

454 Pyrosequencing analyses of forest soils reveal an unexpectedly high fungal diversity

M. Buée^{1*}, M. Reich^{1*}, C. Murat¹, E. Morin¹, R. H. Nilsson², S. Uroz¹ and F. Martin¹

¹INRA, UMR 1136 INRA/Nancy Université Interactions Arbres/Microorganismes, INRA-Nancy, 54280 Champenoux, France; ²Department of Plant and Environmental Sciences, University of Gothenburg, Box 461, 405 30 Gothenburg, Sweden

Summary

Author for correspondence:

Marc Buée

Tel: +33 383 39 40 72

Email: buee@nancy.inra.fr

Received: 27 May 2009

Accepted: 14 July 2009

New Phytologist (2009)

doi: 10.1111/j.1469-8137.2009.03003.x

Key words: 454 pyrosequencing, community structure, ectomycorrhizal fungi, environmental samples, fungal diversity, nuclear ribosomal internal transcribed spacers.

- Soil fungi play a major role in ecological and biogeochemical processes in forests. Little is known, however, about the structure and richness of different fungal communities and the distribution of functional ecological groups (pathogens, saprobes and symbionts).
- Here, we assessed the fungal diversity in six different forest soils using tag-encoded 454 pyrosequencing of the nuclear ribosomal internal transcribed spacer-1 (ITS-1). No less than 166 350 ITS reads were obtained from all samples. In each forest soil sample (4 g), approximately 30 000 reads were recovered, corresponding to around 1000 molecular operational taxonomic units.
- Most operational taxonomic units (81%) belonged to the Dikarya subkingdom (Ascomycota and Basidiomycota). Richness, abundance and taxonomic analyses identified the Agaricomycetes as the dominant fungal class. The ITS-1 sequences (73%) analysed corresponded to only 26 taxa. The most abundant operational taxonomic units showed the highest sequence similarity to *Ceratobasidium* sp., *Cryptococcus podzolicus*, *Lactarius* sp. and *Scleroderma* sp.
- This study validates the effectiveness of high-throughput 454 sequencing technology for the survey of soil fungal diversity. The large proportion of unidentified sequences, however, calls for curated sequence databases. The use of pyrosequencing on soil samples will accelerate the study of the spatiotemporal dynamics of fungal communities in forest ecosystems.

Introduction

Fungi represent an essential functional component of terrestrial ecosystems as decomposers, mutualists and pathogens, and are one of the most diverse groups of the Eukarya (Mueller *et al.*, 2007). Studying the ecological factors that underlie the dynamics of fungal communities remains a challenge because of this high taxonomic and ecological diversity. PCR-based molecular methods and sequencing of ribosomal DNA have been used successfully to identify subsets of this species' richness (Vandenkoornhuys *et al.*, 2002), and have provided insights into the ecological processes that affect the structure and diversity of fungal communities (Gomes *et al.*, 2003; Schadt *et al.*, 2003; Artz *et al.*, 2007). These advances are particularly noteworthy in below-ground studies of ectomycorrhizal (EM) fungi as a

result of the combination of morphological and molecular identifications of EM root tips (Horton & Bruns, 2001; Martin & Slater, 2007). Spatial and temporal variations of fungal communities in forest soils are affected by numerous biotic and abiotic factors, including seasons, soil characteristics, stand age and host tree species (Nordén & Paltto, 2001; Peter *et al.*, 2001; Dickie *et al.*, 2002; Buée *et al.*, 2005; Genney *et al.*, 2005; Koide *et al.*, 2007; Tedersoo *et al.*, 2008).

The internal transcribed spacer (ITS) region is now widely used as a validated DNA barcode marker for the identification of many fungal species (Seifert, 2008). With improvements in sequencing techniques and dedicated DNA databases (Kõljalg *et al.*, 2005), recent studies have demonstrated the potential of large-scale Sanger sequencing of ITS for quantifying and characterizing soil fungal diversity (O'Brien *et al.*, 2005). To our knowledge, the species' richness of communities of soil fungi

*These authors contributed equally to this work.

has not yet been assessed using high-throughput pyrosequencing.

In this article, we present the use of high-throughput tag-encoded FLX amplicon pyrosequencing (Acosta-Martinez *et al.*, 2008) to assess the fungal diversity in six soil samples from a French temperate forest site. We show that the abundance and diversity of fungi in the six soil samples were much higher than hypothesized previously. A few fungal taxa account for most of the species' abundance, whereas the majority of species are only rarely retrieved. There is reason to believe that the spatial diversity and difference in fungal richness among the six soil samples could be explained partly by forest management, that is, plantation tree species. The use of pyrosequencing on soil samples will accelerate the study of the spatiotemporal dynamics of fungal communities in forest ecosystems.

Materials and Methods

Study site and sampling

The experimental site of Breuil-Chenue forest is situated in the Morvan Mountains, Burgundy, France (latitude 47°18'10", longitude 4°4'44"). The elevation is 640 m, the annual rainfall is 1280 mm and the mean annual temperature is 9°C. The parent rock is granite and the soil is an alocrisol, with a pH ranging between 4 and 4.5 (Ranger *et al.*, 2004). The native forest is an old coppice composed of beech (*Fagus sylvatica* L., 90% of the stems), Durmast oak (*Quercus sessiliflora* Smith), sporadic weeping birch (*Betula verrucosa* Ehrh) and hazel trees (*Corylus avellana* L.). In 1976, a part of the native forest was clear-cut and this area was planted with the following six species: beech (*Fagus sylvatica* L.), Durmast oak (*Quercus sessiliflora* Smith), Norway spruce (*Picea abies* Karst), Douglas fir (*Pseudotsuga menziesii* Franco), Corsican pine (*Pinus nigra* Arn. ssp. *laricio* Poiret var. *Corsicana*) and Nordmann fir (*Abies nordmanniana* Spach.). Six plots (1000 m² each), corresponding to these six plantations, were selected for the study. These plots were relatively contiguous, because the six plantations were distributed on a total area of *c.* 14 000 m². The site is surrounded mainly by native forest and Douglas fir plantations. In March 2008, eight soil cores (1 × 1 × 5 cm depth) were sampled independently along two 30 m transects in each of these six plots. After removal of the forest litter, the 48 soil cores were sampled in the organic horizon (depth, 0–5 cm) and transported to the laboratory in an ice chest (8°C). Soil cores from each plot were independently homogenized, and minor woody debris and roots (>2 mm) were eliminated. Finally, 500 mg of the remaining soil was subsampled for DNA extraction from each soil core.

DNA extraction, PCR and pyrosequencing

Genomic DNA was extracted from the 48 subsamples of soil using the 'FastDNA SPIN for Soil Kit' (MP Biomedicals, Illkirch, France), according to the manufacturer's instructions. Amplicon libraries were performed using a combination of tagged primers designed for the variable ITS-1 region, as recommended for the tag-encoded 454 GS-FLX amplicon pyrosequencing method (Acosta-Martinez *et al.*, 2008). The 48 genomic DNA samples were diluted to 1 : 5 and 1 : 100. These 96 diluted genomic DNA samples were amplified separately using the fungal primer pair ITS1F (5'-*Axxx*CTTGATTTAGAGGAAGTAA-3') and ITS2 (5'-*BGCTGCGTTCTTCATC-GATGC*-3') to generate PCR ITS rRNA fragments of *c.* 400 bp, where *A* and *B* represent the two pyrosequencing primers (GCCTCCCTCGCGCCATCAG and GCCTTGCCAGCCCGCTCAG) and *xxx* was designed for the sample identification barcoding key. The PCR conditions used were 94°C for 4 min, 30 cycles of 30 s at 94°C (denaturation), 50°C for 1 min (annealing) and 72°C for 90 s (extension), followed by 10 min at 72°C. The 96 PCR products were purified using the Multiscreen-PCR plate system (Millipore Corporation, Billerica, MA, USA), and then pooled to obtain six amplicon libraries corresponding to the six different forest soils. The amplicon length and concentration were estimated, and an equimolar mix of all six amplicon libraries was used for pyrosequencing. Pyrosequencing of the six amplicon libraries (from the ITS1F primer) on the Genome Sequencer FLX 454 System (454 Life Sciences/Roche Applied Biosystems, Nutley, NJ, USA) at Cogenics (Meylan, France) resulted in 180 200 reads that satisfied the sequence quality criteria employed (cf. Droege & Hill, 2008). Tags were extracted from the FLX-generated composite FASTA file into individual sample-specific files based on the tag sequence by the proprietary software COGENICS v.1.14 (Cogenics Genome Express FLX platform, Grenoble, France).

Sequence editing and analysis of the reads by operational taxonomic unit (OTU) clustering

The filtered sequences were trimmed using the *trimseq* script from the EMBOSS package (Rice *et al.*, 2000). Sequences shorter than 100 bp after quality trimming were not considered. The average length of the 166 350 edited reads was 252 bp. The resultant individual sample FASTA files were assembled in tentative consensus sequences using BLASTclust v.2.2.1.6 (Altschul *et al.*, 1997), with the requirement that at least 97% similarity be obtained over at least 90% of the sequence length (-S 97 -L 0.9). In order to identify OTUs, a random sequence was compared with the nonredundant GenBank database and a custom-curated database (C-DB, described below). The OTUs defined at

97% sequence similarity (O'Brien *et al.*, 2005) were used to perform rarefaction analysis and to calculate the richness (Shannon) and diversity (Chao1) indices. The rarefaction analysis was performed using ANALYTIC RAREFACTION v.1.4 (Hunt Mountain Software, Department of Geology, University of Georgia, Athens, GA, USA). Calculation of the richness (Shannon) and diversity (Chao1) indices was performed using the ESTIMATE S software package (version 8.00, R. K. Colwell; <http://viceroy.eeb.uconn.edu/EstimateSPages/Est-SUsersGuide/EstimateSUsersGuide.htm>).

Phylogenetic assignment of the ITS-1 reads with MEGAN

In a second analysis, the individual sample FASTA files were evaluated using NCBI-BLASTn (Altschul *et al.*, 1997) against the nonredundant GenBank database (Benson *et al.*, 2008) and C-DB derived from the GenBank and UNITE (Kõljalg *et al.*, 2005; <http://unite.ut.ee/index.php>) databases. To construct C-DB, all fully identified fungal ITS sequences in GenBank and UNITE, as of November 2007, were screened for appropriate length (300–1500 bp), IUPAC DNA ambiguity content (less than five symbols) and taxonomic reliability, as established by Nilsson *et al.* (2006). A maximum of five sequences per species was selected at random, resulting in a total of 23 390 sequences representing 9678 Latin binomials. A post-processing Perl script generated best-hit files comprising the top 10 best BLAST hits with an *E*-value < 10e⁻³ for tentative species' identification.

According to their best matches, the rDNA ITS sequences were phylogenetically assigned using MEGAN v. 3.0.2 (MEtaGenome ANalyzer, Center for Bioinformatics, Tübingen, Germany) (Huson *et al.*, 2007), which provides unique names and IDs for over 350 000 taxa from the NCBI taxonomic database. The output files obtained from the nonredundant GenBank and C-DB databases were then processed. All parameters of MEGAN, including the lowest common ancestor (LCA) assignment, were kept at default values, except for the 'min support' option (regulating the minimum number of sequence reads that must be assigned to a taxon), which was set to either unity or five depending on the analysis (cf. Wu & Eisen, 2008).

Results

Analysis of the reads by OTU clustering

A total of 166 350 ITS-1 sequences passed the quality control, and the number of reads per sample (i.e. pools of eight soil cores per plantation) ranged from 25 700 to 35 600. A maximum of 1000 OTUs (including 594 singletons) was identified in the soil samples collected in the oak plantation of the experimental site, whereas only 590 OTUs (including 333 singletons) were identified in the same volume of soil

collected in the beech plantation. From 4 g of forest soil and a mean of 30 000 reads, the number of OTUs obtained was *c.* 830 (± 73). The number of OTUs increased with the number of reads, and a plot of OTUs vs the number of ITS-1 sequences resulted in rarefaction curves that did not approach a plateau (Fig. 1), in spite of the large number of reads. At 97% similarity, the nonparametric Chao1 estimator (Chao *et al.*, 2005) predicted that the maximum number of OTUs probably ranges from 1350 to 3400 (data not shown) depending on soil sample, with a mean estimated OTU richness close to 2240 (± 360).

To identify the most frequent fungal taxa present in the organic soils from the Breuil-Chenue forest site, OTUs were clustered with all the 166 350 reads. The 26 most abundant OTUs represented 73% of the total reads (Table 1). The most frequent OTU was assigned to an 'uncultured fungus' in GenBank, but Menkis *et al.* (2006) suggested that it corresponds to the root plant pathogen *Ceratobasidium* sp. The six most abundant OTUs were distributed in three phyla and six distinct orders: Cantharellales, Mortierellales, Helotiales, Tremellales, Agaricales and Boletales (Table 1).

Analysis of reads with MEGAN

The set of individual DNA reads was also compared against the nonredundant GenBank database of known ITS sequences using BLASTn. MEGAN was used to compute the taxonomic content of the dataset, employing NCBI taxonomy to order and cluster the results (Huson *et al.*, 2007). Most of the sequences (71.5%) lack an explicit taxonomic annotation (Fig. 2a). To obtain a better assessment of the taxonomic diversity of the known species, sequences were queried against C-DB, containing only ITS sequences from known fungal species (see Materials and Methods section). After this curation, only 11% of the OTUs remained in the 'unclassified fungi' category, 81% in the Dikarya subkingdom and 8% in the Mortierellaceae family (Fig. 2b). With 43.7% of the remaining ITS sequences, the Basidiomycota represented the predominant fungal phylum in the pooled results from soils of the Breuil-Chenue plantations.

Comparative analysis of the six plantation soil samples revealed a distinct distribution of fungal phyla (Fig. 3). For instance, Basidiomycota accounted for 65% of OTUs in the soil cores collected in the oak plot, whereas this phylum accounted for only 28% of OTUs in the soil cores sampled in the spruce plot. Alternatively, soil samples from the spruce plot were characterized by a relatively high percentage (*c.* 17%) of species from the order Mortierellales, parasitic or saprobic fungi belonging to the Mucoromycotina (Hibbett *et al.*, 2007), with the number of ITS reads two- to five-fold higher than the five other forest soil samples (only 3% of Mortierellales in the soil cores from the oak plantation).

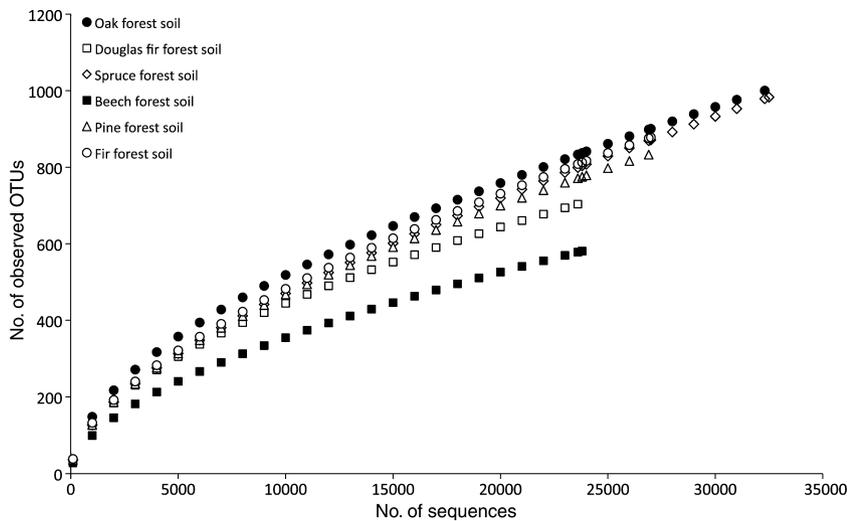


Fig. 1 Rarefaction curves depicting the effect of internal transcribed spacer (ITS) sequence number on the number of operational taxonomic units (OTUs) identified from the six soil samples. Between 25 680 and 35 600 sequences, depending on soil core, were generated, corresponding to 580–1000 OTUs (at 3% sequence dissimilarity). Forest soil (x, number of sequences; y, number of observed OTUs): oak forest soil (32330; 1001); Douglas fir forest soil (23624; 704); Norway spruce forest soil (32555; 983); Corsican pine forest soil (26908; 833); Nordmann fir forest soil (27054; 833); beech forest soil (23878; 581).

Closest NCBI database match	Closest accession number (NCBI)	Identities, length (%)	No. 454 reads
Uncultured fungus (<i>Ceratobasidium</i> sp.) ¹	DQ093748.1	189/190 (99)	20067
Uncultured <i>Cryptococcus</i>	FM866335.1	222/222 (100)	19452
<i>Lactarius</i> spp. ² (<i>quietus</i> , <i>tabidus</i> and <i>Lactarius</i> spp.)	EF493299.1	212/214 (99)	11302
<i>Scleroderma</i> sp. (<i>citrinum</i>)	EU784414.1	274/285 (96)	8733
Uncultured Dermateaceae	FJ554441.1	258/268 (96)	5617
Uncultured Mortierellaceae	FJ475737.1	259/268 (96)	5253
Uncultured fungus sp. 2	EF521220.1	217/282 (76)	5099
Uncultured soil fungus sp. 1	EU806458.1	189/203 (93)	4562
<i>Inocybe</i> sp. (uncultured ectomycorrhiza)	FN393147.1	228/235 (97)	4476
<i>Russula</i> sp. (<i>parazurea</i>)	DQ422007.1	251/271 (92)	4390
Uncultured soil fungus sp. 2	DQ421207.1	219/227 (96)	4271
Uncultured <i>cryptococcus</i>	FJ554344.1	257/260 (98)	4041
Uncultured fungus (<i>Cyllumyces</i> sp.)	AM260910.1	82/85 (96)	3759
Uncultured soil fungus sp. 3	FJ553866.1	259/262 (98)	3403
Uncultured <i>Sebacinales</i>	DQ421200.1	245/297 (82)	2936
Uncultured basidiomycete	FJ475793.1	275/285 (96)	2703
Uncultured soil fungus sp. 4 (<i>Mortierellaceae</i>)	FJ554362.1	122/126 (96)	2087
Uncultured dothideomycete (<i>Cenococcum</i> sp.)	DQ273316.1	246/258 (95)	1443
Uncultured Helotiales	FJ552732.1	272/281 (96)	1370
<i>Tylospora asterophora</i>	AF052557.1	269/276 (97)	1220
Uncultured basidiomycete (<i>Cortinarius</i> sp.)	AM902090.1	231/248 (93)	1184
<i>Amanita</i> sp. (<i>spissa</i>)	AJ889924.1	237/242 (97)	1049
<i>Pseudotomentella</i> sp. (<i>tristis</i>)	AJ889968.1	232/237 (97)	1048
Uncultured Helotiales	FJ475783.1	239/264 (90)	1016
Uncultured soil fungus sp. 5 (<i>Mortierellaceae</i>)	EU807054.1	249/253 (98)	989
<i>Boletus</i> sp. (<i>pruinatus</i>)	AJ889931.1	239/244 (97)	974

Table 1 List of the 26 most abundant fungal operational taxonomic units (OTUs) found in the forest soil of the Breuil-Chenue site

The 166 350 reads were assembled into tentative consensus sequences with the requirement that at least 97% similarity be obtained over at least 90% of the sequence length (-S 97 -L 0.9). To identify OTUs, a random sequence was compared with the nonredundant GenBank database.

¹This OTU was assigned to 'Uncultured fungus' in GenBank. It corresponds to *Ceratobasidium* sp. (Menkis *et al.*, 2006).

²The 11 302 reads of this OTU (*Lactarius* spp.) correspond to a complex of 6952 reads of *L. quietus*, 3489 reads of *L. tabidus* and 757 reads of other *Lactarius* species (as *L. theiogalus* and *L. rufus*) with ITS-1 sequences showing >97% homology.

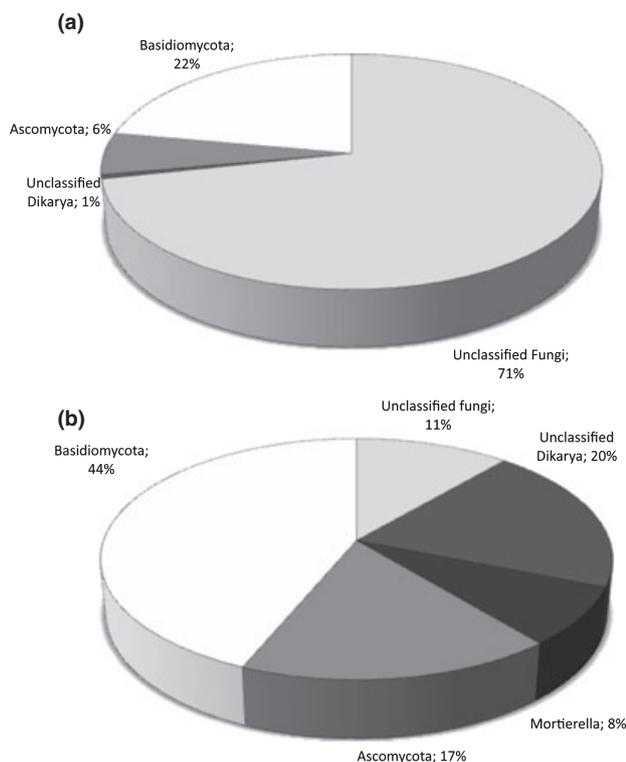


Fig. 2 Proportional distribution of different phyla and fungal groups in the sequenced internal transcribed spacer (ITS) clone libraries. The half-plate pyrosequencing produced 180 213 fungal sequences (166 350 after trimming procedures). Results obtained after BLASTn of sequences performed against GenBank and UNITE (a) or filtered database (b) (16 987 sequences, representing 9 678 fungal species), containing well-identified sequences and excluding all 'uncultured fungi' sequences and 'environmental sample' sequences (see Materials and Methods section for more details). Taxonomic clustering was performed in MEGAN (Metagenome Analysis) with the following lowest common ancestor (LCA) parameter values: 1, 35, 10 and 0 for Min support, Min score, Top percent and Win score, respectively.

At the family and genus levels, the fungal communities showed similar taxonomic distribution across all soil samples (Tables S1 and S2). Among saprotrophic, parasitic

and mycorrhizal fungi, the genera *Ceratobasidium*, *Cryptococcus*, *Lactarius*, *Mortierella*, *Russula*, *Scleroderma*, *Neofabraea*, *Inocybe* and *Cenococcum* were the most prominent genera found in this study (Table 1). Moreover, numerous Agaromycotina families were common to all soil samples, with a strong representation of the EM species from the Boletales, Agaricales, Thelephorales, Russulales, Cantharellales and Sebaciniales (according to Rinaldi *et al.*, 2008). Other EM genera, such as *Lactarius* and *Tylospora*, were mainly identified in the oak and spruce plots, respectively. EM fungi represented more than 50% of the 30 most abundant genera (Table S1).

At the species' level, the fungal community composition also revealed similar taxa between different soils (Table S2). The two yeast species, *Cryptococcus podzolicus* and *C. terricola*, occurring on the surface of roots and in the rhizosphere (Golubtsova *et al.*, 2006), were the most abundant Dikarya found in all organic forest soils of the Breuil-Chenue site. The plant pathogen fungus *Ceratobasidium* sp. was also dominant in all soil samples. EM species, such as *Cenococcum geophilum* and *Cortinari* sp. (*saturninus* from C-DB), were also ubiquitous. Other species, such as *Scleroderma* sp. (*citrinum* or *bovista*), were very abundant in most plantations (between 1000 and 2000 reads), except under oak (*c.* 100 reads). By contrast, the oak-specific EM symbiont *Lactarius quietus* was restricted to the soil collected under the oak plantation. *Russula puellaris* was only identified in the Douglas fir forest soil samples and Corsican pine plantation soil samples, whereas *Russula vesca* was found in the soils from Corsican pine and Nordmann fir plots.

Discussion

This pilot study used 454 pyrosequencing to evaluate the fungal diversity in six distinct and spatially distant soil samples from a temperate forest. By sequencing a total of 166 350 PCR-amplified ITS-1 sequences, we identified

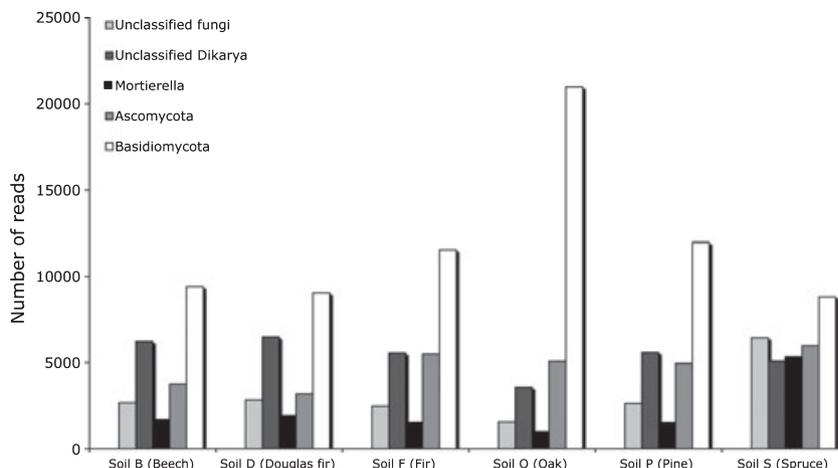


Fig. 3 Relative abundance of fungal phyla for each soil library. ITS-1 sequences were classified according to their BLASTn similarity by MEGAN (Metagenome Analysis). After using BLASTn against the curated database, taxonomic clustering was performed in MEGAN with the lowest common ancestor (LCA) parameter values set to 5, 35, 10 and 0 for Min support, Min score, Top percent and Win score, respectively.

600–1000 OTUs in each of the forest soil samples. The nonparametric Chao1 estimator (Chao *et al.*, 2005) predicted that the mean number of OTUs in 4 g of forest soil was *c.* 2240 (\pm 360). Interestingly, 73% of the DNA reads corresponded to 26 taxa only, and a detailed analysis showed that the three most abundant OTUs were supported by 25–55% of reads whatever soil was considered. Using a cloning/Sanger sequencing approach, Fierer *et al.* (2007) have estimated a similar number of OTUs in rain-forest soil samples (1 g), ranging between 1000 and 2000 OTUs in each community, depending on the parametric model used. Although we found 600–1000 OTUs in each of the forest soil samples, we highlighted between 249 and 408 taxonomic groups from these soil samples, supported by a minimum of two reads. Therefore, the number of singletons, which were close to only 1.8% of the total number of reads, corresponded to approximately 60% of the observed OTUs. This large proportion of OTUs, supported by unique reads, suggests that these sequences result from the sequencing of the numerous individuals isolated in the samples. This low abundance of numerous fungal taxa should be correlated with the inconspicuous nature of fungi and their dispersal ability. Hyphae and spores present in litter, leaves, pollen or needles, or the microscopic propagules, probably favour the spread of fungal species in diverse ecosystems. These species constitute a microbial reservoir (Finlay, 2002), which may play important functions in forest ecosystems facing environmental stresses.

At the present time, the ITS regions have been validated as the best DNA barcode marker for fungal species' identification (Seifert, 2008). In the present pyrosequencing experiment, and as reported in other studies (Liu *et al.*, 2008; Nilsson *et al.*, 2009), an average length of 252 bp for the ITS-1 sequences is long and sufficiently polymorphic to allow the identification of the majority of fungal OTUs at the species' or genus levels. A large part of the sequenced ITS regions belonged to unclassified fungi from incompletely annotated environmental samples. Lack of taxonomic annotation and errors in taxonomic assignments of ITS sequences deposited in the international DNA databases (Vilgalys, 2003) are major limitations to the survey of fungal species, and have hampered such efforts (Nilsson *et al.*, 2006; Bidartondo *et al.*, 2008; Horton *et al.*, 2009). For these reasons, the use of a curated ITS database (Nilsson *et al.*, 2005, 2006) should provide more pertinent taxonomic information. Using a curated database, we found that the majority of fungal sequences recovered belonged to the Dikarya (Ascomycota and Basidiomycota), which account for 81% of the OTUs. Basidiomycota was the most abundant phylum (43.7% of OTUs), whereas Ascomycota accounted for a much smaller percentage of the community (17.3%). These results are very similar to those of a large-scale survey of temperate forest soils carried out using Sanger sequencing (O'Brien *et al.*, 2005). By contrast, Schadt

et al. (2003) found a large proportion of Ascomycota in 125 cloned fungal sequences from tundra soils. The Glomeromycota and Chytridiomycota were probably underestimated in our ITS-1 libraries as we have amplified ITS from soil DNA using primers designed for Dikarya (ITS-1F/ITS-2). In addition, these fungal taxa and several genera, including Glomeromycota, were underestimated in our survey, as poorly annotated ITS sequences from GenBank were excluded from C-DB (Vilgalys, 2003; Nilsson *et al.*, 2008; Ryberg *et al.*, 2009).

Lindahl *et al.* (2007) reported that saprotrophic fungi were confined to the surface of the boreal forest floor. This functional ecological group of fungi seems to be under-represented in our topsoil samples. Ryberg *et al.* (2009) reported that numerous saprotrophic species are also poorly represented in the sequence databases compared with mycorrhizal sequences, and this imbalance may explain the apparent bias. Moreover, the season of sampling can influence the pattern of fungal richness and the under-representation of some species in our current samples (Taylor, 2002; Koide *et al.*, 2007). However, several saprotrophic species were found in all six soil samples. For instance, the two ubiquitous anamorphic Basidiomycota yeast species (Fonseca *et al.*, 2000), *Cryptococcus podzolicus* and *C. terricola*, showed a large number of reads in the six forest soils, and three *Mortierella* species were also very abundant in the six soil samples (Table S2). Interestingly, the three functional ecological fungal groups (parasitic, saprotrophic and mutualistic) were represented by the three most abundant OTUs, belonging to *Ceratobasidium*, *Cryptococcus* and *Lactarius* genera, respectively.

Owing to the large proportion of unclassified fungi found in the present and other soil surveys, a collection of curated sequences for fungal identification is urgently needed. Nevertheless, several of these unclassified fungal sequences seem to correspond to a well-supported clade of Ascomycota, equivalent to a subphylum, and referred to as soil clone group I (Porter *et al.*, 2008).

Amongst the taxonomically assigned species, EM species from the Boletales, Agaricales, Thelephorales, Russulales, Cantharellales and Sebaciniales were predominant in the six soil samples from different plantations (Tables S1 and S2), supporting recent results on EM community structure (Tedersoo *et al.*, 2008). These authors reported a host preference of EM fungi in wet sclerophyll forest, but revealed that the lineages of *Cortinarius*, *Tomentella*–*Thelephora*, *Russula*–*Lactarius*, *Clavulina*, *Descolea* and *Laccaria* prevailed in the total community studied. The wide distribution of these fungi is likely to favour their dissemination (Baker, 1966; Lockwood *et al.*, 2005), as are their resistance to environmental stresses and their capacity for invasiveness (Desprez-Loustau *et al.*, 2007).

The diversity and OTU richness between the six different forest soils suggest a strong spatial heterogeneity. Numerous

factors could explain this diversity, including the influence of the host tree or the impact of the soil organic matter. Moreover, a difference in organic matter composition and functioning has been reported in previous topsoil analyses from three plantations of this site (Moukoui *et al.*, 2006). The taxonomic information obtained in the present high-throughput survey shows an unexpectedly high richness of fungal species in forest soils. Additional 454 pyrosequencing-based surveys of fungal diversity will shed light on the factors that have the largest impact on the fungal communities.

Acknowledgements

We thank three anonymous referees for their comments and suggestions on an earlier draft of the manuscript. This project was funded by grants from the European Network of Excellence EVOLTREE, the INRA ECOGER programme and the Région Lorraine Council (Project FORBOIS) to F.M. M.R. was funded by a PhD scholarship from the Marie Curie Actions Host fellowships for Early Stage Research Training (TraceAM project). We would like to thank Dr Jacques Ranger and Dominique Gelhaye (INRA Nancy) for their support and for providing access to the Breuil-Chenu site. We would like to thank Dr Jonathan Plett for critical reading of the manuscript.

References

- Acosta-Martinez V, Dowd S, Sun Y, Allen V. 2008. Tag-encoded pyrosequencing analysis of bacterial diversity in a single soil type as affected by management and land use. *Soil Biology and Biochemistry* 40: 2762–2770.
- Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research* 25: 3389–3402.
- Artz RR, Anderson IE, Chapman SJ, Hagn A, Schlöter M, Potts JM, Campbell CD. 2007. Changes in fungal community composition in response to vegetational succession during the natural regeneration of cutover peatlands. *Microbial Ecology* 54: 508–522.
- Baker GE. 1966. Inadvertent distribution of fungi. *Canadian Journal of Microbiology* 12: 109–112.
- Benson DA, Karsch-Mizrachi I, Lipman DJ, Ostell J, Wheeler DL. 2008. GenBank. *Nucleic Acids Research* 36: D25–D30.
- Bidartondo M, Bruns T, Blackwell M, Edwards I, Taylor A, Horton T, Zhang N, Koljalg U, May G, Kuyper TW *et al.* 2008. Preserving accuracy in GenBank. *Science* 319: 1616.
- Buée M, Vairalles D, Garbaye J. 2005. Year-round monitoring of diversity and potential metabolic activity of the ectomycorrhizal community in a beech (*Fagus sylvatica*) forest subjected to two thinning regimes. *Mycorrhiza* 15: 235–245.
- Chao A, Chazdon RL, Colwell RK, Shen T-J. 2005. A new statistical approach for assessing similarity of species composition with incidence and abundance data. *Ecology Letters* 8: 148–159.
- Desprez-Loustau M-L, Robin C, Buée M, Courtecuisse R, Garbaye J, Suffert F, Sache I, Rizzo DM. 2007. The fungal dimension of biological invasions. *Trends in Ecology and Evolution* 22: 472–480.
- Dickie IA, Xu B, Koide RT. 2002. Vertical niche differentiation of ectomycorrhizal hyphae in soil as shown by T-RFLP analysis. *New Phytologist* 156: 527–535.
- Droege M, Hill B. 2008. The genome sequencer FLX™ System – longer reads, more applications, straight forward bioinformatics and more complete data sets. *Journal of Biotechnology* 136: 3–10.
- Fierer N, Breitbart M, Nulton J, Salamon P, Lozupone C, Jones R, Robeson M, Edwards RA, Felts B, Rayhawk S *et al.* 2007. Metagenomic and small-subunit rRNA analyses reveal the genetic diversity of bacteria, archaea, fungi, and viruses in soil. *Applied and Environmental Microbiology* 73: 7059–7066.
- Finlay BJ. 2002. Global dispersal of free-living microbial eukaryote species. *Science* 296: 1061–1063.
- Fonseca A, Scorzetti G, Fell JW. 2000. Diversity in the yeast *Cryptococcus albidus* and related species as revealed by ribosomal DNA sequence analysis. *Canadian Journal of Microbiology* 46: 7–27.
- Genney DR, Anderson IC, Alexander IJ. 2005. Fine-scale distribution of pine ectomycorrhizas and their extrametrical mycelium. *New Phytologist* 170: 381–390.
- Golubtsova YV, Glushakova AM, Chernov IY. 2006. The seasonal dynamics of yeast communities in the rhizosphere of soddy-podzolic soils. *Eurasian Soil Science* 40: 978–983.
- Gomes NC, Fagbola O, Costa R, Rumjanek NG, Buchner A, Mendonça-Hagler L, Smalla K. 2003. Dynamics of fungal communities in bulk and maize rhizosphere soil in the tropics. *Applied and Environmental Microbiology* 69: 3758–3766.
- Hibbett DS, Binder M, Bischoff JF, Blackwell M, Cannon PF, Eriksson OE, Huhndorf S, James T, Kirk PM, Lücking R *et al.* 2007. A higher-level phylogenetic classification of the fungi. *Mycological Research* 111: 509–547.
- Horton TR, Bruns TD. 2001. The molecular revolution in ectomycorrhizal ecology: peeking into the black-box. *Molecular Ecology* 10: 1855–1871.
- Horton T, Arnold AE, Bruns TD. 2009. FESIN workshops at ESA – the mycelial network grows. *Mycorrhiza* 19: 283–285.
- Huson DH, Auch AF, Qi J, Schuster SC. 2007. MEGAN analysis of metagenomic data. *Genome Research* 17: 377–386.
- Koide RT, Shumway DL, Xu B, Sharda JN. 2007. On temporal partitioning of a community of ectomycorrhizal fungi. *New Phytologist* 174: 420–429.
- Köljalg U, Larsson KH, Abarenkov K, Nilsson RH, Alexander IJ, Eberhardt U, Erland S, Hoiland K, Kjoller R, Larsson E *et al.* 2005. UNITE: a database providing web-based methods for the molecular identification of ectomycorrhizal fungi. *New Phytologist* 166: 1063–1068.
- Lindahl BD, Ihrmark K, Boberg J, Trumbore SE, Högberg P, Stenlid J, Finlay RD. 2007. Spatial distribution of litter decomposition and mycorrhizal nitrogen uptake in a boreal forest. *New Phytologist* 173: 611–620.
- Liu Z, DeSantis TZ, Andersen GL, Knight G. 2008. Accurate taxonomy assignments from 16S rRNA sequences produced by highly parallel pyrosequencers. *Nucleic Acids Research* 36: e120.
- Lockwood JL, Cassey P, Blackburn T. 2005. The role of propagule pressure in explaining species invasions. *Trends in Ecology and Evolution* 20: 23–28.
- Martin F, Slater H. 2007. New Phytologist – an evolving host for ectomycorrhizal research. *New Phytologist* 174: 225–228.
- Menkis A, Vasiliauskas R, Taylor AFS, Stenström E, Stenlid J, Finlay R. 2006. Fungi in decayed roots of conifer seedlings in forest nurseries, afforested clear-cuts and abandoned farmland. *Plant Pathology* 55: 117–129.
- Moukoui J, Munier-Lamy C, Berthelin J, Ranger J. 2006. Effect of tree species substitution on organic matter biodegradability and mineral nutrient availability in a temperate topsoil. *Annals of Forest Sciences* 63: 763–771.

- Mueller GM, Schmit JP, Leacock PR, Buyck B, Cifuentes J, Desjardin DE, Halling RE, Hjortstman K, Iturriaga T, Larsson K *et al.* 2007. Global diversity and distribution of macrofungi. *Biodiversity Conservation* 16: 37–48.
- Nilsson RH, Kristiansson E, Ryberg M, Larsson KH. 2005. Approaching the taxonomic affiliation of unidentified sequences in public databases – an example from the mycorrhizal fungi. *BMC Bioinformatics* 6: 178.
- Nilsson RH, Ryberg M, Kristiansson E, Abarenkov K, Larsson KH. 2006. Taxonomic reliability of DNA sequences in public sequence databases: a fungal perspective. *PLoS ONE* 1: e59.
- 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, Sjökvist 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.
- Nordén B, Paltto H. 2001. Wood-decay fungi in hazel wood: species richness correlated to stand age and dead wood features. *Biological Conservation* 101: 1–8.
- O'Brien HE, Parrent JL, Jackson JA, Moncalvo JM, Vilgalys R. 2005. Fungal community analysis by large-scale sequencing of environmental samples. *Applied and Environmental Microbiology* 71: 5544–5550.
- Peter M, Ayer F, Egli S. 2001. Nitrogen addition in a Norway spruce stand altered macrofungi sporocarp production and below-ground ectomycorrhizal species composition. *New Phytologist* 149: 311–325.
- Porter TM, Schadt CW, Rizvi L, Martin AP, Schmidt SK, Scott-Denton L, Vilgalys R, Moncalvo JM. 2008. Widespread occurrence and phylogenetic placement of a soil clone group adds a prominent new branch to the fungal tree of life. *Molecular Phylogenetics and Evolution* 46: 635–644.
- Ranger J, Andreux BSF, Berthelin BP, Boudot JP, Bréchet C, Buée M, Calmet JP, Chaussod R, Gelhaye D, Gelhaye L *et al.* 2004. *Effet des substitutions d'essence sur le fonctionnement organo-minéral de l'écosystème forestier, sur les communautés microbiennes et sur la diversité des communautés fongiques mycorrhiziennes et saprophytes (cas du dispositif expérimental de Breuil – Morvan)*. Final report of contract INRA-GIP Ecofor 2001–24. Champenoux, France: N° INRA 1502A, INRA BEF Nancy.
- Rice P, Longden I, Bleasby A. 2000. EMBOSS: the European molecular biology open software suite. *Trends in Genetics* 16: 276–277.
- Rinaldi AC, Comandini O, Kuyper TW. 2008. Ectomycorrhizal fungal diversity: separating the wheat from the chaff. *Fungal Diversity* 33: 1–45.
- Ryberg M, Kristiansson E, Sjökvist E, Nilsson RH. 2009. An outlook on the fungal internal transcribed spacer sequences in GenBank and the introduction of a web-based tool for the exploration of fungal diversity. *New Phytologist* 181: 471–477.
- Schadt CW, Martin AP, Lipson DA, Schmidt SK. 2003. Seasonal dynamics of previously unknown fungal lineages in tundra soils. *Science* 301: 1359–1361.
- Seifert KA. 2008. Integrating DNA barcoding into the mycological sciences. *Persoonia* 21: 162–166.
- Taylor AFS. 2002. Fungal diversity in ectomycorrhizal communities: sampling effort and species detection. *Plant and Soil* 244: 19–28.
- Tedersoo L, Jairus T, Horton BM, Abarenkov K, Suvi T, Saar I, Kõljalg U. 2008. Strong host preference of ectomycorrhizal fungi in a Tasmanian wet sclerophyll forest as revealed by DNA barcoding and taxon-specific primers. *New Phytologist* 180: 479–490.
- Vandenkoornhuysse P, Baldauf SL, Leyval C, Straczek J, Young JPW. 2002. Extensive fungal diversity in plant roots. *Science* 295: 2051.
- Vilgalys R. 2003. Taxonomic misidentification in public DNA databases. *New Phytologist* 160: 4–5.
- Wu M, Eisen JA. 2008. A simple, fast, and accurate method of phylogenomic inference. *Genome Biology* 9: R151.

Supporting Information

Additional supporting information may be found in the online version of this article.

Table S1 Fungal community composition, at the genus level, in the six forest soils

Table S2 Fungal community composition, at various taxonomic levels, in the six forest soil samples

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting information supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.