Assessment of soil fungal diversity in different alpine tundra habitats by means of pyrosequencing

Guillaume Lentendu · Lucie Zinger · Stéphanie Manel · Eric Coissac · Philippe Choler · Roberto A. Geremia · Christelle Melodelima

Abstract Studying fungal diversity is vital if we want to shed light on terrestrial ecosystem functioning. However, there is still poor understanding of fungal diversity and variation given that *Fungi* are highly diversified and that most of fungal species remain uncultured. In this study we explored diversity with 454 FLX sequencing technology by using the Internal Transcribed Spacer 1 (ITS1) as the fungal barcode marker in order to evaluate the effect of 11 environmental conditions on alpine soil fungal diversity, as well as the consistency of those results by taking into account rare or unidentified Molecular Operational Taxonomic Units (MOTUs). In total we obtained 205131 ITS1 reads corresponding to an estimated fungal gamma diversity of between 5100 and 12 000 MOTUs at a 98% similarity threshold when considering respectively only identified fungal and all MOTUs. Fungal beta-diversity patterns were significantly explained by the environmental conditions, and were very consistent for abundant/rare and fungal/unidentified MOTUs confirming the ecological significance of rare/unidentified MOTUs, and therefore the existence of a fungal rare biosphere. This study shows that a beta-diversity estimation based on pyrosequencing is robust enough to support ecological studies. Additionally, our results suggest that rare MOTUs harbour ecological

Guillaume Lentendu and Lucie Zinger equally contributed to this paper.

G. Lentendu · L. Zinger · S. Manel · E. Coissac · P. Choler · R. A. Geremia (✉) · C. Melodelima
Laboratoire d'Ecologie Alpine, CNRS UMR 5553,
Université Joseph Fourier, Grenoble 1,
BP 53, 38041 Grenoble Cedex 09, France
e-mail: roberto.geremia@ujf-grenoble.fr

S. Manel
Laboratoire Population Environnement Développement,
UMR 151 UP/IRD, Université de Provence,
3 place Victor Hugo,
13331 Marseille Cedex 03, France

P. Choler
Station Alpine J. Fourier,
Grenoble Univ, CNRS UMS 2925,
BP 53X, 38041 Grenoble Cedex 9, France

Present Address:
G. Lentendu
Department Soil Ecology, UFZ—Helmholtz Centre for Environmental Research,
Theodor-Lieser-Straße 4,
06120 Halle/Saale, Germany

Present Address:
L. Zinger
Max Plank Institute for Marine Microbiology,
28453 Bremen, Germany

Present Address:
S. Manel
Laboratoire d'Ecologie Alpine, CNRS UMR 5553,
Université de Grenoble,
BP 53, 38041 Grenoble Cedex 09, France

information. Thus the fungal rare biosphere may be important for ecosystem dynamics and resilience.

Introduction

Studying fungal diversity is vital if we are to shed light on terrestrial ecosystem functioning. Numerous soil *Fungi* facilitate plant nutrient uptake through mycorrhizal associations with the majority of plants, and/or significantly contribute to litter decomposition through saprophytic activities (Wardle et al. 2004; De Deyn and Van der Putten 2005; van der Heijden et al. 2008). To date, their diversity was estimated to be 0.712 to 1.500 10^6 species (Hawksworth 2001; Schmit and Mueller 2007), but this number has been questioned over the last decades with the emergence of molecular-based approaches that have enabled numerous uncultivable species to be uncovered. These methods also enabled microbiologists to begin the exploration of fundamental ecological issues such as: (a) do fungal assemblages vary between different environments, (b) what is the relative impact of environment and historical contingencies in microbial diversity patterns (c) does microbial diversity correlate with macro-organism diversity, and finally (d) what are the implications of fungal diversity for ecosystem resilience. Although numerous studies deciphered the influence of land use types (Oehl et al. 2003; Kasel et al. 2008), soil properties (Lauber et al. 2008), plant cover (Waldrop et al. 2006; Wallenstein et al. 2007) and isolation by distance (Green et al. 2004; Taylor et al. 2006) on fungal diversity and community assemblages, a better characterization of these patterns is still needed in order to infer global biogeographical rules for *Fungi*.

The advent of next generation sequencing technologies such as pyrosequencing (Margulies et al. 2005) opens up interesting prospects for studying microorganism communities, diversity and biogeography, due to their high resolution. These techniques are increasingly used for this purpose, mostly for bacteria (Sogin et al. 2006; Huber et al. 2007; Roesch et al. 2007) and more recently for *Fungi* (Buée et al. 2009; Opik et al. 2009; Jumpponen and Jones 2009; Jumpponen et al. 2010). All these studies have reported very high numbers of molecular operational taxonomic units (MOTUs), i.e. high microbial species diversity, supporting what was hitherto only presumed (Mueller and Schmit 2007). For Bacteria, these studies revealed a change of diversity between different environments (Roesch et al. 2007) and a significant number of low-abundance species, the so-called "rare biosphere" (Sogin et

al. 2006). However, the pyrosequencing approach is even more sensitive to PCR amplification biases and sequencing errors due to greater sequencing depth, which enhances the detection of artificial fragments compared to classic sequencing methods, resulting in an overestimation of diversity (Quince et al. 2009; Kunin et al. 2010). On the other hand, these studies also revealed that the sequencing depth of pyrosequencing was still insufficient to uncover all bacterial diversity. Furthermore, it is now admitted that the poor representativeness of DNA sequence databases precludes accurate identification of the majority of MOTUs, especially for *Fungi* (Nilsson et al. 2006, 2009). To date, the effect of those biases on observed fungal diversity has been completely ignored.

In this study, we explored the diversity of alpine soil *Fungi* with 454 FLX sequencing technology by using the Internal Transcribed Spacer 1 (ITS1) as a fungal barcode marker. We used robust bioinformatics and statistical approaches to estimate fungal diversity and its variation across an alpine landscape marked by drastic plant community shifts. Comparing the results obtained from defined fungal MOTUs and rare or non-identified ones, we then assessed the weight of putative methodological biases on the observed diversity patterns.

Material and methods

Sample description

Soil sampling and location characterization are described in Table 1. Briefly, the sampling area is located in the Southern French Alps, (Col du Galibier, 1,900–2,800 m elevation) in a watershed displaying strong habitat fragmentation, shaped by the rapid turnover of plant community composition along a mesotopographical gradient (Körner 1995). In the studied area, 11 environmental conditions (EC) were defined on the basis of plant cover in accordance with previous observations (Choler and Michalet 2002). Three locations per EC were defined, except for two EC (ES and TR) for which only two locations were analysed. In each of the 31 locations, three soil samples were collected in sterile conditions from the top 10 cm of soil.

Molecular analysis

For each soil sample, DNA extraction was carried out in triplicate. Soil DNA extraction was performed as follows: 0.25 g (wet mass) of each soil sample was extracted with a PowerSoil-htp™ 96 Well Soil DNA Isolation Kit (MO BIO Laboratoires, Ozyme, St Quentin en Yvelines, France) in accordance with the manufacturer's instructions. DNA

Table 1 Description of the 11 environmental conditions, and their dominant plant species

Plant community designation	Replicates	Plant community description	Dominant species
CF	31, 32, 33	Chionophilous alpine meadow	*Carex foetida, Alchemilla pentaphyllea, Salix herbacea*
CTR	11, 12, 13	Subalpine/alpine meadow	*Carex sempervirens, Trifolium alpinum*
EN	51, 52, 53	Scree community with long lasting snow cover	*Ranunculus glacialis*
ES	56, 57	Scree community on southern exposed slopes	*Crepis pygmeae, Doronicum grandiflorum*
FG	61, 62, 63	Mesophilous subalpine/alpine grassland	*Festuca violacea, Alchemilla filicaulis, Geum montanum*
FP	1, 2, 3	Mesophilous subalpine grassland	*Festuca paniculata*
HS	21, 22, 23	Open subalpine meadow on screes	*Helictotrichon sedenense, Festuca violacea*
KD	41, 42, 43	Fellfield	*Kobresia myosuroides, Dryas octopetala*
KS	46, 47, 48	Thermic alpine meadow	*Kobresia myosuroides, Sesleria coerulea, Carex rosae*
SR	36, 37, 38	Psychrophilic dwarf willows community	*Salix retusa, Salix reticulata*
TR	6, 7	Subalpine tall herb community	*Trifolium pratense, Geranium sylvaticum*

extract concentrations were quantified using the NanoDrop ND-1000 (NanoDrop technologies). DNA extracts of spatial location replicates were pooled in order to limit the effects of soil spatial heterogeneity (Schwarzenbach et al. 2007). The Internal Transcribed Spacer 1 (ITS1) was used as a molecular marker as previously suggested (Nilsson et al. 2009). The ITS1 was amplified using the fungal specific primers ITS5 and ITS2 (White et al. 1990). Although less specific to fungi, ITS5 was chosen here because it matches with a larger part of fungal diversity compared to ITS1-F (Bellemain et al. 2010). The 454 FLX adaptors A and B were attached at the 5′ end of the forward and reverse primers respectively. Tags of 6 nucleotides (nt) were designed using oligoTag (http://www.grenoble.prabi.fr/trac/OBITools) intercalated 454 adaptor and primers in order to identify the original location of each sequence after pyrosequencing.

The PCR reactions were performed as previously described (Zinger et al. 2008), using 20 ng of DNA template/reaction and AmpliTaqGold DNA polymerase (Applied Biosystems, Courtaboeuf, France). To reduce PCR biases, 8 PCR reactions were carried out for each location and then pooled per location. PCR products were purified with the QIAquick kit in accordance with the manufacturer's instructions (Qiagen, Courtaboeuf, France) and DNA was quantified using the Bioanalyser (Agilent Technologies, Inc., Santa Clara, CA). The 31 amplicons were pooled for the subsequent pyrosequencing by using equivalent molarities for each location. Pyrosequencing of this mixture was performed at the Centre National du Séquençage (Genoscope, Paris) with the 454 Life Sciences GS FLX systems (Brandford, CT, USA).

Bioinformatic analysis

The reads were filtered as follows: firstly, sequences containing ambiguous nucleotides (nt) in the primers were

automatically removed, making for 254,837 reads that were then filtered by quality and size. Similarly, reads containing ambiguous nucleotides were removed. Secondly, the sequencing error rate per nucleotide has been reported to increase when read length diverged from the predicted one (Huber et al. 2007); the median ITS1 length being 252 ± 58 nt, we therefore defined a confidence interval length of 116 nt. Then, and despite the size polymorphism of ITS1, we removed the reads that were shorter than 196 nt (35,053) and longer than 310 nt (126). A total of 205,131 reads were recovered after filtering. The redundant reads were then dereplicated using FastaUniq from OBITools, producing a set of 57,609 dissimilar reads, which were used for subsequent analyses.

For MOTU definition, we aligned dereplicated sequences using a global alignment algorithm (Needleman and Wunsch 1970) running in 'borneo' software (available on request at eric.coissac@inrialpes.fr). Alignment scores were normalized on the shortest sequence of each couple. MOTUs were created from the distance matrix by using the non-hierarchical clustering method MCL (Markov Clustering, (van Dongen 2000)). This clustering step was applied at ten similarity thresholds from 90% to 99% and produced a presence/absence matrix of MOTUs per locations for each threshold (hereafter $MOTUs_{0.98}$ to $MOTUs_{0.90}$), which were then used for subsequent diversity analyses.

A reference database was created from the Eukaryote nucleotide sequences available in EMBL (release embl_102, January 2010, (Cochrane et al. 2009)). Environmental sequences as well as sequences with ambiguous nucleotide and/or with a poor taxonomic definition were removed from this database. A custom fungal database was built using ecoPCR (OBITools) and by selecting sequences flanked by our two primers (three mismatches allowed) with lengths between 100 and 700 nt. This database

resulted in 20,515 entries corresponding to 9,117 defined species. The dereplicated reads were then aligned against this custom database using the "fasta35" algorithm (Pearson 2000) included in ecoTag (OBITools). This script uses the path along the EMBL-Bank taxonomic tree to assign reads to a given taxon. During the assignment process, each read is bound to a group of up to 100 database matches included between the best alignment score (I_{max}) and I_{max}^{S} (S being a shapeness index), with a minimum of 90% similarity between the read and query on database matches. The decrease of S helps to narrow the similarity interval, resulting in a finer taxonomic resolution. Given the poor quality of fungal databases (Nilsson et al. 2006, 2009), S was fixed here at 2 to reduce assignment stringency. A first taxonomic assignment of each read was defined as the longest common path in the taxonomic tree from the database matches. The taxonomic assignment of the MOTUs$_{0.90}$ was then performed at phylum, kingdom and domain level. MOTUs were only assigned to phylum level if there were no contradictions. The taxonomic assignation of MOTUs$_{0.90}$ was then re-attributed to the corresponding reads, which were then used for MOTU assignment at other similarity thresholds (91–99%). Raw reads are available on the Station Alpine Joseph Fourier website (http://station-alpine-joseph-fourier.ujf-grenoble.fr/IMG/zip/454_all_Lentendu-et-al_FUDI.fasta.zip).

Diversity analysis

We successively derived gamma-, alpha- and beta-diversities. Gamma-diversity was estimated using non-parametric Chao1 and Abundance Coverage Estimator (ACE) estimators. Alpha-diversity was calculated as the number of observed/estimated MOTUs per location.

The MOTUs composition between locations was compared using the Jaccard similarity index. A multivariate analysis of variance (Anderson 2001) was performed to evaluate the effect of ECs on Jaccard similarities; significance of the analysis was assessed by permutation (999). Rare MOTUs were defined as those with fewer than 11 reads (i.e. the rare biosphere) and their influence on our results was explored using datasets either without or only with rare MOTUs. Finally, we analysed the correlation between beta-diversity patterns obtained from datasets only with unidentified/eukaryal MOTUs and only with fungal identified MOTUs by using Procrustes analysis (Peres-Neto and Jackson 2001) on the two corresponding non-metric multidimensional scaling (NMDS) ordinations based on Jaccard distances, with 1,000 Monte Carlo permutations. All the diversity analyses were performed using R (The R Development Core Team 2009).

Results

We obtained an average of 8,221 reads per location. After quality filtering, this average was reduced to 6,617 reads per location (from 2,233 to 8,734 reads), which constituted the data set analysed below.

Taxonomic assignment

A first taxonomic screening revealed the presence of reads assigned to Viridiplantae (1.3%, all *Lilliopsida*), Metazoa (0.01%) and other Eukaryota (0.02%, *Alveolata*, *Rhizaria* and *Rhodophyta*), which were removed from the initial dataset prior to the alignment and clustering stages of the process. Figure 1 displays the taxonomic assignment of reads on the basis of MOTUs$_{0.98}$ occurrence (Fig. 1a) and MOTUs$_{0.98}$ abundance (Fig. 1b) (see below for threshold choice). Most of the reads were successfully assigned to *Fungi*. The assignation rate to fungal taxa (phyla, subkingdom and kingdom) is higher in abundant MOTUs$_{0.98}$ (84.8%) than in rare MOTUs$_{0.98}$ (53.3%) (Fig. 1a). In all cases, *Ascomycota* and *Basidiomycota* were the most abundant taxa. Nevertheless, diversity analysis was performed with either all MOTUs (subsequently referred to as the full dataset) or a sub-dataset containing fungal MOTUs only.

Sampling and sequencing effort

To assess the coverage of sampling and sequencing, we computed the accumulation of MOTUs$_{0.98}$ according to the locations and their rarefaction according to the sequence number per location (Fig. 2a–c and b–d respectively). When considering fungal MOTUs only, the accumulation of MOTUs at regional level almost reached a plateau (Fig. 2a), although it did not saturate when we consider the full dataset (Fig. 2c), whatever the similarity threshold. Likewise, fungal MOTUs$_{0.98}$ seemed to be closer to saturation whatever the location (Fig. 2b), while the rarefaction curves obtained from the full dataset displayed no plateaux (Fig. 2d). The sequencing depth therefore almost covered the full local diversity of "known *Fungi*", as defined by the composition of databases, but was still insufficient to estimate the full MOTUs diversity. In any case, the insufficient sequencing and sampling coverage precluded accurate assessment of overall diversity for all MOTUs$_{0.98}$ and only fungal at regional level.

Profile of fungal diversity

Gamma-diversity indices steadily increased with increasing similarity thresholds and showed an increase in this rise at 98% of similarity, whatever the diversity index used (Fig. 3). This was especially true when considering fungal
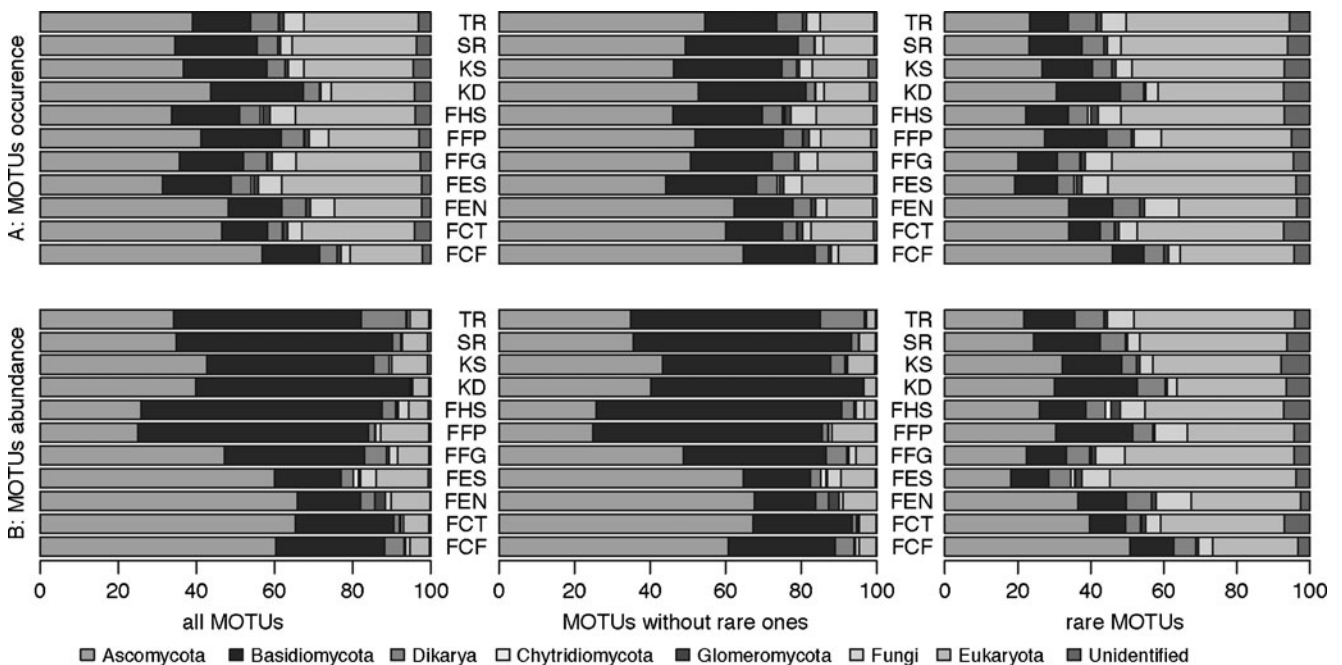
Fig. 1 Relative taxonomic composition of fungal communities in each environmental condition (EC) presented in terms of MOTUs occurrence (number of MOTUs at 98% similarity threshold, **a**) and in terms of read abundance (**b**)

MOTUs only and therefore led us to keep this threshold for subsequent analyses. When considering fungal MOTUs only, the regional diversity represented 2,847 $MOTUs_{0.98}$ of which 28.8 and 69.5% were singletons and rare $MOTUs_{0.98}$ respectively. On the other hand, singletons and rare fungal $MOTUs_{0.98}$ represented only 0.44 and 3.18% of fungal reads. Analysis of the full dataset rendered 5,550 $MOTUs_{0.98}$ from which 38.4% were singletons and 80% were rare. However, singletons and rare $MOTUs_{0.98}$ represented only 1 and 5.6% of the total number of reads. Regional diversity was estimated by the ACE and Chao1 methods as at least 5,100 and 12,000 $MOTUs_{0.98}$ for the fungal and full datasets respectively. Notably, we also observed a convergence of observed and estimated richness for fungal MOTUs whereas this would require lower similarity thresholds for the full dataset.

The average alpha-diversity for fungal $MOTUs_{0.98}$ and in the full dataset was 267 ($\pm$ 56) and 417 ($\pm$ 118) respectively, varying from 142 to 210 (EN52) to 432 and 750 (FG63) (Online Resource 1). The alpha diversity increased linearly with the number of reads. On average, 18 and 26% of MOTUs occurred in only one location (from which 47 and 51% were rare) while 7 and 4% were found in at least two locations of one EC. The remaining 78 and 70% were shared by at least two locations from different EC.

Assessing microbial diversity patterns across different habitats requires the differences between the local communities, i.e. beta diversity to be characterised. We first compared beta diversity between two EC, KD and CF (Table 2). The Jaccard similarity patterns were significantly

affected by ECs ($P=0.001$). For all samples (31 locations for 11 EC), we also observed a significant difference between EC ($P=0.001$) in MOTUs composition. It is worth mentioning that both results (2 and 11 EC) were similar for (a) the full dataset or only fungal $MOTUs_{0.98}$, and (b) abundant $MOTUs_{0.98}$ or rare $MOTUs_{0.98}$ (Table 2). Although the analysis was conducted with occurrence data, identical results were observed with abundance community matrix (data not shown). Finally, NMDS ordinations based on Jaccard distance matrices from unidentified/eukaryal $MOTUs_{0.98}$ and fungal identified $MOTUs_{0.98}$ displayed a correlation of 71.6% (Procrustes correlation $r=0.001$), thus implying highly similar beta-diversity patterns (Fig. 4).

Discussion

The advent of next generation sequencing technologies opens the door for systematic and comprehensive studies of fungal diversity, but only six reports, to our knowledge, have dealt to date with 454 pyrosequencing data to evaluate fungal diversity (Buée et al. 2009; Opik et al. 2009; Amend et al. 2010; Jumpponen and Jones 2009; Ovaskainen et al. 2010; Tedersoo et al. 2010). Here, MOTUs were constructed using an ab initio approach based on a pairwise alignment of pyrosequencing reads and MOTUs construction using MCL, which has been deemed more appropriate for pyrosequencing datasets of highly size polymorphic DNA fragments (Zinger et al. 2009a). The approach was used to investigate the feasibility of estimating global
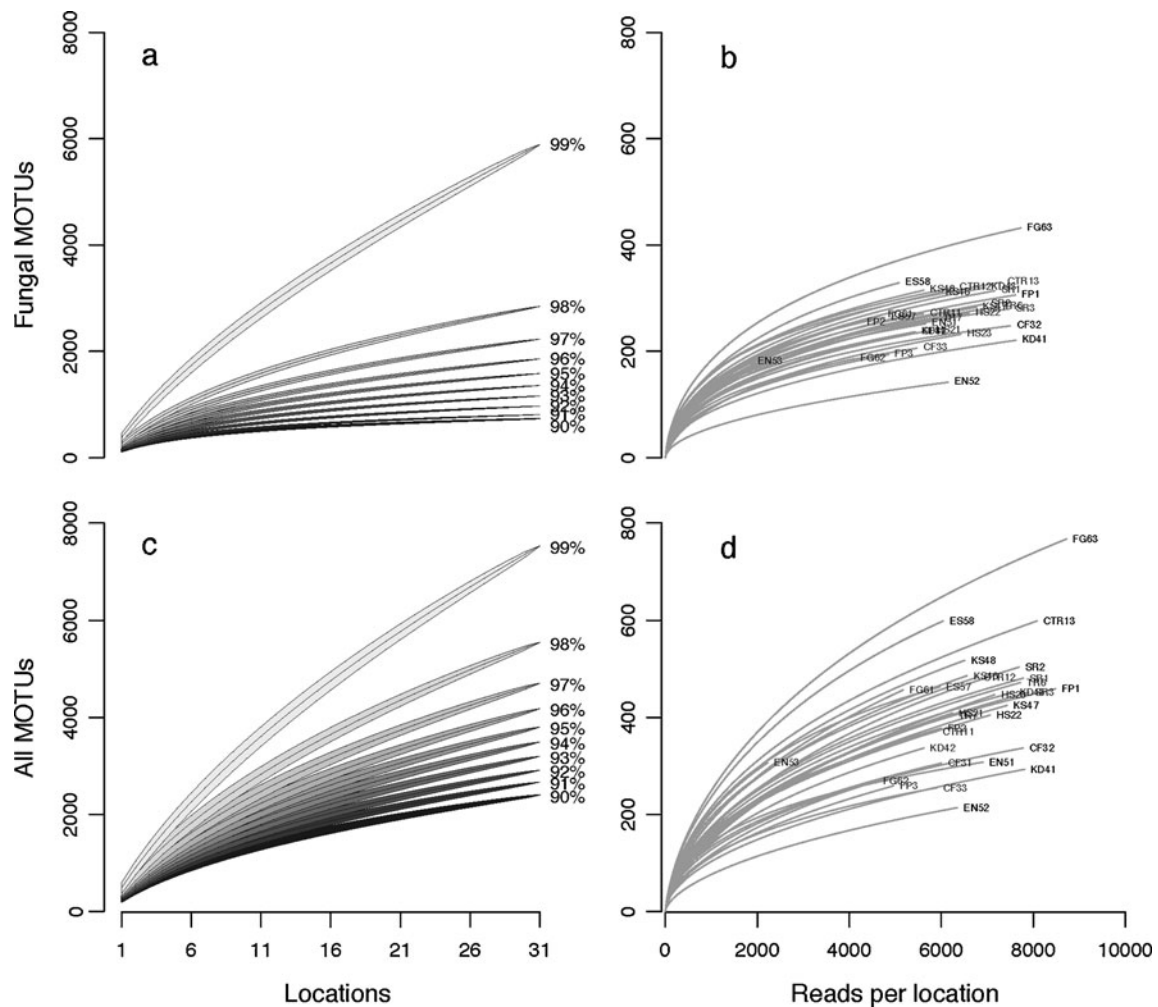
Fig. 2 Sampling (a–c) and sequencing (b–d) sufficiency as represented by accumulation and rarefaction curves respectively. Results are displayed for both fungal (a–b) and all MOTUs (c–d) at 98% of similarity threshold. Mean and confidence intervals at 95% are displayed for accumulation curves and confidence intervals were obtained after 100 permutations of the location's position. *Grey shadowing* indicates the similarity threshold

diversity in the studied area, assessing the effect of environmental conditions on fungal communities, and gave hints of a potential fungal "rare biosphere".

Consideration of diversity estimation with pyrosequencing

The assessment of diversity by molecular methods presents several experimental and informatics hitches that introduce uncertainties into diversity estimation (Huse et al. 2007; Quince et al. 2009; Kunin et al. 2010) and which are further complicated by the huge amount of unknown fungal taxa likely to be found in soils (Schmit and Mueller 2007; Buée et al. 2009). From an experimental point of view, PCR and sequencing errors, as well as alignment and clustering methods based on approximate algorithms, may create artificial MOTUs which may inflate the diversity estimation (Huse et al. 2007, 2010; Quince et al. 2009; Kunin et al. 2010).

The algorithm for taxonomic assignation used here includes an innovative filter, the shapeness index S, which enables the stringency of assignment to be selected. Here, we used $S=2$ which favours robust assignment at a high taxonomic level, resulting in 43.4% of unidentified/eukaryal $MOTUs_{0.98}$ that however accounted only for 8% of reads and were mostly rare MOTUs. Two main questions arise: do these unidentified/eukaryal MOTUs result from experimental artefacts or do they actually represent unknown species? Assuming that the latter is true, do they belong to *Fungi*? First, a negligible proportion of MOTUs corresponded to plants, Metazoan and other micro-eukaryotes (fewer than 2% of all reads), confirming a good but not total fungal specificity for the ITS5-ITS2 primer pair, as recently suggested (Bellemain et al. 2010). Because DNA sequence databases are not representative of overall fungal diversity (Nilsson et al. 2008), the assignment of eukaryal MOTUs to *Fungi* would inevitably fail if they
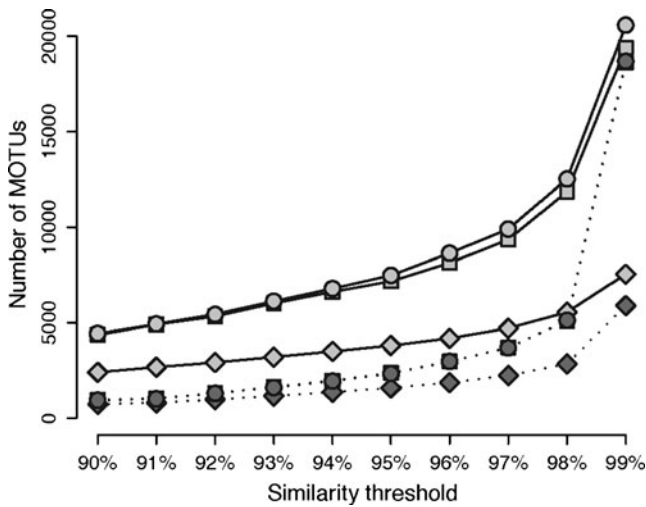
Fig. 3 Gamma diversity of the studied area for increasing similarity thresholds (from 90% to 99%) was calculated for both all (*light-grey* and *solid line*) and only fungal identified MOTUs (*dark-grey* and *dashed line*). Observed gamma-diversity (*diamonds*) was first calculated as the total number of MOTUs over the incidence matrix. Non-parametric Chao 1 (*squares*) and ACE (*circles*) estimations were computed based on the number of occurrences

belong to unknown species. The question remains about the unidentified MOTUs, for which considerable efforts should be made towards isolation and characterisation of new fungal strains, DNA sequence databases curation and improvements in molecular procedures and bioinformatics to reduce the amount of artifactual amplicons. Due to the

ambiguity of unidentified/eukaryal MOTUs, we analysed fungal diversity patterns using both identified fungal MOTUs and the full dataset, respectively providing minimal and maximal estimations of empirical fungal diversity.

Relevance of the species threshold

Although the consensus value for species discrimination using ITS is usually set at 97% (O'Brien et al. 2005; Buée et al. 2009; Amend et al. 2010; Tedersoo et al. 2010), it has been shown recently that the ITS intra-specific similarity varied from 99% to 76% depending on the species in question (Nilsson et al. 2008). For instance, it has recently been shown that the ITS of isolates belonging to *Laetiporus* genus display an intragenomic similarity ≤95% (Lindner and Banik 2011). However, defining a mean for an artificial species threshold is still needed for fungal diversity studied at community scale. Given that phylogeny fails to indicate an inter-specific threshold for hyper-size-polymorphic regions, and that some species display higher ITS similarity than others, the use of breakpoints in gamma-diversity is a good alternative way of placing a global "species threshold" for *Fungi*. Here, the MOTUs gamma-diversity curves along the similarity threshold levels (Fig. 3) showed a break at 98% similarity whatever the diversity index used, similarly to what was previously observed when studying phylloplane fungal communities (Jumpponen and Jones 2009). This suggests a shift from intra- to inter-specific

Table 2 Effect of the environmental conditions on beta-diversity patterns

	Source	Df	Similarity in MOTUs composition[a]			Without rare MOTUs[a,b]			Rare MOTUs only[a,b]		
			F.Model	R2[c]	Pr	F.Model	R2[c]	Pr	F.Model	R2[c]	Pr
All MOTUs	2 EC (KD, CF)										
	EC	1	1.83	0.31	0.001 ***	2.71	0.4	0.001***	1.23	0.24	0.001***
	Residuals	4		0.69			0.6			0.76	
	Total	5		1			1			1	
	11 EC										
	EC	10	1.49	0.43	0.001 ***	2.1	0.51	0.001***	1.17	0.37	0.001***
	Residuals	20		0.57			0.49			0.63	
	Total	30		1			1			1	
Fungal MOTUs	2 EC (KD, CF)										
	EC	1	2.12	0.35	0.001***	2.98	0.43	0.018*	1.3	0.25	0.001***
	Residuals	4		0.65			0.57			0.75	
	Total	5		1			1			1	
	11 EC										
	EC	10	1.66	0.45	0.001***	2.15	0.52	0.001***	1.23	0.38	0.001***
	Residuals	20		0.55			0.48			0.62	
	Total	30		1			1			1	

[a] Permutational multivariate analysis of variance based on the Jaccard similarity matrices, using 999 Monte Carlo permutations. This analysis was applied separately to the six locations of the EC "CF" (late snowmelt) and "KD" (early snowmelt), which had previously been reported as harbouring distinct fungal communities (52), and for all the 31 locations of the 11 EC

[b] Rare MOTUs were defined as containing one to ten reads in all the locations

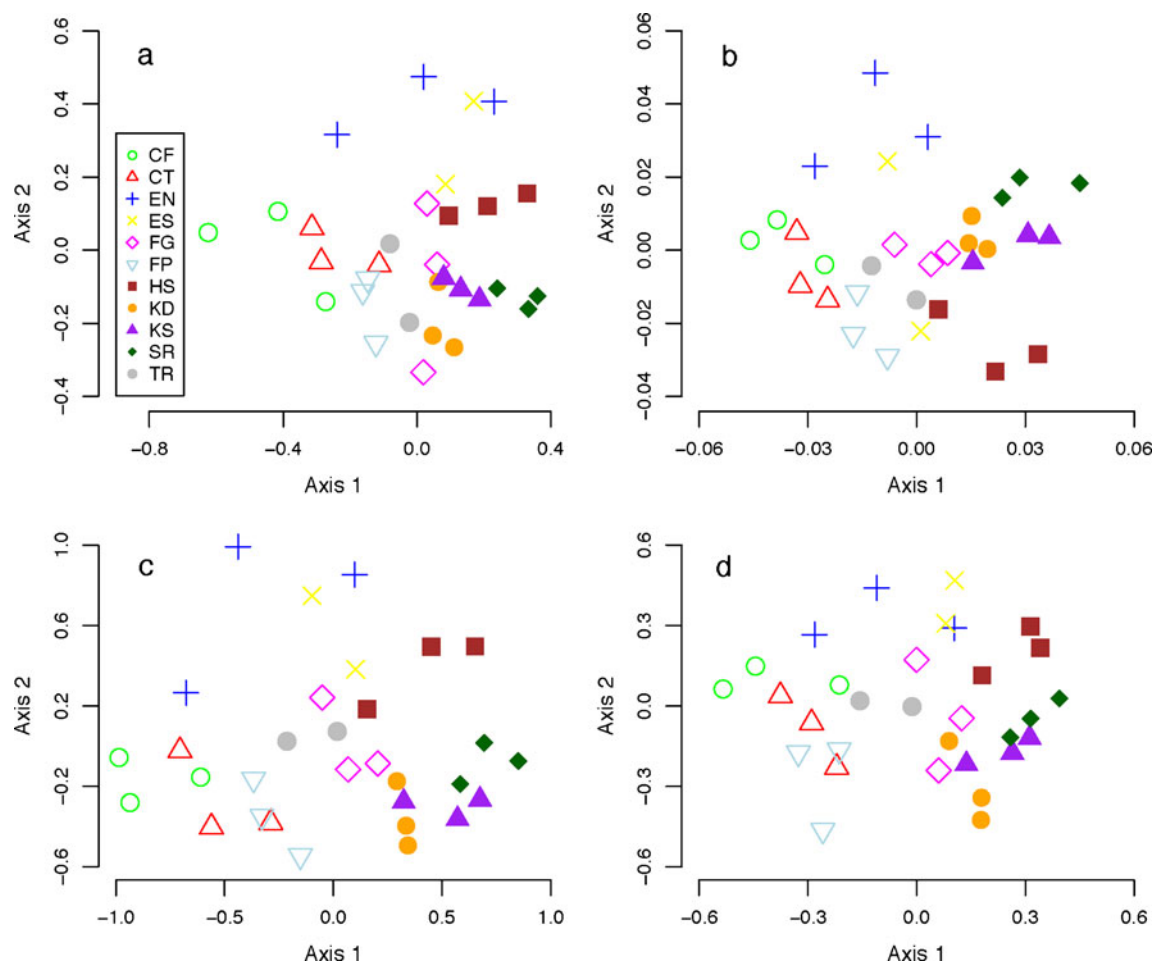[c] Amount of variance explain by the model

Fig. 4 Non-metric multidimensional scaling (NMDS) based on Jaccard distance matrix of MOTUs$_{0.98}$ occurrence. All the MOTUs$_{0.98}$ are one time split between non identified/Eukaryote identified MOTUs$_{0.98}$ (**a**) and fungal identified MOTUs$_{0.98}$ (**b**) which displayed a correlation of 71.6% (Procrustes correlation $r=0.001$), and a second time between abundant MOTUs$_{0.98}$ (more than 10 reads, **c**) and rare MOTUs$_{0.98}$ (less than 11 reads, **d**), which displayed a correlation of 94.7% (Procrustes correlation $r=0.001$)

variability, which supports the in silico study of Nilsson and colleagues who defined an average threshold at 97.5% (Nilsson et al. 2008).

Diversity profiling of alpine soil Fungi

The number of MOTUs$_{0.98}$ per location ranged from 142 to 768 (Online Resource 1), for 2.25 g (wet mass) of soil. This observed diversity is twice that of previous estimates reported in previous fungal diversity studies based on cloning-sequencing (Wardle et al. 2004; O'Brien et al. 2005; Wallenstein et al. 2007), but fits with the estimations obtained with pyrosequencing (Buée et al. 2009). The non-parametrical estimation of regional diversity is 5,100–12,000 MOTUs$_{0.98}$. A recent study (Schmit and Mueller 2007), estimated the lower limit of global fungal species diversity at 7.12×10^5 species. If a fungal MOTUs$_{0.98}$ is considered as a species, we found between 0.7% and 1.7% of the estimated global fungal diversity in ~70 g of soil.

The most surprising result obtained in our study was the absence of saturation in observed alpha and gamma diversities at 98% of similarity (Fig. 2), but also at lower similarity levels considering the full dataset (Fig. 2c,d), suggesting that our sequencing depth and sampling effort were insufficient to cover the entire fungal diversity both at regional and local scale. Furthermore, we detected high fungal species richness even in constraining environments such as scree soils, which displayed a similar fungal MOTUs$_{0.98}$ diversity to those of dense plant coverage areas (Table 1, Fig. 2b–d, Online Resource 1). Our results suggest that an important part of fungal diversity may still remain uncovered (Fig. 2), especially if the unidentified/eukaryal MOTUs are actually fungal species (Figs. 1 and 2). Estimating the true fungal diversity would therefore require a greater sequencing depth, as previously suggested for soil bacterial diversity (Quince et al. 2008).

Fungal beta-diversity patterns were strongly related to plant communities (EC, Table 2). This trend was not only

significant for the abundant MOTUs$_{0.98}$, but also for rare ones, and for both fungal MOTUs$_{0.98}$ and the full dataset. This confirms the previous conclusions on the studied area based on fingerprinting methods (Zinger et al. 2011). Such differences have already been reported in other tundra soils (Bissett and Parkinson 1979; Wallenstein et al. 2007; Zinger et al. 2009b), in grasslands (Singh et al. 2007) and tree plantations (Kasel et al. 2008) and could result from plant-*Fungi* relationship for resources assessment or specific soil structuring by roots (Christensen 1989; van der Heijden et al. 1998; Berg and Smalla 2009).

Beta-diversity patterns also provided new clues about the putative fungal origin of unidentified/eukaryal MOTUs$_{0.98}$ as their ecological signature strongly correlated to that of fungal MOTUs$_{0.98}$ (Fig. 4) and to a higher extent than previously observed between different groups of organisms (Zinger et al. 2011). This result strongly suggests that unidentified/eukaryal MOTUs$_{0.98}$ have a fungal origin. Indeed, they may correspond to unknown species that are still not listed in molecular databases, but also to experimental errors, which, if originating from fungal fragments, will keep an ecological fungal signature. Although this does not enable us to say if unidentified/eukaryal MOTUs actually belong to *Fungi*, it clearly shows that their consideration in diversity analysis will have dramatic effects on alpha and gamma diversity estimation, but would not affect the conclusions that can be inferred from a beta-diversity study.

Ecological consistency of a possible fungal rare biosphere

Singletons were abundantly present at regional and local scale, as 38% of MOTUs$_{0.98}$ were singletons (1% of reads). These results approximately fit with previous findings (Buée et al. 2009) in which 60% of the total number of MOTUs$_{0.98}$ were singletons (1.8% of reads). The variation in singleton numbers between the two studies may arise from a different sequencing depth, but also from the intrinsic discrepancies between forest vs. alpine tundra soils, the first known to harbour higher fungal diversity due to higher microhabitat fragmentation (Ettema and Wardle 2002). Such rank-abundance profile, with low number of abundant species and a high number of rare species, was commonly observed in culture-based studies (Bissett and Parkinson 1979; Christensen 1989; Bills et al. 2004), but also for plants (Murray et al. 1999) and tropical insects (Novotny and Basset 2000). Here the high sequencing depth obviously increase the proportion of rare MOTUs detected compare to earlier studies. Moreover, a significant part of rare MOTUs were successfully assigned to *Fungi* (53.3% per EC), meaning that their systematic assimilation to artefacts is not justified. This supports the notion of a "long tail" rank-abundance curve, or in other words the

existence of a rare biosphere, for *Fungi*, as reported for bacteria (Sogin et al. 2006; Galand et al. 2009).

Additionally, each EC displayed distinct communities when considering only rare MOTUs (Table 2), which can arise from either artefacts or from environmental and historical processes. On one hand, if rare MOTUs arise from artefacts, the cumulated error should differ between communities of different phylogenetic composition, as explained for unidentified/eukaryal MOTUs. On the other hand, the existence of a fungal rare biosphere may mirror the history of the studied area, as a fingerprint of the past plant cover composition and earlier succession stages (Bergero et al. 2003), and the high dispersal potential of fungal spores (Fröhlich-Nowoisky et al. 2009). Indeed, these rare MOTUs may represent a fungal "seed bank", i.e. pool of dormant cells (Lennon and Jones 2011), that could enhance the resilience in the face of disturbances. Studying this "seed bank" would be an interesting way of predicting the evolution of soil fungal communities with different climatic *scenari*. In this context, determining the part of the rare biosphere being functionally active and the dormant MOTUs by using metagenomic or metatranscriptomic approaches, coupled with culture-based approaches, would be of great importance when evaluating ecosystem resilience.

This study deals with the unidentified problematic and rare taxa that are nevertheless common in 454 datasets but are often ignored in other studies. We showed that the analysis of pyrosequencing data using an ab-initio approach coupled with bioinformatics and statistics provides robust results for the study of soil fungal beta-diversity regardless of rare taxa, but still need improvement in order to estimate alpha and gamma-diversity. Nevertheless, comparing both the full dataset and fungal MOTUs, we defined intervals for alpine soil fungal diversity, provided evidence of habitat determinism for fungal community distribution, and showed the existence of a fungal rare biosphere. This study opens the door to more comprehensive studies of fungal diversity at different spatial scales to test classic biogeographical hypotheses.

References

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

Anderson MJ (2001) A new method for non-parametric multivariate analysis of variance. Aust Ecol 26:32–46

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

Berg G, Smalla K (2009) Plant species and soil type cooperatively shape the structure and function of microbial communities in the rhizosphere. FEMS Microbiol Ecol 68:1–13

Bergero R, Girlanda M, Bello F, Luppi AM, Perotto S (2003) Soil persistence and biodiversity of ericoid mycorrhizal fungi in the absence of the host plant in a Mediterranean ecosystem. Mycorrhiza 13:69–75

Bills GF, Christensen M, Powell M, Thorn G (2004) Saprobic soil fungi. In: Mueller GM, Bills GF, Foster MS (eds) Biodiversity of fungi: inventory and monitoring methods. Academic Press, p 777 pages

Bissett J, Parkinson D (1979) Distribution of fungi in some alpine soils. Can J Bot 57:1609–1629

Buée 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

Choler P, Michalet R (2002) Niche differentiation and distribution of Carex curvula along a bioclimatic gradient in the southwestern Alps. J Veg Sci 13:851–858

Christensen M (1989) A View of Fungal Ecology. Mycologia 81:1–19

Cochrane G, Akhtar R, Bonfield J et al (2009) Petabyte-scale innovations at the European Nucleotide Archive. Nucleic Acids Res 37:D19–D25

De Deyn GB, Van der Putten WH (2005) Linking aboveground and belowground diversity. Trends Ecol Evol 20:625–633

Ettema CH, Wardle DA (2002) Spatial soil ecology. Trends Ecol Evol 17:177–183

Fröhlich-Nowoisky J, Pickersgill DA, Despres VR, Poschl U (2009) High diversity of fungi in air particulate matter. Proc Natl Acad Sci USA 106:12814–12819

Galand PE, Casamayor EO, Kirchman DL, Lovejoy C (2009) Ecology of the rare microbial biosphere of the Arctic Ocean. Proc Natl Acad Sci USA 106:22427–22432

Green JL, Holmes AJ, Westoby M, Oliver I, Briscoe D, Dangerfield M, Gillings M, Beattie AJ (2004) Spatial scaling of microbial eukaryote diversity. Nature 432:747–750

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

Huber JA, Mark Welch DB, Morrison HG, Huse SM, Neal PR, Butterfield DA, Sogin ML (2007) Microbial population structures in the deep marine biosphere. Science 318:97–100

Huse SM, Huber JA, Morrison HG, Sogin ML, Welch DM (2007) Accuracy and quality of massively parallel DNA pyrosequencing. Genome Biol 8:R143

Huse SM, Welch DM, Morrison HG, Sogin ML (2010) Ironing out the wrinkles in the rare biosphere through improved OTU clustering. Environ Microbiol 12:1889–1898

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

Jumpponen A, Jones KL, David Mattox J, Yaege C (2010) Massively parallel 454-sequencing of fungal communities in Quercus spp. ectomycorrhizas indicates seasonal dynamics in urban and rural sites. Mol Ecol 19(Suppl 1):41–53

Kasel S, Bennett LT, Tibbits J (2008) Land use influences soil fungal community composition across central Victoria, south-eastern Australia. Soil Biol Biochem 40:1724–1732

Körner C (1995) Towards a better experimental basis for upscaling plant responses to elevated CO2 and climate warming. Plant Cell Environ 18:1101–1110

Kunin V, Engelbrektson A, Ochman H, Hugenholtz P (2010) Wrinkles in the rare biosphere: pyrosequencing errors can lead to artificial inflation of diversity estimates. Environ Microbiol 12:118–123

Lauber CL, Strickland MS, Bradford MA, Fierer N (2008) The influence of soil properties on the structure of bacterial and fungal communities across land-use types. Soil Biol Biochem 40:2407–2415

Lennon JT, Jones SE (2011) Microbial seed banks: the ecological and evolutionary implications of dormancy. Nat Rev Microbiol 9:119–130

Lindner DL, Banik MT (2011) Intra-genomic variation in the ITS rDNA region obscures phylogenetic relationships and inflates estimates of operational taxonomic units in genus Laetiporus. Mycologia. doi:103852/10-331

Margulies M, Egholm M, Altman WE et al (2005) Genome sequencing in microfabricated high-density picolitre reactors. Nature 437:376–380

Mueller GM, Schmit JP (2007) Fungal biodiversity: what do we know? What can we predict? Biodivers Conserv 16:1–5

Murray BR, Rice BL, Keith DA, Myerscough PJ, Howell J, Floyd AG, Mills K, Westoby M (1999) Species in the tail of rank-abundance curves. Ecology 80:1806–1816

Needleman SB, Wunsch CD (1970) A general method applicable to the search for similarities in the amino acid sequence of two proteins. J Mol Biol 48:443–453

Nilsson RH, Ryberg M, Kristiansson E, Abarenkov K, Larsson KH, Koljalg U (2006) Taxonomic reliability of DNA sequences in public sequence databases: a fungal perspective. PLoS ONE 1

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. Evol Bioinform 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 Microbiol Lett 296:97–101

Novotny V, Basset Y (2000) Rare species in communities of tropical insect herbivores: pondering the mystery of singletons. Oikos 89:564–572

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

Oehl F, Sieverding E, Ineichen K, Mader P, Boller T, Wiemken A (2003) Impact of land use intensity on the species diversity of arbuscular mycorrhizal fungi in agroecosystems of Central Europe. Appl Environ Microbiol 69:2816–2824

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

Pearson WR (2000) Flexible sequence similarity searching with the FASTA3 program package. Meth Mol Biol 132:185–219

Peres-Neto P, Jackson D (2001) How well do multivariate data sets match? The advantages of a Procrustean superimposition approach over the Mantel test. Oecologia 129:169–178

Quince C, Curtis TP, Sloan WT (2008) The rational exploration of microbial diversity. ISME J 2:997–1006

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

Roesch LF, Fulthorpe RR, Riva A, Casella G, Hadwin AKM, Kent AD, Daroub SH, Camargo FAO, Farmerie WG, Triplett EW (2007) Pyrosequencing enumerates and contrasts soil microbial diversity. ISME J 1:283–290

Schmit JP, Mueller GM (2007) An estimate of the lower limit of global fungal diversity. Biodivers Conserv 16:99–111

Schwarzenbach K, Enkerli J, Widmer F (2007) Objective criteria to assess representativity of soil fungal community profiles. J Microbiol Meth 68:358–366

Singh B, Munro S, Potts J, Millard P (2007) Influence of grass species and soil type on rhizosphere microbial community structure in grassland soils. Appl Soil Ecol 36:147–155

Sogin ML, Morrison HG, Huber JA, Mark Welch D, Huse SM, Neal PR, Arrieta JM, Herndl GJ (2006) Microbial diversity in the deep sea and the underexplored "rare biosphere". Proc Natl Acad Sci USA 103:12115–12120

Taylor JW, Turner E, Townsend JP, Dettman JR, Jacobson D (2006) Eukaryotic microbes, species recognition and the geographic limits of species: examples from the kingdom Fungi. Philos Trans R Soc Lond B Biol Sci 361:1947–1963

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

The R Development Core Team (2009) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna

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

van der Heijden MGA, Bardgett RD, van Straalen NM (2008) The unseen majority: soil microbes as drivers of plant diversity and productivity in terrestrial ecosystems. Ecol Lett 11:296–310

van Dongen S (2000) Graph clustering by flow simulation. University of Utrecht

Waldrop MP, Zak DR, Blackwood CB, Curtis CD, Tilman D (2006) Resource availability controls fungal diversity across a plant diversity gradient. Ecol Lett 9:1127–1135

Wallenstein MD, McMahon S, Schimel J (2007) Bacterial and fungal community structure in Arctic tundra tussock and shrub soils. FEMS Microbiol Ecol 59:428–435

Wardle DA, Bardgett RD, Klironomos JN, Setala H, van der Putten WH, Wall DH (2004) Ecological linkages between aboveground and belowground biota. Science 304:1629–1633

White T, Bruns T, Lee S, Taylor J (1990) Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. In: Innis M, Gelfand D, Shinsky J, White T (eds). Academic Press, pp 315–322

Zinger L, Gury J, Alibeu O, Rioux D, Gielly L, Sage L, Pompanon F, Geremia RA (2008) CE-SSCP and CE-FLA, simple and high-throughput alternatives for fungal diversity studies. J Microbiol Meth 72:42–53

Zinger L, Coissac E, Choler P, Geremia RA (2009a) Assessment of microbial communities by graph partitioning in a study of soil fungi in two alpine meadows. Appl Environ Microbiol 75:5863–5870

Zinger L, Shahnavaz B, Baptist F, Geremia RA, Choler P (2009b) Microbial diversity in alpine tundra soils correlates with snow cover dynamics. ISME J 3:850–859

Zinger L, Lejon DPH, Baptist F, Bouasria A, Aubert S, Geremia RA, Choler P (2011) Contrasting diversity patterns of crenarchaeal, bacterial and fungal soil communities in an alpine landscape. PLoS ONE In Press