. (2012)
Soil (G)	18S	NS31/AML2	170		YES					YES	YES	BLASTN	OOO	Davison et al. (2012)
Soil	ITS	ITS3/ITS4	ITS2			MOTHUR	CRUNCHCLUST	97			YES	MOTHUR	OOO	Hartmann et al. (2012)

Table 1 (continued)

Sample[a]	Target	Primer pair(s)	Length[b]	Denoising	Chimera removal	Data cleanup[c]	Clustering[d]	Percentage	Rep. sequence[e]	Singleton removal	Diversity[f]	Annotation[g]	Size[h]	Reference
Soil, plant	ITS	ITS1f/ITS4, fITS7/ITS4, gITS7/ITS4, fITS9/ITS4	ITS2			SCATA	SCATA	97.5	C			PlutoF	OOO	Ihrmark et al. (2012)
Roots	ITS	ITS5/ITS2	150			CLOTU	BLASTCLUST	97	L			BLASTN	OO	Kauserud et al. (2012)
Wood	ITS	ITS1f/ITS4	200			SCATA	SCATA	98.5	A	YES		SCATA	OOO	Kubartova et al. (2012)
Sediment	ITS	ITS1f/ITS4	200		YES	MOTHUR	CRUNCHCLUST	97				PlutoF	OOO	La Duc et al. (2012)
Roots (G)	28S	glo454/NDL22	150			RDP	Other	97		YES		BLASTN	OOO	Lekberg et al. (2012)
Soil (G)	18S	AMV4.5NF/ AMDGR	200			QIIME	Other	97	A		YES (R)	BLASTN	OOO	Lin et al. (2012)
Soil	18S	SSU817f/ SSU1196r	?			QIIME	Other	97			YES (R)	BLASTN	OO	McGuire et al. (2012)
Soil, roots	ITS	ITS9/ITS4	200			SEQMAN PRO	Other	97			YES	BLASTN	O	Menkis et al. (2012)
Soil	ITS	ITS1/ITS4	300	YES	YES	MOTHUR	CD-HIT (S)	97	C		YES (R)	PlutoF	O	Štursová et al. (2012)
Roots	ITS	ITS1f/ITS2	150			CLOTU	BLASTCLUST	97	L	YES	YES	BLASTN	OO	Yu et al. (2012a)
Roots	ITS	ITS1f/ITS4	150			CLOTU	BLASTCLUST	97		YES	YES	BLASTN	OO	Yu et al. (2012b)
Soil, plant	ITS	ITS1/ITS4	380	YES	YES	MOTHUR	CD-HIT (S)	97	C		YES (R)	PlutoF	OO	Voříšková and Baldrian (2013)

[a] Soil also includes litter; (G)—only Glomeromycota

[b] ITS2: the length of the entire ITS2 region

[c] CLC Gen. W. (http://www.clcbio.com/products/clc-genomics-workbench/), CLOTU (http://www.bioportal.uio.no/appinfo/show.php?app=CLOTU), EMBOSS (http://emboss.sourceforge.net/), MOTHUR (http://www.mothur.org/), NEWBLER (http://en.wikipedia.org/wiki/Newbler), PANGEA (http://pangea-16s.sourceforge.net/), QIIME (http://qiime.org/), RDP (http://rdp.cme.msu.edu/), SCATA (http://scata.mykopat.slu.se/), SEQMAN PRO (http://www.dnastar.com/t-sub-products-lasergene-seqmanpro.aspx), Sequencher (http://genecodes.com/)

[d] BLASTCLUST (http://www.ncbi.nlm.nih.gov/Web/Newsltr/Spring04/blastlab.html), CAP3 (http://seq.cs.iastate.edu/), CD-HIT (http://weizhong-lab.ucsd.edu/cdhit_suite/cgi-bin/index.cgi), CRUNCHCLUST (http://code.google.com/p/crunchclust/), SCATA (http://scata.mykopat.slu.se/), UCLUST (http://www.drive5.com/usearch/). (S)—sequences resized to identical lengths

[e] A—most abundant, C—cluster consensus, L—longest, R—random

[f] R—resampling to the same depth

[g] BLASTN (blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastn), CLOTU (http://www.bioportal.uio.no/appinfo/show.php?app=CLOTU), MEGAN (http://ab.inf.uni-tuebingen.de/software/megan/), MOTHUR (http://www.mothur.org/), PlutoF (http://elurikkus.ut.ee/plutof.php?lang=eng), SCATA (http://scata.mykopat.slu.se/)

[h] O<10,000 sequences, OO>10 000 sequences, OOO>100 000 sequences

Voříšková and Baldrian 2013). In the last paper, the data analysis workflow recommended here was used for data processing.

The SEED pipeline is a workbench that runs in the Microsoft Windows environment with internal functions and functions performed by external programmes that must be installed for full functionality. The removal of pyrosequencing noise is performed using Pat Schloss's translation of Chris Quince's PyroNoise algorithm implemented within the Mothur package (Schloss et al. 2009). The removal of chimeras created during PCR amplification is performed using Uchime (Edgar et al. 2011), and Usearch (Edgar 2010) is used for sequence clustering. Sequence alignment is performed by calling MAFFT (Katoh et al. 2009), and BLAST searching and the creation of local databases are dependent on the National Center for Biotechnology Information (NCBI) tools (http://www.ncbi.nlm.nih.gov/; Altschul et al. (1990)). Internet connection is required for searching online databases, e.g. the NCBI nucleotide database.

The SEED pipeline is freely available for non-commercial use and can be downloaded along with documentation from the SEED project webpage: http://www.biomed.cas.cz/mbu/lbwrf/seed/main.html. The installation of external programmes may require the consent of their authors: more information can be found at the web pages of these projects, accessible by hyperlink from the above address.

Results

In total, 42 published studies were analysed (Table 1). The number of papers using amplicon pyrosequencing to analyse fungal communities increased rapidly from 3 in 2009 to 22 in 2012. Soil fungal communities were the most common target of amplicon pyrosequencing (21 studies), along with fungal communities in plant roots (12 papers). Other environments (sediments, aboveground plant tissues, corals or wood) were only rarely addressed. Although most studies were designed to cover the entire fungal community, six papers targeted specifically arbuscular mycorrhizal fungi. In addition to analysing entire fungal communities, amplicon sequencing was also applied to analyse the diversity of the fungal *cbhI* exocellulase gene, a proxy for the community of cellulose-decomposing fungi (Baldrian et al. 2012; Štursová et al. 2012; Voříšková and Baldrian 2013). There is only one single study to date in which RNA-derived amplicons were used to specifically analyse metabolically active fungal taxa (Baldrian et al. 2012; Purahong and Krüger 2012).

The ITS region was by far the most frequently analysed region of fungal rDNA: only four and three papers analysed various regions of the 18S and 28S rRNA genes, respectively (Fig. 1). Within the ITS, ITS1 was mainly targeted with several primer pairs to amplify only this region; in additional studies, both the ITS1 and ITS2 regions were amplified, but because of the limiting lengths of pyrosequencing-derived sequences and the fact that sequencing mostly occurred from primers within the 18S, the sequence data also covered predominantly the ITS1 region. Only recently, studies analysing the ITS2 region specifically have been conducted (Davey et al. 2012; Hartmann et al. 2012; Ihrmark et al. 2012; Menkis et al. 2012).

The initial steps of sequence data processing typically consisted of sequence quality filtering and reduction of PCR or sequencing errors. A wide set of tools was used for data cleanup, which resulted in the removal of sequences of insufficient length or quality, but the minimal length of sequences retained in the cleaned dataset varied considerably (Table 1 and Fig. 2). Typically, between 10 and 40 % of sequences were removed in this step. Pyrosequencing-derived errors, typically the variable lengths of longer homopolymer regions, were corrected by clustering pyrosequencing flowgrams, termed "denoising", and PCR-derived errors were removed by chimera-cleaning tools. Despite the high rate of occurrence of both types of errors, only <30 % of all studies used one of these approaches and only 12 % used both (Fig. 2).

Sequences that passed filtering steps were used to create virtual taxa, i.e. the sequence similarity clusters most often termed operational taxonomic units. Despite the inconsistency of clustering sequences of variable lengths, only a handful of studies truncated sequences to identical lengths or extracted particular DNA regions before clustering. CD-HIT, BLASTCLUST and CAP3 were most frequently used for clustering. For annotation, OTUs were represented either by a randomly selected sequence or by the longest or most abundant sequence. In six studies, consensus sequences were constructed to represent OTUs (Table 1). Approximately one half of the studies only considered non-singleton sequences for community analysis (Fig. 2), and BLAST against the NCBI database was the most frequent approach to assign taxonomic identity to OTUs. In 55 % of studies, diversity parameters were calculated for individual samples. Among these, only 26% performed resampling to the same depth before calculating diversity (Fig. 2).

After considering the previous data analysis protocols, we suggest the following workflow (Table 2). Quality trimming should first exclude sequences of low base quality and length. The minimal length of sequences to be analysed should be at least above 150 bases because both the ITS1 and ITS2 are longer than that for many fungi. The quality of taxonomic assignments based on the 18S or 28S region analyses also greatly increases with sequence length. Both denoising and chimera removal should be performed to reduce the sequence error rate to a minimum. Because clustering algorithms compare sequences in a pairwise manner, the regions to be clustered should optimally be defined as the same DNA region, i.e. with defined primer positions at

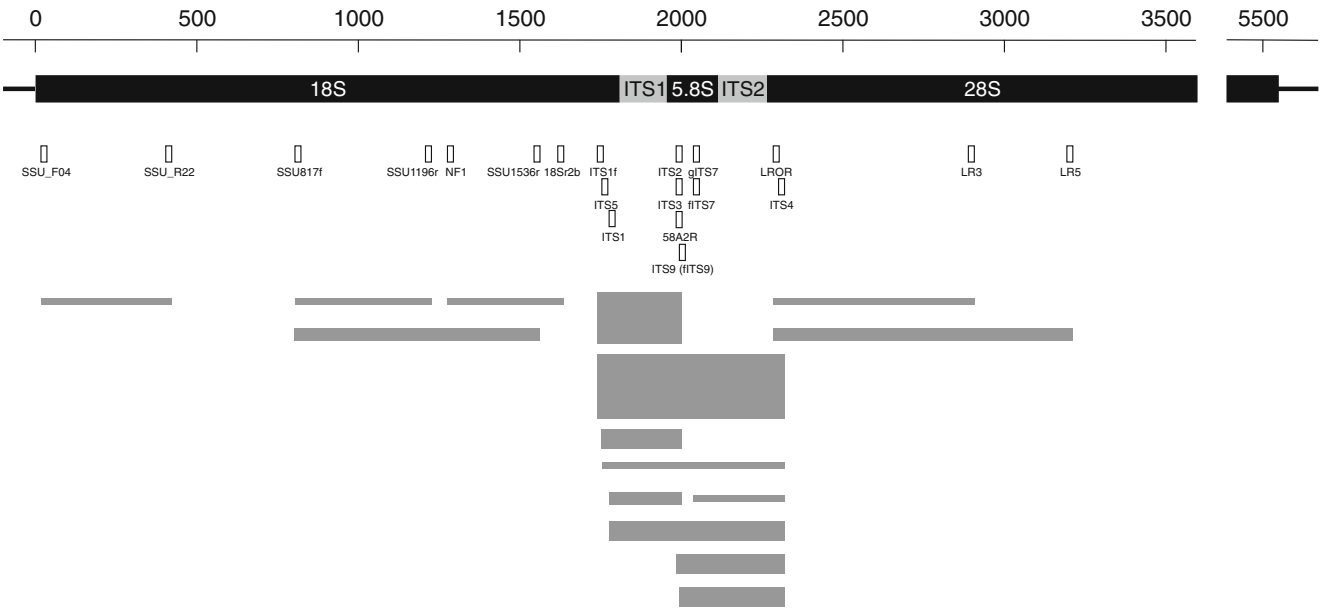


Fig. 1 Primers and PCR amplicons used in amplicon pyrosequencing analyses of the community composition of general fungi. The thickness of grey bars indicates the number of studies using the respective amplicons. Numbers indicate the positions in the rDNA of *Fusarium oxysporum*

both ends (if the amplicons are shorter than pyrosequencing read length), or using a defined sequence (e.g. ITS1, ITS2, ITS1+5.8S+ITS2) that can be extracted easily (Nilsson et al. 2010), or at least defined by the same length of all sequences. Consensus sequences best represent individual sequences within an OTU.

Depending on the aim of the study, sequence identification may be requested either based on the identity of the closest database hit or through multiple alignment of OTU sequences with known sequences. In the studies targeting the diversity of fungal communities, community richness, evenness or other parameters may also be derived. To obtain comparable data, the sequence database has to be randomly resampled to obtain identical numbers of sequences from each sample.

Clustering of sequences truncated at 380 bases gave lower OTU counts, numbers of singletons and Chao estimates of total community richness than the clustering of

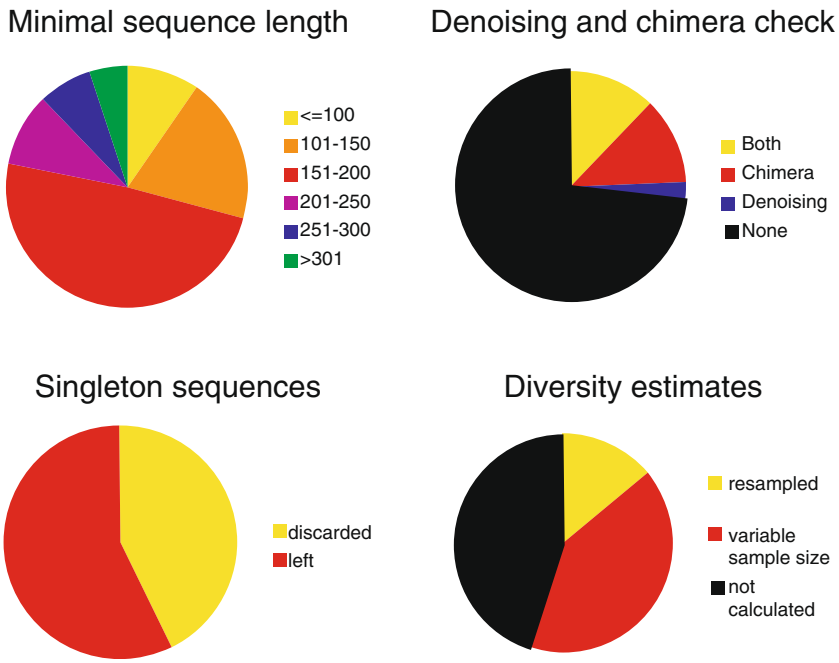


Fig. 2 Overview of approaches used to analyse sequences derived by amplicon pyrosequencing of fungal rDNA based on 42 recently published studies

Table 2 Workflow of the analysis of sequences derived by amplicon pyrosequencing of fungal communities

Step	Description and comments
Quality trimming	• Removal of sequences of inferior quality (e.g. quality score <20) or length (e.g. length <100)
	Trimming before denoising helps reduce the size of the dataset and, thus, the length of the denoising procedure.
Sample identification	• Distribution of sequences into samples based on barcode sequences
Denoising	• Removal of pyrosequencing noise
	Denoising helps reduce the amount of method-dependent errors (e.g. \homopolymers, point mistakes) by comparing pyrosequencing flow grams between sets of highly similar sequences.
Removal of chimeric sequences	• Deletion of potentially chimeric sequences from the dataset
	Chimeric sequences arise during PCR at various rates (often >10 %) depending on PCR conditions, such as the number of cycles or template concentrations. The large numbers of sequences from each PCR reaction allow the detection of chimeric sequences based on their multiple comparisons.
Selection of region for clustering	• Trimming of sequences to contain the same part of the template
	For clustering, sequences should contain identical regions of DNA because length differences make clustering algorithms unreliable because of the uneven similarity of short and long sequences. This can be achieved by trimming to identical lengths (setting a lower sequence length limit, e.g. 300 bases, and truncating long sequences) or by the identification of sequence boundaries within rDNA (e.g. the start and end of ITS1 or ITS2). It may be desirable to exclude primer sequence(s).
Clustering	• OTUs are created by grouping sequences based on their similarity
	Typically, clustering is based on sequential comparison of individual sequences with sequences used for cluster establishment (seed sequences, i.e. those sequences that show similarity lower than the defined threshold with all seed sequences of clusters established so far).
Creation of consensus sequences	• Consensus sequences for each OTU are created by sequence alignment.
	Consensus sequences better represent the OTU than individual sequences because consensus creation removes random sequencing errors that survived denoising. As a result, closer hits to known taxa are found for consensus sequences than for individual sequences. Consensus sequences can also be used to represent OTUs in phylogenetic analyses.
Sequence identification	• Best-identified hits are retrieved for each OTU consensus sequence. The full taxonomy of the best hit may also be retrieved.
	The best hits are retrieved along with the similarity values (*E*, per cent similarity) that help assign the OTU to a specific taxon.
Community composition analysis	• Based on the full taxonomic placement of best hits, the abundance of taxa of various taxonomic levels can be calculated.
	The abundance of individual taxa (e.g. OTUs, genera, orders, phyla) in each sample can be expressed as a percentage of all sequences. Community composition data are usually used for statistical purposes (comparison of samples by correlation, analysis of variance, multivariate methods or sample similarity clustering).
Estimation of diversity parameters	• Based on the sequence counts for each OTU of a sample, diversity and evenness parameters of the community are calculated.
	Because diversity estimates tend to scale up with increasing sampling depth, it is essential to randomly resample the sequence database to include identical numbers of sequences from each sample. For this subsampled database, OTUs must be newly created.

sequences of their original lengths of 380–560 bases. The numbers of OTUs, singletons and Chao estimates were lower by 13.4±1.6, 12.7±2.4 and 7.4±3.3 %, respectively, all differences being statistically significant at $P<0.003$. This shows that the OTU counts are inflated when sequences of different lengths are clustered together. The

application of chimera removal on sequences truncated to 380 bases decreased the numbers of OTUs, singletons and Chao estimates further by 20.1±1.2, 16.7±2.0 and 17.5±3.4 %, respectively, all differences being statistically significant at $P<0.001$. This shows that a significant part of the apparent diversity in the dataset may be due to the presence of chimeric sequences. Consensus sequences of the 150 most abundant OTUs in the dataset showed significantly higher ($P<0.0001$) sequence similarity to the closest BLAST hit in GenBank than random sequences, with 69 % consensus sequences showing higher similarity, 27 % showing the same similarity and 4 % showing lower similarity. Moreover, 13 % OTU consensus sequences showed 100 % similarity to the GenBank sequence, whilst the corresponding random sequences were less similar. On average, consensus sequences were by 0.29±0.04 % more similar to the closest GenBank hits than randomly selected sequences.

The SEED pipeline makes it possible to perform all steps of the sequence analysis workflow from a single, user-friendly interface (Fig. 3). The features of the pipeline are summarised in Table 3, and more information can be found on the project webpage (http://www.biomed.cas.cz/mbu/lbwrf/seed/main.html) that contains full documentation of the functions and a step-by-step introduction to the data processing workflow. Importantly, in addition to sequence grouping, SEED makes it possible to perform batch operations with groups, such as chimera removal from individual samples, calculation of consensus sequences for individual OTUs, resampling of all samples at a specific depth, etc. SEED can be used to analyse PCR amplicons of any type, e.g. bacterial

Table 3 Features of the amplicon pyrosequencing pipeline SEED

Sequence editing and sorting
Extraction of sequences and sequence qualities from *.sff files
Quality trimming
Grouping sequences based on sequence motifs or sequence titles
Sequence batch processing
Sequence denoising (using the PyroNoise algorithm translation within Mothur)[a]
Chimera removal (using Uchime)[a]
Sequence alignment (using MAFFT)[a]
Sequence clustering (using Usearch)[a]
Construction of consensus sequences
Searching for best hits in a local database or the NCBI (using nucleotide BLAST)[a]
Retrieval of taxonomical classification of best hits from the NCBI
Creation of local databases for searching by nucleotide BLAST
Calculation of diversity parameters

[a] Mothur (Schloss et al. 2009), Uchime (Edgar et al. 2011), MAFFT (Katoh et al. 2009), Usearch (Edgar 2010), BLAST (Altschul et al. 1990), NCBI (http://www.ncbi.nlm.nih.gov/)

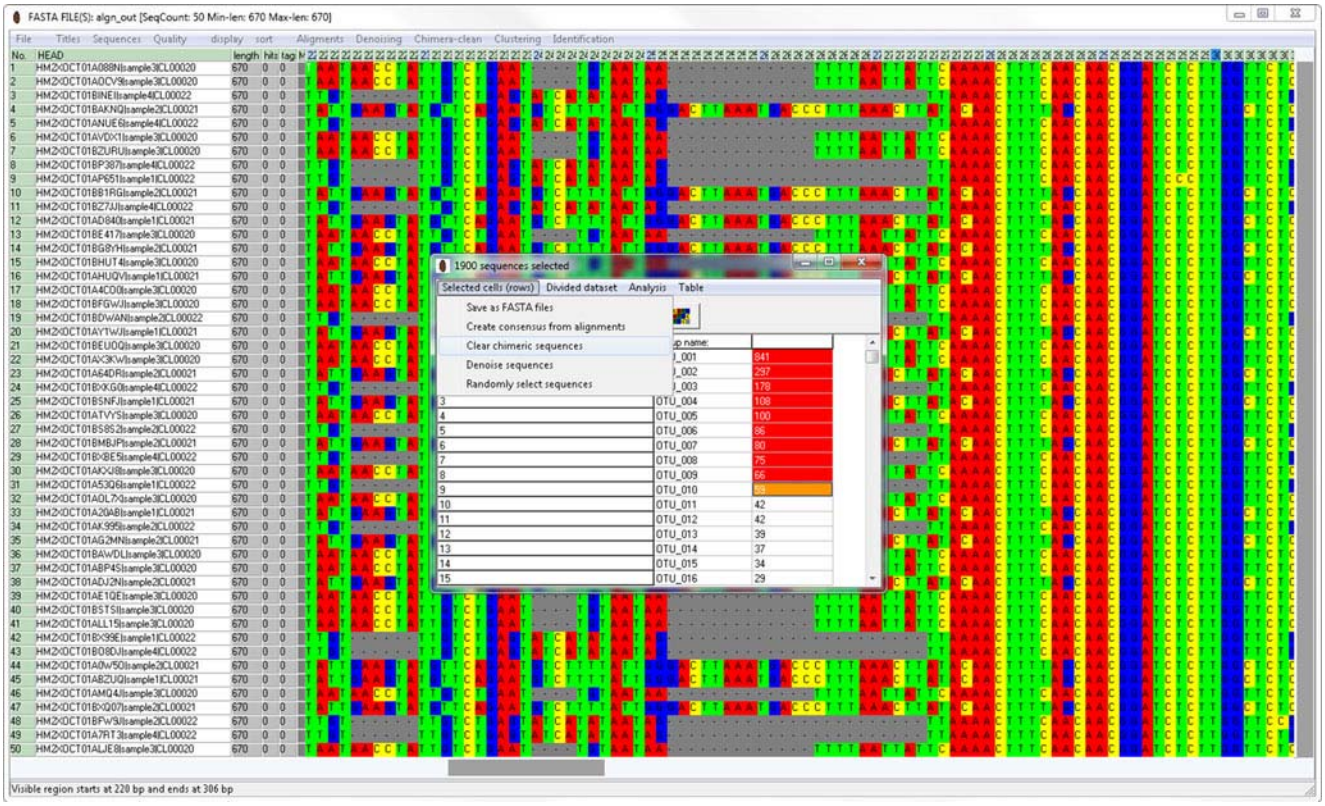


Fig. 3 Screenshot of the amplicon pyrosequencing pipeline SEED

16S rDNA or functional genes, or to analyse gene sequences obtained by other means (e.g. batch download from the NCBI nucleotide or genome database).

Discussion

The methods of next-generation sequencing have revolutionised microbial ecology, allowing researchers to explore complex communities at unprecedented depths. Despite the first applications of the Illumina (Caporaso et al. 2012) or Ion Torrent (Whiteley et al. 2012) technologies to explore bacterial communities, pyrosequencing remains the method of choice for fungal and bacterial amplicon sequencing, offering the advantages of reasonable sequence length, easy multiplexing and sufficient sequencing depth for most studies (Glenn 2011). Nevertheless, successful applications of pyrosequencing approaches are dependent on a number of methodological considerations, including sampling strategies and metadata collection, the choice of suitable molecular marker and approaches for data analysis. Because of the diversity of all of the above methodologies in the published studies, it is extremely difficult to use the wealth of information derived by pyrosequencing for inter-study comparisons or meta-studies. Furthermore, published papers differ widely in the level of method descriptions and data availability. We strongly agree with the previous paper by Nilsson et al. (2011) in that full description of the experimental procedures and public data availability should be a standard.

Here we show that despite some general preferences, many different molecular targets are used to study both general fungi and arbuscular mycorrhizal fungi. Without exception, fungal rDNA was targeted despite widely varying relationships between its copy numbers and fungal cell counts or biomasses (Amend et al. 2010a; Baldrian et al. 2013). The ITS region amplified using various sets of primers was the preferred target, consistent with the dominant current opinion (Schoch et al. 2012).

Although ITS1 was frequently sequenced, it is notable that the results obtained with various primers cannot be easily compared because of their variable coverage of the fungal tree of life (Anderson et al. 2003). Unfortunately, there are only a few papers in which various primers were compared. The recent paper by Ihrmark et al. (2012) demonstrates that PCR amplification can be highly uneven among primer pairs as well as diversity estimates. More work is still required in this direction.

The data analysis procedures used in past amplicon pyrosequencing studies indicate many potential limitations of data quality. Studies using sequences of <150 bases length covered less than the entire ITS1 or ITS2 regions of certain fungi because of the differences in the regions' lengths, and this seems to be unsuitable. In our in silico study considering the region between the ITS1/ITS4 primers, fungal sequence assignment quality increased with increasing sequence length up to the length of 350–380 bases (data not shown). Such sequence lengths are easily available with current technologies and may be desirable when reliable OTU classification is required. Furthermore, clustering algorithms work best with sequences of identical boundaries (or lengths), a fact that is usually not considered. Here, we show that clustering of sequences of uneven length significantly increases the diversity estimates.

PCR and pyrosequencing have been shown to cause method-dependent sequencing errors (Quince et al. 2009; Tedersoo et al. 2010). In PCR amplification, chimeric sequences are formed with frequencies at or above 3 %, depending on the number of cycles (Taylor and Houston 2011). Because these sequences are most often singletons, the presence of chimeric sequences may result in an overestimation of diversity. This was also clearly demonstrated here in the comparison of diversity estimates among the original and chimera-cleaned dataset. Chimera-cleaning procedures should therefore always be applied. When choosing a minimal length, one should also consider that the probability of detecting chimeric sequences rapidly increases with sequence length, and shorter sequences are more likely to contain undiscovered chimeras. In addition, the increase of sequence error counts associated with increasing sequence lengths and the frequency of sequencing errors in homopolymeric regions that stem from the techniques of pyrosequencing should be reduced by applying denoising (i.e. error correction) procedures (Quince et al. 2011). Unfortunately, error-correcting procedures have been rarely applied so far. Given the error rate of pyrosequencing-derived reads and the random distribution of such errors, the creation of OTU consensus sequences should further improve the representation of an OTU. This was demonstrated here by the fact that the consensus sequences are significantly more similar to the Sanger sequences deposited in GenBank than individual OTU sequences.

To explore fungal diversity, the analysis of identical numbers of sequences from all samples is essential because diversity estimates always scale up with sampling depth. This fact has also been frequently neglected in past studies.

We here outline a workflow of data analysis that aims to reflect all of the considerations required for obtaining high-quality data for community analysis and offer the SEED pipeline to accomplish this task. We hope that the unification of data analysis procedures represents an important step towards better comparability of individual studies and justification of their conclusions. The SEED pipeline should offer ecologists a tool that is easy to use, even for those with no preliminary experience with amplicon pyrosequencing, the method that will likely continue to dominate microbial community analysis in the coming years.

Acknowledgments This work was supported by the Ministry of Education, Youth and Sports of the Czech Republic (LD12048, LD12050), by the Czech Science Foundation (P504/12/0709) and by the Research concept of the Institute of Microbiology ASCR (RVO61388971).

References

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215:403–410

Amend AS, Barshis DJ, Oliver TA (2012) Coral-associated marine fungi form novel lineages and heterogeneous assemblages. ISME J 6:1291–1301

Amend AS, Seifert KA, Bruns TD (2010a) Quantifying microbial communities with 454 pyrosequencing: does read abundance count? Mol Ecol 19:5555–5565

Amend AS, Seifert KA, Samson R, Bruns TD (2010b) Indoor fungal composition is geographically patterned and more diverse in temperate zones than in the tropics. Proc Nat Acad Sci U S A 107:13748–13753

Anderson IC, Campbell CD, Prosser JI (2003) Potential bias of fungal 18S rDNA and internal transcribed spacer polymerase chain reaction primers for estimating fungal biodiversity in soil. Environ Microbiol 5:36–47

Arfi Y, Buéé M, Marchand C, Levasseur A, Record E (2012a) Multiple markers pyrosequencing reveals highly diverse and host-specific fungal communities on the mangrove trees *Avicennia marina* and *Rhizophora stylosa*. FEMS Microbiol Ecol 79:433–444

Arfi Y, Marchand C, Wartel M, Record E (2012b) Fungal diversity in anoxic–sulfidic sediments in a mangrove soil. Fungal Ecol 5:282–285

Baldrian P, Kolarik M, Stursova M, Kopecky J, Valaskova V, Vetrovsky T, Zifcakova L, Snajdr J, Ridl J, Vlcek C, Voriskova J (2012) Active and total microbial communities in forest soil are largely different and highly stratified during decomposition. ISME J 6:248–258

Baldrian P, Větrovský T, Cajthaml T, Dobiášová P, Petránková M, Šnajdr J, Eichlerová I (2013) Estimation of fungal biomass in forest litter and soil. Fungal Ecol 6:1–11

Becklin KM, Hertweck KL, Jumpponen A (2012) Host identity impacts rhizosphere fungal communities associated with three alpine plant species. Microb Ecol 63:682–693

Bellemain E, Carlsen T, Brochmann C, Coissac E, Taberlet P, Kauserud H (2010) ITS as an environmental DNA barcode for fungi: an in silico approach reveals potential PCR biases. BMC Microbiol 10:189

Bik HM, Sung W, De Ley P, Baldwin JG, Sharma J, Rocha-Olivares A, Thomas WK (2012) Metagenetic community analysis of microbial eukaryotes illuminates biogeographic patterns in deep-sea and shallow water sediments. Mol Ecol 21:1048–1059

Blaalid R, Carlsen T, Kumar S, Halvorsen R, Ugland KI, Fontana G, Kauserud H (2012) Changes in the root-associated fungal communities along a primary succession gradient analysed by 454 pyrosequencing. Mol Ecol 21:1897–1908

Buéé M, Reich M, Murat C, Morin E, Nilsson RH, Uroz S, Martin F (2009) 454 Pyrosequencing analyses of forest soils reveal an unexpectedly high fungal diversity. New Phytol 184:449–456

Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Huntley J, Fierer N, Owens SM, Betley J, Fraser L, Bauer M, Gormley N, Gilbert JA, Smith G, Knight R (2012) Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. ISME J 6:1621–1624

Cordier T, Robin C, Capdevielle X, Desprez-Loustau ML, Vacher C (2012) Spatial variability of phyllosphere fungal assemblages: genetic distance predominates over geographic distance in a European beech stand (*Fagus syluatica*). Fungal Ecol 5:509–520

Davey ML, Heegaard E, Halvorsen R, Ohlson M, Kauserud H (2012) Seasonal trends in the biomass and structure of bryophyte-associated fungal communities explored by 454 pyrosequencing. New Phytol 195:844–856

Davison J, Öpik M, Zobel M, Vasar M, Metsis M, Moora M (2012) Communities of arbuscular mycorrhizal fungi detected in forest soil are spatially heterogeneous but do not vary throughout the growing season. PLoS One 7:e41938

Dumbrell AJ, Ashton PD, Aziz N, Feng G, Nelson M, Dytham C, Fitter AH, Helgason T (2011) Distinct seasonal assemblages of arbuscular mycorrhizal fungi revealed by massively parallel pyrosequencing. New Phytol 190:794–804

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

Edgar RC, Haas BJ, Clemente JC, Quince C, Knight R (2011) UCHIME improves sensitivity and speed of chimera detection. Bioinformatics 27:2194–2200

Glenn TC (2011) Field guide to next-generation DNA sequencers. Mol Ecol Res 11:759–769

Gottel NR, Castro HF, Kerley M, Yang ZM, Pelletier DA, Podar M, Karpinets T, Uberbacher E, Tuskan GA, Vilgalys R, Doktycz MJ, Schadt CW (2011) Distinct microbial communities within the endosphere and rhizosphere of *Populus deltoides* roots across contrasting soil types. Appl Environ Microbiol 77:5934–5944

Hartmann M, Howes CG, Vaninsberghe D, Yu H, Bachar D, Christen R, Nilsson HR, Hallam SJ, Mohn WW (2012) Significant and persistent impact of timber harvesting on soil microbial communities in Northern coniferous forests. ISME J 6:2199–2218

Hui N, Jumpponen A, Niskanen T, Liimatainen K, Jones KL, Koivula T, Romantschuk M, Strommer R (2011) EcM fungal community structure, but not diversity, altered in a Pb-contaminated shooting range in a boreal coniferous forest site in Southern Finland. FEMS Microbiol Ecol 76:121–132

Ihrmark K, Bödeker IT, Cruz-Martinez K, Friberg H, Kubartova A, Schenck J, Strid Y, Stenlid J, Brandstrom-Durling M, Clemmensen KE, Lindahl BD (2012) New primers to amplify the fungal ITS2 region—evaluation by 454-sequencing of artificial and natural communities. FEMS Microbiol Ecol 82:666–677

Jumpponen A, Jones KL (2009) Massively parallel 454 sequencing indicates hyperdiverse fungal communities in temperate *Quercus macrocarpa* phyllosphere. New Phytol 184:438–448

Jumpponen A, Jones KL (2010) Seasonally dynamic fungal communities in the *Quercus macrocarpa* phyllosphere differ between urban and nonurban environments. New Phytol 186:496–513

Jumpponen A, Jones KL, Blair J (2010a) Vertical distribution of fungal communities in tallgrass prairie soil. Mycologia 102:1027–1041

Jumpponen A, Jones KL, Mattox D, Yaege C (2010b) Massively parallel 454-sequencing of fungal communities in *Quercus* spp. ectomycorrhizas indicates seasonal dynamics in urban and rural sites. Mol Ecol 19:41–53

Katoh K, Asimenos G, Toh H (2009) Multiple alignment of DNA sequences with MAFFT. In: Posada D (ed) Bioinformatics for DNA sequence analysis. Humana, Totowa, pp 39–64

Kauserud H, Kumar S, Brysting AK, Nordén J, Carlsen T (2012) High consistency between replicate 454 pyrosequencing analyses of ectomycorrhizal plant root samples. Mycorrhiza 22:309–315

Krüger D, Kapturska D, Fischer C, Daniel R, Wubet T (2012) Diversity measures in environmental sequences are highly dependent on alignment quality-data from ITS and new LSU primers targeting basidiomycetes. PLoS One 7:e32139

Kubartova A, Ottosson E, Dahlberg A, Stenlid J (2012) Patterns of fungal communities among and within decaying logs, revealed by 454 sequencing. Mol Ecol 21:4514–4532

La Duc MT, Vaishampayan P, Nilsson HR, Torok T, Venkateswaran K (2012) Pyrosequencing-derived bacterial, archaeal, and fungal diversity of spacecraft hardware destined for Mars. Appl Environ Microbiol 78:5912–5922

Lekberg Y, Schnoor T, Kjøller R, Gibbons SM, Hansen LH, Al-Soud WA, Sorensen SJ, Rosendahl S (2012) 454-Sequencing reveals stochastic local reassembly and high disturbance tolerance within arbuscular mycorrhizal fungal communities. J Ecol 100:151–160

Lentendu G, Zinger L, Manel S, Coissac E, Choler P, Geremia RA, Melodelima C (2011) Assessment of soil fungal diversity in different alpine tundra habitats by means of pyrosequencing. Fungal Divers 49:113–123

Lin XG, Feng YZ, Zhang HY, Chen RR, Wang JH, Zhang JB, Chu HY (2012) Long-term balanced fertilization decreases arbuscular mycorrhizal fungal diversity in an arable soil in North China revealed by 454 pyrosequencing. Environ Sci Technol 46:5764–5771

Lumini E, Orgiazzi A, Borriello R, Bonfante P, Bianciotto V (2010) Disclosing arbuscular mycorrhizal fungal biodiversity in soil through a land-use gradient using a pyrosequencing approach. Environ Microbiol 12:2165–2179

McGuire KL, Fierer N, Bateman C, Treseder KK, Turner BL (2012) Fungal community composition in neotropical rain forests: the influence of tree diversity and precipitation. Microb Ecol 63:804–812

Mello A, Napoli C, Murat C, Morin E, Marceddu G, Bonfante P (2011) ITS-1 versus ITS-2 pyrosequencing: a comparison of fungal populations in truffle grounds. Mycologia 103:1184–1193

Menkis A, Burokiene D, Gaitnieks T, Uotila A, Johannesson H, Rosling A, Finlay RD, Stenlid J, Vasaitis R (2012) Occurrence and impact of the root-rot biocontrol agent *Phlebiopsis gigantea* on soil fungal communities in *Picea abies* forests of northern Europe. FEMS Microbiol Ecol 81:438–445

Moora M, Berger S, Davison J, Öpik M, Bommarco R, Bruelheide H, Kuhn I, Kunin WE, Metsis M, Rortais A, Vanatoa A, Vanatoa E, Stout JC, Truusa M, Westphal C, Zobel M, Walther GR (2011) Alien plants associate with widespread generalist arbuscular mycorrhizal fungal taxa: evidence from a continental-scale study using massively parallel 454 sequencing. J Biogeogr 38:1305–1317

Nilsson RH, Tedersoo L, Lindahl BD, Kjoller R, Carlsen T, Quince C, Abarenkov K, Pennanen T, Stenlid J, Bruns T, Larsson KH, Koljalg U, Kauserud H (2011) Towards standardization of the description and publication of next-generation sequencing datasets of fungal communities. New Phytol 191:314–318

Nilsson RH, Veldre V, Hartmann M, Unterseher M, Amend A, Bergsten J, Kristiansson E, Ryberg M, Jumpponen A, 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 Ecol 3:284–287

Öpik M, Metsis M, Daniell TJ, Zobel M, Moora M (2009) Large-scale parallel 454 sequencing reveals host ecological group specificity of arbuscular mycorrhizal fungi in a boreonemoral forest. New Phytol 184:424–437

Ovaskainen O, Nokso-Koivista J, Hottola J, Rajala T, Pennanen T, Ali-Kovero H, Miettinen O, Oinonen P, Auvinen P, Paulin L, Larsson KH, Mäkipää R (2010) Identifying wood-inhabiting fungi with 454 sequencing—what is the probability that BLAST gives the correct species? Fungal Ecol 3:274–283

Purahong W, Krüger D (2012) A better understanding of functional roles of fungi in the decomposition process: using precursor rRNA containing ITS regions as a marker for the active fungal community. Ann Forest Sci 69:659–662

Quince C, Lanzén A, Curtis TP, Davenport RJ, Hall N, Head IM, Read LF, Sloan WT (2009) Accurate determination of microbial diversity from 454 pyrosequencing data. Nature Meth 6:639–641

Quince C, Lanzén A, Davenport RJ, Turnbaugh PJ (2011) Removing noise from pyrosequenced amplicons. BMC Bioinformatics 12:38

Shokralla S, Spall JL, Gibson JF, Hajibabaei M (2012) Next-generation sequencing technologies for environmental DNA research. Mol Ecol 21:1794–1805

Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, Lesniewski RA, Oakley BB, Parks DH, Robinson CJ, Sahl JW, Stres B, Thallinger GG, Van Horn DJ, Weber CF (2009) Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. Appl Environ Microbiol 75:7537–7541

Schoch CL, Seifert KA, Huhndorf S, Robert V, Spouge JL, Levesque CA et al (2012) Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for fungi. Proc Nat Acad Sci U S A 109:6241–6246

Štursová M, Žifčáková L, Leigh MB, Burgess R, Baldrian P (2012) Cellulose utilization in forest litter and soil: identification of bacterial and fungal decomposers. FEMS Microbiol Ecol 80:735–746

Taylor DL, Houston S (2011) A bioinformatics pipeline for sequence-based analyses of fungal biodiversity. In: Xu JR, Bluhm BH (eds) Fungal genomics. Humana, Totowa, pp 141–155

Tedersoo L, Nilsson RH, Abarenkov K, Jairus T, Sadam A, Saar I, Bahram M, Bechem E, Chuyong G, Koljalg U (2010) 454 Pyrosequencing and Sanger sequencing of tropical mycorrhizal fungi provide similar results but reveal substantial methodological biases. New Phytol 188:291–301

Větrovský T, Baldrian P (2013) The variability of the 16S rRNA gene in bacterial genomes and its consequences for bacterial community analyses. PLoS One 8:e57923

Voříšková J, Baldrian P (2013) Fungal community on decomposing leaf litter undergoes rapid successional changes. ISME J 7:477–486

Wallander H, Johansson U, Sterkenburg E, Durling MB, Lindahl BD (2010) Production of ectomycorrhizal mycelium peaks during canopy closure in Norway spruce forests. New Phytol 187:1124–1134

Whiteley AS, Jenkins S, Waite I, Kresoje N, Payne H, Mullan B, Allcock R, O'Donnell A (2012) Microbial 16S rRNA Ion Tag and community metagenome sequencing using the Ion Torrent (PGM) Platform. J Microbiol Methods 91:80–88

Xu LH, Ravnskov S, Larsen J, Nicolaisen M (2011) Influence of DNA extraction and PCR amplification on studies of soil fungal communities based on amplicon sequencing. Can J Microbiol 57:1062–1066

Yu L, Nicolaisen M, Larsen J, Ravnskov S (2012a) Molecular characterization of root-associated fungal communities in relation to health status of *Pisum sativum* using barcoded pyrosequencing. Plant Soil 357:395–405

Yu L, Nicolaisen M, Larsen J, Ravnskov S (2012b) Succession of root-associated fungi in *Pisum sativum* during a plant growth cycle as examined by 454 pyrosequencing. Plant Soil 358:216–224

VI

The ISME Journal (2012) 6, 248–258

www.nature.com/ismej

ORIGINAL ARTICLE

Active and total microbial communities in forest soil are largely different and highly stratified during decomposition

Petr Baldrian[1], Miroslav Kolařík[1], Martina Šťursová[1], Jan Kopecký[2], Vendula Valášková[1], Tomáš Větrovský[1], Lucia Žifčáková[1], Jaroslav Šnajdr[1], Jakub Rídl[3], Čestmír Vlček[3] and Jana Voříšková[1]

[1]*Laboratory of Environmental Microbiology, Institute of Microbiology of the ASCR, v.v.i., Vídeňská, Praha, Czech Republic;* [2]*Crop Research Institute, Drnovská, Praha, Czech Republic and* [3]*Institute of Molecular Genetics of the ASCR, v.v.i., Vídeňská, Praha, Czech Republic*

Soils of coniferous forest ecosystems are important for the global carbon cycle, and the identification of active microbial decomposers is essential for understanding organic matter transformation in these ecosystems. By the independent analysis of DNA and RNA, whole communities of bacteria and fungi and its active members were compared in topsoil of a *Picea abies* forest during a period of organic matter decomposition. Fungi quantitatively dominate the microbial community in the litter horizon, while the organic horizon shows comparable amount of fungal and bacterial biomasses. Active microbial populations obtained by RNA analysis exhibit similar diversity as DNA-derived populations, but significantly differ in the composition of microbial taxa. Several highly active taxa, especially fungal ones, show low abundance or even absence in the DNA pool. Bacteria and especially fungi are often distinctly associated with a particular soil horizon. Fungal communities are less even than bacterial ones and show higher relative abundances of dominant species. While dominant bacterial species are distributed across the studied ecosystem, distribution of dominant fungi is often spatially restricted as they are only recovered at some locations. The sequences of *cbhI* gene encoding for cellobiohydrolase (exocellulase), an essential enzyme for cellulose decomposition, were compared in soil metagenome and metatranscriptome and assigned to their producers. Litter horizon exhibits higher diversity and higher proportion of expressed sequences than organic horizon. Cellulose decomposition is mediated by highly diverse fungal populations largely distinct between soil horizons. The results indicate that low-abundance species make an important contribution to decomposition processes in soils.

The ISME Journal (2012) **6,** 248–258; doi:10.1038/ismej.2011.95; published online 21 July 2011
Subject Category: microbial population and community ecology
Keywords: bacteria; cellulose decomposition; forest soil; fungi; RNA; transcription

Introduction

Most terrestrial ecosystem functions occur in the soil, which possesses the greatest amount of biodiversity on Earth. Yet, the understanding of how ecosystem functions are influenced by soil biodiversity is far behind our understanding of how aboveground organisms contribute to these functions (Bowker *et al.*, 2010). Soil microorganisms represent a considerable fraction of the living biomass on Earth, with 10^3–10^4 kg of microbial biomass per hectare of surface soils (Fierer *et al.*, 2007). In addition, microbial community composition is now recognised as an important determinant of ecosystem process rates (Reed and Martiny 2007; Strickland *et al.*, 2009). Understanding the structure and function of soil microbial communities is thus central to predicting how ecosystems will respond to future environmental conditions.

While several recent studies have used deep sequencing approaches to assess the diversity of soil bacterial components (Roesch *et al.*, 2007; Lauber *et al.*, 2009), the number of such studies addressing fungal diversity is still limited. This is true despite the fact that fungi comprise a large proportion of soil microbial biomass and have a dominant role in decomposition and nutrient cycling in soil (Bailey *et al.*, 2002; Buée *et al.*, 2009). Only a minor fraction of the estimated 1.5 million fungal species worldwide have been described (Hawksworth, 2001), and the ecological roles of most fungal taxa are poorly understood since the complexity of fungal communities has

Correspondence: P Baldrian, Laboratory of Environmental Microbiology, Institute of Microbiology of the ASCR, v.v.i., Vídeňská 1083, 14220 Praha 4, Czech Republic.
E-mail: baldrian@biomed.cas.cz
Received 17 February 2011; revised 16 June 2011; accepted 20 June 2011; published online 21 July 2011

		L	H	
pH		3.7 ± 0.0	3.8 ± 0.1	
dry mass	%	73.9 ± 1.4	46.0 ± 6.5	***
organic matter	%	94.5 ± 0.5	23.5 ± 4.6	***
C	%	51.6 ± 0.3	14.1 ± 4.4	***
N	%	1.93 ± 0.08	0.62 ± 0.22	***
C : N	molar	26.8 ± 1.2	23.0 ± 1.1	***
C : H	molar	0.77 ± 0.02	1.00 ± 0.06	***
PLFA bacteria	nmol g^{-1}	486.6 ± 106.1	205.3 ± 48.8	**
PLFA fungi	nmol g^{-1}	189.0 ± 32.3	26.7 ± 7.1	***
fungi / bacteria (PLFA)		0.39 ± 0.03	0.13 ± 0.03	***
ergosterol	µg g^{-1}	0.509 ± 0.257	0.020 ± 0.012	***
rDNA bacteria	10^3 x copy / ng DNA	18.4 ± 3.3	11.3 ± 1.0	***
rDNA fungi	10^3 x copy / ng DNA	19.5 ± 4.4	3.4 ± 0.7	***
fungi / bacteria (rDNA)		1.11 ± 0.27	0.30 ± 0.06	***
rRNA bacteria	10^5 x copy / ng cDNA	58.4 ± 23.6	3.8 ± 1.8	***
ITS fungi	10^3 x copy / ng cDNA	9.61 ± 0.80	0.11 ± 0.06	***
β-glucosidase	µmol min^{-1} g^{-1} DM	26.5 ± 10.4	11.0 ± 6.6	***
α-glucosidase	µmol min^{-1} g^{-1} DM	3.2 ± 1.1	1.0 ± 0.7	***
cellobiohydrolase	µmol min^{-1} g^{-1} DM	5.3 ± 2.2	2.4 ± 2.2	**
β-xylosidase	µmol min^{-1} g^{-1} DM	6.7 ± 1.8	8.1 ± 1.9	
N-acetylglucosaminidase	µmol min^{-1} g^{-1} DM	11.4 ± 5.4	4.9 ± 5.6	*
arylsulfatase	µmol min^{-1} g^{-1} DM	0.6 ± 0.1	1.3 ± 0.6	**
phosphomonoesterase	µmol min^{-1} g^{-1} DM	114.4 ± 24.7	90.3 ± 22.4	*

* Significant differences among soil horizons (one-way ANOVA followed by Tukey post-hoc test, * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$).

Figure 1 Properties of *Picea abies* forest soil, abundance of microorganisms and activity of extracellular enzymes involved in organic matter decomposition in the L and H horizons. The data represent mean values and s.d. from four studied sites.

so far limited our ability to estimate diversity and distinguish individual taxa (McGuire and Treseder, 2010). In order to understand the soil ecosystem processes, it is essential to address the fungal and bacterial community at the same time. In addition, the ecology of total microbial communities is so far largely derived from the studies on DNA and no information is available on the relationships between the diversity of this total community and community of active microbes assessed, for example, by targeting the RNA molecules (Anderson and Parkin, 2007; Urich *et al.*, 2008).

From the global viewpoint, the understanding of fungal and bacterial diversity is highly important in the biomes of coniferous forests, where fungi quantitatively dominate bacteria in decomposing litter material, while the importance of bacteria increases with soil depth (Bååth and Anderson 2003). Coniferous forest ecosystems have a prominent role in the global carbon cycle (Myneni *et al.*, 2001), and knowledge of microbially mediated soil functions is thus required for estimating global C fluxes and their potential future changes. Forests dominated by spruce (*Picea* spp.) constitute large ecosystems in boreal forest biomes and are also widely distributed in higher altitude forests and plantation forests in the northern temperate zone.

It was recently proposed that the analyses of soil microbial community composition should be based on direct analysis of total RNA to avoid PCR bias (Urich *et al.*, 2008). While this approach may be feasible for bacterial community analyses, the sequence information contained in fungal rRNA molecules is insufficient for species discrimination; thus, internal transcribed spacer (ITS) regions of the rRNA are used instead. Because there are 10^3–10^4 times fewer fungal ITS sequences than bacterial 16S rRNA gene sequences in soil (Figure 1), amplification of fungal ITS is inevitable to achieve reasonable sampling depth. Here, we combined the analysis of DNA-derived bacterial 16S rRNA gene sequences representing all bacteria present and the RNA-derived sequences representing the content of bacterial ribosomes reflecting thus the active part of the total community. The analysis of fungal ITS1 and ITS2 sequences offers a unique opportunity to target the precursor rRNA molecules with fast turnover, thus identifying these species synthesising ribosomes at a given moment and thus likely metabolically active (Anderson and Parkin, 2007). The comparison of the DNA and RNA communities can also help to answer the question how well are the metabolically active microbial taxa represented in the common studies using the sequencing of soil DNA.

The aim of this work was to demonstrate how the DNA and RNA communities differ and what part of the total community is metabolically active at a given moment. The study was performed at the beginning of winter under freshly fallen snow to target the period when decomposition processes in soil prevail. In a mountainous *Picea abies* forest in central Europe where mycorrhizal fungi have a major role, the winter period without photosynthate flow is expected to show increased activity of decomposer fungal species. Litter and the organic horizons were studied separately because fungal and bacterial communities were previously found to differ between these horizons due to the differences in nutrient availability and the presence of root-associated microorganisms (O'Brien *et al.*, 2005; Lindahl *et al.*, 2007; Šnajdr *et al.*, 2008).

We expected that during the decomposition period, decomposer microorganisms will be transcriptionally

active and a large proportion of decomposition-related genes will be expressed by dominant taxa of microbial decomposers. In order to specifically target an important decomposition process, the gene and transcriptome pools of the fungal cellobiohydrolase (exocellulase) gene *cbhI* sequences were compared, as its gene product catalyses the rate-limiting step in the decomposition of cellulose, the most abundant biopolymer in the ecosystem (Baldrian and Valášková, 2008). The *cbhI* gene occurs in both *Ascomycota* and *Basidiomycota*, and it is also common in the genomes of saprotrophic fungi (Edwards *et al.*, 2008; Ohm *et al.*, 2010). While some recent studies showed that expression of eukaryotic decomposition-related genes in soils can be analysed (Luis *et al.*, 2005; Kellner and Vandenbol, 2010), only the DNA/RNA approach can answer the questions on the diversity of decomposer communities and the proportion of expressed genes. Since cellulose is present in both the litter and soil organic horizon (Šnajdr *et al.*, 2011), the same cellulose-decomposing microorganisms can be present and potentially active in both horizons. However, the higher amount of cellulose in litter likely supports higher diversity of cellulose decomposers.

Materials and methods

Study site, sample collection and soil analysis
Study area was located in the highest altitudes (1170–1200 m) of the Bohemian Forest mountain range (Central Europe) and was covered by an unmanaged spruce (*P. abies*) forest (49°02.64 N, 13°37.01 E). Sampling was performed in late October 2009 under freshly fallen snow (8–12 cm, 3 days after the snowfall) at −5 °C. At four sites, located 250 m from each other, six topsoil samples located around the circumference of a 4-m-diameter circle were collected. Litter horizon (L) and organic (humic) horizon (H) material were separately pooled. After removal of roots, L material was cut into 0.5 cm pieces and mixed; H material was passed through a 5-mm sterile mesh and mixed. Aliquots for nucleic acids extraction were immediately frozen and stored in liquid nitrogen. Samples for phospholipid fatty acid and ergosterol analysis were frozen and stored at −45 °C until analysis. Enzyme assays were performed within 48 h in samples kept at 4 °C in soil homogenates (Štursová and Baldrian, 2011). Dry mass content was measured after drying at 85 °C, organic matter content after burning at 650 °C and pH was measured in distilled water (1:10). Soil C and N content was measured using an elemental analyser.

Quantification of microbial biomass
Phospholipid fatty acid was extracted by chloroform–methanol–phosphate buffer, subjected to alkaline methanolysis and free methyl esters were analysed by GC-MS (Šnajdr *et al.*, 2008). Fungal biomass was quantified based on 18:2ω6,9 content, and bacterial biomass as the sum of bacteria-specific phospholipid fatty acid (Bååth and Anderson, 2003). Total ergosterol was extracted with 10% KOH in methanol and analysed by HPLC (Šnajdr *et al.*, 2008). Partial bacterial and fungal rDNAs were quantified by qPCR using 1108f and 1132r primers for bacteria (Wilmotte *et al.*, 1993; Amann *et al.*, 1995) and ITS1/qITS2* primers for fungi (White *et al.*, 1990; Šnajdr *et al.*, 2011).

Nucleic acid extraction and reverse transcription
RNA and DNA were co-extracted using the RNA PowerSoil Total RNA Isolation Kit and DNA Elution Accessory Kit (MoBio Laboratories, Carlsbad, CA, USA) combined with the OneStep PCR Inhibitor Removal Kit (Zymo Research, Irvine, CA, USA). Three soil aliquots (3 × 3 g of material) were extracted per sample. Extracted RNA was treated with DNase I and 1 μg was reverse transcribed using M-MLV Reverse Transcriptase (Invitrogen, Carlsbad, CA, USA) and random hexamer primers. Samples were designated as LD = litter DNA, LR = litter cDNA, HD = humic horizon DNA and HR = humic horizon cDNA.

Tag-encoded amplicon pyrosequencing and sequence analysis
The eubacterial primers eub530F/eub1100aR (modified from Dowd *et al.*, 2008) were used to amplify the V4–V6 region of bacterial 16S rDNA and the fungi-specific primers ITS1/ITS4 (White *et al.*, 1990) were used to amplify the ITS1, 5.8S rDNA and ITS2 regions of fungal rDNA. Primers *cbhI*F and *cbhI*R (Edwards *et al.*, 2008) were used to amplify a partial sequence of fungal cellobiohydrolase I. Primers for tag-encoded 454-Titanium pyrosequencing contained in addition sample tags separated from primers by spacers and Titanium A or B adaptors (Roche, Basel, Switzerland). Spacer sequences were designed to contain a trinucleotide, absent in all GenBank sequences at this position to avoid preferential amplification of some targets (Parameswaran *et al.*, 2007). Primer pairs were designed using OligoCalc (http://www.basic.northwestern.edu/biotools/oligocalc.html) and tested by cloning/sequencing. Tags and spacer sequences of all composite primers used for tag-encoded amplicon pyrosequencing in this study can be found in Supplementary Information.

PCR amplifications were performed in two steps. In the first step, each of three independent 50 μl reactions per DNA/cDNA sample contained 5 μl of 10 × polymerase buffer, 3 μl of 10 mg ml^{-1} bovine serum albumin, 2 μl of each primer (0.01 mM), 1 μl of PCR Nucleotide Mix (10 mM), 1.5 μl polymerase (2 U μl^{-1}; Pfu DNA polymerase:DyNAZyme II DNA polymerase, 1:24) and 2 μl of template DNA. Cycling conditions were 94 °C for 5 min; 35 cycles of 94 °C for 1 min, 62 °C for 50 s, 72 °C for 30 s, followed by

72 °C for 10 min for primers eub530F/eub1100aR; 94 °C for 5 min; 35 cycles of 94 °C for 1 min, 55 °C for 1 min, 72 °C for 1 min, followed by 72 °C for 10 min for primers ITS1/ITS4; 94 °C for 3 min; 40 cycles of 94 °C for 30 s, 51 °C for 45 s, 72 °C for 1 min 30 s, followed by 72 °C for 15 min for primers *cbhI*F/*cbhI*R. Pooled PCR products were purified using the Wizard SV Gel and PCR Clean-Up System (Promega, Madison, WI, USA). In all, 100 ng DNA was used as template in the second PCR performed under the same conditions except that fusion primers were used and cycle number was 10. PCR products were separated by electrophoresis and gel purified using the Wizard SV Gel and PCR Clean-Up System. DNA was quantified using ND1000 (NanoDrop, Wilmington, DE, USA), an equimolar mix of PCR products from all samples was made for each primer pair and the pooled products were mixed in a molar ratio of 12:4:1 (bacterial:fungal: *cbhI* amplicons). The mixture was subjected to sequencing on a GS FLX Titanium platform (Roche).

The pyrosequencing resulted in 329 820 reads of sufficient quality and a length >200 bases. Pyrosequencing noise reduction was performed using the Denoiser 0.851 (Reeder and Knight, 2010) and chimeric sequences were detected using UCHIME (Edgar, 2010) and deleted. In fungal community analyses, sequences >380 bases were used that contained the ITS1 region, 5.8S rDNA and a significant part of the ITS2 region. These sequences were truncated to 380 bases, clustered using CD-HIT (Li and Godzik, 2006) at 97% similarity (O'Brien *et al*., 2005) to yield Operational Taxonomic Units (OTUs) and consensus sequences were constructed for all OTUs. PlutoF pipeline (Tedersoo *et al*., 2010) was used to generate best species hits. In bacterial analysis, sequences of 350 bases were clustered at a 97% similarity and Ribosomal Database Project (Cole *et al*., 2009) as well as BLASTn hits against GenBank were used to generate best hits (Altschul *et al*., 1997). DNA/RNA ratio was calculated as sum of sequences derived from DNA divided by the sum of all sequences, and the L/H ratio was calculated similarly. Clusters of *cbhI* sequences were constructed using 400-base sequences at 96% similarity. Intron positions were recorded and introns removed from the DNA-derived sequences and DNA and cDNA-derived clusters were merged. For identification, *cbhI* sequences were retrieved from GenBank and also obtained by the analysis of isolates or cultured strains from the studied ecosystem by cloning/sequencing. Nucleotide sequences of OTUs with abundances over 0.3% were translated into amino-acid sequences in Bionumerics 7.0 (Applied Maths, Sint-Martens-Latem, Belgium).

Rarefaction and diversity analyses on OTUs/clusters were performed at 8500 bacterial, 1000 fungal or 350 *cbhI* sequences per sample, to eliminate the effect of sampling effort and used for clustering as described above. Richness and diversity indices were calculated using EstimateS 8.00 (http://viceroy.eeb.uconn.edu/estimates) and quality of Chao1 estimates was evaluated according to Kemp and Aller (2004).

One-way analysis of variance (ANOVA) with the Fisher's LSD *post hoc* test was used to analyse the statistical differences among treatments. To analyse the differences in bacterial and fungal communities and the *cbhI* sequences, principal component analysis was run with abundance data of all OTUs or clusters with >0.3% abundance. PC1 and PC2 loads were subjected to ANOVA with the Fisher's LSD *post hoc* test. Differences at $P<0.05$ were regarded as statistically significant.

Results and discussion

Microbial communities in P. abies *topsoil are diverse and vertically stratified*

The topsoil of the *P. abies* forest was strongly acidic (pH 3.7–3.8) and consisted of a 1–4-cm-thick litter horizon (L) and a 2–4-cm-thick organic (humic) horizon (H). The horizons were significantly different with respect to organic matter and C and N contents, and decreasing nutrient availability was reflected by a decrease in both bacterial and fungal biomass contents with depth. The results of qPCR showed a decrease of fungal-to-bacterial rDNA copy number ratio from 1.11 in the L horizon to 0.30 in the H horizon. The cDNA contained 10^5–10^6 copies of the bacterial 16S rRNA gene per nanogram cDNA, but only 10^2–10^4 copies per nanogram cDNA of fungal ITS region (Figure 1). More rapid organic matter transformation occurred in the L horizon than in the H horizon, as documented by higher activities of several extracellular enzymes, especially those hydrolysing glucans (α- and β-glucosidase and cellobiohydrolase; Figure 1).

Bacterial communities analysed at 8500 sequences per sample showed about 1500 OTUs per sample in the L and H horizon-derived DNA (LD and HD) and in the L-derived RNA (LR) samples, about 1200 OTUs were identified in the H-derived RNA (HR). Also, the Chao1 estimator predictions were lower for HR. The RNA-derived communities were less diverse and less even than the DNA-derived communities, particularly in the H horizon (Supplementary Table 1). Principal component analysis followed by ANOVA showed significant differences in community composition among LD, HD, LR and HR ($P<0.00001$ for differences among L and H as well as among DNA/RNA).

In the DNA community, *Steroidobacter* (OTU4) was the most abundant in all L samples. In the H horizon, Gp1 Acidobacterium (OTU1) was the most abundant at three sites and Gp2 Acidobacterium (OTU5) at one site (Supplementary Table 2; Figure 1). Members of 505 genera were found in the entire pooled community, with most of the sequences belonging to Gp1, Gp3 and Gp2 *Acidobacteria* (on average 17.2%, 11.2% and 8.4%,

respectively), the *Actinobacteria Actinoallomurus* (7.1%), *Conexibacter* (1.3%) and *Iamia* (1.1%), and *Proteobacteria Steroidobacter* (5.4%), *Rhodoplanes* (3.3%), *Phenylobacterium* (2.1%), *Desulfomonile* (1.7%) and *Burkholderia* (1.5%; Supplementary Table 2). Of the most abundant OTUs, 33 (*Chondromyces*), 53 and 20 (*Phenylobacterium*) and 39 (*Caulobacteraceae*) were identified as highly enriched in the RNA-derived community and several taxa showed preferential association with either the L or the H horizon (Figure 2; Supplementary Table 2).

Bacterial sequences belonged to 21 phyla, but only 8 were recorded with abundances over 0.1%. In both horizons, *Proteobacteria*, *Acidobacteria* and *Actinobacteria* were dominant, comprising 80–90% of all sequences; this dominance was even stronger in the RNA (Figure 3). In the L horizon, the RNA community was enriched in *Acidobacteria* and *Firmicutes*, while most of the minor phyla were less represented. In the H horizon, *Actinobacteria* were more abundant in the RNA community (Supplementary Table 2; Supplementary Figure 1).

Bacterial abundance and diversity have been reported to decrease with decreasing soil pH (Lauber *et al.*, 2009; Rousk *et al.*, 2010). Despite this, a highly diverse bacterial community was found in our strongly acidic soil. Compared with other soils with pH <4, in which a high degree of dominance by *Acidobacteria*, around 63%, was previously reported (Lauber *et al.*, 2009), the phyla *Proteobacteria*, *Actinobacteria*, *Firmicutes* and *Verrucomicrobia* were more represented in the *P. abies* forest. The bacterial community in the soil of the study area is specific in several aspects. The genus *Chitinophaga*, which was abundant in a previous study that compared different soils (Fulthorpe *et al.*, 2008), was found at a frequency of only 0.1% in this study; the genera *Actinoallomurus* and *Steroidobacter*, ranked among the five most abundant genera in the *P. abies* forest, were not recovered in the previous study.

Deep sequencing analyses of bacterial communities associated with litter have not previously been reported. Here, we show that the litter horizon exhibited higher phylogenetic diversity and a

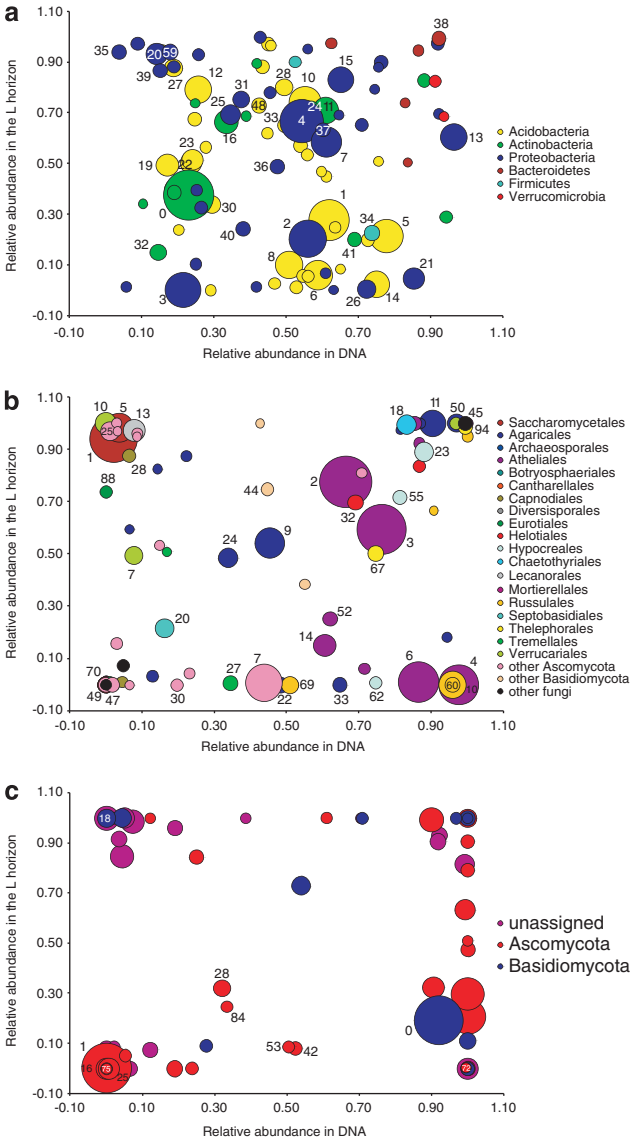


Figure 2 Distribution of major bacterial and fungal OTUs and *cbhI* clusters from *Picea abies* forest topsoil between the L and H horizons and between DNA and RNA. The data represent mean values from four sampling sites. Symbol areas correspond to relative abundance in the combined set of DNA and RNA sequences from both horizons. (**a**) Bacteria, identifications: 0 = *Actinoallomurus*; 1 = Gp1 Acidobacterium; 2 = *Rhodoplanes*; 3 = Rhodospirillales; 4 = *Steroidobacter*; 5 = Gp2 Acidobacterium; 6 = Gp1 Acidobacterium; 7 = *Rhizobiales*; 8 = Gp2 Acidobacterium; 10 = Gp1 Acidobacterium; 11 = Frankineae; 12 = Gp3 Acidobacterium; 13 = *Afipia*; 14 = Gp2 Acidobacterium; 15 = *Burkholderia*; 16 = Actinomycetales; 19 = Gp3 Acidobacterium; 20 = *Phenylobacterium*; 21 = *Desulfomonile*; 22 = Gp3 Acidobacterium; 23 = Gp3 Acidobacterium; 24 = *Ferrithrix*; 25 = Acetobacteraceae; 26 = Rhizobiales; 27 = Gp3 Acidobacterium; 28 = Gp1 Acidobacterium; 30 = Gp3 Acidobacterium; 31 = *Acidisphaera*; 32 = *Actinoallomurus*; 33 = Gp1 Acidobacterium; 34 = *Sporomusa*; 35 = *Chondromyces*; 36 = Acetobacteraceae; 37 = *Steroidobacter*; 38 = Chitinophagaceae; 39 = Caulobacteraceae; 40 = Rhizobiales; 41 = *Mycobacterium*; 48 = Gp1 Acidobacterium; 59 = *Phenylobacterium*. (**b**) Fungi, OTU identifications: 1 = Ascomycete; 2 = *Tylospora fibrillosa*; 3 = *Piloderma*; 4 = *Piloderma*; 5 = Ascomycete; 6 = *Tylospora asterophora*; 7 = *Cenococcum geophilum*; 8 = *Verrucaria*; 9 = *Hygrophorus olivaceoalbus*; 10 = *Russula cyanoxantha*; 11 = *Cortinarius biformis*; 13 = *Lecanora*; 14 = *Tylospora fibrillosa*; 18 = *Cladophialophora minutissima*; 20 = *Auriculoscypha*; 22 = *Inocybe*; 23 = Ascomycete; 24 = Basidiomycete; 25 = Ascomycete; 27 = *Cryptococcus podzolicus*; 28 = *Mycocentrospora acerina*; 29 = Ascomycete; 30 = Ascomycete; 32 = *Meliniomyces vraolstadiae*; 33 = *Amanita spissa*; 44 = *Phellopilus*; 45 = Chytridiomycete; 47 = *Alternaria alternata*; 49 = *Cenococcum geophilum*; 50 = *Cortinarius gentilis*; 52 = *Piloderma*; 55 = *Neofusicoccum*; 60 = *Russula cyanoxantha*; 62 = *Trichosporon porosum*; 67 = *Pseudotomentella*; 69 = *Russula cyanoxantha*; 70 = *Mycoarthris*; 88 = *Elaphocordyceps*; 94 = *Tomentella sublilacina*. (**c**) *cbhI*, clusters with sequence similarities to genes of known *cbhI* producers: 0 = *Mycena*; 1 = *Xylariales* spp.; 16 = *Phacidium*; 18 = *Mycena*; 25 = *Phacidium*; 28 = *Xylariales* spp.; 42 = *Phialophora*; 53 = *Xylariales* spp.; 72 = *Phialophora*; 75 = *Ceuthospora* and 84 = *Phialophora* sp.

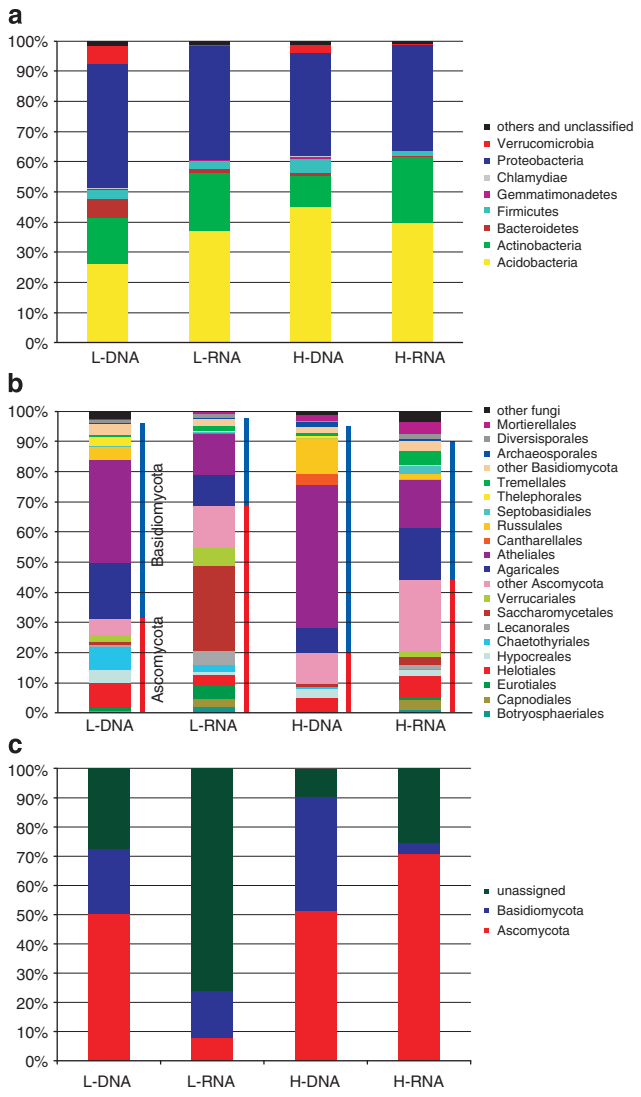


Figure 3 Phylogenetic assignment of bacterial, fungal and *cbhI* sequences from *Picea abies* forest topsoil. The data represent mean values from four study sites. (**a**) Bacteria, (**b**) fungi and (**c**) *cbhI* sequences.

reduced proportion of *Acidobacteria* in the total community. In the most abundant phylum, *Acidobacteria*, the members of Gp1, Gp2 and Gp3 detected were also reported to occur in other low pH soils (Jones *et al.*, 2009). Genome sequencing of Gp1 and Gp3 *Acidobacteria* and isolate culturing showed that these bacteria are able to decompose a variety of polysaccharides, including cellulose, xylan and chitin, and thus may be involved in decomposition (Ward *et al.*, 2009). Indeed, in our study, the Gp3 *Acidobacteria* were preferentially detected in the RNA, which indicates their activity during the decomposition period. The ecology of the other most abundant bacterial genera is unclear. Members of the genus *Actinoallomurus* (formerly belonging to *Actinomadura*) were repeatedly isolated from soils or litter, and some species from this genus are root endophytes. The genus *Phenylobacterium* contains bacteria from upper aerobic soil horizons capable of phenolic compound degradation.

Fungal communities sampled at 1000 randomly selected sequences per sample had between 150 and 220 OTUs with no significant differences between the DNA and RNA samples. The Chao1 estimates predicted higher diversity in LD and HD than in LR and HR (Supplementary Table 1). Principal component analysis followed by ANOVA showed significant differences in community composition among LD, HD, LR and HR ($P<0.0005$ for differences among L and H and $P<0.044$ for DNA/RNA).

In the DNA community, *Piloderma* sp. (OTU3) was dominant at two sites and *Tylospora fibrillosa* (OTU2) and *Cortinarius biformis* (OTU11) were each dominant at one site in the L horizon. In the H horizon, each of the sites was dominated by a different OTU (*Tylospora fibrillosa*, OTU2; *Tylospora asterophora,* OTU6; *Russula cyanoxantha*, OTU10; and *Piloderma* sp., OTU4). OTUs with the closest similarity to 422 different genera were recorded, the most abundant being *Tylospora* (14.8% of all sequences), *Piloderma* (12.8%), *Russula* (4.4%), *Cenococcum* (4.2%), *Cortinarius* (3.9%), *Hygrophorus* (2.9%), *Cladophialophora* (2.4%), *Amanita* (1.8%), *Cadophora* (1.7%), *Mortierella* (1.6%) and *Verrucaria* (1.6%; Supplementary Table 3). The distribution of the abundant OTUs among the L and H horizons and the DNA and RNA communities shows a strict confinement of many OTUs to either the L or the H horizon (Figure 2; Supplementary Table 3). Several of the most abundant OTUs were highly enriched in the RNA community (Supplementary Table 3).

Fungal sequences belonged mainly to Dikarya (53.5% *Basidiomycota* and 41.1% *Ascomycota*). *Glomeromycota* were represented by 2.24% of the sequences, *Mucoromycotina* by 1.77%, and *Chytridiomycota* by 0.73 (Supplementary Table 3). For several groups of phylogenetically related OTUs abundant in the ecosystem, no close sequence of an isolated strain was available. Some of these, including also the putative members of basal fungal lineages, exhibited high abundance in the RNA samples (Supplementary Figure 2). Members of the orders *Atheliales*, *Agaricales*, *Helotiales*, *Chaetothyriales* and *Russulales* were most abundant in the soil DNA; several minor orders, including *Botryosphaeriales*, *Lecanorales* and *Eurotiales* in the L horizon and *Tremellales* and *Capnodiales* in the H horizon, were infrequent in the DNA communities but highly abundant among the RNA sequences (Figure 3).

The composition of fungal communities has been previously shown to differ substantially between litter and organic horizons, while deeper soil horizons showed greater similarity (O'Brien *et al.*, 2005; Lindahl *et al.*, 2007). In several forest types, this is due to the higher abundance of saprotrophic fungi in litter and the dominance of ectomycorrhizal species

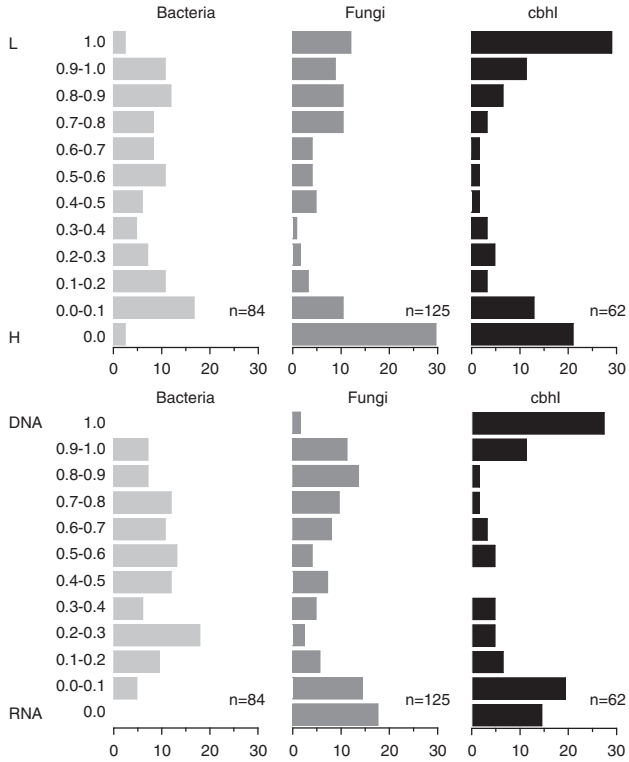


Figure 4 Distribution of bacterial and fungal OTUs and *cbhI* clusters from *Picea abies* forest soil among the L and H horizons and among DNA and RNA, in percents. Based on the data for OTUs/clusters with abundance >0.1%. The *y* axis represents relative share of transcripts in the L horizon or in the DNA. The values 0.0 and 1.0 represent OTUs/clusters present only in either the H or L horizon, in RNA or in the DNA.

in deeper soil (Lindahl *et al*., 2007; Edwards and Zak, 2010). Although our results showed that ectomycorrhizal fungi were highly dominant in both horizons (most likely due to the shallow rooting of *P. abies*), we also confirmed profound differences between the two horizons; 42% of abundant species were only recovered from either the L or the H horizon (Figure 4). The most abundant genera of ectomycorrhizal fungi found in this study, *Piloderma* and *Tylospora* spp., are also the most abundant in the boreal *P. abies* forests in Finland and Sweden (Rosling *et al*., 2003; Korkama *et al*., 2006; Wallander *et al*., 2010).

Ecological functions could be reliably assigned to 64–94% of the abundant members of the fungal community (Supplementary Table 3). Among these, ectomycorrhizal fungi dominated all communities, representing 83% of the sequences in LD, 95% in HD, 66% in LR and 69% in HR (Supplementary Figure 2C). During a period when decomposition processes prevail, a substantial reduction in the activity of ectomycorrhizal fungi compared with saprotrophs is expected (Yarwood *et al*., 2009; Lindahl *et al*., 2010). Saprotrophic and parasitic species were indeed significantly more represented in the RNA communities, the ratio of mycorrhizal/saprotrophic + parasitic fungi being 6.9 and 2.9 in the LD and LR and 21.2 and 2.3 in the HD and HR, respectively. Notably, we found many fungal sequences belonging to lichen-forming fungi and detected these preferentially in the RNA community (9.3% in LR).

Ecology of soil bacteria and fungi is largely different

In agreement with previous studies on the vertical stratification of soil decomposition processes (Wittmann *et al*., 2004; Šnajdr *et al*., 2008), we show that the L and H horizons differ significantly in both the total and relative amounts of bacterial and fungal biomasses (Figure 1). According to the phospholipid fatty acid/biomass C conversion factors (Anderson and Parkin 2007), the L horizon contained 2.9 times more fungal than bacterial biomass (10.7 and 3.7 mg biomass C per gram, respectively). In the H horizon, fungal and bacterial biomasses were equal (1.50 and 1.58 $mg\,g^{-1}$). The diversity of bacterial populations was considerably higher than that of fungal populations. When 1000 randomly selected sequences were analysed, 302–366 OTUs were found in bacteria, compared with 141–236 in fungi.

While the diversity estimates for DNA- and RNA-derived communities of bacteria were similar, the Chao1 estimators for fungal communities surprisingly showed that a more diverse community is revealed when RNA is analysed (Supplementary Table 1). The DNA- and RNA-derived communities of bacteria largely overlapped, and among the abundant OTUs none was found exclusively either in DNA or RNA. By contrast, 18% of fungal OTUs were found only in the RNA community, and 2% were found exclusively in the DNA community. Among *cbhI* sequences, indicating the presence of cellulolytic members of the fungal community, 27% were found only in DNA and 15% only in RNA (Figure 4). These data show that the DNA sequencing approaches miss a significant and functionally relevant part of microbial communities and our current knowledge largely based on this approach is incomplete. The high RNA/DNA ratios for some microbial taxa show that species with low abundance can be highly active.

Abundances of the major bacterial and fungal OTUs in different soils have been reported to be 2–3% and 7–17%, respectively (Fierer *et al*., 2007; Buée *et al*., 2009). In this study, the first to report on bacteria and fungi from the same soil, bacterial communities also showed higher evenness than fungal communities. The most abundant bacterial OTU accounted for 5–7% of all sequences, while the dominant fungal OTU in our ectomycorrhiza-dominated ecosystem represented up to >30% of all sequences. Between 30 and 60 of the most abundant bacterial and 6 and 22 of the most abundant fungal OTUs made up 50% of their respective communities (Supplementary Table 1).

Most bacterial OTUs with abundance >0.3% were recovered from all study sites. In contrast,

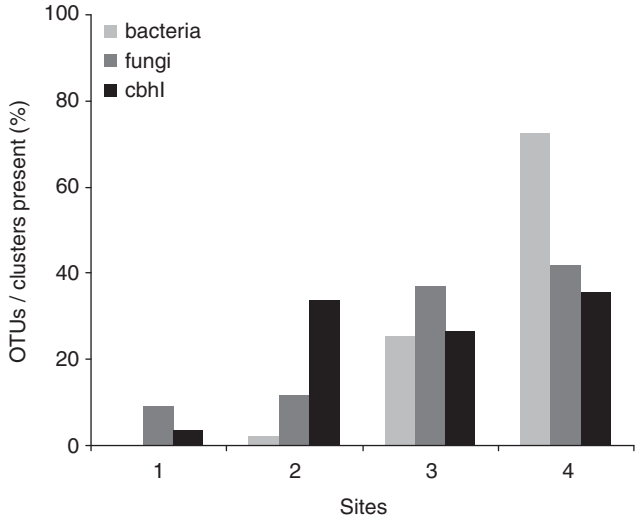


Figure 5 Number of sites in the *Picea abies* forest soil where DNA of dominant bacterial and fungal OTUs and *cbhI* clusters were detected. Only OTUs/clusters with abundance >0.3% in the DNA were considered; $n=50$ for bacteria, $n=54$ for fungi and $n=56$ for *cbhI*. To correct for the same sampling depth, 350 sequences of each target were randomly selected from each sample.

majority of fungal OTUs and *cbhI* clusters were found only at one to three sites (Figure 5). Also, the abundance of bacteria across study sites varied considerably less than that of fungi; the mean coefficient of variation was 0.56 for bacteria and 1.30 for fungi. This applied even in the case of the most abundant fungal taxa, for example, the second-most abundant fungal OTU was absent at one study site. The uneven spatial distribution of fungi in the ecosystem is likely a consequence of a combination of factors including the size of organisms, mobility and the association of many taxa with large nutrient patches or plant roots. This is supported by the fact that fungi forming large mycelial systems (for example, *Russulales* and *Agaricales*) showed higher variation in abundance than species with limited mycelia (for example, *Mortierellales* and *Archaeosporales*).

The affinity of microbial taxa and *cbhI* clusters for either the L or H horizon and their differential abundance in DNA versus RNA points to their different niches and ecological roles (Supplementary Tables 2–4). Over 60% of dominant fungal OTUs and as much as 74% of *cbhI* clusters showed 10-fold enrichment in either the L or H horizons. In contrast, vertical stratification was less distinct in bacterial OTUs (Figure 4). This observation further stresses the importance of fungi in shaping the spatial structure of the forest floor.

Cellobiohydrolase genes exhibit high diversity in the soil metagenome and metatranscriptome

Total RNA extracted from soils contains <10% mRNA (Urich *et al.*, 2008); even after mRNA enrichment it still contains both rRNA and transcripts of abundant genes such as those encoding ribosomal proteins (Bailly *et al.*, 2007). This leads to low recovery of targeted sequences of functional genes by shotgun sequencing approaches. For example, only nine gene clusters of denitrification genes were obtained from 77 000 metagenome-derived clones (Demaneche *et al.*, 2009). Amplification of target sequences is thus the only way to assess the diversity of functional genes.

Here, we show for the first time that a single eukaryotic functional gene can be analysed at a depth that allows diversity estimation; a reliable Chao1 diversity prediction of 46 ± 9 *cbhI* clusters per sample was obtained for the HR sample. When samples from all sites were analysed together, a total of 456 clusters were predicted for LD, 344 for HD, 201 for LR and 99 for HR, with sufficient sampling effort for LR and HR. Approximately 40% and 25% of sequences present in the DNA were being transcribed in the L and H horizons, respectively. Because most of the analysed fungi harbour more than one *cbhI* gene (Baldrian and Valášková, 2008; Edwards *et al.*, 2008), the diversity of cellulolytic fungi in forest topsoil can be estimated only roughly at 50–300. This means that a considerable proportion of the fungal community transcribes or at least harbours the *cbhI* gene. Almost all *cbhI* clusters showed distinct association with either the L or H horizons, indicating only a minor overlap of cellulolytic fungal communities between horizons. Principal component analysis followed by ANOVA showed significant differences in *cbhI* pool composition among LD, HD, LR and HR ($P<0.0019$ for differences among L and H and $P<0.0054$ for DNA/RNA).

Current attempts to assign fungal producers to the sequences of functional genes derived from soil metagenomes suffer from a lack of sequence information in public databases. Because of this, the closest hits for most *cbhI* clusters in public databases were rather distant, with only three clusters showing >96% similarity. By sequencing *cbhI* genes from fungi occurring in the soil, we were able to identify the taxonomic affiliation of producers for 13 additional clusters (Supplementary Table 4). Some of the most abundant *cbhI* sequences were transcribed by fungi with low abundances in the ecosystem (for example, *Mycena* sp. and *Xylariales*), showing the importance of low-abundance species for cellulose hydrolysis. Phylogeny trees constructed using cDNA and peptide sequences (Supplementary Figure 3) allowed coarse taxonomic placement of producers for 24–90% of dominant *cbhI* clusters into either *Ascomycota* or *Basidiomycota* (Figure 3). In the H horizon, where 75–90% of sequences were assigned, genes of both taxa were equally present, but the transcripts were mainly of ascomycetous origin (>95% of assigned sequences).

To match the sequences of genes and transcripts, introns were removed from DNA sequences (Supplementary Figure 3). There were 28 DNA clusters

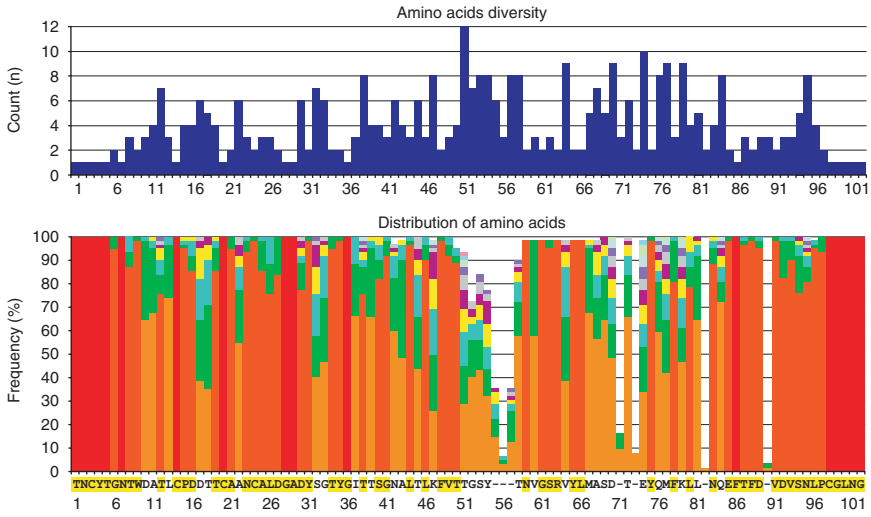


Figure 6 Diversity of amino-acid composition and frequency of alternative amino acids over the length of a *cbhI* internal peptide obtained by translation of the 62 most abundant *cbhI* sequences detected in *Picea abies* forest soil. Abundance of alternative amino acids at each position is colour coded; the identity of the most abundant amino acid at each position is indicated, and consensus amino acids (>75%) are highlighted.

containing one intron and one containing four introns. The internal peptide of *cbhI* obtained after intron removal varied in length and contained both conserved and highly variable regions (Figure 6). The consensus sequence derived from this study differed from the one based on published *cbhI* sequences from isolated fungal strains at 4 amino-acid positions out of 101 (Edwards *et al.*, 2008). Thus, the depth of environmental amplicon sequencing may contribute to the construction and evaluation of better primers and qPCR probes for targeted functional genes.

Conclusions

Much of what is currently known about the ecology of soil microbial communities has been inferred from studies targeting DNA. Despite similar diversity of microbial communities based on DNA and RNA analysis, the fact that several major fungal OTUs were found exclusively in the RNA pool and that several active bacterial OTUs exhibited low abundance in the DNA pool demonstrates the limitations of DNA-based surveys, which likely miss considerable portions of active microbial populations. In the soil ecosystem, bacterial and fungal communities differ in their spatial distributions with fungal taxa more distinctly confined to either the litter or the organic horizon of soil and more heterogeneously distributed in the ecosystem. The diversity and distribution of functional genes responsible for important biogeochemical processes and consequently of their producers can be efficiently targeted by amplicon sequencing. Low abundance of several fungal taxa highly expressing the *cbhI* gene suggests that these species are highly important for decomposition.

Acknowledgements

This work was supported by the Czech Science Foundation (526/08/0751), by the Ministry of Education, Youth and Sports of the Czech Republic (LC06066) and by the Institutional Research Concept of the Institute of Microbiology of the ASCR in Prague (AV0Z50200510).

References

Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W *et al.* (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25**: 3389–3402.

Amann RI, Ludwig W, Schleifer KH. (1995). Phylogenetic identification and *in situ* detection of individual microbial cells without cultivation. *Microbiol Rev* **59**: 143–169.

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

Bååth E, Anderson TH. (2003). Comparison of soil fungal/bacterial ratios in a pH gradient using physiological and PLFA-based techniques. *Soil Biol Biochem* **35**: 955–963.

Bailey VL, Smith JL, Bolton H. (2002). Fungal-to-bacterial ratios in soils investigated for enhanced C sequestration. *Soil Biol Biochem* **34**: 997–1007.

Bailly J, Fraissinet-Tachet L, Verner M-C, Debaud J-C, Lemaire M, Wesolowski-Louvel M *et al.* (2007). Soil eukaryotic functional diversity, a metatranscriptomic approach. *ISME J* **1**: 632–642.

Baldrian P, Valášková V. (2008). Degradation of cellulose by basidiomycetous fungi. *FEMS Microbiol Rev* **32**: 501–521.

Bowker MA, Maestre FT, Escolar C. (2010). Biological crusts as a model system for examining the biodiversity-ecosystem function relationship in soils. *Soil Biol Biochem* **42**: 405–417.

Buée M, Reich M, Murat C, Morin E, Nilsson RH, Uroz S *et al.* (2009). 454 Pyrosequencing analyses of forest soils reveal an unexpectedly high fungal diversity. *New Phytol* **184**: 449–456.

Cole JR, Wang Q, Cardenas E, Fish J, Chai B, Farris RJ *et al.* (2009). The Ribosomal Database Project: improved alignments and new tools for rRNA analysis. *Nucleic Acids Res* **37**: D141–D145.

Demaneche S, Philippot L, David MM, Navarro E, Vogel TM, Simonet P. (2009). Characterization of denitrification gene clusters of soil bacteria via a metagenomic approach. *Appl Environ Microbiol* **75**: 534–537.

Dowd S, Callaway T, Wolcott R, Sun Y, McKeehan T, Hagevoort R *et al.* (2008). Evaluation of the bacterial diversity in the feces of cattle using 16S rDNA bacterial tag-encoded FLX amplicon pyrosequencing (bTEFAP). *BMC Microbiol* **8**: 125.

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

Edwards IP, Upchurch RA, Zak DR. (2008). Isolation of fungal cellobiohydrolase I genes from sporocarps and forest soils by PCR. *Appl Environ Microbiol* **74**: 3481–3489.

Edwards IP, Zak DR. (2010). Phylogenetic similarity and structure of Agaricomycotina communities across a forested landscape. *Mol Ecol* **19**: 1469–1482.

Fierer N, Breitbart M, Nulton J, Salamon P, Lozupone C, Jones R *et al.* (2007). Metagenomic and small-subunit rRNA analyses reveal the genetic diversity of bacteria, archaea, fungi, and viruses in soil. *Appl Environ Microbiol* **73**: 7059–7066.

Fulthorpe RR, Roesch LFW, Riva A, Triplett EW. (2008). Distantly sampled soils carry few species in common. *ISME J* **2**: 901–910.

Hawksworth DL. (2001). The magnitude of fungal diversity: the 1.5 million species estimate revisited. *Mycol Res* **105**: 1422–1432.

Jones RT, Robeson MS, Lauber CL, Hamady M, Knight R, Fierer N. (2009). A comprehensive survey of soil acidobacterial diversity using pyrosequencing and clone library analyses. *ISME J* **3**: 442–453.

Kellner H, Vandenbol M. (2010). Fungi unearthed: transcripts encoding lignocellulolytic and chitinolytic enzymes in forest soil. *PLoS One* **5**: e10971.

Kemp PF, Aller JY. (2004). Estimating prokaryotic diversity: when are 16S rDNA libraries large enough? *Limnol Oceanogr Methods* **2**: 114–125.

Korkama T, Pakkanen A, Pennanen T. (2006). Ectomycorrhizal community structure varies among Norway spruce (*Picea abies*) clones. *New Phytol* **171**: 815–824.

Lauber CL, Hamady M, Knight R, Fierer N. (2009). Pyrosequencing-based assessment of soil pH as a predictor of soil bacterial community structure at the Continental Scale. *Appl Environ Microbiol* **75**: 5111–5120.

Li W, Godzik A. (2006). Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* **22**: 1658–1659.

Lindahl BD, Ihrmark K, Boberg J, Trumbore SE, Högberg P, Stenlid J *et al.* (2007). Spatial separation of litter decomposition and mycorrhizal nitrogen uptake in a boreal forest. *New Phytol* **173**: 611–620.

Lindahl BD, de Boer W, Finlay RD. (2010). Disruption of root carbon transport into forest humus stimulates fungal opportunists at the expense of mycorrhizal fungi. *ISME J* **4**: 872–881.

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

McGuire KL, Treseder KK. (2010). Microbial communities and their relevance for ecosystem models: decomposition as a case study. *Soil Biol Biochem* **42**: 529–535.

Myneni RB, Dong J, Tucker CJ, Kaufmann RK, Kauppi PE, Liski J *et al.* (2001). A large carbon sink in the woody biomass of Northern forests. *Proc Natl Acad Sci USA* **98**: 14784–14789.

O'Brien HE, Parrent JL, Jackson JA, Moncalvo JM, Vilgalys R. (2005). Fungal community analysis by large-scale sequencing of environmental samples. *Appl Environ Microbiol* **71**: 5544–5550.

Ohm RA, de Jong JF, Lugones LG, Aerts A, Kothe E, Stajich JE *et al.* (2010). Genome sequence of the model mushroom *Schizophyllum commune*. *Nat Biotechnol* **28**: 957–963.

Parameswaran P, Jalili R, Tao L, Shokralla S, Gharizadeh B, Ronaghi M *et al.* (2007). A pyrosequencing-tailored nucleotide barcode design unveils opportunities for large-scale sample multiplexing. *Nucleic Acids Res* **35**: e130.

Reed HE, Martiny JBH. (2007). Testing the functional significance of microbial composition in natural communities. *FEMS Microbiol Ecol* **62**: 161–170.

Reeder J, Knight R. (2010). Rapid denoising of pyrosequencing amplicon data: exploiting the rank-abundance distribution. *Nat Methods* **7**: 668–669.

Roesch LFW, Fulthorpe RR, Riva A, Casella G, Hadwin AKM, Kent AD *et al.* (2007). Pyrosequencing enumerates and contrasts soil microbial diversity. *ISME J* **1**: 283–290.

Rosling A, Landeweert R, Lindahl BD, Larsson KH, Kuyper TW, Taylor AFS *et al.* (2003). Vertical distribution of ectomycorrhizal fungal taxa in a podzol soil profile. *New Phytol* **159**: 775–783.

Rousk J, Baath E, Brookes PC, Lauber CL, Lozupone C, Caporaso JG *et al.* (2010). Soil bacterial and fungal communities across a pH gradient in an arable soil. *ISME J* **4**: 1340–1351.

Strickland MS, Lauber C, Fierer N, Bradford MA. (2009). Testing the functional significance of microbial community composition. *Ecology* **90**: 441–451.

Šnajdr J, Dobiášová P, Větrovský T, Valášková V, Alawi A, Boddy L *et al.* (2011). Saprotrophic basidiomycete mycelia and their interspecific interactions affect spatial distribution of extracellular enzymes in soil. *FEMS Microbiol Ecol*, DOI:10.1111/j.1574-6941.2011.01123.x (in press).

Šnajdr J, Valášková V, Merhautová V, Herinková J, Cajthaml T, Baldrian P. (2008). Spatial variability of enzyme activities and microbial biomass in the upper layers of *Quercus petraea* forest soil. *Soil Biol Biochem* **40**: 2068–2075.

Štursová M, Baldrian P. (2011). Effects of soil properties and management on the activity of soil organic matter transforming enzymes and the quantification of soil-bound and free activity. *Plant Soil* **338**: 99–110.

Tedersoo L, Nilsson RH, Abarenkov K, Jairus T, Sadam A, Saar I *et al.* (2010). 454 Pyrosequencing and Sanger sequencing of tropical mycorrhizal fungi provide similar results but reveal substantial methodological biases. *New Phytol* **188**: 291–301.

Urich T, Lanzen A, Qi J, Huson DH, Schleper C, Schuster SC. (2008). Simultaneous assessment of soil microbial

community structure and function through analysis of the meta-transcriptome. *PLoS One* **3**: e2527.

Wallander H, Johansson U, Sterkenburg E, Durling MB, Lindahl BD. (2010). Production of ectomycorrhizal mycelium peaks during canopy closure in Norway spruce forests. *New Phytol* **187**: 1124–1134.

Ward NL, Challacombe JF, Janssen PH, Henrissat B, Coutinho PM, Wu M *et al.* (2009). Three genomes from the phylum acidobacteria provide insight into the lifestyles of these microorganisms in soils. *Appl Environ Microbiol* **75**: 2046–2056.

White TJ, Bruns T, Lee S, Taylor J. (1990). Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. In: Innins MA, Gelfand DH, Sninsky JJ, White TJ (eds). *PCR Protocols*. Academic Press: San Diego, pp 315–322.

Wilmotte A, Van der Auwera G, De Wachter R. (1993). Structure of the 16 S ribosomal RNA of the thermophilic cyanobacterium chlorogloeopsis HTF ('mastigocladus laminosus HTF') strain PCC7518, and phylogenetic analysis. *FEBS Lett* **317**: 96–100.

Wittmann C, Kahkonen MA, Ilvesniemi H, Kurola J, Salkinoja-Salonen MS. (2004). Areal activities and stratification of hydrolytic enzymes involved in the biochemical cycles of carbon, nitrogen, sulphur and phosphorus in podsolized boreal forest soils. *Soil Biol Biochem* **36**: 425–433.

Yarwood SA, Myrold DD, Hogberg MN. (2009). Termination of belowground C allocation by trees alters soil fungal and bacterial communities in a boreal forest. *FEMS Microbiol Ecol* **70**: 151–162.

Supplementary Information accompanies the paper on The ISME Journal website (http://www.nature.com/ismej)

Biol Fertil Soils
DOI 10.1007/s00374-015-1029-9

ORIGINAL PAPER

An in-depth analysis of actinobacterial communities shows their high diversity in grassland soils along a gradient of mixed heavy metal contamination

Tomáš Větrovský[1] · **Petr Baldrian**[1]

Received: 18 March 2015 / Revised: 14 May 2015 / Accepted: 6 June 2015

Abstract Several previous studies indicated that Actinobacteria may be enriched in soils with elevated content of heavy metals. In this study, we have developed a method for the in-depth analysis of actinobacterial communities in soil through phylum-targeted high-throughput sequencing and used it to address this question and examine the community composition in grassland soils along a gradient of heavy metal contamination (Cu, Zn, Cd, Pb). The use of the 16Sact111r primer specific for Actinobacteria resulted in a dataset obtained by pyrosequencing where over 98 % of the sequences belonged to Actinobacteria. The diversity within the Actinobacterial community was not affected by the heavy metals, but the contamination was the most important factor affecting community composition. The most significant changes in community composition were due to the content of Cu and Pb, while the effects of Zn and Cd were relatively minor. For the most abundant actinobacterial taxa, the abundance of taxa identified as members of the genera *Acidothermus*, *Streptomyces*, *Pseudonocardia*, *Janibacter* and *Microlunatus* increased with increasing metal content, while those belonging to *Jatrophihabitans* and *Actinoallomurus* decreased. The genus *Ilumatobacter* contained operational taxonomic units (OTUs) that responded to heavy metals both positively and negatively. This study also confirmed that Actinobacteria appear to be less affected by heavy metals than other bacteria. Because several Actinobacteria were also identified in playing a significant role in cellulose and lignocellulose decomposition in soil, they potentially represent important decomposers of organic matter in such environments.

Keywords Actinobacterial · 16S rDNA · Pyrosequencing · Heavy metal contamination · Soil microbiology · Microbial ecology

Introduction

Recent reports show that in addition to fungi, also bacteria play a significant role in cellulose and lignocellulose decomposition in soil environments (Baldrian et al. 2012; Berlemont and Martiny 2013). Among bacteria, the members of the phylum Actinobacteria show the highest abundance of genes involved in cellulose decomposition in their genomes (Berlemont and Martiny 2013), and many isolates have been shown to be able to efficiently degrade polysaccharides and polyphenols (Abdulla and El-Shatoury 2007; Anderson et al. 2012; Chater et al. 2010; Enkhbaatar et al. 2012; Větrovský et al. 2014; Yin et al. 2010). Bulk lignocellulose materials are more accessible for decomposition by filamentous microorganisms such as fungi and Actinobacteria (de Boer et al. 2005), and efficient growth on straw was indeed demonstrated for environmental isolates of Actinobacteria (Větrovský et al. 2014). Filamentous mycelial growth, the ability to form spores and a powerful secondary metabolism also represent suitable traits for the survival of Actinobacteria under stress conditions such as heavy metal-contaminated soils, for which these bacteria have also developed several resistance mechanisms (Bajkic et al. 2013; Ivshina et al. 2013; Schmidt et al. 2005).

Electronic supplementary material The online version of this article (doi:10.1007/s00374-015-1029-9) contains supplementary material, which is available to authorized users.

✉ Tomáš Větrovský
vetrovsky@biomed.cas.cz

[1] Laboratory of Environmental Microbiology, Institute of Microbiology of the ASCR, v. v. i., Vídeňská 1083, 14220 Prague 4, Czech Republic

Published online: 20 June 2015

This unique combination of traits makes Actinobacteria good candidates for the efficient decomposition of lignocellulose in heavy metal-contaminated areas where they may potentially represent the major metabolically active group. While the abundance of other bacteria phyla typically decreases with increasing heavy metal bioavailability, previous studies demonstrated that the abundance of the Actinobacteria is similar or higher than in clean soils of similar properties (Berg et al. 2012; Gremion et al. 2003).

Recent high-throughput sequencing-based studies of bacterial communities in soil revealed that Actinobacteria can be highly abundant in certain soils, such as those of arid ecosystems, where they often represent over 45 % of the bacterial community (Bachar et al. 2010; Dunbar et al. 1999), while in typical soils, they represent only between <5 and 20 % of the bacterial community (Baldrian et al. 2012; Dominguez-Mendoza et al. 2014; Lauber et al. 2009). Because of the importance of Actinobacteria in soil processes and their varying abundance, previous attempts have been made to analyse the composition of their communities by phylum-targeted approaches. Studies published so far have unfortunately been limited to fingerprinting methods, such as DGGE and T-RFLP, in combination with other molecular identification techniques such as clone sequencing (Liu et al. 2013; Steger et al. 2007; Wang et al. 2013) and do not offer sufficient depth for a detailed analysis of the composition of their communities in soils. However, the massive parallel sequencing that is currently available promises to overcome this limitation.

Through massive parallel sequencing, the aims of this study were to develop a method for an in-depth analysis of actinobacterial communities in soil and to provide an effective tool for the detailed examination of their community composition. This approach was used to describe the actinobacterial community in grassland soils developed at a gradient of heavy metal contamination from the vicinity of a polymetallic smelter. The study sites for this study were selected to represent a long-term gradient of total heavy metal content (Cu, Zn, Cd, Zn, Pb and As) which was due to the variable distance from a local polymetallic smelter. Although the available concentrations of heavy metals may seem to be more relevant for studying biological effects, we have decided to base the analyses on the total heavy metal content for several reasons: (i) the metal availability can be variable in time both short-term and long-term and there is not sufficient information on the history of heavy metal availability in the area; (ii) availability of heavy metals is highly affected by the method used for extraction, for example, in the area of study, two routinely used methods delivered estimates differing almost by one order of magnitude (Mühlbachová et al. 2015), and this makes comparisons to other studies where heavy metal availability was estimated difficult; and (iii) for the area of this study, the correlations of total and available heavy metal contents were found to be both linear and highly significant. Also, the effects of total and available heavy metals on soil functions (respiration, dehydrogenase activity) were both significant (Mühlbachová et al. 2015).

The analysis was aimed to verify the hypothesis that the relative abundance of Actinobacteria increases with the increasing heavy metal availability and to identify the effects of heavy metal content on the diversity and composition of the actinobacterial community. We hypothesised that the increasing heavy metal toxicity only allows the existence of a limited number of resistant actinobacterial taxa, decreasing their overall diversity in soil. This can be potentially of critical importance for soil functioning: several actinobacterial taxa from the area of this study were previously demonstrated to decompose multiple polysaccharides and phenolics and to grow on complex lignocellulose compounds (Větrovský et al. 2014). This indicates their involvement in the decomposition of plant-derived biomass in the studied soil which can be potentially harmed by heavy metal toxicity.

Materials and methods

Study site, soil sample collection and analysis

The study area was located near Příbram, Czech Republic (49° 42′ 22.207″ N, 13° 58′ 27.296″ E). The area was covered by a mixed grassland. The soil was a Cambisol with a pH ranging from 5.5 to 6.5 and a clay/silt/sand ratio of approximately 40:30:30 %. The smelter was in operation from 1786, and it originally mostly worked lead ores. Metal mining ceased in 1972, but the smelter still processes secondary lead and cadmium sources. The present levels of heavy metal contamination are reported from the soils over the period >30 years which makes this area suitable for the exploration of long-term heavy metal contamination (Kalac et al. 1991; Mühlbachová 2011; Mühlbachová et al. 2015).

The sampling was performed in June 2012 and November 2012 to cover potential seasonal differences in community composition. The same five sites, located along a gradient of contamination represented by distances of 300 to 2500 m from the polymetallic smelter, were sampled on both occasions.

At each site, four samples were collected using plastic cores of a 4.5-cm diameter. The samples were transported to the laboratory and stored at 4 °C until processing, which was performed within 24 h. The top 5 cm of each soil core were collected, the roots and stones were removed, and the soil was sieved through a 5-mm sterile sieve and mixed. Sieved samples were freeze-dried and kept at −40 °C until analysis.

DNA was extracted from 300 mg of freeze-dried soil of each sample using the SV method (Sagova-Mareckova et al. 2008) and cleaned using the GeneClean Turbo Kit (MP Biomedicals, Solon, OH, USA). The sample pH was measured in soil water extract (1 g soil/10 ml deionised water), and the soil

moisture content was calculated by estimating the soil dry mass before and after freeze-drying. Oxidizable C (C_{ox}) and total N (N_{tot}) content was measured using an elemental analyser in an external laboratory (Research Institute for Soil and Water Conservation, Prague, Czech Republic). C_{ox} was measured using sulfochromic oxidation (ISO 14235), and nitrogen content was estimated by sulphuric acid mineralisation with the addition of selenium and sodium sulphate and conversion to ammonium ions (ISO 11261), which were measured by the segmented flow analyser (SFA) Skalar. The content of heavy metals (Cd, Cu, Pb, Zn) was measured after decomposition by aqua regia on the atomic absorption spectrometer Varian 240.

Quantification of microbial biomass

Two sets of specific PCR primers were used to quantify the relative amounts of actinobacterial and bacterial DNA: Actino235 (CGCGGCCTATCAGCTTGTTG) (Stach et al. 2003) and Eub518 (ATTACCGCGGCTGCTGG) (Muyzer et al. 1993) for *Actinobacteria* and 1108 F (ATGGYTGTCG TCAGCTCGTG) (Amann et al. 1995) and 1132R (GGGTTG CGCTCGTTGC) (Wilmotte et al. 1993) for bacteria. Real-time PCR assays were based on the method of Leigh et al. (Leigh et al. 2007). Amplifications were performed on a Step One Plus cycler (Applied Biosystems) using optical grade 96-well plates. Each 15 μl reaction mixture contained 7.5 μl SYBR Green Master Mix (Applied Biosystems), 0.6 μl BSA (10 mg/ml), 0.9 μl of each primer, 1.0 μl of template and 4.1 μl of water. The PCR cycling protocol was the same for actinobacterial and bacterial DNA quantification: incubation at 56 °C for 2 min and at 95 °C for 10 min, followed by 40 cycles of amplification consisting of 15 s at 95 °C and 1 min at 60 °C. Cloned fragments of the 16S rRNA gene from *Streptomyces lincolnensis* DNS 40335 were used as standards.

Tag-encoded amplicon pyrosequencing of actinobacterial community

The combination of a universal primer, eub530F (GTGCCA GCMGCNGCGG (Dowd et al. 2008)) and an Actinobacteria-specific primer 16Sact1114r (GAGTTGACCCCGGCRGT, (Kyselkova et al. 2008)) was used to selectively amplify partial 16S sequences of the Actinobacteria. As a proxy of primer coverage of the Actinobacteria-specific primer (i.e. the percentage of actinobacterial sequences in a dataset from which the primer set generates amplicons) and specificity (i.e. the percentage of non-actinobacterial sequences amplified by the primers in other bacterial phyla), the percentages of sequences with primer hits with zero to three mismatches was tested in silico on a 16S rRNA dataset derived from RDP and GenBank covering the primer region (185,400 sequences belonging to 2316 genera), allowing up to three mismatches.

Tag-encoded samples for 454 pyrosequencing were obtained in two steps. In the first step, three PCR reactions were performed independently for each extracted DNA sample using the above primer pair to selectively amplify part of the 16S rRNA gene from Actinobacteria. Each 25 μl reaction mixture contained 16.75 μl H_2O, 2.5 μl 10× buffer for DyNAzyme II DNA polymerase, 1.5 μl 10 mg/ml BSA, 1 μl forward primer (final concentration 10 pmol/μl), 1 μl reverse primer (final concentration 10 pmol/μl), 1 μl template DNA, 0.75 μl 4 % Pfu polymerase/DyNAzyme DNA Polymerase (final concentration 2 U/μl) and 0.5 μl PCR Nucleotide Mix (10 mM). The cycling parameters were 5 min at 94 °C, 35 cycles of 60 s at 94 °C, 60 s at 57 °C and 60 s at 72 °C, followed by 10 min final extension at 72 °C. The size and quality of the PCR products were verified on an agarose gel. The three primary PCR reactions from each sample were pooled, cleaned using the MinElute PCR Purification kit (Qiagen, Hilden, Germany) and concentrated to 20 μl.

Secondary PCR was performed using composite eub530f/eub1100br primers modified from Dowd et al. (2008) as described previously (Baldrian et al. 2012). Each 50 μl reaction mixture contained 38.2 μl H_2O, 5 μl 10× buffer for DyNAzyme II DNA polymerase, 1.5 μl DMSO, 0.4 μl tagged forward primer (final concentration, 10 pmol/μl), 0.4 μl reverse primer (final concentration, 10 pmol/μl), 2 μl template DNA, 1.5 μl 4 % Pfu polymerase/DyNAzyme DNA Polymerase (final concentration, 2 U/μl) and 1 μl PCR Nucleotide Mix (10 mM). Cycling parameters were 5 min at 94 °C, 15 cycles of 60 s at 94 °C, 60 s at 64 °C, 60 s at 72 °C, followed by 10 min final extension at 72 °C. PCR products were checked on an agarose gel and purified using AMPure Beads (Agencourt, Beverly, MA, USA) to remove short fragments. The concentration of the purified PCR product was quantified using the Quant–iT™ PicoGreen ds DNA kit (Invitrogen, Grand Island, NY, USA), and an equimolar mixture of all samples was prepared. The mixture was subjected to electrophoresis, and after separation, the band with appropriate size was excised, purified using the Wizard SV Gel and PCR Clean-Up System (Promega, Madison, WI, USA), denatured (2 min. 95 °C) and again cleaned using AMPure Beads. This product was cleaned using the MinElute PCR Purification kit (Qiagen, Hilden, Germany) and quantified by the Kapa Library Quantification kit (Kapa Biosystems, Woburn, MA, USA). The 454 pyrosequencing was performed on the GS Junior instrument (Roche, Basel, Switzerland).

Bioinformatic analysis

Sequencing data were filtered and trimmed using the pipeline SEED v 1.2.1 (Větrovský and Baldrian 2013a). All of the sequences with mismatches in tags were removed from the dataset. Pyrosequencing noise reduction was performed using the PyroNoise algorithm translation within Mothur v 1.28.0

(Schloss et al. 2009); chimeric sequences were detected using the Uchime implementation in USEARCH v 7.0.1090 (Edgar et al. 2011) and deleted. All sequences were trimmed to 380 bp and clustered into operational taxonomic units (OTUs) using UPARSE implementation in USEARCH 7.0.1090 (Edgar 2013) with a 97 % similarity threshold. The consensus from each OTU was constructed from an MAFFT alignment (Katoh et al. 2009) based on the most abundant nucleotide at each position. For sample comparison and diversity analyses, non-actinobacterial sequences were removed, and the dataset was randomly resampled at the same sampling depth of 400 sequences per sample. The sequence data were deposited in the MG-RAST public database (http://metagenomics.anl.gov/, dataset numbers 4638524.3, 4638526.3, 4638527.3, 4638521.3, 4638519.3, 4638523.3, 4638518.3, 4638525.3, 4638522.3, 4638520.3).

OTU identification, phylogenetic analyses and statistics

The identification of OTUs was performed by BLASTn against a local database derived from the Ribosomal Database Project (Cole et al. 2005) from 25.2.2014. Each OTU was assigned to the taxonomic level of genus (or nearest lower level), comparing the BLASTn best hits with the taxonomic information from the NCBI taxonomy server. A dataset containing only actinobacterial sequences was used for OTU table construction. To reduce the bias caused by different numbers of rRNA gene copies in bacterial genomes, read abundances of each OTU were divided by the copy number of the rRNA genes in the genome of the closest taxon with a complete genome sequence as described previously (Větrovský and Baldrian 2013b).

Phylogenetic analysis and tree construction was performed in MEGA 6 (Tamura et al. 2013), and a circular graphical tree representation was prepared using the program GraPhlAn (http://huttenhower.sph.harvard.edu/graphlan). Statistical tests and principal component analysis were conducted using the software package Statistica 7 (StatSoft, USA) and PAST 3.03 (http://folk.uio.no/ohammer/past/). Differences in site properties were tested by a one-way analysis of variance (ANOVA) followed by the Fisher post hoc test. Because the actinobacterial genome abundance data did not show a normal distribution, the nonparametric Mann-Whitney *U* test was used to compare the abundances of taxa among sites. A Mantel test with 9999 permutations was used to explore the relationships between community similarity and environmental parameters, either the heavy metal content or the soil C_{ox}, N_{tot} and pH values. A Bray-Curtis similarity was used as a measure of similarity in the actinobacterial abundance matrix, while Euclidean distances were used for soil chemistry and heavy metal content data. A Pearson *r* was used as a measure of linear correlation between variables. In all cases, differences of $P<0.05$ were regarded as statistically significant.

Results

In this study, we analysed actinobacterial community composition in five soil samples from heavy metal-contaminated sites near a polymetallic smelter. The sites were located along a gradient of contamination by multiple heavy metals, most importantly, Cd, Cu, Zn and Pb. The concentrations of these metals at site S1, located 2500 m from the smelter, were 1.2±0.1, 9.3±0.1, 84.1±3.4 and 160±2 mg kg^{-1}, respectively, and represented background values for the larger area of several tens of km around the smelter. At the most polluted site S5, located 300 m from the metal smelter, the concentrations of the above heavy metals were 8.3±0.1, 22.0±1.0, 272±15 and 1714±62 mg kg^{-1}, respectively, i.e. 7-, 2.3-, 3.2- and 11-fold increases compared with S1 (Table 1). The other environmental properties were rather similar across the sites: pH values ranged from 5.9 to 6.1, N_{tot} from 0.23 to 0.43 % and C_{ox} from 2.42 to 3.99 %. Soil moisture content measured at the times of sampling ranged between 18 and 32 % and exhibited high variation both within sites and on a seasonal basis. The background site S1 and the most contaminated site S5 did not show significant differences in any of these environmental variables (Table 1).

Quantitative PCR demonstrated that bacterial rRNA gene copy numbers ranged from 0.4 to 3.8×10^9 per gram of soil, and actinobacterial counts were between 0.9 and 8.2×10^8. The lowest mean values were recorded at the two most contaminated sites, S4 and S5, but due to the high variation in the rRNA gene copy numbers within sites, the differences in bacterial and actinobacterial biomasses were not significant among most pairs of sites (Table 1).

To specifically amplify partial 16S sequences of Actinobacteria, nested PCR was performed using the Actinobacteria-specific primer 16Sact1114r. The in silico test preceding the amplification showed that primer 16Sact1114r was highly specific to Actinobacteria, as in total only 1.8 % of matches came from bacterial genera belonging to other phyla when mismatches were not allowed, 3.1 % in the case of one mismatch, 4.7 % for two mismatches and 10.7 % when three mismatches were allowed. The bacterial phyla with the most false positive hits with up to three mismatches were Proteobacteria (8.3 %; 83 sequences), Firmicutes (18.7 %; 72 sequences) and Cyanobacteria (14 %; 19 sequences). This primer also had high coverage, as it amplified over 97.7 % of the actinobacterial genera present in the dataset with three mismatches allowed, 92.1 % for two mismatches, 90.4 % for three mismatches and 85.1 % when no mismatch was allowed (Supplementary Fig. 1).

The high specificity of primer 16Sact1114r towards Actinobacteria was subsequently confirmed by the results of 454 pyrosequencing, where 16Sact1114r was used in combination with the universal bacterial primer eub530f. In total, 36,381 sequences were obtained by 454 pyrosequencing after the

Table 1 Soil properties and diversity of Actinobacteria in grassland soils across a gradient of heavy metal concentrations

Variable	S1	S2	S3	S4	S5
N_{tot} (%)	0.3 ab	0.4 a	0.4 a	0.2 b	0.3 ab
	0.0	*0.0*	*0.1*	*0.0*	*0.0*
C_{ox} (%)	3.8 a	3.6 ab	4.0 a	2.4 b	3.4 ab
	0.5	*0.1*	*0.4*	*0.1*	*0.2*
C/N	10.8 a	8.7 b	9.6 ab	10.7 a	10.9 a
	0.5	*0.3*	*0.4*	*0.1*	*0.2*
Cd (mg/kg)	1.2 a	1.6 a	4.2 b	4.7 b	8.3 c
	0.1	*0.0*	*0.5*	*0.2*	*0.2*
Cu (mg/kg)	9.3 a	10.1 a	14.2 b	21.8 c	22.0 c
	0.1	*1.4*	*1.4*	*1.1*	*2.0*
Pb (mg/kg)	160.5 a	269.5 a	829.0 b	1085.0 c	1713.5 d
	3.9	*11.7*	*187.4*	*32.5*	*123.4*
Zn (mg/kg)	84.1 a	111.8 a	168.2 b	190.5 b	272.4 c
	6.9	*9.2*	*13.6*	*3.5*	*31.3*
pH	6.1	5.9	5.9	5.9	6.0
	0.0	*0.0*	*0.1*	*0.0*	*0.0*
Bacteria (million copies/g of soil)	1682.5 a	2267.1 ab	3789.4 b	410.9 a	458.2 a
	526.6	*681.0*	*632.1*	*100.6*	*112.9*
Actinobacteria (million copies/g of soil)	210.8 ab	486.5 ab	824.0 b	79.6 a	86.5 a
	37.6	*171.7*	*325.6*	*16.5*	*18.3*
Shannon-Wiener diversity index	3.9 a	3.6 b	3.8 a	3.9 a	3.9 a
	0.0	*0.1*	*0.1*	*0.0*	*0.1*
OTU richness	94.5 ab	85.4 b	87.9 ab	96.6 a	98.1 a
	1.9	*4.8*	*4.7*	*1.9*	*2.5*
Evenness	0.86 a	0.80 b	0.85 a	0.86 a	0.85 a
	0.0	*0.0*	*0.0*	*0.0*	*0.0*
Chao-1	150.7 ab	142.1 ab	128.6 a	160.1 ab	164.6 b
	7.4	*14.8*	*10.9*	*9.3*	*8.9*

For each site, the data are expressed as the means from eight samples; standard errors are shown in italics. Statistically significant differences among sites (ANOVA followed by Fisher post hoc test, $P=< 0.05$) are indicated by different letters

removal of low-quality sequences shorter than 380 bases and potentially chimeric sequences. The clustering of filtered sequences at a 97 % similarity threshold yielded 1305 bacterial OTUs (778 singletons, 59.6 % of all OTUs), where the majority were represented by 1155 actinobacterial OTUs (695 singletons, 60.2 %) that gave the best hits in 147 genera. Only 433 sequences (1.7 % of the total) were identified to belong to other bacterial phyla, mostly to Proteobacteria (55 OTUs) and Firmicutes (52 OTUs). An overview of the taxonomy of the sequences and the OTU richness and abundance at the genus level is shown in Fig. 1.

The analysis of the diversity of soil Actinobacteria showed that even at a relatively shallow depth of 400 sequences per sample, a substantial portion of the diversity was recovered: while the Chao-1 estimates of total OTU richness ranged between 130 and 160, the observed species richness was 85–98 so that between 60 and 70 % of expected diversity was recovered (Table 1). Interestingly, neither the diversity nor the evenness of the actinobacterial community differed significantly between the control plot S1 and the most polluted plot S5 (Table 1). In contrast to the predicted decrease of diversity with the increasing content of heavy metals, the Shannon-Wiener diversity index showed a positive correlation with the concentrations of Cu ($P<0.01$), Cd ($P<0.05$) and Pb ($P<0.05$).

In the most abundant actinobacterial OTUs representing at least 1 % of the estimated number of genomes within a sample for at least three samples, the following genera dominated: *Ilumatobacter*, *Nocardioides*, *Mycobacterium*, *Aciditerrimonas*, *Microlunatus*, *Actinoallomurus* and *Pseudonocardia*. Genus *Ilumatobacter* and *Aciditerrimonas* were also represented by the largest number of OTUs within

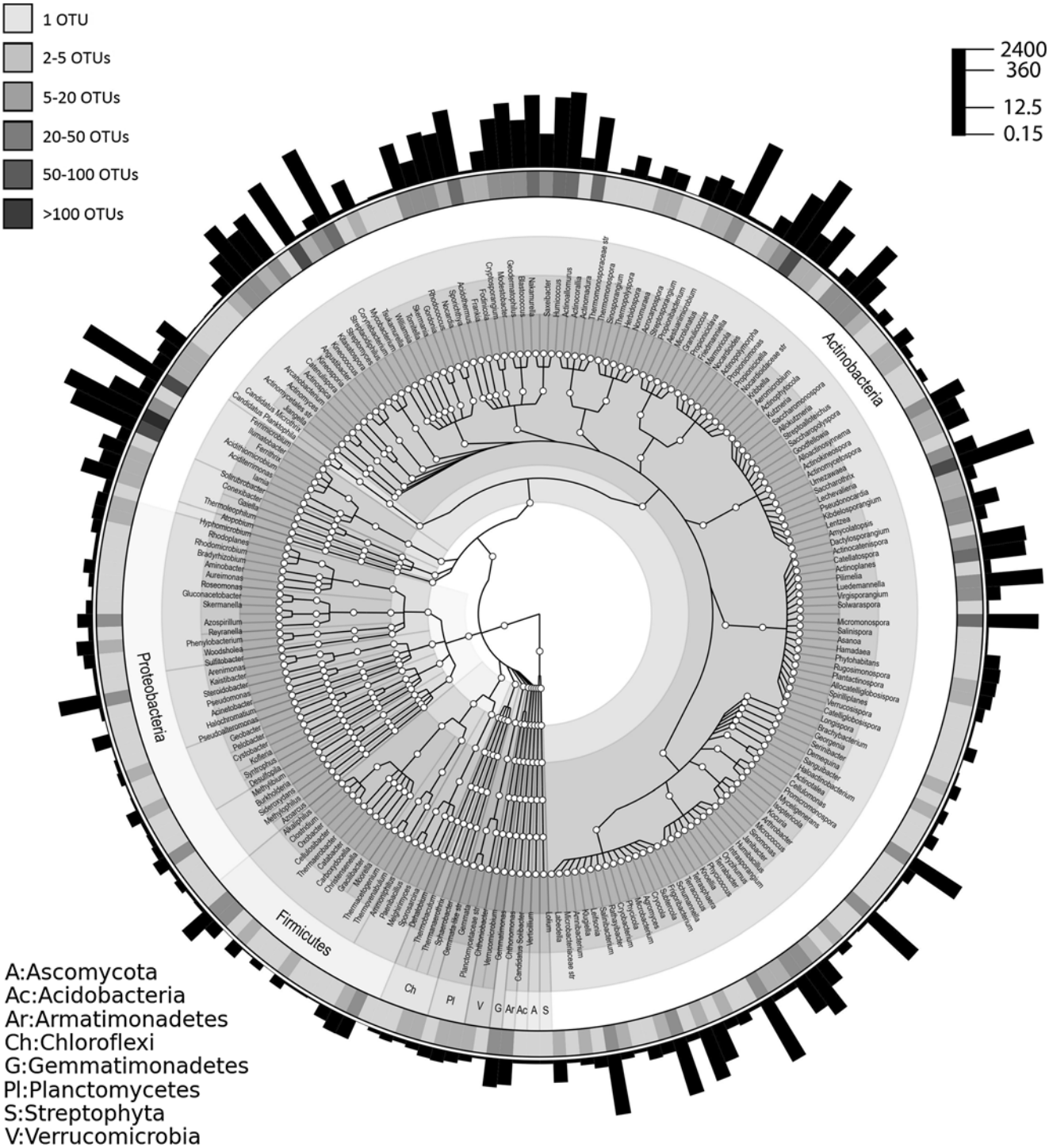


Fig. 1 Overview of abundances and OTU richness of microbial genera from grassland soils across a gradient of heavy metal concentrations obtained by 454 pyrosequencing. Number of OTUs per genus is depicted in *greyscale* in the inner circle, and a log scale of the relative abundance of genomes per genus is shown by *black bars*

the whole dataset. These most abundant actinobacterial OTUs and their closest related species are listed in Supplementary Table 1.

The quantitative analysis of actinobacterial and bacterial relative abundance in samples based on qPCR showed that the amount of bacteria negatively correlated with the increasing Cu concentration ($P<0.01$) and Zn concentration ($P<0.05$), but there was no significant correlation for abundance of Actinobacteria. The Actinobacteria-bacteria ratio appeared to be rising from sites with low contamination to sites

with high contamination, but there was high variation in the samples from site S2.

Principal component analysis showed that heavy metals had a slight effect on shaping the actinobacterial community composition (Fig. 2). Despite Pb having the highest concentrations of all of the measured metals at all sites, the most significant effect on the actinobacterial community composition appeared to have been the Cu concentration, as revealed by its strong correlation to most abundant OTUs (Fig. 3). Furthermore, a Mantel test with 9999 permutations indicated a clear relationship between community similarity and heavy metal content (R=0.3409, P=0.0001), indicating that actinobacterial community similarity decreased with the increasing difference in heavy metal contents among the samples. The most significant relationships were found between community composition and the content of Cu and Pb, while Zn and Cd showed less of an effect on the actinobacterial community composition. No significant relationship was observed when the actinobacterial community composition was tested against basic soil chemistry data (C_{ox}, N_{tot}, pH; R=−0.05596, P=0.8035) or any combination thereof. Among the most abundant actinobacterial taxa, OTUs identified as the members of the genera *Acidothermus*, *Streptomyces*, *Pseudonocardia*, *Janibacter* and *Microlunatus* increased in abundance with increasing metal content, while the members of *Jatrophihabitans* and *Actinoallomurus* decreased. The genus *Ilumatobacter* contained several OTUs that responded to heavy metals both positively and negatively (Fig. 3, Supplementary Table 1).

Discussion

Because heavy metals constitute the most widespread pollutants in a wide range of environmental contexts, including industrial areas and their surroundings, and because of their cumulative toxicity effects (Brumelis et al. 2002; Hernandez et al. 2003), an investigation of the response of microorganisms to their presence in the environment is important for understanding the potential effects of metal contamination on ecosystem functions. High concentrations of heavy metals are known to negatively affect enzyme activity and are toxic to most microorganisms (Giller et al. 1998). However, there is some evidence that various metals influence the composition of bacterial communities to various extents. Exposure of soil bacterial communities to Cr and As ions significantly lowered their diversity and species richness and changed their community composition (Sheik et al. 2012), while Cd, Pb and especially Zn reduced OTU diversity in polluted soils (Golebiewski et al. 2014; Hur et al. 2011; Moffett et al. 2003). In contrast, long-term Cu exposure was shown to change the bacterial community composition, but not its diversity or species richness (Berg et al. 2012). The results of other studies remain inconclusive because the effects of heavy metals are difficult to separate from the effects of other environmental factors (Chodak et al. 2013).

This study considered soils with a range of heavy metal concentrations which were previously demonstrated to affect soil functions, namely respiration and dehydrogenase activity, which both significantly decreased with the increasing total

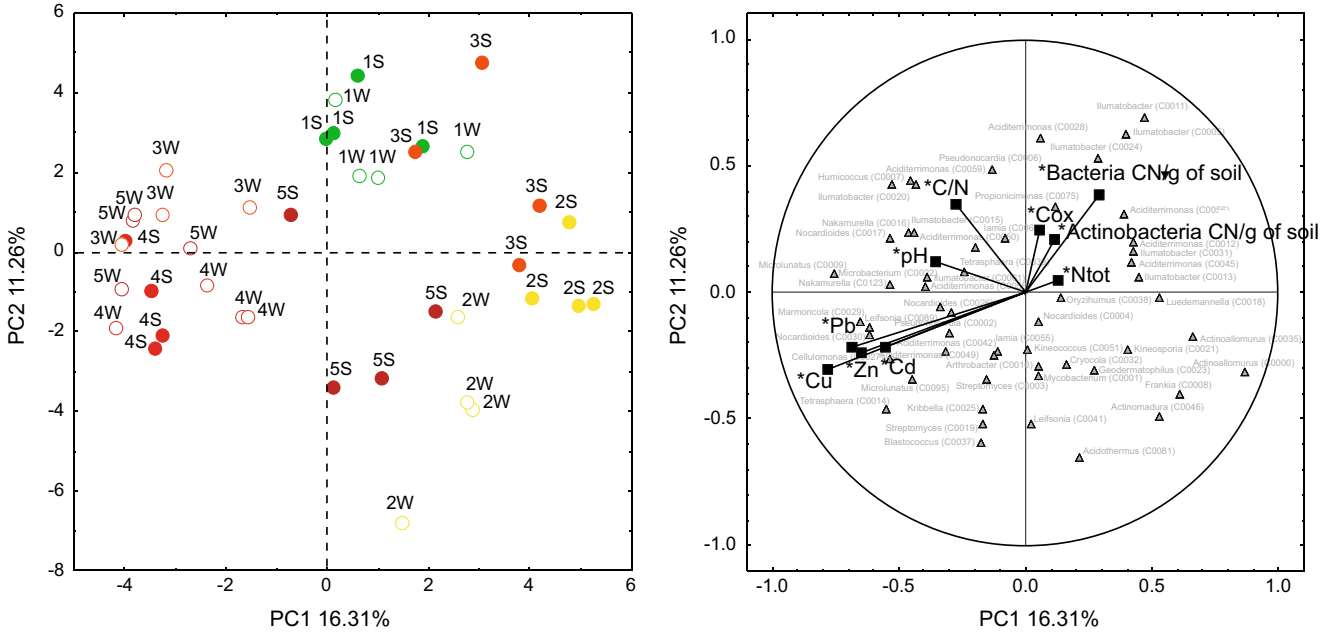
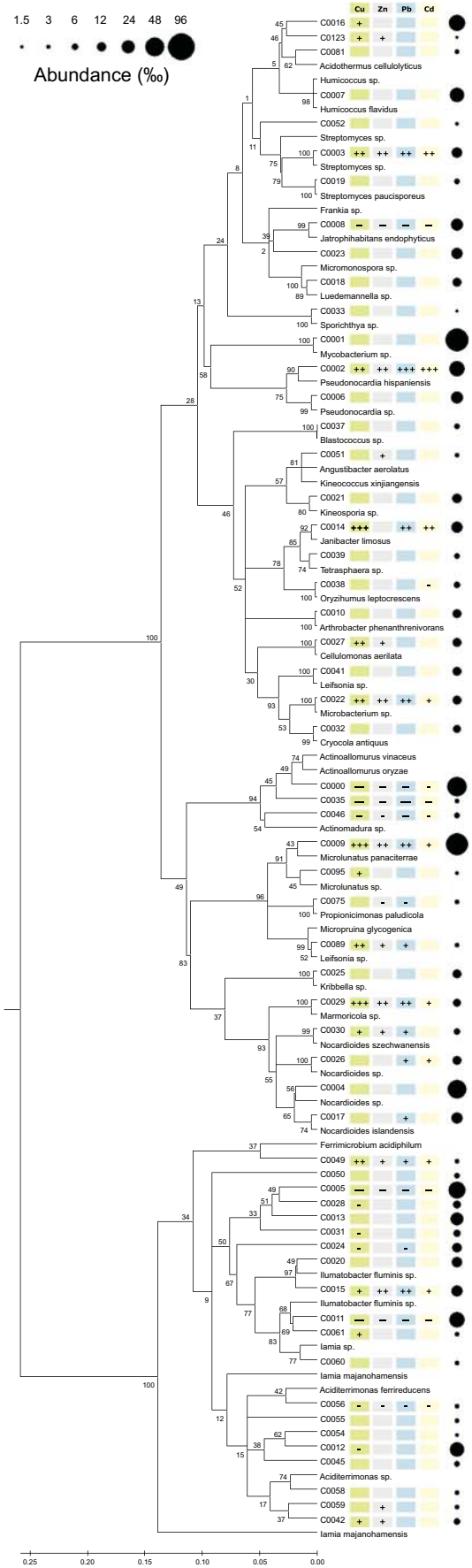


Fig. 2 Principal component analysis of relative abundances of genomes of those actinobacterial OTUs representing at least 1 % of total abundance within a sample for at least three samples. Heavy metal concentrations, environmental variables and relative bacterial and actinobacterial abundance are plotted as additional variables



◀ **Fig. 3** Phylogenetic tree of all actinobacterial OTUs from grassland soils across a gradient of heavy metal concentrations, representing at least 1 % of genome abundance within a sample for at least three samples and their closest identified hits. Relative abundances within the sampling sites are shown by *filled circles*. Statistically significant differences in correlation between abundance and heavy metal concentration are indicated by +++ for $P<0.001$, ++ for $P<0.01$ and + for $P<0.1$ for positive correlation and similar for negative correlation. (Tree was computed in MEGA6 using Maximum Likelihood method using the parameters as follows. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbour-Join and BioNJ algorithms to a matrix of pairwise distances estimated using the maximum composite likelihood (MCL) approach and then selecting the topology with superior log likelihood value. A discrete Gamma distribution was used to model evolutionary rate differences among the sites (5 categories (+*G*, parameter=0.5976)). The rate variation model allowed for some sites to be evolutionarily invariable ([+*I*], 50.3 % sites). The analysis involved 106 nucleotide sequences. All positions containing gaps and missing data were eliminated. There were a total of 349 positions in the final dataset. Evolutionary analyses were conducted in MEGA6 (Tamura et al. 2013).)

amount of heavy metals or their extractable fraction (Mühlbachová et al. 2015). We have now demonstrated that the abundance of bacteria estimated using qPCR decreased significantly with the increasing concentrations of Cu and Zn in the soil, while the abundance of Actinobacteria was not affected and the Actinobacteria/bacteria ratio tended to increase (Supplementary Table 2). This result may indicate a lower sensitivity of the Actinobacteria towards metal content. Previous studies on the response of Actinobacteria towards heavy metals remain controversial: despite the fact that some studies using the analyses of phospholipid fatty acid signatures to estimate actinobacterial abundance in soils tended to show that they are less resistant to heavy metals than other bacteria or fungi (Kelly et al. 2003; Zelles 1997), whereas sequencing studies strongly suggest the opposite. The relative abundance of Actinobacteria-dominating soils across the Cu gradient remained stable with increasing bioavailable Cu (Berg et al. 2012), and the sites with increasing heavy metal content that were revegetated using *Paulownia fortune* showed a higher richness and diversity of Actinobacteria (Liu et al. 2013). Gremion et al. (2003) have shown evidence that the actinomycetes species dominate over all prokaryote diversity in heavy metal-contaminated soil by a combined analysis of 16S rDNA and 16S rRNA, indicating that they might be a dominant part of the metabolically active bacteria in heavy metal-contaminated soils.

In the light of the uncertainty of the response of Actinobacteria to metal toxicity, the actinobacterial communities are an interesting target for exploration. Furthermore, the analysis of the response of individual taxa of the Actinobacteria, which had not been performed to date, should also be able to identify whether the response of Actinobacteria is a phylum-specific trait or whether there are differences on the level of individual taxa. To address this task, we have tested the use of Actinobacteria-specific primers to explore

their community composition using 454 pyrosequencing at a high resolution. In agreement with the in silico analyses, the amplicon pool obtained with the 530 F/16Sact1114r primers was greatly dominated by actinobacterial sequences that represented over 98 % of the total. Even with a relatively shallow sampling, more than 60 % of the total estimated actinobacterial diversity was recovered from the study sites.

The results of this study showed that the diversity of Actinobacteria was unaffected by heavy metal content, and a similar diversity was recorded in control plots as well as in those with the highest heavy metal content (Table 1). This finding is in contrast to the theoretical expectation that only a limited number of taxa would survive at highly contaminated sites. Furthermore, the results indicated that major actinobacterial taxa are rarely missing from the highly contaminated sites, although their abundance was affected (Supplementary Table 1).

Despite the fact that total diversity of Actinobacteria was not affected by metal contamination in the studied grassland soils, heavy metals represented the most important factor in shaping their community composition. Although Pb content reached the highest values in the studied sites, it was Cu that most affected the community composition. The abundance of several genera positively correlated with its concentration, especially members of *Acidothermus*, *Streptomyces*, *Pseudonocardia*, *Janibacter* and *Microlunatus*. Members of the genera *Streptomyces* and *Janibacter* were previously studied for their Cu and Sb resistance (Shi et al. 2013), and the heavy metal resistance of certain *Streptomyces* spp. is known (Schmidt et al. 2005). To survive in contaminated areas, *Streptomyces* use heavy metal-protective mechanisms such as a lead-absorbing superoxide dismutase described in *Streptomyces subrutilus*, allowing them to carry up to 1100 lead atoms per subunit, thus providing highly effective protection against Pb toxicity (So et al. 2001). Probably other members of this genus use similar mechanisms to survive in Pb-contaminated soils, notably because lead is often the main contaminant in mining and smelting areas, as in the case of this study. There are also multiple reports on their ability to degrade lignocellulose (Crawford 1978; Chater et al. 2010; Petrosyan et al. 2002; Větrovský et al. 2014), and together with members from genus *Acidothermus*, which also includes efficient degraders of cellulose, e.g. *Acidothermus cellulolyticus* (Blumer-Schuette et al. 2014; Parales et al. 2014) and *Microlunatus*, which are also involved in the decomposition of plant biomass (Fan et al. 2014), these taxa could be potentially responsible for important organic matter degrading processes in heavy metal-contaminated soils. In addition, the abundance of *Pseudonocardia* was high in the study area and positively correlated with the heavy metal content. The members of this genus are known as degraders of various organic compounds; *Pseudonocardia dioxanivorans* is able to use 1,4-dioxane as a sole carbon source, and its growth is stimulated by the addition of Zn (Pornwongthong et al. 2014). In contrast, *Jatrophihabitans* and *Actinoallomurus* represented the genera with a significant negative response to heavy metal content. Members of both genera were described as endophytes of plants, which were isolated mainly from plant roots (Madhaiyan et al. 2013; Matsumoto et al. 2012), and thus, the effects of metals on their occurrence might potentially be mediated by the effects on their plant hosts.

This study demonstrated the potential of specific amplification combined with next-generation sequencing to analyse the actinobacterial community composition and its response to mixed metal contamination. The actinobacterial community composition appeared not to be affected by metals in terms of diversity, but the contamination significantly affected community composition. Still, Actinobacteria seem to be less affected by heavy metals than other bacteria, and considering their metabolic potential, they can provide important soil functions, such as a contribution to organic matter decomposition. In this respect, these heavy metal-resistant taxa represent an interesting target for selective exploration.

Acknowledgments This work was supported by the Academy of Sciences of the Czech Republic (IAA603020901), by the Ministry of Education, Youth and Sports of the Czech Republic (LD12048, LD12050) and by the research concept of the Institute of Microbiology of the ASCR, v.v.i. (RVO61388971). This publication was also supported by the project 'BIOCEV–Biotechnology and Biomedicine Centre of the Academy of Sciences and Charles University'

(CZ.1.05/1.1.00/02.0109) from the European Regional Development Fund. The authors are thankful to the anonymous reviewers and the editor of the manuscript whose comments helped to improve the content of the paper.

References

Abdulla HM, El-Shatoury SA (2007) Actinomycetes in rice straw decomposition. Waste Manag 27:850–853

Amann RI, Ludwig W, Schleifer KH (1995) Phylogenetic identification and in-situ detection of individual microbial-cells without cultivation. Microbiol Rev 59:143–169

Anderson I, Abt B, Lykidis A, Klenk HP, Kyrpides N, Ivanova N (2012) Genomics of aerobic cellulose utilization systems in Actinobacteria. PLoS One 7, e39331

Bachar A, Al-Ashhab A, Soares MIM, Sklarz MY, Angel R, Ungar ED, Gillor O (2010) Soil microbial abundance and diversity along a low precipitation gradient. Microb Ecol 60:453–461

Bajkic S, Narancic T, Dokic L, Dordevic D, Nikodinovic-Runic J, Moric I, Vasiljevic B (2013) Microbial diversity and isolation of multiple metal-tolerant bacteria from surface and underground pits within the copper mining and smelting complex bor. Arch Biol Sci 65:375–386

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

Berg J, Brandt KK, Al-Soud WA, Holm PE, Hansen LH, Sorensen SJ, Nybroe O (2012) Selection for Cu-tolerant bacterial communities

with altered composition, but unaltered richness, via long-term Cu exposure. Appl Environ Microbiol 78:7438–7446
Berlemont R, Martiny AC (2013) Phylogenetic distribution of potential cellulases in bacteria. Appl Environ Microbiol 79:1545–1554
Blumer-Schuette SE, Brown SD, Sander KB, Bayer EA, Kataeva I, Zurawski JV, Conway JM, Adams MWW, Kelly RM (2014) Thermophilic lignocellulose deconstruction. FEMS Microbiol Rev 38:393–448
Brumelis G, Lapina L, Nikodemus O, Tabors G (2002) Use of the O horizon of forest soils in monitoring metal deposition in Latvia. Water Air Soil Pollut 135:291–309
Chater KF, Biro S, Lee KJ, Palmer T, Schrempf H (2010) The complex extracellular biology of *Streptomyces*. FEMS Microbiol Rev 34: 171–198
Chodak M, Golebiewski M, Morawska-Ploskonka J, Kuduk K, Niklinska M (2013) Diversity of microorganisms from forest soils differently polluted with heavy metals. Appl Soil Ecol 64:7–14
Cole JR, Chai B, Farris RJ, Wang Q, Kulam SA, McGarrell DM, Garrity GM, Tiedje JM (2005) The Ribosomal Database Project (RDP-II): sequences and tools for high-throughput rRNA analysis. Nucleic Acids Res 33:D294–D296
Crawford DL (1978) Lignocellulose decomposition by selected streptomyces strains. Appl Environ Microbiol 35:1041–1045
de Boer W, Folman LB, Summerbell RC, Boddy L (2005) Living in a fungal world: impact of fungi on soil bacterial niche development. FEMS Microbiol Rev 29:795–811
Dominguez-Mendoza CA, Bello-Lopez JM, Navarro-Noya YE, de Leon-Lorenzana AS, Delgado-Balbuena L, Gomez-Acata S, Ruiz-Valdiviezo VM, Ramirez-Villanueva DA, Luna-Guido M, Dendooven L (2014) Bacterial community structure in fumigated soil. Soil Biol Biochem 73:122–129
Dowd SE, Callaway TR, Wolcott RD, Sun Y, McKeehan T, Hagevoort RG, Edrington TS (2008) Evaluation of the bacterial diversity in the feces of cattle using 16S rDNA bacterial tag-encoded FLX amplicon pyrosequencing (bTEFAP). BMC Microbiol 8:125
Dunbar J, Takala S, Barns SM, Davis JA, Kuske CR (1999) Levels of bacterial community diversity in four arid soils compared by cultivation and 16S rRNA gene cloning. Appl Environ Microbiol 65: 1662–1669
Edgar RC (2013) UPARSE: highly accurate OTU sequences from microbial amplicon reads. Nat Methods 10:996–998
Edgar RC, Haas BJ, Clemente JC, Quince C, Knight R (2011) UCHIME improves sensitivity and speed of chimera detection. Bioinformatics 27:2194–2200
Enkhbaatar B, Temuujin U, Lim JH, Chi WJ, Chang YK, Hong SK (2012) Identification and characterization of a xyloglucan-specific family 74 glycosyl hydrolase from *Streptomyces coelicolor* A3(2). Appl Environ Microbiol 78:607–611
Fan FL, Yin C, Tang YJ, Li ZJ, Song A, Wakelin SA, Zou J, Liang YC (2014) Probing potential microbial coupling of carbon and nitrogen cycling during decomposition of maize residue by C-13-DNA-SIP. Soil Biol Biochem 70:12–21
Giller KE, Witter E, McGrath SP (1998) Toxicity of heavy metals to microorganisms and microbial processes in agricultural soils: a review. Soil Biol Biochem 30:1389–1414
Golebiewski M, Deja-Sikora E, Cichosz M, Tretyn A, Wrobel B (2014) 16S rDNA pyrosequencing analysis of bacterial community in heavy metals polluted soils. Microb Ecol 67:635–647
Gremion F, Chatzinotas A, Harms H (2003) Comparative 16S rDNA and 16S rRNA sequence analysis indicates that Actinobacteria might be a dominant part of the metabolically active bacteria in heavy metal-contaminated bulk and rhizosphere soil. Environ Microbiol 5:896–907
Hernandez L, Probst A, Probst JL, Ulrich E (2003) Heavy metal distribution in some French forest soils: evidence for atmospheric contamination. Sci Total Environ 312:195–219
Hur M, Kim Y, Song HR, Kim JM, Choi YI, Yi H (2011) Effect of genetically modified poplars on soil microbial communities during the phytoremediation of waste mine tailings. Appl Environ Microbiol 77:7611–7619
Ivshina IB, Kuyukina MS, Kostina LV (2013) Adaptive mechanisms of nonspecific resistance to heavy metal ions in alkanotrophic actinobacteria. Russ J Ecol 44:123–130
Kalac P, Burda J, Staskova I (1991) Concentrations of lead, cadmium, mercury and copper in mushrooms in the vicinity of lead smelter. Sci Total Environ 105:109–119
Katoh K, Asimenos G, Toh H (2009) Multiple alignment of DNA sequences with MAFFT. In: Posada D (ed) Bioinformatics for DNA sequence analysis, vol 537. Methods Mol Biol. p 39–64
Kelly JJ, Haggblom MM, Tate RL (2003) Effects of heavy metal contamination and remediation on soil microbial communities in the vicinity of a zinc smelter as indicated by analysis of microbial community phospholipid fatty acid profiles. Biol Fertil Soils 38:65–71
Kyselkova M, Kopecky J, Felfoldi T, Cermak L, Omelka M, Grundmann GL, Moenne-Loccoz Y, Sagova-Mareckova M (2008) Development of a 16S rRNA gene-based prototype microarray for the detection of selected actinomycetes genera Antonie Van Leeuwenhoek. Int J Gen Mol Microbiol 94:439–453
Lauber CL, Hamady M, Knight R, Fierer N (2009) Pyrosequencing-based assessment of soil pH as a predictor of soil bacterial community structure at the continental scale. Appl Environ Microbiol 75: 5111–5120
Leigh MB, Pellizari VH, Uhlik O, Sutka R, Rodrigues J, Ostrom NE, Zhou JH, Tiedje JM (2007) Biphenyl-utilizing bacteria and their functional genes in a pine root zone contaminated with polychlorinated biphenyls (PCBs). ISME J 1:134–148
Liu W, Wang J, Zhang CB (2013) Evaluation of soil chemical properties and actinomycete community structure following a temporal sequence of revegetation through *Paulownia fortunei* in the heavy metal-contaminated soil. Water Air Soil Pollut 224:1730
Madhaiyan M, Hu CJ, Kim SJ, Weon HY, Kwon SW, Ji LH (2013) Jatrophihabitans endophyticus gen. nov., sp nov., an endophytic actinobacterium isolated from a surface-sterilized stem of *Jatropha curcas* L. Int J Syst Evol Microbiol 63:1241–1248
Matsumoto A, Fukuda A, Inahashi Y, Omura S, Takahashi Y (2012) *Actinoallomurus radicium* sp. nov., isolated from the roots of two plant species. Int J Syst Evol Microbiol 62:295–298
Moffett BF, Nicholson FA, Uwakwe NC, Chambers BJ, Harris JA, Hill TCJ (2003) Zinc contamination decreases the bacterial diversity of agricultural soil. FEMS Microbiol Ecol 43:13–19
Mühlbachová G (2011) Soil microbial activities and heavy metal mobility in long-term contaminated soils after addition of EDTA and EDDS. Ecol Eng 37:1064–1071
Mühlbachová G, Ságová-Marečková M, Omelka M, Szákova J, Tlustoš P (2015) The influence of soil organic carbon on interactions between microbial parameters and metal concentrations at a long-term contaminated site. Sci Total Environ 502:218–223
Muyzer G, Dewaal EC, Uitterlinden AG (1993) Profiling of complex microbial populations by denaturing gradient gel electrophoresis analysis of polymerase chain reaction-amplified genes coding for 16S rRNA. Appl Environ Microbiol 59:695–700
Parales RE, Berry AM, Parales JV, Barabote RD (2014) Use of acidothermus cellulolyticus xylanase for hydrolyzing lignocellulose. US Patent No 08778649 B2
Petrosyan P, Luz-Madrigal A, Huitron C, Flores ME (2002) Characterization of a xylanolytic complex from Streptomyces sp. Biotechnol Lett 24:1473–1476
Pornwongthong P, Mulchandani A, Gedalanga PB, Mahendra S (2014) Transition metals and organic ligands influence biodegradation of 1, 4-dioxane. Appl Biochem Biotechnol 173:291–306
Sagova-Mareckova M, Cermak L, Novotna J, Plhackova K, Forstova J, Kopecky J (2008) Innovative methods for soil DNA purification

tested in soils with widely differing characteristics. Appl Environ Microbiol 74:2902–2907

Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, Lesniewski RA, Oakley BB, Parks DH, Robinson CJ, Sahl JW, Stres B, Thallinger GG, Van Horn DJ, Weber CF (2009) Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. Appl Environ Microbiol 75:7537–7541

Schmidt A, Haferburg G, Sineriz M, Merten D, Buchel G, Kothe E (2005) Heavy metal resistance mechanisms in actinobacteria for survival in AMD contaminated soils. Chem Erde-Geochem 65: 131–144

Sheik CS, Mitchell TW, Rizvi FZ, Rehman Y, Faisal M, Hasnain S, McInerney MJ, Krumholz LR (2012) Exposure of soil microbial communities to chromium and arsenic alters their diversity and structure. PLoS One 7, e40059

Shi ZJ, Cao Z, Qin D, Zhu WT, Wang Q, Li MS, Wang GJ (2013) Correlation models between environmental factors and bacterial resistance to antimony and copper. PLoS One 8:78533

So NW, Rho JY, Lee SY, Hancock IC, Kim JH (2001) A lead-absorbing protein with superoxide dismutase activity from *Streptomyces subrutilus*. FEMS Microbiol Lett 194:93–98

Stach JEM, Maldonado LA, Ward AC, Goodfellow M, Bull AT (2003) New primers for the class Actinobacteria: application to marine and terrestrial environments. Environ Microbiol 5:828–841

Steger K, Sjogren AM, Jarvis A, Jansson JK, Sundh I (2007) Development of compost maturity and Actinobacteria populations during full-scale composting of organic household waste. J Appl Microbiol 103:487–498

Tamura K, Stecher G, Peterson D, Filipski A, Kumar S (2013) MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. Mol Biol Evol 30:2725–2729

Větrovský T, Baldrian P (2013a) Analysis of soil fungal communities by amplicon pyrosequencing: current approaches to data analysis and the introduction of the pipeline SEED. Biol Fertil Soils 49:1027–1037

Větrovský T, Baldrian P (2013b) The variability of the 16S rRNA gene in bacterial genomes and its consequences for bacterial community analyses. PLoS One 8:57923

Větrovský T, Steffen KT, Baldrian P (2014) Potential of cometabolic transformation of polysaccharides and lignin in lignocellulose by soil Actinobacteria. PLoS One 9:89108

Wang YX, Liu Q, Yan L, Gao YM, Wang YJ, Wang WD (2013) A novel lignin degradation bacterial consortium for efficient pulping. Bioresour Technol 139:113–119

Wilmotte A, Vanderauwera G, Dewachter R (1993) Structure of the 16 S ribosomal RNA of the thermophilic cyanobacterium *Chlorogloeopsis* HTF (‘*Mastigocladus laminosus* HTF’) strain PCC7518, and phylogenetic analysis. FEBS Lett 317:96–100

Yin LJ, Huang PS, Lin HH (2010) Isolation of cellulase-producing bacteria and characterization of the cellulase from the isolated bacterium Cellulomonas Sp YJ5. J Agric Food Chem 58:9833–9837

Zelles L (1997) Phospholipid fatty acid profiles in selected members of soil microbial communities. Chemosphere 35:275–294

OPEN ACCESS Freely available online

PLOS ONE

The Variability of the 16S rRNA Gene in Bacterial Genomes and Its Consequences for Bacterial Community Analyses

Tomáš Větrovský, Petr Baldrian*

Laboratory of Environmental Microbiology, Institute of Microbiology of the Academy of Sciences of the Czech Republic, Praha, Czech Republic

Abstract

16S ribosomal RNA currently represents the most important target of study in bacterial ecology. Its use for the description of bacterial diversity is, however, limited by the presence of variable copy numbers in bacterial genomes and sequence variation within closely related taxa or within a genome. Here we use the information from sequenced bacterial genomes to explore the variability of 16S rRNA sequences and copy numbers at various taxonomic levels and apply it to estimate bacterial genome and DNA abundances. In total, 7,081 16S rRNA sequences were *in silico* extracted from 1,690 available bacterial genomes (1–15 per genome). While there are several phyla containing low 16S rRNA copy numbers, in certain taxa, e.g., the Firmicutes and Gammaproteobacteria, the variation is large. Genome sizes are more conserved at all tested taxonomic levels than 16S rRNA copy numbers. Only a minority of bacterial genomes harbors identical 16S rRNA gene copies, and sequence diversity increases with increasing copy numbers. While certain taxa harbor dissimilar 16S rRNA genes, others contain sequences common to multiple species. Sequence identity clusters (often termed operational taxonomic units) thus provide an imperfect representation of bacterial taxa of a certain phylogenetic rank. We have demonstrated that the information on 16S rRNA copy numbers and genome sizes of genome-sequenced bacteria may be used as an estimate for the closest related taxon in an environmental dataset to calculate alternative estimates of the relative abundance of individual bacterial taxa in environmental samples. Using an example from forest soil, this procedure would increase the abundance estimates of Acidobacteria and decrease these of Firmicutes. Using the currently available information, alternative estimates of bacterial community composition may be obtained in this way if the variation of 16S rRNA copy numbers among bacteria is considered.

Citation: Větrovský T, Baldrian P (2013) The Variability of the 16S rRNA Gene in Bacterial Genomes and Its Consequences for Bacterial Community Analyses. PLoS ONE 8(2): e57923. doi:10.1371/journal.pone.0057923

Editor: Josh Neufeld, University of Waterloo, Canada

Received November 5, 2012; **Accepted** January 28, 2013; **Published** February 27, 2013

Funding: This work was supported by the Ministry of Education, Youth and Sports of the Czech Republic (LD12048, LD12050) and by the research concept of the Institute of Microbiology of the Academy of Sciences of the Czech Republic (RVO61388971). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: baldrian@biomed.cas.cz

Introduction

rRNA sequences and especially the 16S rRNA represent the most important current targets of study in bacterial evolution and ecology, including the determination of phylogenetic relationships among taxa, the exploration of bacterial diversity in the environment and the quantification of the relative abundance of taxa of various ranks [1]. The 16S rRNA is suitable for this purpose for several reasons. The gene is universally distributed, allowing the analysis of phylogenetic relationships among distant taxa. As a functionally indispensable part of the core gene set, the 16S rRNA gene is expected to be only weakly affected by horizontal gene transfer [2], which further supports its use for phylogenetic studies. Despite the above, 16S rRNA is still subject to variation, especially in certain variable regions. While the presence of variable regions allows sufficient diversification to provide a tool for classification, the presence of conserved regions enabled the design of suitable PCR primers or hybridization probes for various taxa at different taxonomic levels ranging from individual strains to whole phyla [3].

Despite the wide use of 16S rRNA, there are several aspects that limit the interpretation of 16S rRNA-derived results. The most important is the fact that its copy numbers per genome vary from 1 up to 15 or more copies [4]. Copy numbers seem to be taxon-specific to some extent, but variation among strains of the same species has also been recorded [5]. The numbers of rRNA copies have been put into context with the life strategy of bacteria because the rRNA copy number of some taxa are correlated with their ability to respond to favorable growth conditions. Taxa with low copy numbers have been assumed to be more oligotrophic [6,7].

It is assumed that copies of rRNA genes within an organism are subject to homogenization through gene conversion [8]. Nevertheless, 16S sequences from the same species or even the same genome are often different. Consequently, the amount of 16S rRNA variants was estimated to be 2.5-fold greater than the number of bacterial species [5], and highly dissimilar 16S rRNA sequences were observed in some bacterial taxa [9,10]. Bacterial species with sequences that differ by >1% are quite common [11]. An even greater variability of 16S rRNA sequences was detected in thermophilic bacteria. In this particular case, the higher

incidence of horizontal gene transfer was proposed to be the potential cause [5].

Because the current analysis of bacterial communities most often relies on the construction of similarity clusters (or operational taxonomic units, OTUs) of 16S rRNA gene PCR amplicons, both the multiplicity and variability represent problems for the assessment of bacterial diversity and community structure, i.e., the relative abundance of individual taxa. While the former skews the abundance estimates of individual taxa, the latter affects diversity estimates, and it was recognized early that the lack of information on 16S rRNA copy numbers and genome sizes make the relative abundance estimates in complex bacterial populations unreliable [12]. These problems are not restricted to bacteria because the linking of cell abundance and PCR amplicon abundance is also limited by the multicopy nature and intragenomic variability of the most common molecular marker of fungi, the ITS region of rRNA [13,14]. In contrast to the PCR amplification-based analyses, shotgun metagenomic data are not immediately affected by the 16S rRNA copy numbers; however, their use for the estimation of bacterial community composition is limited by the variation in genome sizes.

Until recently, it was difficult to draw conclusions about the extent of the biases caused by the use of 16S rRNA analyses; this was primarily due to the limited amount of sequenced bacterial genomes and the absence of these data from certain common phyla [5]. However, recent advances in bacterial genomics allow to analyze the existing dataset in order to answer important questions about the potential of 16S rRNA in bacterial community ecology: Are related taxa more similar in genome sizes/16S rRNA copy numbers? If there are more copies per genome, how similar are they? Are they so distant that it affects diversity estimates using OTU construction? How does this affect the analyses of bacterial community composition using 16S rRNA amplicon analysis? The aim of this paper was to answer the above questions. In addition, we propose the use of currently available genome data for the improvement of the estimates of the relative abundance of bacterial taxa and verify this approach *in silico*.

Materials and Methods

In silico analysis of sequenced bacterial genomes

For the *in silico* analysis, 1,690 publicly available complete genomes (i.e., those designated as "Complete" with known gene count and genome size) of identified bacterial species were downloaded from GenBank (http://www.ncbi.nlm.nih.gov/genbank/) in February 2012 (Dataset S1). The 16S rRNA sequences were identified by locating the 16S rRNA-targeting reverse and forward primers eub530F and eub1100aR targeting the V4–V6 regions [15,16] within a distance of 1000 bases on both the plus and minus strands. Regions 900 bases upstream or downstream from this region were retrieved for further processing. In addition, the consensus sequence of the rDNA between the above primers (obtained by the alignment of full-length bacterial 16S rRNA sequences selected from GenBank to represent all phyla with genome sequences available) was searched in the genomes by local BLAST. In the rare cases where the 16S rRNA copy numbers per genome obtained using the latter approach were higher (<0.5% of genomes), the regions around the blast hit were also retrieved. All sequences obtained were aligned with the GenBank 16S rRNA dataset using the RDP pyrosequencing aligner (http://pyro.cme.msu.edu/spring/align.spr) with default settings, and incomplete sequences were manually removed. In total, 7,081 complete 16S rRNA sequences were retrieved. The FASTA file containing these sequences is available as Dataset S2 (unaligned) and Dataset S3 (aligned).

GenBank taxonomy was used to group bacterial species into higher taxonomic ranks-families, classes, and phyla. These assignments were used to characterize the variation in genome sizes and 16S rRNA copy numbers per genomes at various taxonomic levels. At a species level, all genomes of individual strains were analyzed. At higher taxonomic levels, to avoid the effect of variable numbers of sequenced genomes per bacterial species, calculations were performed with mean values for all species belonging to the taxon.

Pairwise similarities of whole 16S rRNA sequences were calculated with a gap opening considered to be equal to mismatch and the gap extension disregarded in BioNumerics 6.5 (Applied Maths, Belgium). Mean pairwise similarities were calculated among all 16S rRNA sequences within each genome, among all pairs of 16S rRNA sequences belonging to different genomes within the same species and among all pairs of 16S rRNA sequences belonging to different species within the same bacterial genus. The within-genome similarity matrices were screened for genomes with at least one pair of 16S rRNA sequences with a pairwise similarity lower than 97%. A tree was constructed from these sequences using the neighbor-joining method with the nucleotide substitution type, the Number of differences model and complete deletion of gaps in the software MEGA 5.05 [17].

Operational taxonomic units (OTUs), the groups of DNA sequences that share defined level of similarity, are often constructed to estimate the diversity of the bacterial community at certain taxonomic levels. We performed OTU construction with the entire 16S rRNA sequences from the Dataset S2 to analyze the relationships between the taxonomic identity of bacterial taxa and these mathematical constructs. OTUs were constructed using CD-HIT [18] at sequence similarity levels of 90%–100%. At all similarity levels, the total number of OTUs was calculated, and the number of species and genera within each OTU were calculated. The taxa denoted as *Escherichia* and *Shigella* were treated as belonging to the same genus.

Assessment of bacterial genome copy numbers and DNA abundances in amplicon pyrosequencing data

The information on the genome size and 16S rRNA copy numbers per genome were used to estimate the abundance of bacterial genomes and DNA using the previously generated 454-pyrosequencing dataset that contained partial 16S rRNA sequences of the DNA and cDNA from forest topsoil. The sequences originated from a previous study [15] and represented sequences derived from the litter horizon DNA, litter horizon cDNA, soil DNA and soil cDNA (four replicates each). The amplicons were generated using the primers eub530F/eub1100aR and sequenced from the eub530F primer; in total, 204,826 sequences with a length above 300 bp were obtained [15]. Pyrosequencing noise reduction was performed using a Denoiser 0.851 [19], and chimeric sequences were detected using UCHIME [20] and deleted. Sequences were truncated to 300 bases, clustered at a 97% similarity using CD-HIT [18], and consensus sequences were constructed for OTUs with >2 sequences (or representative sequences of OTUs with <3 sequences). For each OTU consensus sequence, the best database hit (this with the lowest E value) was obtained using BLASTn against GenBank [21], while the best genome hit was obtained using the BLASTn against the 16S rRNA dataset derived from bacterial genomes (i.e., Dataset S2). The quality of hits was compared between the best database hit and the best genome hit for each sequence considering the percentage of similarity for hits to >90% of the query sequence

length. Phylum assignments were also compared for the two best hits.

The best genome hits were used to calculate the relative abundance of 16S rRNA sequences among bacterial phyla in the samples. For the calculation of the relative abundance of bacterial genomes by phyla, the abundance of sequences belonging to each OTU was divided by the count of 16S rRNA sequences per genome of the best genome hit. To obtain the relative abundance of bacterial DNA by phyla, these values were further multiplied by the sizes of the genomes of the best genome hit.

To analyze the efficiency of the above approach for approximating the genome counts and sizes in the environmental samples, the Dataset S2 was processed as follows: a subset was created that contained one genome from each bacterial species. From this subset containing 909 genomes, 100 genomes were randomly selected. From these genomes, 300-bp sequences, starting with the eub530F primer, were generated, and these sequences were searched in the set of the remaining 809 genomes. For each of the 100 selected genomes, the actual genome size and 16S rRNA copy number were compared to the genome size estimates and 16S rRNA copy number estimates obtained (1) as the values of the genomes with the closest hit; (2) as the average values of the bacterial phylum to which the genome belonged; and (3) as a mean of all genomes (i.e., considering equal size and 16S rRNA copy numbers for all genomes). The sets of these estimates were compared.

One-way ANOVA with the Fisher LSD post-hoc test was used to analyze the significant differences among datasets (e.g., 16S rRNA gene copy numbers or the abundance of fungal genomes in soil among phylogenetic groups). Differences of $P<0.05$ were regarded as statistically significant.

Results

In silico analysis of sequenced bacterial genomes

Of a total of 1,690 available bacterial genomes that belonged to 909 identified bacterial species and 454 genera, most available genomes belonged to the Proteobacteria (especially Gammapro-

Table 1. Overview of bacterial genomes and their properties at the phylum level (class level for the Proteobacteria).

Phylum	Genera	Species	Genomes	16S rRNA/genome	Genome size (Mb)
Acidobacteria	3	4	4	1.0±0.0	5.24±0.88
Actinobacteria	66	117	201	3.1±1.7	5.03±2.53
Aquificae	6	6	6	2.0±0.6	1.63±0.11
Bacteroidetes	37	52	59	3.5±1.5	4.51±1.83
Caldiserica	1	1	1	1	1.56
Chlamydiae	5	11	72	1.4±0.5	1.54±0.72
Chlorobi	5	10	11	1.7±0.7	2.59±0.41
Chloroflexi	9	10	10	2.2±1.2	4.12±1.73
Cyanobacteria	10	11	23	2.3±1.2	5.66±2.45
Deferribacteres	4	4	4	2.0±0.0	2.63±0.42
Deinococcus-Thermus	6	13	16	2.7±1.0	3.12±0.67
Dictyoglomi	1	2	2	2.0±0.0	1.91±0.07
Elusimicrobia	1	1	1	1	1.64
Fibrobacteres	1	1	2	3	3.84
Firmicutes	69	186	395	5.8±2.8	3.09±1.18
Fusobacteria	5	5	5	5.0±0.7	2.79±1.09
Gemmatimonadetes	1	1	1	1	4.64
Ignavibacteria	1	1	1	1	3.66
Nitrospirae	2	2	2	2.0±1.4	2.28±0.39
Planctomycetes	5	6	6	1.7±0.8	5.71±1.07
Proteobacteria					
Alphaproteobacteria	58	112	179	2.2±1.3	3.58±2.01
Betaproteobacteria	39	66	101	3.3±1.6	5.09±2.18
Deltaproteobacteria	25	37	43	2.7±1.4	4.95±2.56
Epsilonproteobacteria	9	23	70	3.0±1.1	1.97±0.41
Gammaproteobacteria	78	157	379	5.8±2.8	4.23±1.25
Spirochaetes	5	20	31	2.4±1.0	3.16±0.95
Synergistetes	4	4	4	2.5±1.0	1.88±0.15
Tenericutes	4	25	43	1.6±0.5	0.93±0.20
Thermodesulfobacteria	1	1	1	2	2.32
Thermotogae	6	13	13	1.8±1.0	2.02±0.17
Verrucomicrobia	4	4	4	1.8±1.0	3.66±1.65

doi:10.1371/journal.pone.0057923.t001

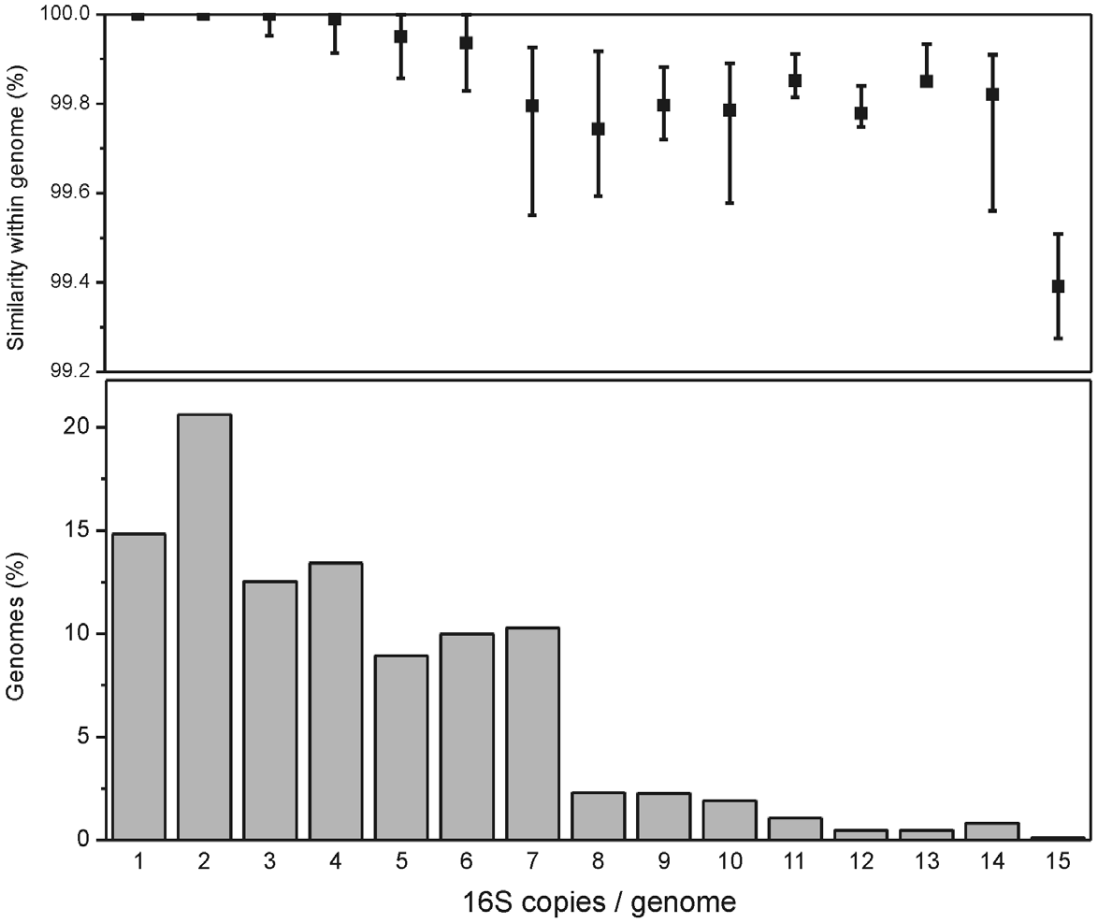


Figure 1. 16S rRNA within-genome similarity and copy numbers in bacterial genomes. Upper panel: the similarity of genomes with various copy numbers: the values indicated represent the first, the second and the third quartile. Lower panel: distribution of 16S rRNA copy numbers per genome in 1,690 sequenced bacterial genomes.
doi:10.1371/journal.pone.0057923.g001

teobacteria, Alphaproteobacteria and Betaproteobacteria), Firmicutes, and Actinobacteria with >100 genomes per phylum; the phyla Caldiserica, Elusimicrobia, Gemmatimonadetes, Ignavibacteria, and Thermodesulfobacteria were represented by only one genome each (Table 1). A total of 7,081 16S rRNA copies were identified in, with an average of 4.2 copies per genome. Fifteen percent of genomes contained a single 16S rRNA copy, while 21% contained two copies, and 3–7 16S rRNA copies per genome were frequently found. Copy numbers above 7 were relatively rare, and the maximum value was 15 (Figure 1).

The 16S rRNA copy numbers per genome were taxon-specific at several taxonomical levels (Figure 2, Figure S1). They were especially high in the Firmicutes, Gammaproteobacteria and Fusobacteria, with averages per genome of above five. However, in the former two groups, copy numbers varied widely between 1 and 15. One 16S rRNA copy per genome was present in Caldiserica, Elusimicrobia, Gemmatimonadetes, Ignavibacteria and Acidobacteria, but only the latter group was represented by more than one genome sequence. Most bacterial phyla showed wide variation in 16S rRNA copy numbers, with most containing at least one representative with a single 16S rRNA copy. At lower taxonomic levels, the variation of 16S rRNA copy numbers was less pronounced, and differences at the level of families, genera and species were often statistically significant (data not shown). Among bacterial families with available genomes from more than eight different species, several exhibited high 16S rRNA copy numbers: all members of the Shewanellaceae, Vibrionaceae, Bacillaceae, and Pasteurellaceae contained more than five copies. The 16S rRNA copy numbers were even more conserved at the level of bacterial genera and species. For the top 36 bacterial taxa (families with >6 species, genera with >3 species, or species with >6 genomes), the coefficients of variation (standard deviations divided by means) of the 16S rRNA copy numbers were 29% on the family level, 22% on the genus level and 9.2% on the species level. However, some species, e.g., *Bacillus amyloliquefaciens*, *Campylobacter jejuni* or *Bifidobacterium longum*, showed high interspecific variability in the 16S rRNA copy numbers (Figure 2).

Furthermore, the sizes of bacterial genomes were taxon-specific at all studied taxonomical levels (Figure 3). Several bacterial phyla (or classes of Proteobacteria) harbored small genomes with low size variation: the smallest genomes were found in Tenericutes (0.93±0.20) and Chlamydiae (1.54±0.72). However, several groups widely differed in genome sizes, e.g., the Actinobacteria, with sizes from 1 to 12 Mb, as well as the Beta- and Deltaproteobacteria (2–13 Mb) and the Cyanobacteria (2–9 Mb). At lower taxonomic levels, the size of genomes was less variable then the 16S rRNA copy numbers: the coefficients of variation were 19% on the family level and 16% on the genus level. Within

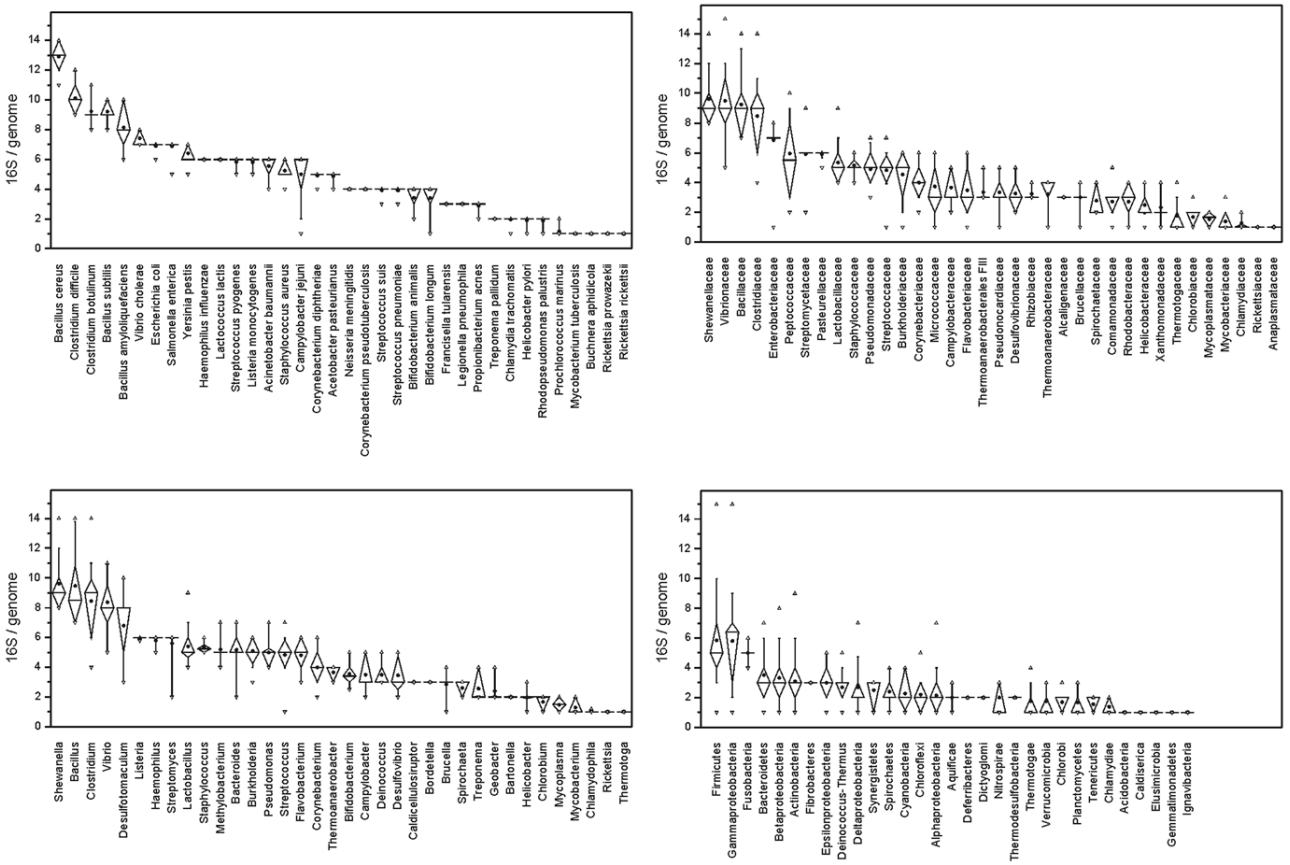


Figure 2. 16S rRNA copy numbers in bacterial genomes. 16S rRNA copy numbers in bacterial phyla (classes), selected families (n>6), genera (n>3), and species (n>6). Open triangles indicate minima and maxima, whiskers the 10th and 90th quantile and boxes the 25th and 75th quantile. Median is indicated as a horizontal line and mean as a dot.
doi:10.1371/journal.pone.0057923.g002

individual bacterial species, genome size was highly conserved, with a variation lower than 3.8% (Figure 3).

16S rRNA is currently the most common target of analysis of the diversity of bacterial communities. Because the most common approach is based on similarity clustering, we explored the variability of the 16S rRNA at various taxonomic levels. At a genome level, 19.8% of all genomes with more than one 16S rRNA copy harbor 2–5 identical 16S rRNA copies, and the average 16S rRNA similarity within a genome is 99.70±0.46%, with 97.6% of genomes showing average 16S rRNA similarity above 99%. The similarity of genomes decreases with 16S rRNA copy numbers, and all genomes with six and more copies harbored at least two different variants of 16S rRNA (Figure 1). Phylogenetically, Firmicutes and Gammaproteobacteria showed high 16S rRNA within-genome variability, while Alphaproteobacteria and Chlamydiae showed low variability; the differences were statistically significant. The level of dissimilarity within a genome can be relatively high: fourteen genomes contained at least one pair of 16S rRNA sequences with a similarity below 97% (Figure 4). These diverse genomes were found in the genera *Clostridium*, *Desulfitobacterium*, *Desulfosporinus*, *Desulfotomaculum*, *Photobacterium*, *Salmonella*, *Selenomonas*, *Syntrophomonas*, *Thermoanaerobacter*, *Thermoanaerobacterium*, and *Thermobispora*. While some of the genomes contained groups of similar 16S rRNA sequences (e.g., the *Thermobispora bispora* DSM 43833 and *Clostridium* sp. BNL 1100), others contained several sequences with low pairwise similarity (Figure S2).

On the level of bacterial species, average 16S rRNA similarity was 99.30±1.38%, and 95.4% of all genomes had mean 16S rRNA similarity over 97%. Among genera, the average 16S rRNA similarity of genomes belonging to different species was 95.56±3.68%. In 12.2% and 41.7% of genera, 16S rRNA similarities among species were higher than 99% and 97%, respectively, while 12.2% of genera contain species with mean pairwise 16S rRNA similarity below 90% (Figure 4).

The fact that the level of 16S rRNA similarity among bacterial species within a genus varied widely poses a question about the reliability of diversity estimates based on the construction of operational taxonomic units by similarity clustering, as well as about the phylogenetic relatedness within the OTU. Ideally, OTUs of certain similarity cutoffs should contain sequences belonging to bacteria of certain, defined taxonomic rank. To test the potential of OTU clustering to meet these criteria, OTUs were constructed at various levels of sequence similarity and analyzed for the number of bacterial species and genera that they contain. Clustering at 99% similarity and 95% similarity gave OTU counts that best corresponded to the species and genus counts, 906 and 471, respectively (Figure 5). However, even at 100% sequence similarity, 6.7% of OTUs contained sequences that belonged to multiple species (up to nine). At a 97% similarity level, 21.3% of OTUs contained sequences of multiple species, and 9.3%

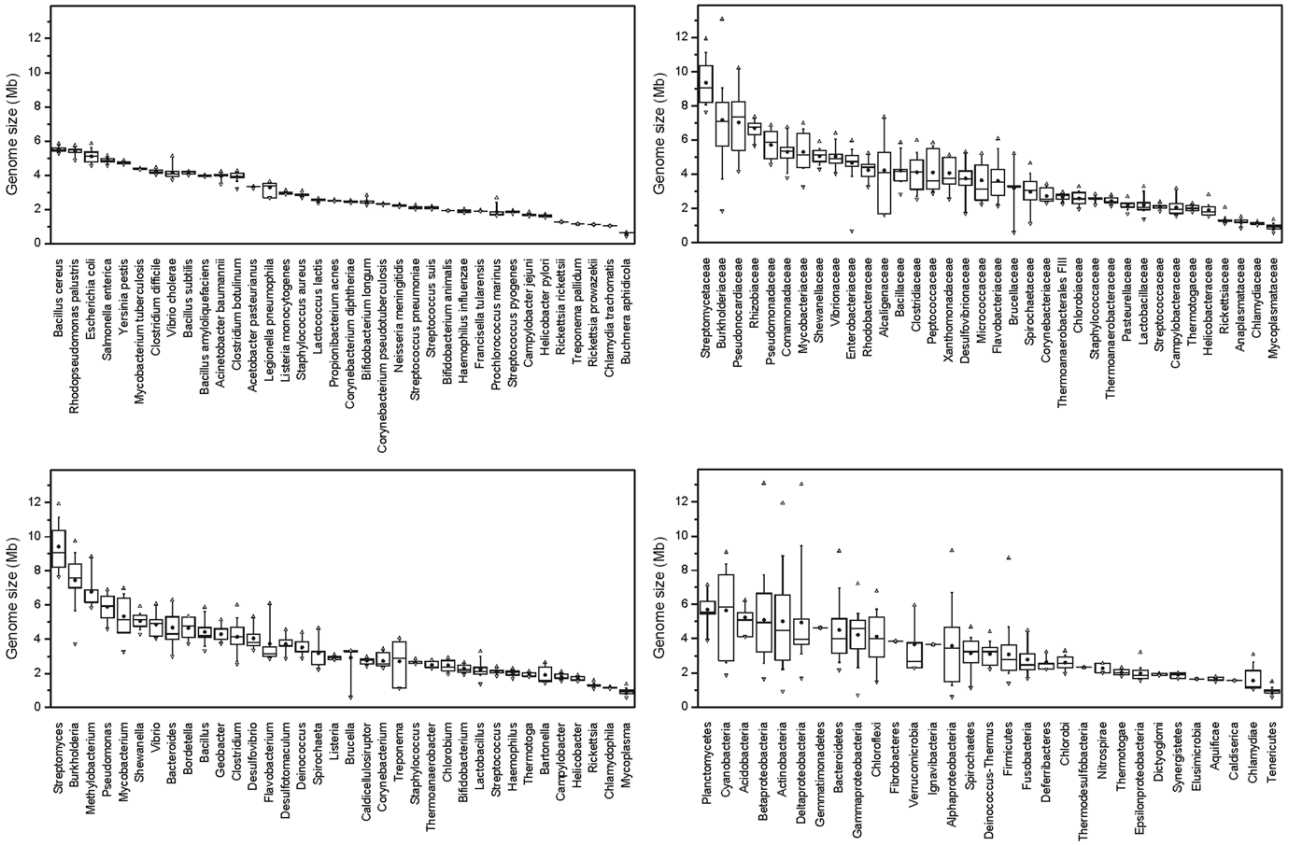


Figure 3. Sizes of bacterial genomes. Sizes of genomes in bacterial phyla (classes), selected families (n>6), genera (n>3), and species (n>6). Open triangles indicate minima and maxima, whiskers the 10th and 90th quantile and boxes the 25th and 75th quantile. Median is indicated as a horizontal line and mean as a dot.
doi:10.1371/journal.pone.0057923.g003

contained sequences of multiple genera (Figure 5). At a 95% similarity level, a large majority of OTUs was still represented by sequences belonging to the same genus (84.6%), while the amount of OTUs containing more than two genera was still relatively low (6.8%).

Assessment of bacterial genome counts and DNA abundances in amplicon pyrosequencing data

The data on the relative abundances of the 16S rRNA sequences in PCR amplicons derived from metagenomic soil DNA were used to estimate the abundance of bacterial genomes (i.e., cells) and DNA. To achieve this, data on 16S rRNA sequence copy numbers of individual OTUs were combined with the predicted 16S rRNA copy numbers and genome sizes, i.e., the values found in the closest known relatives with sequenced genomes. The dataset contained 93,806 sequences that clustered into 15,421 OTUs at a 97% similarity level. When best hits were generated for each OTU, the best hits obtained by comparison with GenBank showed an average 93.2±3.6% similarity with the query, while hits against the genome-derived sequences showed 90.3±3.9% similarity. A different phylum (or class within the Proteobacteria) was assigned to 4.2% of OTUs, most of which has a low similarity to database entries.

In the original dataset, the 16S rRNA sequences of the Acidobacteria, Alphaproteobacteria and Actinobacteria were the most abundant, representing 28.1%, 20.9% and 15.7%, respectively, of sequences in litter and 50.1%, 18.2% and 11.8%, respectively, of sequences in soil. The estimated genome (cell) abundances showed higher share of the Acidobacteria (42.1% in litter and 66.0% in soil) and a lower share of the Actinobacteria (11.3% and 7.6%). The Acidobacteria were even more represented in the total DNA (46.6% and 71.0%). In addition to Acidobacteria, Chlamydiae, Gemmatimonadetes, and Elusimicrobia also showed substantially higher representation among genomes than among 16S rRNA sequences, while the Bacteroidetes, Betaproteobacteria and Firmicutes were less represented (Figure 6). Although the Firmicutes represented 3.2% of 16S rRNA sequences in soil, they were predicted to compose only 1.1% of genomes. In addition, the relative abundance and ranks of individual OTUs also differed substantially when expressed as 16S rRNA or genome abundances (Table S1).

To explore whether the best hits in the database of sequenced genomes provide reliable estimates of 16S rRNA copy numbers and genome sizes, the best hits for the subset of 100 randomly chosen 16S rRNA sequences were retrieved from the dataset containing 16S rRNA sequences from other bacterial species. The best hits gave significantly better estimates of 16S rRNA copy numbers compared to the means of all genomes ($P = 0.0002$, better estimate in 75 cases) and also compared to phylum averages ($P = 0.0058$, 63 cases). This approach also gave better estimates of genome size compared to the grand average ($P = 0.0011$, 67 cases) or phylum average ($P = 0.0443$, better estimate in 63 cases). Using best hits, 16S rRNA copy numbers and genome sizes were estimated with average difference of 24% and 25%; the errors

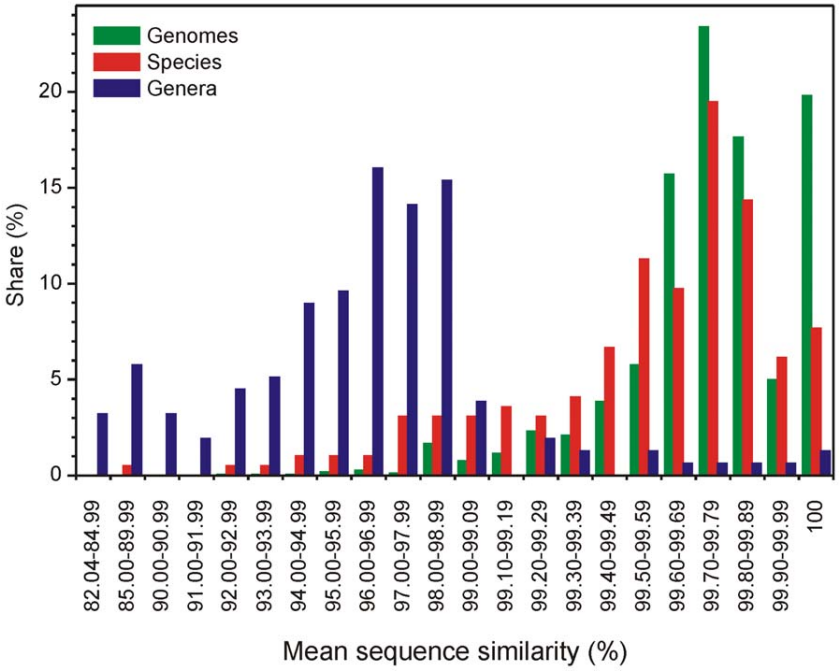


Figure 4. Mean pairwise sequence similarity of 16S rRNA sequences. Similarity within bacterial genomes, among genomes belonging to the same species and among species belonging to the same genus.
doi:10.1371/journal.pone.0057923.g004

were substantially higher when the phylum average was used as the estimate (38% and 32%) or when the grand average was used (48% and 41%). It can therefore be concluded that the use of the best hit in sequenced genomes significantly and substantially improves the estimates of bacterial genome and DNA abundances compared to the abundances in the original 16S rRNA dataset.

Discussion

Although some 15% of bacterial genomes contain only a single 16S rRNA copy, it seems that most bacterial phyla may contain bacteria with more than one copy and one half of the currently analyzed genomes harbor five or more copies. 16S rRNA copy numbers have been found to show both narrow and wide variation within the tested bacterial genera [22,23]. Here, we show that the variability in copy numbers remains high even at the level of bacterial families and genera, but despite some exceptions, especially the species with high 16S rRNA copy numbers, they seem to be conserved at the species level. No statistically significant relationship between genome size and 16S rRNA copy numbers were recorded, but, interestingly, the genome size is much better conserved at all levels of phylogenetic classification.

Only a minority of bacterial genomes with multiple 16S rRNA copies carry identical copies of the gene, although the variation is usually minute. Unsurprisingly, the genomes with more 16S rRNA copies tend to carry more diverse variants of the gene. Among the 568 bacterial species analyzed by [11], 16S rRNA sequences with >1% difference were found in 10% of genomes. Here, we show that 2.4% of genomes have 16S rRNA sequences with <99% mean similarity. Genomes with higher dissimilarity are thus relatively rare: although some distant variants may theoretically evolve into pseudogenes, highly dissimilar 16S rRNA sequences from the same genome were demonstrated to still carry a conserved secondary structure [11] and thus seem to be functional and as such kept in the genomes. The existence of highly dissimilar 16S rRNA sequences in certain genomes and the fact that 16S rRNA similarity within such genome can be lower than among certain bacterial genera (as seen for *Desulfitobacterium youngiae*/*D. hafniense* in the Figure S2) may speak in favor of their evolution by horizontal gene transfer.

The diversity of 16S rRNA within a genome is frequently reported as a factor that can potentially increase diversity estimates obtained by OTU construction; however, that the resolution of the 16S rRNA gene is often too low to allow the differentiation of closely related species [24,25] is often overlooked. Here, we show that this interspecific or intergeneric similarity can also significantly affect OTU construction: in certain genera, member species had exactly the same 16S rRNA sequence (Figure 4). If the standard level of 16S rRNA similarity of 97% is used, a few species, even genomes, can fall to different OTUs due to intragenomic or intraspecific differences. In contrast, species of several genera will not be separated at the same similarity level because in 41.7% of genera, the 16S rRNA differences were lower than 97%. This, however, depends on the part of the 16S rRNA gene used for OTU construction. We should also note that our data were based on bacterial species/genera names that are currently valid. It is unclear how precisely the present definitions of existing bacterial species and genera comply with a biological concept of species definition. This discussion should be opened, and there is also a space for the use of 16S rRNA for this purpose, e.g., for the reclassification of polyphyletic taxa.

Using an example of forest soil bacteria, we show that the recently widely applied 16S rRNA-based abundance estimates provide an imperfect description of bacterial community composition. In general, abundance estimates based on the 16S rRNA sequence counts tend to underestimate the abundance of taxa with low 16S rRNA copy numbers such as the Acidobacteria and to overestimate taxa with high 16S rRNA copy numbers such as

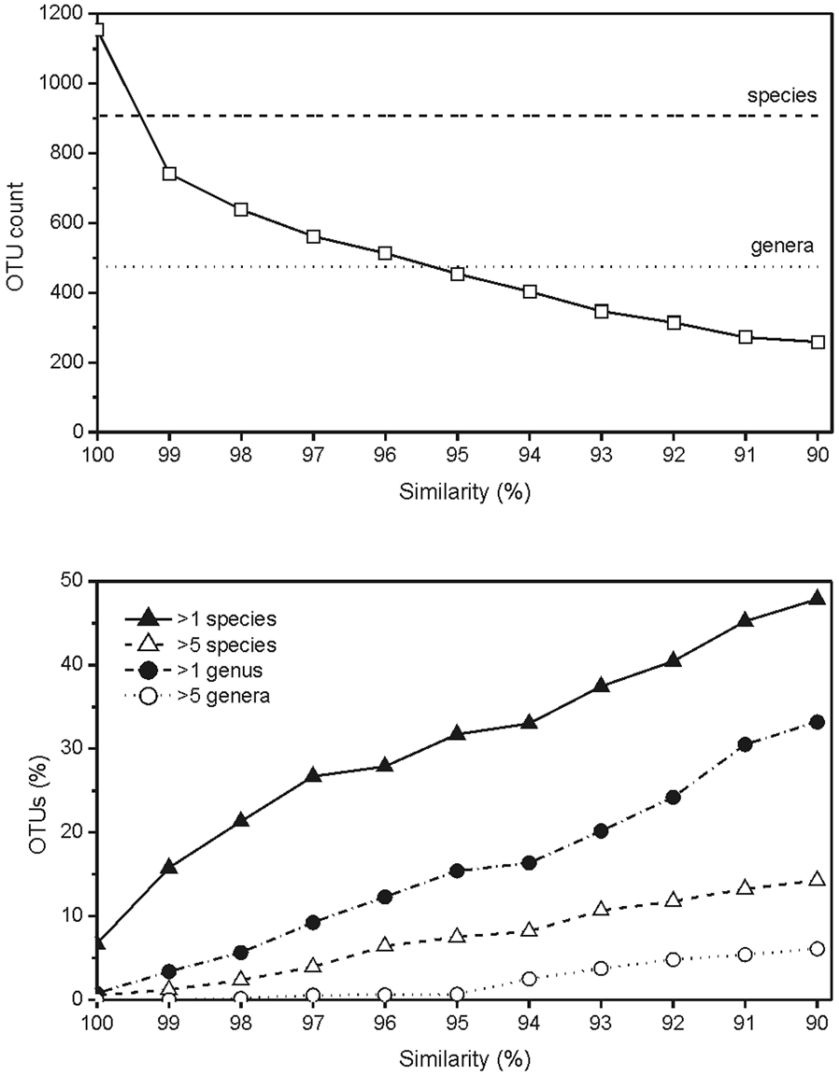


Figure 5. OTU counts and the percentage of OTUs harboring multiple bacterial taxa at various levels of 16S rRNA similarity. Upper panel: OTU counts at various levels of 16S rRNA similarity; the dashed line and the dotted line indicate the number of bacterial species and genera in the Dataset S2. Bottom panel: the percentage of OTUs harboring 16S rRNA sequences belonging to multiple bacterial species or genera.
doi:10.1371/journal.pone.0057923.g005

Gammaproteobacteria and Firmicutes. This fact should be carefully considered when interpreting the 16S rRNA abundance results. For example, the abundance of 16S rRNA sequences of Acidobacteria in certain organic-rich low-pH soils was found to be over 60% [26]. Based on our results, it is possible that their dominance in genome counts is even greater than this, and they may largely dominate in such environments. We also demonstrate that the currently available genome data can substantially change our estimates of bacterial community composition in environmental samples by providing estimates of genome counts that are ecologically more relevant measures of abundance. Caution must be taken, however, when interpreting estimates such as the counts of bacterial cells. Several bacteria, e.g., the hyperthermophile *Thermus thermophilus* or the vegetative cells of filamentous cyanobacteria, were demonstrated to be polyploidic, with 4–18 genome copies per cell [27,28]. In some cases, genome counts per cell can be even considerably higher-over 100 in the cyanobacterial akinetes [27] or the large marine species of *Epulopiscium* [29]. Moreover, even if the abundance estimates are improved as demonstrated, they will be still affected by a wide set of potential biases associated with other steps of the experimental procedures including the variation of DNA extraction efficiency among taxa, differences in PCR amplification or random PCR errors (for more information, see e.g. [13]). More research is definitely needed to reduce these sources of uncertainity.

In the past, several single copy housekeeping genes, e.g., those coding for RNA polymerase, ribosomal proteins or amino-acyl synthetases, or the 60 kDa chaperonin have been proposed as potential phylogenetic markers [30–34], theoretically avoiding the problems with multiple and variable 16S rRNA copies within bacterial genomes. Indeed, it has been documented that the phylogenetic resolution of *cpn60* or *rpoB* genes can be comparable or even better than that of the 16S rRNA [34–35]. Protein-coding single copy genes were also successfully used for the assignment of metagenomic sequences, with a better taxonomic resolution than 16S rRNA fragments [36–37]. However, the use of these genes for

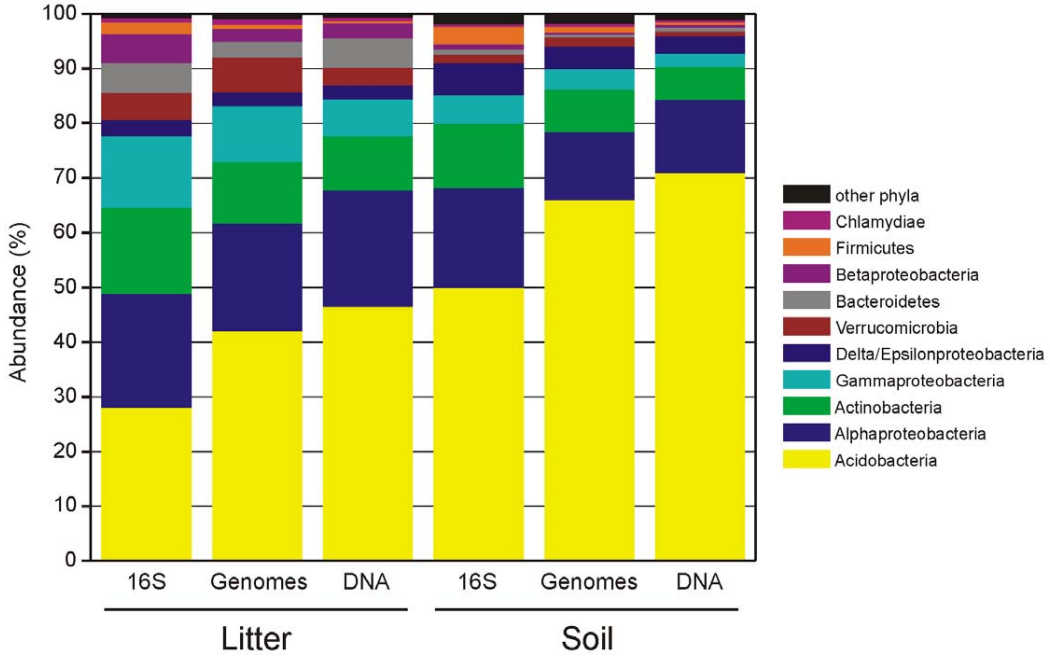


Figure 6. Abundance of bacterial 16S rRNA sequences, genomes and DNA in forest litter and soil. Relative abundance of bacterial 16S rRNA sequences in the amplicon pool from *Picea abies* litter and soil (Baldrian et al., 2012), and estimates of the relative abundance of bacterial genomes and DNA. The estimates were calculated using the values of 16S rRNA copy numbers and genome sizes of the closest hits to each bacterial OTU.
doi:10.1371/journal.pone.0057923.g006

the analysis of environmental amplicons is limited by the fact that the degeneracy of protein sequences make the design of suitable primers difficult. Although universal primers have been designed for the *cpn60* gene [34], the abundance of 16S rRNA gene sequences in the sequence databases greatly exceeds those of other bacterial genes, which still makes it use preferable by increasing the probability of finding a close hit for taxon identification.

Although the 16S rRNA remains still the target of choice for studies in bacterial ecology, this paper demonstrates that there are limitations in its use for community studies and shows how the current developments in bacterial genomics helps to make bacterial abundance estimates more accurate. We strongly believe that in the very near future, the increasing amount of sequenced bacterial genomes allows even better predictions and our knowledge of bacterial community ecology will significantly increase.

Supporting Information

Figure S1 Distribution of 16S rRNA copy numbers in bacterial genomes belonging to selected phyla (classes). The numbers in parentheses indicate numbers of genomes in the respective groups.
(PDF)

Figure S2 Neighbor-joining tree of 16S rRNA sequences from bacterial genomes where the pairwise similarity of at least one 16S rRNA pair within a genome was lower than 97%.
(PDF)

Table S1 Top ten most abundant bacterial OTUs in *Picea abies* forest soil according to the 16S rRNA amplicon sequence copy numbers and estimated genome counts.
(XLS)

Dataset S1 Overview of bacterial genome sequences used in this study.
(XLS)

Dataset S2 Fasta file containing all identified 16S rRNA sequences from bacterial genomes used in this study.
(FAS)

Dataset S3 Fasta file containing aligned 16S rRNA sequences from bacterial genomes used in this study.
(FAS)

Author Contributions

Conceived and designed the experiments: TV PB. Performed the experiments: TV PB. Analyzed the data: TV PB. Wrote the paper: TV PB.

References

1. Hugenholtz P, Goebel BM, Pace NR (1998) Impact of culture-independent studies on the emerging phylogenetic view of bacterial diversity. J Bacteriol 180: 4765–4774.
2. Daubin V, Moran NA, Ochman H (2003) Phylogenetics and the cohesion of bacterial genomes. Science 301: 829–832.
3. Head IM, Saunders JR, Pickup RW (1998) Microbial evolution, diversity, and ecology: A decade of ribosomal RNA analysis of uncultivated microorganisms. Microb Ecol 35: 1–21.
4. Klappenbach JA, Saxman PR, Cole JR, Schmidt TM (2001) rrndb: the Ribosomal RNA Operon Copy Number Database. Nuc Acid Res 29: 181–184.
5. Acinas SG, Marcelino LA, Klepac-Ceraj V, Polz MF (2004) Divergence and redundancy of 16S rRNA sequences in genomes with multiple rrn operons. J Bacteriol 186: 2629–2635.
6. Eichorst SA, Breznak JA, Schmidt TM (2007) Isolation and Characterization of Soil Bacteria That Define Terriglobus gen. nov., in the Phylum Acidobacteria. Appl Environ Microbiol 73: 2708–2717.

7. Klappenbach JA, Dunbar JM, Schmidt TM (2000) rRNA Operon Copy Number Reflects Ecological Strategies of Bacteria. Appl Environ Microbiol 66: 1328–1333.
8. Hashimoto JQ, Stevenson BS, Schmidt TM (2003) Rates and consequences of recombination between rRNA operons. J Bacteriol 185: 966–972.
9. Wang Y, Zwang ZS, Ramanan N (1997) The actinomycete *Thermobispora bispora* contains two distinct types of transcriptionally active 16S rRNA genes. J Bacteriol 179: 3270–3276.
10. Yap WH, Zhang ZS, Wang Y (1999) Distinct types of rRNA operons exist in the genome of the actinomycete *Thermomonospora chromogena* and evidence for horizontal transfer of an entire rRNA operon. J Bacteriol 181: 5201–5209.
11. Pei AY, Oberdorf WE, Nossa CW, Agarwal A, Chokshi P, et al. (2010) Diversity of 16S rRNA Genes within Individual Prokaryotic Genomes. Appl Environ Microbiol 76: 3886–3897.
12. Farrelly V, Rainey FA, Stackebrandt E (1995) Effect of Genome Size and rrn Gene Copy Number on PCR Amplification of 16S Ribosomal RNA Genes from a Mixture of Bacterial Species. Appl Environ Microbiol 61: 2798–2801.
13. Amend AS, Seifert KA, Bruns TD (2010) Quantifying microbial communities with 454 pyrosequencing: does read abundance count? Mol Ecol 19: 5555–5565.
14. Baldrian P, Větrovský T, Cajthaml T, Dobiášová P, Petránková M, et al. (2013) Estimation of fungal biomass in forest litter and soil. Fungal Ecol in press, doi: 10.1016/j.funeco.2012.10.002.
15. Baldrian P, Kolarik M, Stursova M, Kopecky J, Valaskova V, et al. (2012) Active and total microbial communities in forest soil are largely different and highly stratified during decomposition. ISME J 6: 248–258.
16. Dowd S, Callaway T, Wolcott R, Sun Y, McKeehan T, et al. (2008) Evaluation of the bacterial diversity in the feces of cattle using 16S rDNA bacterial tag-encoded FLX amplicon pyrosequencing (bTEFAP). BMC Microbiol 8: 125.
17. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, et al. (2011) MEGA5: Molecular Evolutionary Genetics Analysis Using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. Mol Biol Evol 28: 2731–2739.
18. Li W, Godzik A (2006) Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics 22: 1658–1659.
19. Reeder J, Knight R (2010) Rapidly denoising pyrosequencing amplicon reads by exploiting rank-abundance distributions. Nat Meth 7: 668–669.
20. Edgar RC (2010) Search and clustering orders of magnitude faster than BLAST. Bioinformatics: btq461.
21. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, et al. (1997) Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. Nuc Acid Res 25: 3389–3402.
22. Candela M, Vitali B, Matteuzzi D, Brigidi P (2004) Evaluation of the rrn operon copy number in *Bifidobacterium* using real-time PCR. Lett Appl Microbiol 38: 229–232.
23. Lee CM, Sieo CC, Abdullah N, Ho YW (2008) Estimation of 16S rRNA gene copy number in several probiotic *Lactobacillus* strains isolated from the gastrointestinal tract of chicken. FEMS Microbiol Lett 287: 136–141.
24. Fox GE, Wisotzkey JD, Jurtshuk P (1992) How close is close: 16S rRNA sequence identity may not be sufficient to guarantee species identity. Int J Systematic Bacteriol 42: 166–170.
25. Stackebrandt E, Goebel BM (1994) A Place for DNA-DNA Reassociation and 16S Ribosomal RNA Sequence Analysis in the Present Species Definition in Bacteriology. Int J Syst Bacteriol 44: 846–849.
26. Lauber CL, Hamady M, Knight R, Fierer N (2009) Pyrosequencing-Based Assessment of Soil pH as a Predictor of Soil Bacterial Community Structure at the Continental Scale. Appl Environ Microbiol 75: 5111–5120.
27. Sukenik A, Kaplan-Levy RN, Welch JM, Post AF (2012) Massive multiplication of genome and ribosomes in dormant cells (akinetes) of *Aphanizomenon ovalisporum* (Cyanobacteria). ISME J 6: 670–679.
28. Ohtani N, Tomita M, Itaya M (2010) An Extreme Thermophile, *Thermus thermophilus*, Is a Polyploid Bacterium. J Bacteriol 192: 5499–5505.
29. Bresler V, Fishelson L (2003) Polyploidy and polyteny in the gigantic eubacterium Epulopiscium fishelsoni. Marine Biol 143: 17–21.
30. Dahllof I, Baillie H, Kjelleberg S (2000) rpoB-based microbial community analysis avoids limitations inherent in 16S rRNA gene intraspecies heterogeneity. Appl Environ Microbiol 66: 3376–3380.
31. Santos SR, Ochman H (2004) Identification and phylogenetic sorting of bacterial lineages with universally conserved genes and proteins. Environ Microbiol 6: 754–759.
32. Konstantinidis KT, Ramette A, Tiedje JM (2006) Toward a more robust assessment of intraspecies diversity, using fewer genetic markers. Appl Environ Microbiol 72: 7286–7293.
33. von Mering C, Hugenholtz P, Raes J, Tringe SG, Doerks T, et al. (2007) Quantitative phylogenetic assessment of microbial communities in diverse environments. Science 315: 1126–1130.
34. Schellenberg J, Links MG, Hill JE, Hemmingsen SM, Peters GA, et al. (2011) Pyrosequencing of Chaperonin-60 (cpn60) Amplicons as a Means of Determining Microbial Community Composition. In: Kwon YM, Ricke SC, editors. High-Throughput Next Generation Sequencing: Methods and Application. Totowa: Humana Press Inc. pp. 143–158.
35. Case RJ, Boucher Y, Dahllof I, Holmstrom C, Doolittle WF, et al. (2007) Use of 16S rRNA and rpoB genes as molecular markers for microbial ecology studies. Appl Environ Microbiol 73: 278–288.
36. Roux S, Enault F, Bronner G, Debroas D (2011) Comparison of 16S rRNA and protein-coding genes as molecular markers for assessing microbial diversity (Bacteria and Archaea) in ecosystems. FEMS Microbiol Ecol 78: 617–628.
37. Vos M, Quince C, Pijl AS, de Hollander M, Kowalchuk GA (2012) A Comparison of rpoB and 16S rRNA as Markers in Pyrosequencing Studies of Bacterial Diversity. PLOS One 7: e30600.

MOLECULAR ECOLOGY
RESOURCES
Molecular Ecology Resources (2015) doi: 10.1111/1755-0998.12456

The *rpb2* gene represents a viable alternative molecular marker for the analysis of environmental fungal communities

TOMÁŠ VĚTROVSKÝ,[1] MIROSLAV KOLAŘÍK,[1] LUCIA ŽIFČÁKOVÁ, TOMÁŠ ZELENKA and PETR BALDRIAN
Institute of Microbiology of the ASCR, v.v.i., Vídeňská 1083, 14220, Praha 4, Czech Republic

Abstract

Although the commonly used internal transcribed spacer region of rDNA (ITS) is well suited for taxonomic identification of fungi, the information on the relative abundance of taxa and diversity is negatively affected by the multicopy nature of rDNA and the existence of ITS paralogues. Moreover, due to high variability, ITS sequences cannot be used for phylogenetic analyses of unrelated taxa. The part of single-copy gene encoding the second largest subunit of RNA polymerase II (*rpb2*) was thus compared with first spacer of ITS as an alternative marker for the analysis of fungal communities in spruce forest topsoil, and their applicability was tested on a comprehensive mock community. In soil, *rpb2* exhibited broad taxonomic coverage of the entire fungal tree of life including basal fungal lineages. The gene exhibited sufficient variation for the use in phylogenetic analyses and taxonomic assignments, although it amplifies also paralogues. The fungal taxon spectra obtained with *rbp2* region and ITS1 corresponded, but sequence abundance differed widely, especially in the basal lineages. The proportions of OTU counts and read counts of major fungal groups were close to the reality when *rpb2* was used as a molecular marker while they were strongly biased towards the Basidiomycota when using the ITS primers ITS1/ITS4. Although the taxonomic placement of *rbp2* sequences is currently more difficult than that of the ITS sequences, its discriminative power, quantitative representation of community composition and suitability for phylogenetic analyses represent significant advantages.

Keywords: community composition, diversity, Fungi, internal transcribed spacer, phylogeny, RNA polymerase II

Received 13 August 2014; revision received 28 July 2015; accepted 14 August 2015

Introduction

The advent of next-generation sequencing (NGS) of environmental samples has broadened the exploration of fungal diversity and ecology. One of the most widely used applications of NGS is metabarcoding, which uses a selected DNA sequence to catalogue sequence reads into operational taxonomic units (OTUs) and helps to identify the closest known species. Despite its technical feasibility, this approach to diversity exploration is still limited by several factors, including the method of DNA isolation, selection of marker and PCR primers, and analysis of the data (Quince *et al.* 2008; Youssef *et al.* 2009; Bellemain *et al.* 2010; Engelbrektson *et al.* 2010; Huse *et al.* 2010; Kunin *et al.* 2010; Schloss 2010; Gihring *et al.* 2012). Since the first mycological study applying NGS (Öpik *et al.* 2009), parts of the rDNA cluster have been exclusively used as a molecular marker (Větrovský & Baldrian 2013). The ITS rDNA region, accessible with universal primers, can sufficiently distinguish between most fungal species and is also the most abundant fungal marker in public databases (Nilsson *et al.* 2008; Schoch *et al.* 2012). The rDNA cluster is a multicopy marker present in fungal genomes in 1–200 copies (Raidl *et al.* 2005; Debode *et al.* 2009; Herrera *et al.* 2009; Baldrian *et al.* 2013); although the multicopy nature of this sequence facilitates obtaining ITS amplicons from low-quality DNA, it also seriously limits its value for the quantification of the relative abundance of fungal taxa. The multicopy nature also results in intraspecies and intragenomic variability. Comprehensive studies mapping the distribution of intragenomic ITS rDNA variability among fungi are lacking, although its extent is generally considered to be low (Lindahl *et al.* 2013; Lindner *et al.* 2013) with the exception of the *Glomeromycota* (Stockinger *et al.* 2010; Kruger *et al.* 2012). Most of the studies on the *Ascomycota* and the *Basidiomycota* (Wang & Yao 2005; Smith *et al.* 2007; Simon & Weiss 2008; Connell *et al.* 2010; Lindner *et al.* 2013) have reported a within-genome variability lower than 3%, which corresponds to the

Correspondence: Petr Baldrian, Fax: +420 241 062 384; E-mail: baldrian@biomed.cas.cz

[1]These authors contributed equally to this work.

typical intraspecies variation in this group. In addition to these shallow levels of paralogy, however, deep paralogues reaching 3.6% (Simon & Weiss 2008) or even 10% dissimilarity (Lindner & Banik 2011; Lindner *et al.* 2013) have also been detected. Even more variability could be hidden in potentially unrecognized rDNA pseudogenes (reviewed in Glass *et al.* 2013).

ITS rDNA is a quantitatively dominant marker in public databases, followed by *β*-tubulin (*tub2*), translation elongation factor 1-*α* (*tef1α*), and the second largest subunit of ribosomal polymerase II (*rpb2*) (Feau *et al.* 2011), and the last two also represent the markers included in the Assembling the Fungal Tree of Life (AFTOL) project (Lutzoni *et al.* 2004; James *et al.* 2006). Of them, *tef1α* is not universally present in fungi and together with *tub2* is known to have paralogous copies in certain fungal genomes (Keeling & Inagaki 2004; James *et al.* 2006; Hubka & Kolařík 2012). Thus, the *rpb2* gene is a suitable alternative marker characterized in 6378 from 30 780 species deposited in NCBI GenBank (Table S1, Supporting information). ITS rDNA was proposed as the universal fungal barcode (Schoch *et al.* 2012) for its interspecific variability and the availability of conserved primer sites. However, its intragenomic variability could represent a potential limitation for the study of fungi from environmental DNA (Kiss 2012). Cloning or massively parallel sequencing, unlike Sanger sequencing from genomic DNA, yields sequences derived from single alleles, and fungal diversity may be substantially overestimated if these sequences are sufficiently different within a fungal individual or taxon (e.g. when deep paralogues or pseudogenes are recovered) (Lindner & Banik 2011; Lucking *et al.* 2014). The variable length of the ITS region represents another important problem because there is a strong PCR bias against species with longer amplicons, an issue that was found to largely affect the results of community studies (Ihrmark *et al.* 2012). Lastly, the low conservancy of the ITS region precludes its use in phylogenetic studies on higher taxonomic ranks. The protein-encoding genes are suitable for this, as the translated amino acid sequences can be utilized for high-quality alignments of unrelated fungi. This could be exceptionally valuable for the proper placement of unknown higher taxonomic level lineages, which are often encountered in environmental samples (Baldrian *et al.* 2012; Glass *et al.* 2013).

For the above reasons, the use of single-copy protein-encoding genes as alternative markers could be a solution to the multiple problems associated with rDNA-based markers. Although it is recognized that protein-encoding markers, including the *rpb2* gene, have a better species-resolving power than rDNA markers, the absence of universal primers was believed to limit their use as potential universal barcodes (Schoch *et al.* 2012). In this study, we describe the use of the *rpb2* gene as an alternative marker for fungal community analyses in comparison with the ITS region for investigating fungi inhabiting a coniferous forest floor. Forest litter is dominated by mycorrhizal and saprotrophic basidiomycetes (Baldrian *et al.* 2012; Voříšková & Baldrian 2013), both of which are groups with potentially high intragenomic rDNA variation (Kåuserud & Schumacher 2003; Smith *et al.* 2007; Lindner & Banik 2011). In addition, the in vitro-assembled mock community consisting of the DNA of 130 species was analysed for comparison. Our aim was to explore the potential of the *rpb2* gene as an alternative molecular barcode for the study of fungal diversity and ecology in environmental DNA samples.

Materials and methods

Study site and sample collection

The study area was located in the highest altitudes (1170–1200 m) of the Bohemian Forest mountain range (Central Europe) and was covered by an unmanaged spruce (*Picea abies*) forest (49°02.64 N, 13°37.01 E), an area previously used for the study of active and total microbial community composition (Baldrian *et al.* 2012). To reduce seasonal effects on the fungal community composition that was previously observed for forest soils (Voříšková *et al.* 2014), the sampling was performed in September 2010 (autumn) and March 2011 (spring); the same three sites, located 250 m from each other, were sampled on both occasions. At each site, six topsoil samples (4.5-cm diameter) were collected around the circumference of a 4-m-diameter circle.. The litter horizon (L) and organic soil horizon (S) materials were separately pooled for each site (12 samples in total). After the removal of roots, the L material was cut into 0.5-cm pieces and mixed; the S material was passed through a 5-mm sterile mesh and mixed. Samples to be used for DNA extraction were immediately frozen in liquid nitrogen and stored on dry ice.

DNA was extracted in triplicate from 0.300-g portions of each sample using the SV method (Sagova-Mareckova *et al.* 2008) and cleaned using the Geneclean Turbo Kit (MP Bioiomedicals, Solon, OH, USA). A solution (1 M HEPES/1 M $CaCl_2$, pH 7) was added prior to the cleaning procedure, the sample was left standing for 5 min, and the manufacturer's instructions were then followed.

Tag-encoded amplicon pyrosequencing of soil fungal community

PCRs were performed independently for each extracted DNA sample. The volume of each PCR sample was 50 μL. PCR primers ITS1 (TCCGTAGGTGAACCTG

CGG)/ITS4 (TCCTCCGCTTATTGATATGC) (White *et al.* 1990) were used to amplify the ITS region of the fungal rDNA, and primers bRPB2-6F (TGGGGYAT GGTNTGYCCYGC)/bRPB2-7R (GAYTGRTTRTGRTCR GGGAAVGG) (Matheny 2005) were used to amplify a fragment of the *rpb2* gene (500–800 bp). As there are no commonly used universal primers for *rpb2*, we chose a pair that is assumed to be basidiomycete-specific to preferentially amplify basidiomycetes, which potentially have higher intragenome rDNA variation and dominate in the forest litter and soil of the study area (Baldrian *et al.* 2012). Each 50 μL reaction mixture contained 34.5 μL H_2O, 5 μL 10x buffer for DyNAzyme II DNA polymerase, 3 μL 10 mg/mL BSA, 2 μL forward primer (final concentration, 10 pmol/μL), 2 μL reverse primer (final concentration, 10 pmol/μL), 1 μL template DNA, 1.5 μL 4% Pfu polymerase/DyNAzyme DNA Polymerase (final concentration, 0.06 U/μL) and 1 μL PCR Nucleotide Mix (10 mM). The program for PCR amplification of fungal ITS consisted of initial denaturation at 94 °C, 5 min, 20 cycles (94 °C 30 sec, 51 °C 45 sec, 72 °C 90 sec) and a 15-min final extension at 72 °C. For *rpb2* amplification, the program consisted of initial denaturation at 95 °C, 1 min, 34 cycles (95 °C 1 min, 61 °C 1 min, 72 °C 1 min) and a 10-min final extension at 72 °C. Three parallel PCRs were run per sample. PCR products from the same sample were pooled and cleaned using the Wizard SV Gel and PCR Clean-Up System (Promega, Madison, WI, USA). The concentration of the purified PCR product was quantified using the Quant-iT™ PicoGreen ds DNA kit (Invitrogen, Grand Island, NY, USA). For each sample, 500 ng of PCR product was again cleaned using the MinElute PCR Purification kit (Qiagen, Hilden, Germany). The tagging of individual samples was performed using the Roche MID adaptors and GS Junior Rapid Library Preparation Kit (Roche, Basel, Switzerland), following the manufacturer's instructions. The DNA content in each sample was quantified using the Kapa Library Quantification kit (KapaBiosystems, Woburn, MA, USA) following the manufacturer's instructions, and a equimolar mixture of all ITS samples and all *rpb2* samples was prepared. Pyrosequencing was run on the GS Junior (Roche, Basel, Switzerland).

Bioinformatic analysis

Pyrosequencing of the environmental samples yielded a total of 25 037 raw ITS forward sequences starting with the ITS1 primer and 26 712 raw *rpb2* forward sequences starting with the bRPB2-6F primer. The data were filtered and trimmed using the pipeline SEED (Větrovský & Baldrian 2013). All of the sequences with mismatches in tags were removed from the data set. Pyrosequencing noise reduction was performed using the PyroNoise algorithm translation within Mothur (Schloss *et al.* 2009), chimeric sequences were detected using Uchime (Edgar *et al.* 2011) and deleted. A total of 11 866 ITS and 11 295 *rpb2* sequences were retained after the removal low-quality sequences (mean < 25), sequences shorter than 380 bases, and potentially chimeric sequences. Proportion of chimeric sequences was 24.3% in the ITS data set and 0.3% in the *rpb2* data set (Table S2, Supporting information). The nonfungal sequences, that is, those with a best hit in the NCBI database (Benson *et al.* 2000) to a nonfungal taxon, were removed (105 reads in the ITS data set; 6838 reads in the *rpb2* data set). *rpb2* sequences were truncated to 400 bp and shorter than 400 bp were removed, and the ITS sequences were truncated to contain the entire ITS1 and 5.8S rDNA regions using a fungal ITS extractor (Nilsson *et al.* 2009). Details of preprocessing workflow are shown in Table S2, Supporting information. The data set was randomly resampled at the same sampling depth of 4457 sequences for both molecular targets prior to the diversity analyses. Double clustering (with a 98 and 97% similarity threshold (Lundberg *et al.* 2012)) was performed separately for the *rpb2* and ITS sequences to yield operational taxonomic units using USEARCH (Edgar 2010). The consensus from each OTU was constructed from a MAFFT alignment (Katoh *et al.* 2009) based on the most abundant nucleotide at each position, an approach that enabled us to obtain sequences translatable into proteins. As the intraspecies variability of the *rpb2* gene should be similar to ITS (Schoch *et al.* 2012), we used the same similarity threshold for both genes. Additionally, we also compared the diversity and richness estimates for the data set clustered on the level of 96, 97 and 98% sequence nucleotide similarity in SEED (Větrovský & Baldrian 2013).

Mock community composition and analysis

One hundred and thirty fungal species (83 from Basidiomycota, 42 from Ascomycota and five from other fungal groups), representing common wood and litter degrading and mycorrhizal fungi, were selected to generate an in vitro-assembled mock community. DNA was isolated from axenic cultures or from freshly collected fruiting bodies (Table S3, Supporting information) using ArchivePure DNA Yeast & Gram-+ Kit (5 PRIME, Hamburg). DNA isolated from individual strains was mixed in equimolar proportion. The pyrosequencing was performed using the same methods as in the soil samples. Data processing yielded in total 10 802 raw ITS sequences and 21 831 *rpb2* raw sequences. Those sequences in the *rpb2* data set that were not attributable to this gene (6%) were excluded from analysis. After denoising, quality check, and chimera removal, 7668 sequences remained in the ITS data set and the *rpb2* data

set was resampled to the same size. Proportion of chimeric sequences was 7.4% in the ITS data set and 1.3% in the *rpb2* data set. Double clustering and further processing were the same as in environmental sequences (Table S2, Supporting information).

OTU identification, phylogenetic analyses and statistics

In soil data sets, NCBI tBLASTX was used to generate the best species hits for the OTU consensus sequences of *rpb2*, and BLASTN and the PLUTOF pipeline (Tedersoo *et al*. 2010) were used in the case of the ITS data. Each OTU was assigned to the taxonomic level of class (or nearest lower or higher level when the class was not specified) using the taxonomic information from the NCBI taxonomy server. A tBLASTX search using the *rpb2* gene often showed little similarity to identified sequences, preventing reliable identification (mostly in OTUs belonging to the basal fungal lineages). Thus, phylogeny-based taxonomic assignments of the *rpb2* OTUs were also performed. Lastly, the taxonomic spectra of the *rpb2* and ITS data sets were compared to identify differences in primer selectivity.

A data matrix for the phylogenetic analyses was constructed using a pooled data set consisting of the consensus sequences of 340 *rpb2* OTUs, the alignment published by Lutzoni *et al*. (2004), and the best BLAST matches from the NCBI GenBank. Introns were detected based on the alignment obtained using the multiple alignment tool MAFFT server (http://mafft.cbrc.jp/alignment/server/), and the nucleotide sequences were translated into amino acid sequences. The sequences that showed errors (single-base indels), preventing their translation, were excluded from all of the analyses. The relatedness of the protein sequences was inferred using the maximum-likelihood phylogenetic analysis computed with the GTR substitutions model and the Fast likelihood-based method (aLTR SH-like) of branch support estimation in PhyMLOnline (http://atgc.lirmm.fr.phyml/; Guindon *et al*. 2010). The tree was edited in FIGTREE v1.3.1 (http://tree.bio.ed.ac.uk/software/figtree). The amino acid alignment consisted of 587 sequences in 146 positions of which 139 were variable and 117 parsimony informative. OTUs with an uncertain position within the basal fungal lineages were suspected to represent potential paralogues and pseudogenes. We confirmed their orthologous nature using a similarity search against the annotated SWISS-PROT database using a BLASTX search in NCBI GenBank and on the basis of indices measuring codon usage bias, which were calculated using the methods described by Hubka & Kolařík (2012) .

In the mock data sets, OTUs were identified using BLASTN similarity search against the GenBank NCBI database and, where necessary, against local databases of sequences obtained by Sanger sequencing. Only species that were represented at least by 30 sequences in both data sets were used in comparison of intragenomic variability of the ITS and *rpb2* genes and were both resampled at 30 sequences per species. OTU construction was performed as described above.

The software package STAMP (Parks & Beiko 2010) was used for the analysis and comparison of taxonomic unit profiles (Parks & Beiko 2010) using Fisher's exact test with a Bonferroni correction. Differences at $P < 0.01$ were regarded as statistically significant. A sign test was used for the comparison of diversity estimates using the ITS and *rpb2* markers in individual samples. Differences at $P < 0.05$ were regarded as statistically significant.

Results

Clustering of environmental sequences at a 97% similarity threshold yielded 497 OTUs in the ITS data set (273 singletons, 55% of all OTUs, 6.1% of reads) and 340 OTUs in the *rpb2* data set (149 singletons, 44% of all OTUs, 3.3% of all reads). A nucleotide sequence similarity ≥97% with the GenBank hits was found for 37% of ITS OTUs and for 2% of *rpb2* OTUs (Figure S1, Supporting information).

The taxonomic assignments of the ITS and the *rpb2* sequences were compared at the level of class or neighbouring taxonomic units (Fig. 1, Table 1). The relatively long and conserved partial *rpb2* protein sequence enabled the construction of a phylogenetic tree covering all fungi (Fig. 2), which proved to be useful for the taxonomic placement of fungi within the underexplored basal lineages for which tBLASTX did not provide reliable identification. The distribution of OTUs among 35 recognized fungal lineages (ITS, 21 lineages; *rpb2*, 28 lineages) (Fig. 1, Table 1) suggests that both primer sets have broad taxonomic coverage. The nonfungal *rpb2* sequences (1799 OTUs) belonged to all major groups of typical soil organisms, including archaea, bacteria, various protists, soil fauna, algae and plants (Table S4, Supporting information). Although the ITS data set contained more OTUs classified into Ascomycota (ITS: 59%, *rpb2*: 27% of all fungal OTUs) and a similar number assigned to Basidiomycota (ITS: 31%, *rpb2*: 32% of all fugal OTUs), the number of sequences assigned to basal fungal lineages was significantly higher in the *rpb2* data set (ITS: 9%, *rpb2*: 41% of all OTUs). In both data sets, the phylum Basidiomycota was the most dominant, followed by Ascomycota and other fungi, when the number of sequence reads was taken into account.

The dominant part of the fungal communities, as indicated by the ITS marker, was represented by the groups of Agaricomycetes, Leotiomycetes, Eurotiomycetes, Dothideomycetes, Tremellomycetes, Pucciniomycotina

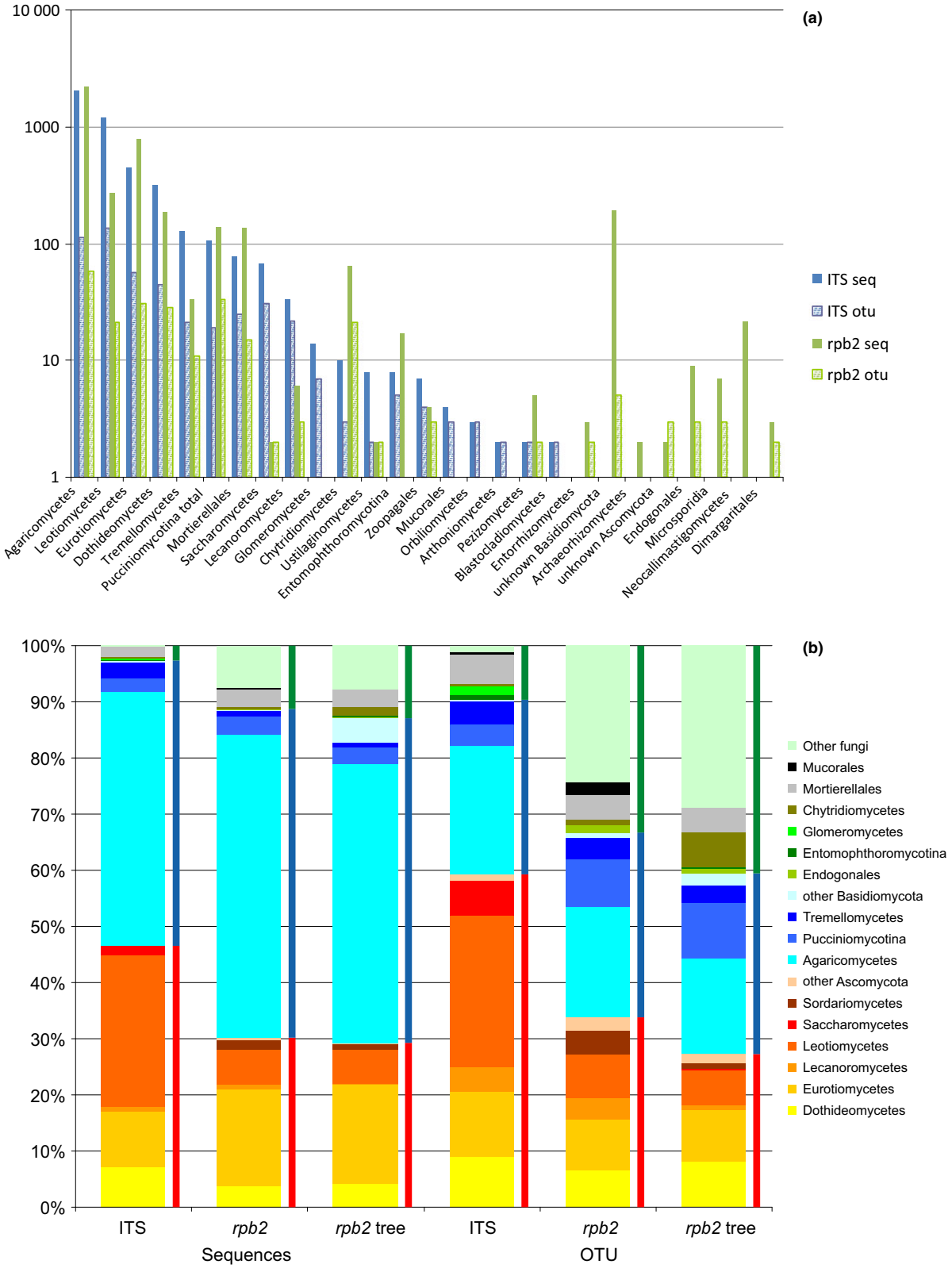


Fig. 1 Abundance of sequences and OTUs at the level of fungal class identified from the ITS and *rpb2* data sets. (a) abundances of fungal classes, (b) relative composition of fungal communities based on sequence and OTU counts. The thin bars in the panel B use red colour to indicate the Ascomycota, blue for the Basidiomycota and green for other fungal lineages.

Table 1 Relative abundance of sequences and OTUs at the level of fungal class identified from the ITS and *rpb2* in spruce forest topsoil

		ITS BLASTN		*rpb2* tBLASTX		*rpb2* tree	
	Fungal class	Reads	OTUs	Reads	OTUs	Reads	OTUs
Basidiomycota	Agaricomycetes	2015**	113	2409	67	2207	58
	Tremellomycetes	127**	21	43	13	33	11
	Agaricostilbomycetes	57**	10	6	1		
	Cystobasidiomycetes	0	0	10	2		
	Microbotryomycetes	24**	6**	127	26	138	33
	Pucciniomycetes	27**	3	0	0		
	Exobasidiomycetes	0	0	3	2	0	0
	Ustilaginomycetes	8	1	0	0	1	1
	Entorrhizomycetes	0	0	3	1	3	1
	Unknown Basidiomycota	–	–	–	–	194	5
	Basidiomycota total	2258**	154	2601	112	2576	109
Ascomycota	Arthoniomycetes	1	1	0	0	0	0
	Dothideomycetes	316**	45	167	22	186	28
	Eurotiomycetes	446**	57	769	31	785	31
	Lecanoromycetes	33	22	33	13	6	3
	Leotiomycetes	1209**	134**	282	27	269	21
	Lichinomycetes	0	0	5	2	0	0
	Orbiliomycetes	3	3	2	1	0	0
	Pezizomycetes	2	2	0	0	5	2
	Archaeorhizomycetes	0	0	2	1	2	1
	Pneumocystidomycetes	0	0	3	1	0	0
	Schizosaccharomycetes	0	0	3	3	0	0
	Sordariomycetes	0**	0**	76	14	52	3
	Saccharomycetes	68**	31**	0	0	1	1
	Unknown Ascomycota	–	–	–	–	2	3
	Ascomycota total	2078**	295**	1342	115	1308	93
Basal lineages	Blastocladiomycetes	1	1	1	1	0	0
	Entomophthoromycotina	8	5	0	0	17	1
	Endogonales	0	0	10	4	9	3
	Mucorales	4	3	12	7	0	0
	Chytridiomycetes	10	3	16	4	64	21
	Microsporidia	0	0	7	3	7	3
	Monoblepharidomycetes	0	0	3	1	0	0
	Mortierellales	77**	25	140	15	137	15
	Glomeromycetes	14**	7	0	0	0	0
	Neocallimastigomycetes	0**	0	22	1	22	1
	Dimargaritales	0	0	0	0	3	2
	Zoopagales	7	4	6	3	4	3
	Basal lineages of uncertain position	0**	0**	297	74	310	89
	Basal lineages total	121**	48**	514	113	573	138
Total		4457	497	4457	340	4457	340

The asterisks indicate significant differences in the relative abundance of sequences and OTUs based on ITS BLASTN and *rpb2* tBLASTX, $*P < 0.01$, $**P < 0.001$.

and Mortierellales (ordered based on sequence abundance). This result is in general agreement with the spectrum identified using tBLASTX within the *rpb2* data set (Table 2), with most sequences belonging to Agaricomycetes, Eurotiomycetes, a group of undefined basal lineages, Leotiomycetes, Dothideomycetes, Pucciniomycotina and Mortierellales. From the more abundant major groups, only Glomeromycetes and Saccharomycetes were found in ITS but were absent or very rare in the *rpb2* data set.

Although the spectrum of the most abundant OTUs identified by both primer sets was remarkably overlapping, their abundance differed largely. An example is the genus *Lactarius*, which was represented by 40% of all sequences in the *rpb2* data set, but only 1.3% sequences in the ITS data set (Table 2). Conversely, *Piloderma*, which dominated among the ITS sequences (10.6%), was rare in the *rpb2* data set (2.4%). The identified spectrum included symbionts of forest trees and ericoid, saprotrophic ascomycetes, basidiomycetes and

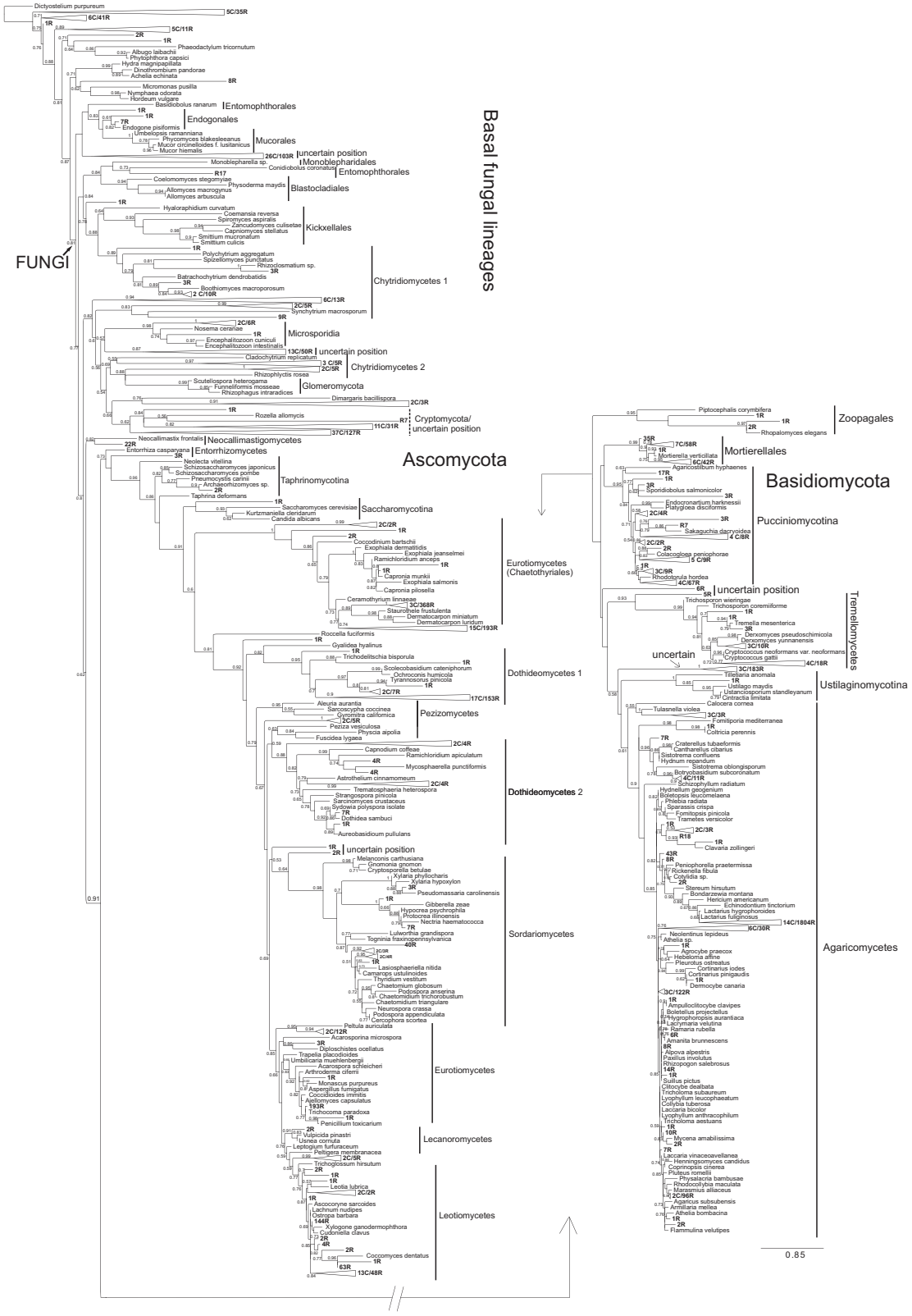


Fig. 2 Maximum-likelihood phylogenetic tree showing the relationships of *rpb2* OTUs, the most related GenBank entries retrieved using tBLASTX, and other representatives of major fungal groups. The clades containing OTUs only were collapsed, and the number of OTUs (O) and sequence reads (R) is marked.

Table 2 Identification of the thirty most abundant OTUs in spruce forest topsoil using ITS and *rpb2* as alternative molecular markers. The identity was assigned using BLASTN (for ITS) or tBLASTX (for *rpb2*). The percentage similarity of protein sequence is given in the case of *rpb2*. The taxonomic classification to *Ascomycota* (A) and *Basidiomycota* (B) is marked

ITS					*rpb2*				
Identification	Accession number	Similarity (%)	Coverage (%)	Abundance (%)	Identification	Accession number	Similarity (%)	Coverage (%)	Abundance (%)
Piloderma (B)	JQ711958	97	100	10.6	*Lactarius* (B)	DQ408128	98	99	22.9
Lachnellula (A)	KC464638	98	100	7.0	*Lactarius* (B)	DQ408128	98	99	13.3
Mycena (B)	EF093152	100	100	6.2	*Ceramothyrium* (A)	AY485617	79	99	8.2
Cenococcum (A)	AM087244	99	100	4.4	*Rasamsonia* (A)	JQ729684	97	99	4.3
Marasmius (B)	FR717227	99	97	4.1	*Lactarius* (B)	DQ408128	99	99	3.6
Cladophialophora (A)	EF016381	97	100	3.8	*Hydropus* (B)	DQ472722	71	62	3.5
Russula (B)	HM189931	99	100	2.9	*Albotricha* sp. (A)	AB481347	98	99	3.2
Hygrophorus (B)	JF908073	100	100	2.5	*Piloderma* (B)	GU187797	99	99	2.4
Meliniomyces (A)	HQ157837	100	100	2.4	*Ceramothyrium* (A)	AY485617	92	90	2.3
Tylospora (B)	JN943896	100	100	2.4	*Gymnopus* (B)	DQ472716	99	97	1.7
Tricholoma (B)	AB036899	99	100	1.8	*Cudonia* (A)	AY641033	92	99	1.4
Uncultured (*Trechispora*) (B)	FJ475683	99	100	1.8	*Tyrannosorus* (A)	DQ470928	74	52	1.4
Cladophialophora (A)	EF016381	96	100	1.5	*Veluticeps* (B)	HM536125	92	100	≤1
Ascomycete (*Scytalidium*) (A)	GU067746	99	100	1.5	*Umbelopsis**	DQ302787	64	96	≤1
Inocybe (B)	AJ889955	99	99	1.5	*Ellisembia* (A)	DQ435090	79	94	≤1
Lactarius (B)	JQ712010	99	100	1.3	*Microbotryum* (B)	DQ789985	90	99	≤1
Rhizoscyphus (A)	JQ711796	96	100	1.3	*Mortierella*	DQ302784	92	100	≤1
Lyophyllum (B)	HE819396	99	45	1.3	*Microthyrium* (A)	GU371734	77	99	≤1
Lachnellula (A)	AB481245	98	100	1.3	*Microbotryum* (B)	DQ789985	90	99	≤1
Xerocomus (B)	HQ207696	100	100	1.1	*Spiromyces**	DQ302790	54	99	≤1
Cadophora (A)	AB543058	97	87	≤1	*Microthyrium* (A)	GU371734	75	99	≤1
Rhizoscyphus (A)	JQ711893	95	100	≤1	*Hydropus* (B)	DQ472722	71	62	≤1
Mortierella	JQ272448	99	100	≤1	*Mortierella*	DQ302784	95	99	≤1
Amanita (B)	EF493271	100	100	≤1	*Mortierella**	DQ302784	78	96	≤1
Meliniomyces (A)	HQ157926	99	100	≤1	*Ceramothyrium* (A)	AY485617	89	99	≤1
Leptodontidium (A)	GU067735	96	100	≤1	*Mortierella*	DQ302784	92	86	≤1
Exophiala (A)	HE605215	87	100	≤1	*Trichopeziza* (A)	AB481360	98	98	≤1
Tylospora (B)	JN943896	99	100	≤1	*Coprinopsis* (B)	XM1829088	100	93	≤1
Uncultured (*Leptodontidium*)	HM488455	100	100	≤1	*Gymnopus* (B)	DQ472716	99	99	≤1
Uncultured (*Rhodosporidium*)	FJ475820	100	62	≤1	*Mortierella*	DQ302784	92	92	≤1

*OTUs with low similarity to known fungal sequences were placed into basal fungal lineages of uncertain position based on the phylogenetic tree.

zygomycetes constituting the core part of the fungal community.

In our study, there was high diversity and abundance of sequences classified into various groups of the basal fungal lineages in the *rpb2* data set (41% of OTUs, 23% of reads). From these, the most abundant OTUs were clustered with *Rozella* (*Cryptomycota*), *Chytridiomycetes* or belonged to several clades with uncertain position between the known lineages of basal fungi (Fig. 2).

The two alternative markers were also used for the exploration of fungal richness. The OTU richness in individual samples was quantified at 209 sequences per sample and ranged from 34 to 80, with an average of 57 in the *ITS* data set, which was significantly higher than in the *rpb2* data set (26–66 OTUs, average of 46). The Chao1 estimates were also significantly higher for *ITS*, although differences in the Shannon–Wiener index and community evenness were not significantly different. Interestingly, the *rpb2* but not *ITS* analysis showed a high level of significance for differences in the species richness and diversity estimates between the litter and soil (Figure S2, Supporting information). The OTU richness and diversity of the entire data set were higher for the *ITS* data set (497 OTUs, Shannon–Wiener Index of 4.45, Simpson diversity index of 0.031 and Chao1 of 940) than the *rpb2* data set (340, 3.63, 0.085 and 527, respectively). The difference in Chao1 estimates was primarily due to the higher number of singletons in the *ITS* data set (270 vs. 149), and the data sets with excluded singletons showed similar diversity estimates (ITS data set: 227 OTUs, Shannon–Wiener Index of 4.13, Simpson diversity index of 0.035 and Chao1 of 227; *rpb2* data set: 191 OTUs, Shannon–Wiener Index of 3.43, Simpson diversity index of 0.091 and Chao1 of 191). We observed a lower effect of nucleotide similarity on the resulting OTU numbers for *rpb2* than the ITS data set: the *rpb2* data set showed only a slight increase in OTU number among 90–97.5% similarity (Figure S3, Supporting information). The similarity used in our study in both genes (98% preclustering, 97% clustering) roughly corresponds to the similarity of 96.5% applied to the original sequence set (without preclustering), showing that the effect of different interspecies variability in both genes does not significantly influence the estimated diversity.

The identification of *rpb2* OTUs using tBLASTX was sufficiently straightforward in most cases when the sequence similarity was sufficiently high (99–85%). For those OTUs that did not have close matches in GenBank, their position within the fungal tree of life had to be inferred using a phylogenetic analysis of their protein sequences. The *rpb2* sequences were found to be conserved enough to allow the construction of a robust phylogenetic tree covering all of Eukaryota. The chosen *rpb2* marker region only rarely contains introns (detected only in three OTUs in this study), which is an another advantage for phylogenetic comparisons. Amplified *rpb2* region can contain up to ten introns as showed by our *in silico* analysis, but these are located out of the sequenced part, and only few fungi from basal lineages contained one or two introns (Table S5, Supporting information).

Our tree was in general agreement with the fungal tree of life (James *et al.* 2006; Hibbett *et al.* 2007; Jones *et al.* 2011). Several OTUs were found to form a new lineage sister to the *Ustilaginomycotina*. Members of the *Chaetothyriales* were not sister with other members of *Eurotiomycetes*, as it is known from multigene phylogenies and were placed close to *Dothideomycetes* in agreement with morphology and *rpb2* based phylogenetic comparisons (Liu & Hall 2004). Three of the most dominant OTUs (Table 2) and several other OTUs belonged to lineages identical or close to the *Cryptomycota*. Other lineages of unknown identity are marked in Fig. 1 and Table 1.

Interpretation of the previous results is limited by the fact that the real species diversity, taxon spectrum and abundance in the analysed soil community are not known. To overcome this limitation, mock community consisting of equimolar DNA concentrations of 130 fungal taxa was studied. Clustering resulted in 275 (120 singletons, 43% of all OTUs, 1.5% of all reads) and 177 OTUs (84 singletons, 47% of all OTUs, 1% of all reads) in the ITS and *rpb2* data sets, respectively. Sequences of all species were recovered using at least one marker: 68 were present in both data sets, whereas 29 and 33 species were only present in the *rpb2* and ITS data sets, respectively (Table S3, Supporting information). All major fungal groups (*Ascomycota*, *Basidiomycota*, *Mortierellomycotina*, *Mucoromycotina*) were present in similar proportions in both data sets, when the number of species was considered. However, the observed diversity (OTU counts) and abundance (read counts) of these higher taxa in the ITS data set differed significantly from expected values (χ^2 test, $P = 0.036$ for OTU number). The same comparison in *rpb2* data set showed that observed OTU number differed from expected value (χ^2 test, $P = 0.027$) but number or reads did not ($P = 0.26$). Most importantly, ITS highly overestimated the OTU counts and relative abundance of the *Basidiomycota* (Table 3). Furthermore, considerable differences in taxon abundance between the two markers were observed which was also the case when comparing the soil community data sets. For example, only seven species ranked among the 20 most abundant species in both data sets.

OTU numbers estimated in the mock community exceeded the real number of species included with both ITS and *rpb2*. When the 15 species most abundant in both data sets were analysed at the same sequencing depth, ITS gave 2.53 ± 1.13 OTU per species and *rpb2* gave

Table 3 Comparison of the representation of major fungal groups in the mock community and in the sequence data sets obtained with ITS and *rpb2* as molecular markers

		ITS			*rpb2*		
	Mock community	Species detected	OTUs	Reads (%)	Species detected	OTUs	Reads (%)
Basidiomycota	83 (63.8%)	64 (63.4%)	205 (74.5%)	90.0	68 (70.1%)	133 (75.1%)	62.9
Ascomycota	42 (32.3%)	32 (31.7%)	62 (22.5%)	8.6	25 (25.8%)	38 (21.5%)	35.3
Mucoromycotina and Mortierellomycotina	5 (3.9%)	5 (4.9%)	8 (3.0%)	1.4	4 (4.1%)	6 (3.4%)	1.8

1.87 ± 0.92 OTU per species, which was marginally less ($P = 0.083$).

Discussion

Universality of degenerated rpb2 primers

The *rpb2* primers used in this study were designated by their authors as basidiomycete-specific (Matheny 2005), and also, our *in silico* analysis indicated their preference for Agaricomycetes and a lower specificity for other groups (Table S6, Supporting information). Unexpectedly, in our study, the *rpb2* gene produced a more taxonomically diverse set of fungal sequences than the universal *ITS* primers. The *rpb2* data set contained much more nonfungal sequences than the ITS data set (61% vs. 1% of reads). The nonfungal organisms amplified using the *ITS* primers primarily belonged to *Ciliata* and *Viridiplantae*, which is in line with the fact that members of fungi, ciliates and plants were used by White *et al.* (1990) for ITS1 and ITS4 primer design, and our data confirmed that these primers are biased towards these eukaryotic groups (Table S4, Supporting information). The provenience of nonfungal sequences in the *rpb2* data set was much complex, containing all major groups of soil biota. In *ITS* and *rpb2* data sets from environmental samples, a proportionally similar number of basidiomycete reads (ITS: 51%, *rpb2*: 58% of all reads) was found in both data sets, showing that both primer sets had similar taxonomic coverage in *Dikarya* (*Ascomycota* and *Basidiomycota*). The broad taxonomic coverage of *rpb2* primers was fully supported by the results of the mock community analysis. It is widely known that use of degenerate primers can greatly reduce the specificity of PCR amplifications, although the mechanisms responsible for such universality of degenerate *rpb2* primers in the amplification of a highly complex DNA template mixture is unknown. The existence of universal primers with sufficient taxonomic resolution is one of the major criteria required for a universal barcode in fungi (Schoch *et al.* 2012). Single-copy protein-coding genes provide good taxonomic resolution but are considered inaccessible using universal primers (Schoch *et al.* 2012). Here, we show that lower primer universality is not necessarily a limitation for the metabarcoding.

Comparison of the ITS and rpb2 data sets at various taxonomic levels

The fungal taxon spectra obtained using the *rbp2* and *ITS* markers corresponded both in the environmental and mock communities, but taxon abundances differed widely. The soil and litter of the coniferous forests of the boreal and temperate zones are considered to be dominated by saprotrophic and mycorrhizal *Dikarya*, which represented typically about 90% of OTUs in the ITS data sets (O'Brien *et al.* 2005; Buéé *et al.* 2009; Baldrian *et al.* 2012) and also in our study. The most notable difference between the ITS and *rpb2* data sets from our study was the high diversity and abundance of sequences classified into various groups of the basal fungal lineages in the *rpb2* data set (41% of OTUs, 23% of reads). Basal fungal lineages are often zoosporic, difficult to cultivate, widespread in water ecosystems or soil; their members live as saprobes, symbionts or parasites (Jones *et al.* 2011; Marano *et al.* 2012). Freeman *et al.* (2009) found that Chytridiomycota constituted over 70% of rDNA sequences in high-elevation soils without vegetation cover. High abundance of chytrids (10% of sequences) was also reported from periodically flooded alpine tundra soil under the snow cover (Freeman *et al.* 2009). Similar conditions, where zoosporic fungi can thrive, occur in the forest from the present study, and thus, they may represent an abundant part of the mycobiota. It is possible that zoosporic fungi were abundant in the studied ecosystem and yet were underestimated by the ITS marker due to the lack of complementarity of ITS primers or possibly due to the proportionally lower rDNA copy numbers per genome in these fungi. Future research should address whether these lesser known fungi constitute a significant portion of the mycobiome of coniferous forests.

Diversity estimation using ITS and rpb2 gene

The published fungal diversity estimates from environmental samples based on ITS analyses are notoriously

high (Blackwell 2011) and should be critically re-evaluated. The presence of chimeras, deeper paralogues in multicopy markers and pseudogenes are the main source of error. In our environmental samples and mock community, the number of singletons was much higher in the ITS data set. The same was observed for the OTU diversity in environmental samples. This is evidently attributable to the higher intragenic variability of the ITS marker as shown by our comparative analysis of the mock community. The ITS data set may theoretically contain deeper intragenomic paralogues, which increased number of singletons and OTUs a in the ITS data set. The difference in the spectrum of fungal groups recovered by the two primer sets may be an alternative explanation because the degree of intragenomic and intraspecies variability is not equal across the fungal kingdom (Nilsson *et al.* 2008).

Our study revealed that multiple sequence clusters phylogenetically distant from known taxa were present in the environmental sample. Such lineages were also reported in several other studies. The one of Jumpponen (2007) indicated that such clusters located at more basal positions in the rDNA phylograms could be chimeric and that the proportion of such chimeric sequences could be very high. The fact that only DNA sequences that are translatable into protein were used in our phylogenetic analysis partly eliminates the risk of chimeras or pseudogenes with the protein-encoding markers. The ITS marker seemed to be much more sensitive to chimera formation than *rpb2* (proportion of chimeric sequences detected by UChime was 25.3% in ITS and 0.30% in *rpb2* sequences, Table S2, Supporting information). Although there is the risk that unrecognized paralogues are present in the *rpb2* data set, such as, for example, in another protein-coding gene, *β*-tubulin (Hubka & Kolařík 2012), genes paralogous to *rpb2* have not been reported thus far. The probability that the phylogenetically distant sequence clusters recovered using *rpb2* represent such paralogues is thus low, and it is possible that these sequences correspond to already known fungal groups, whose *rpb2* sequences are missing in public databases. The unassigned lineages may also represent novel ones that were not previously amplified using ITS primers. The advantages of degenerate primers in the discovery of new uncultured fungi should thus be further evaluated.

As similar levels of OTU inflation were observed for both markers in the soil community as in the defined mock community, this may suggest that the real diversity in the soil is closer to the estimate obtained using the *rpb2*. This inflation is partially attributable to higher intragenomic variability of the ITS region. The polymorphism was found mostly in the *Polyporales* and some members of the *Agaricales*. DNA in our study originated from monokaryotic haploid *Ascomycota* cultures as well as dikaryotic cultures or fruiting bodies of the *Basidiomycota*. The heterokaryotic members of the latter group can theoretically possess divergent alleles that may cause the presence of two separate OTUs after clustering and it seems probable that intraspecies variability in both markers may affect diversity estimates in real samples. It should be also noted that the distribution of sequences among taxa of the mock community was slightly more even in the *rpb2* data set where 80% sequences were represented by 12 species compared to seven species in the ITS. Still, the reason for highly uneven distribution of read counts among species that was also observed in previous studies (Ihrmark *et al.* 2012), remains unclear: theoretically, it can be due to the combination of primer specificity, PCR preference, DNA quality or other factors.

rpb2 as an alternative metabarcoding marker

The ITS gene currently offers an unmatched opportunity for the close identification of a particular OTU. However, the present study shows that the single-copy, protein-encoding gene *rpb2* may be a viable option for fungal metabarcoding. Our results show general agreement in the identity of the fungal genera and fungal classes recovered using this marker and ITS. No such agreement was found in the sequence abundance of the main genera or major fungal groups. Indeed, the single-copy nature of *rpb2* represents an important advantage for proper estimations of diversity and relative abundance; furthermore, the constant length of the *rpb2* amplicon avoids the PCR bias observed in the case of ITS, where length varies largely among taxa (Figure S4, Supporting information). The *rpb2* gene possesses a barcode gap between the inter- and intraspecific variation that is much more clearly defined than in the ITS1 sequence, which corresponds to the results of Schoch *et al.* (2012). The *rpb2* gene has taxonomic sensitivity superior to the ITS (Schoch *et al.* 2012), and our results reveal that this sequence is well suited for the study of basal fungal lineages. The use of a translatable protein-coding gene also enables the identification of potential pseudogenes and the construction of robust phylogenetical trees. Although the precise taxonomic placement of *rbp2* sequences is currently more difficult than it is for ITS due to the lower representation in GenBank, the phylogenetic discriminative power, better quantitative representation of the community composition and suitability for phylogenetic analyses may represent comparative advantages for *rbp2* over the use of ITS and make this molecular marker useful for studies in fungal ecology and diversity.

Acknowledgements

This work was supported by the Czech Science Foundation (13-06763S, P504/12/0709), the Ministry of Education, Youth and Sports of the Czech Republic (LD12050), and by the Research concept of the Institute of Microbiology ASCR (RVO61388971).

References

Baldrian P, Kolařík M, Štursová M *et al.* (2012) Active and total microbial communities in forest soil are largely different and highly stratified during decomposition. *ISME Journal*, **6**, 248–258.

Baldrian P, Větrovský T, Cajthaml T *et al.* (2013) Estimation of fungal biomass in forest litter and soil. *Fungal Ecology*, **6**, 1–11.

Bellemain E, Carlsen T, Brochmann C *et al.* (2010) ITS as an environmental DNA barcode for fungi: an in silico approach reveals potential PCR biases. *BMC Microbiology*, **10**, 189.

Benson DA, Karsch-Mizrachi I, Lipman DJ *et al.* (2000) GenBank. *Nucleic Acids Research*, **28**, 15–18.

Blackwell M (2011) The fungi: 1, 2, 3.. 5.1 million species? *American Journal of Botany*, **98**, 426–438.

Buéé M, Reich M, Murat C *et al.* (2009) 454 Pyrosequencing analyses of forest soils reveal an unexpectedly high fungal diversity. *New Phytologist*, **184**, 449–456.

Connell LB, Redman R, Rodriguez R *et al.* (2010) *Dioszegia antarctica* sp. nov. and *Dioszegia cryoxerica* sp. nov., psychrophilic basidiomycetous yeasts from polar desert soils in Antarctica. *International Journal for Systematic and Evolutionary Microbiology*, **60**, 1466–1472.

Debode J, Van Hemelrijck W, Baeyen S *et al.* (2009) Quantitative detection and monitoring of *Colletotrichum acutatum* in strawberry leaves using real-time PCR. *Plant Pathology*, **58**, 504–514.

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

Edgar RC, Haas BJ, Clemente JC, Quince C, Knight R (2011) UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics*, **27**, 2194–2200.

Engelbrektson A, Kunin V, Wrighton KC *et al.* (2010) Experimental factors affecting PCR-based estimates of microbial species richness and evenness. *ISME Journal*, **4**, 642–647.

Feau N, Decourcelle T, Husson C, Desprez-Loustau ML, Dutech C (2011) Finding single copy genes out of sequenced genomes for multilocus phylogenetics in non-model fungi. *PLoS ONE*, **6**, e18803.

Freeman KR, Martin AP, Karki D *et al.* (2009) Evidence that chytrids dominate fungal communities in high-elevation soils. *Proceedings of the National Academy of Sciences, USA*, **106**, 18315–18320.

Gihring TM, Green SJ, Schadt CW (2012) Massively parallel rRNA gene sequencing exacerbates the potential for biased community diversity comparisons due to variable library sizes. *Environmental Microbiology*, **14**, 285–290.

Glass DJ, Takebayashi N, Olson LE, Taylor DL (2013) Evaluation of the authenticity of a highly novel environmental sequence from boreal forest soil using ribosomal RNA secondary structure modeling. *Molecular Phylogenetics and Evolution*, **67**, 234–245.

Guindon S, Dufayard JF, Lefort V *et al.* (2010) New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Systematic Biology*, **59**, 307–321.

Herrera ML, Vallor AC, Gelfond JA, Patterson TF, Wickes BL (2009) Strain-dependent variation in 18S ribosomal DNA copy numbers in *Aspergillus fumigatus*. *Journal of Clinical Microbiology*, **47**, 1325–1332.

Hibbett DS, Binder M, Bischoff JF *et al.* (2007) A higher-level phylogenetic classification of the Fungi. *Mycological Research*, **111**, 509–547.

Hubka V, Kolařík M (2012) *β*-tubulin paralogue tubC is frequently misidentified as the benA gene in *Aspergillus* section *Nigri* taxonomy: primer specificity testing and taxonomic consequences. *Persoonia*, **29**, 1–10.

Huse SM, Welch DM, Morrison HG, Sogin ML (2010) Ironing out the wrinkles in the rare biosphere through improved OTU clustering. *Environmental Microbiology*, **12**, 1889–1898.

Ihrmark K, Bödeker IT, Cruz-Martinez K *et al.* (2012) New primers to amplify the fungal ITS2 region - evaluation by 454-sequencing of artificial and natural communities. *FEMS Microbiology Ecology*, **82**, 666–677.

James TY, Kauff F, Schoch CL *et al.* (2006) Reconstructing the early evolution of Fungi using a six-gene phylogeny. *Nature*, **443**, 818–822.

Jones MDM, Forn I, Gadelha C *et al.* (2011) Discovery of novel intermediate forms redefines the fungal tree of life. *Nature*, **474**, 200–203.

Jumpponen A (2007) Soil fungal communities underneath willow canopies on a primary successional glacier forefront: rDNA sequence results can be affected by primer selection and chimeric data. *Microbial Ecology*, **53**, 233–246.

Katoh K, Asimenos G, Toh H (2009) Multiple Alignment of DNA Sequences with MAFFT. In: *Bioinformatics for DNA Sequence Analysis* (ed Posada D), pp. 39–46. Humana Press Inc., Totowa.

Kåuserud H, Schumacher T (2003) Ribosomal DNA variation, recombination and inheritance in the basidiomycete *Trichaptum abietinum*: Implications for reticulate evolution. *Heredity*, **91**, 163–172.

Keeling PJ, Inagaki Y (2004) A class of eukaryotic GTPase with a punctate distribution suggesting multiple functional replacements of translation elongation factor 1 alpha. *Proceedings of the National Academy of Sciences, USA*, **101**, 15380–15385.

Kiss L (2012) Limits of nuclear ribosomal DNA internal transcribed spacer (ITS) sequences as species barcodes for Fungi. *Proceedings of the National Academy of Sciences, USA*, **109**, E1811.

Kruger M, Kruger C, Walker C, Stockinger H, Schussler A (2012) Phylogenetic reference data for systematics and phylotaxonomy of arbuscular mycorrhizal fungi from phylum to species level. *New Phytologist*, **193**, 970–984.

Kunin V, Engelbrektson A, Ochman H, Hugenholtz P (2010) Wrinkles in the rare biosphere: pyrosequencing errors can lead to artificial inflation of diversity estimates. *Environmental Microbiology*, **12**, 118–123.

Lindahl BD, Nilsson RH, Tedersoo L *et al.* (2013) Fungal community analysis by high-throughput sequencing of amplified markers – a user's guide. *New Phytologist*, **199**, 288–299.

Lindner DL, Banik MT (2011) Intragenomic variation in the ITS rDNA region obscures phylogenetic relationships and inflates estimates of operational taxonomic units in genus *Laetiporus*. *Mycologia*, **103**, 731–740.

Lindner DL, Carlsen T, Nilsson RH *et al.* (2013) Employing 454 amplicon pyrosequencing to reveal intragenomic divergence in the internal transcribed spacer rDNA region in fungi. *Ecology and Evolution*, **3**, 1751–1764.

Liu YJ, Hall BD (2004) Body plan evolution of ascomycetes, as inferred from an RNA polymerase II, phylogeny. *Proceedings of the National Academy of Sciences, USA*, **101**, 4507–4512.

Lucking R, Lawrey JD, Gillevet PM *et al.* (2014) Multiple ITS haplotypes in the genome of the lichenized basidiomycete *Cora inversa* (Hygrophoraceae): fact or artifact? *Journal of Molecular Evolution*, **78**, 148–162.

Lundberg DS, Lebeis SL, Paredes SH *et al.* (2012) Defining the core *Arabidopsis thaliana* root microbiome. *Nature*, **488**, 86–90.

Lutzoni F, Kauff F, Cox CJ *et al.* (2004) Assembling the fungal tree of life: progress, classification and evolution of subcellular traits. *American Journal of Botany*, **91**, 1446–1480.

Marano AV, Gleason FH, Bärlocher F *et al.* (2012) Quantitative methods for the analysis of zoosporic fungi. *Journal of Microbiological Methods*, **89**, 22–32.

Matheny PB (2005) Improving phylogenetic inference of mushrooms with RPB1 and RPB2 nucleotide sequences (*Inocybe*; *Agaricales*). *Molecular Phylogenetics and Evolution*, **35**, 1–20.

Nilsson RH, Kristiansson E, Ryberg M, Hallenberg N, Larsson KH (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**, 193–201.

Nilsson RH, Ryberg M, Abarenkov K, Sjokvist E, Kristiansson E (2009) The ITS region as a target for characterization of fungal communities

using emerging sequencing technologies. *FEMS Microbiology Letters*, **296**, 97–101.
O'Brien HE, Parrent JL, Jackson JA, Moncalvo J-M, Vilgalys R (2005) Fungal community analysis by large-scale sequencing of environmental samples. *Applied and Environmental Microbiology*, **71**, 5544–5550.
Öpik M, Metsis M, Daniell T, Zobel M, Moora M (2009) Large-scale parallel 454 sequencing reveals host ecological group specificity of arbuscular mycorrhizal fungi in a boreonemoral forest. *New Phytologist*, **184**, 424–437.
Parks DH, Beiko RG (2010) Identifying biologically relevant differences between metagenomic communities. *Bioinformatics*, **26**, 715–721.
Quince C, Curtis TP, Sloan WT (2008) The rational exploration of microbial diversity. *ISME Journal*, **2**, 997–1006.
Raidl S, Bonfigli R, Agerer R (2005) Calibration of quantitative real-time Taqman PCR by correlation with hyphal biomass and its copies in mycelia of *Piloderma croceum*. *Plant Biology*, **7**, 713–717.
Sagova-Mareckova M, Cermak L, Novotna J *et al.* (2008) Innovative methods for soil DNA purification tested in soils with widely differing characteristics. *Applied and Environmental Microbiology*, **74**, 2902–2907.
Schloss PD (2010) The effects of alignment quality, distance calculation method, sequence filtering, and region on the analysis of 16S rRNA gene-based studies. *PLoS Computational Biology*, **6**, e1000844.
Schloss PD, Westcott SL, Ryabin T *et al.* (2009) Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Applied and Environmental Microbiology*, **75**, 7537–7541.
Schoch CL, Seifert KA, Huhndorf S *et al.* (2012) Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for Fungi. *Proceedings of the National Academy of Sciences, USA*, **109**, 6241–6246.
Simon UK, Weiss M (2008) Intragenomic variation of fungal ribosomal genes is higher than previously thought. *Molecular Phylogenetics and Evolution*, **25**, 2251–2254.
Smith ME, Douhan GW, Rizzo DM (2007) Intra-specific and intra-sporocarp ITS variation of ectomycorrhizal fungi as assessed by rDNA sequencing of sporocarps and pooled ectomycorrhizal roots from a *Quercus* woodland. *Mycorrhiza*, **18**, 15–22.
Stockinger H, Krueger M, Schuessler A (2010) DNA barcoding of arbuscular mycorrhizal fungi. *New Phytologist*, **187**, 461–474.
Tedersoo L, Nilsson RH, Abarenkov K *et al.* (2010) 454 Pyrosequencing and Sanger sequencing of tropical mycorrhizal fungi provide similar results but reveal substantial methodological biases. *New Phytologist*, **188**, 291–301.
Větrovský T, Baldrian P (2013) Analysis of soil fungal communities by amplicon pyrosequencing: current approaches to data analysis and the introduction of the pipeline SEED. *Biology and Fertility of Soils*, **49**, 1027–1037.
Voříšková J, Baldrian P (2013) Fungal community on decomposing leaf litter undergoes rapid successional changes. *ISME Journal*, **7**, 477–486.
Voříšková J, Brabcová V, Cajthaml T, Baldrian P (2014) Seasonal dynamics of fungal communities in a temperate oak forest soil. *New Phytologist*, **201**, 269–278.
Wang D-M, Yao Y-J (2005) Intrastrain internal transcribed spacer heterogeneity in *Ganoderma* species. *Canadian Journal of Microbiology*, **51**, 113–121.
White TJ, Bruns T, Lee S, Taylor J (1990) Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. . *PCR Protocols: A guide to methods and applications*, pp. 315–322. Academic Press, New York.
Youssef N, Sheik CS, Krumholz LR *et al.* (2009) Comparison of species richness estimates obtained using nearly complete fragments and simulated pyrosequencing-generated fragments in 16S rRNA gene-based environmental surveys. *Applied and Environmental Microbiology*, **75**, 5227–5236.

T.V., M.K. and P.B. designed and conducted the research. L.Z., P.B. and T.V. collected soils, extracted DNA and analysed the data. T.V., M.K. and T.Z. constructed and analysed the mock community. M.K., T.V. and P.B. wrote the first draft of the manuscript. All authors edited the manuscript.

Data accessibility

Environmental sequencing data have been deposited in the MG-RAST public database (http://metagenomics.anl.gov/, data set numbers 4538645.3 for the fungal ITS region and 4538646.3 for the *rpb2* gene). All other data have been deposited in Dryad (doi:10.5061/dryad.vj3p9).

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Fig. S1 Similarity between the top fungal OTUs from spruce topsoil based on the *ITS* (A) and *rpb2* (B) molecular markers and their best BLASTN hits from Genbank.

Fig S2 Richness and diversity estimates of fungal community in spruce forest topsoil samples estimated using *ITS* and *rpb2* as alternative molecular markers at a sequencing depth of 209 sequences per sample, estimates of richness and diversity in the litter and soil and OTU accummulation curves for the whole *ITS* and *rpb2* datasets (4457 sequences).

Fig. S3 OTU counts of fungi from spruce forest topsoil based on the *ITS* and *rpb2* molecular markers at various levels of similarity.

Fig. S4 Analysis of length of amplicons defined by used primers bRPB2-6F/bRPB2-7R and ITS1/ITS4. ITS amplicons (22553) and *rpb2* amplicons (4848) were obtained from data published in NCBI Genbank (Accessed 24 Apr. 2014) allowing maximum of three mismatches of the primers. Average amplicon length and standard deviation is shown for each class.

Table S1 Number of *rpb2* sequences deposited in NCBI genbank (5.IV.2015) covering 6753 fungal species.

Table S2 Preprocessing workflow of the raw sequence reads from environmental data and mock community.

Table S3 List of species used in the mock community, counts of corresponding OTUs and sequence abundances obtained using ITS and rpb2 as molecular markers.

Table S4 Taxonomic assignment of all OTUs/sequence reads from spruce forest topsoils with the ITS and *rpb2* primers. Identification was done using BLASTN (ITS dataset) or tBLASTX (*rpb2* dataset).

Table S5 Analysis of number and lenght of introns in *rpb2* sequences deposited in NCBI Genbank (5.XII. 2014). Regions

defined by used primers were extracted by SEED pipeline with up to 3 mismathces allowed in primer sequence. Positions of introns were obtained from sequence annotation by custom made script. Only 4368 sequences with annoated intron positions were used.

Table S6 Primer coverage analysis of *rpb2* and ITS primers used in this study. The data published in NCBI Genbank (Accessed 24 Apr. 2014) were used and only sequences of genera from which both genes were available and have maximum of three mismatches at the priming site were used (460 genera).