Graphical abstract

Three forest types

Pinus sylvestris | *Quercus robur* | Mixed

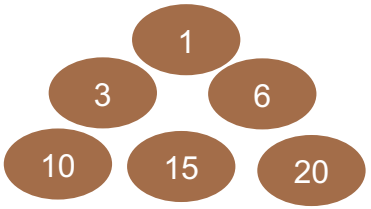
20 soil cores in 100m² plots → Soil sample pooling

Illumina sequencing of fungal ITS2 amplicons

Increasing number of soil sample pools

Species richness

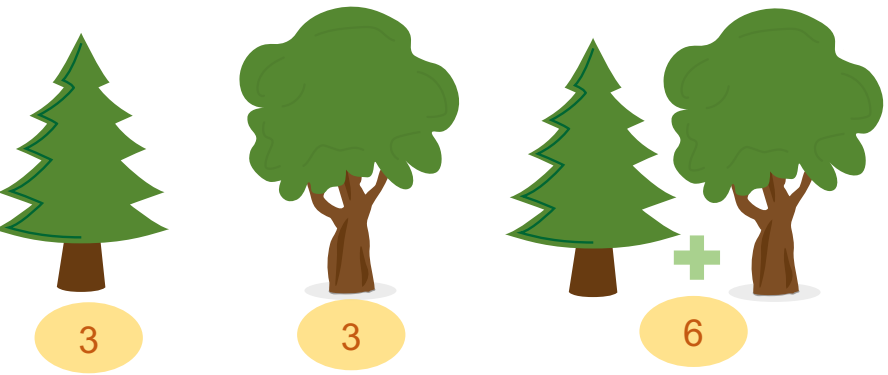
Positive species/area relationship

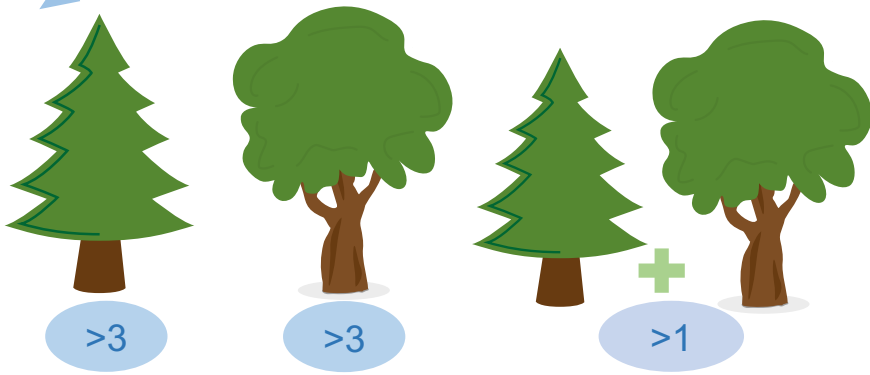


Which optimal sample pooling number?

Fungal ß diversity

Steadily decrease





Highlights

- We identified optimal sampling size in three Mediterranean forests.
- Soil samples were pooled to obtain mixtures of 3, 6, 10, 15, 20 samples.
- Three sample pools in pure, six in mixed stands gave consistent richness estimations.
- ß-diversity decreased with increasing sample pools in monospecific-stands.
- No effect of different number of sample pools on ß-diversity in mixed stands.

Sampling forest soils to describe fungal diversity and composition. Which is the optimal sampling size in Mediterranean pure and mixed pine oak forests?

Irene Adamo[1,2]*, Yasmine Piñuela[1,2,5], José Antonio Bonet[1,2], Carles Castaño[3], Juan Martínez de Aragón[1,5], Javier Parladé[4], Joan Pera[4], Josu G. Alday[1,2]

[1] Joint Research Unit CTFC - AGROTECNIO, Av. Alcalde Rovira Roure 191, E25198 Lleida, Spain

[2] Dep. Crop and Forest Sciences, University of Lleida, Av. Alcalde Rovira Roure 191, E25198 Lleida, Spain.

[3] Swedish University of Agricultural Sciences, Department of Forest Mycology and Plant Pathology, SE-75007 Uppsala, Sweden

[4] Sustainable Plant Protection, IRTA, Centre de Cabrils. Ctra Cabrils km 2, E08348 Cabrils, Barcelona, Spain.

[5]Forest Science and Technology Centre of Catalonia, Ctra. Sant Llorenç de Morunys km 2, 25280 Solsona, Spain

[*]Corresponding author: Irene Adamo

E-mail address: irene.adamo@udl.cat

Abstract

Soil sampling is a critical step affecting perceived fungal diversity, however sampling optimization for high-throughput-DNA sequencing studies have never been tested in Mediterranean forest ecosystems. We identified the minimum number of pooled samples needed to obtain a reliable description of fungal communities in terms of diversity and composition in three different Mediterranean forests (pine, oak, and mixed-pine-oak). Twenty soil samples were randomly selected in each of the three plots per type. Samples were pooled to obtain mixtures of 3, 6, 10, 15, 20 samples, and sequenced using Illumina MiSeq of fungal ITS2 amplicons. Pooling three soil samples in *Pinus* and *Quercus* stands provided consistent richness estimations, while at least six samples were needed in mixed-stands. ß-diversity decreased with increasing sample pools in monospecific-stands, while there was no effect of sample pool size on mixed-stands. Soil sample pooling had no effect over species composition. We estimate that three samples would be already optimal to describe fungal richness and composition in Mediterranean pure stands, while at least six samples would be needed in mixed stands.

Keywords: Fungal communities, DNA metabarcoding, number of pooled samples, mixed forests, Mediterranean forest, beta-diversity

1. Introduction

Soil fungi are drivers of fundamental ecosystems processes (Bardgett and van der Putten, 2014) such as soil carbon cycling and mineral nutrition of plants (Smith and Read, 2008; Bardgett and Wardle, 2010). Due to the enormous diversity of fungi and their fundamental roles as decomposers, mutualists, or pathogens of plants and animals (Mueller and Schmit, 2007; Tedersoo et al., 2014), the characterization of fungal communities has become crucial to disentangle soil microbial community dynamics and related ecological processes (Lindahl et al., 2013). High-throughput sequencing (HTS) methods have become a powerful tool to quantify fungal diversity in soils and have provided new information regarding the ecology of fungi in forests ecosystems (Hibbett et al. 2009; Lindahl et al., 2013; Hibbett et al., 2016; Nilsson et al., 2016). Previous studies have provided laboratory protocols (Clemmensen et al., 2016; De Filippis et al., 2017; Dopheide et al., 2019) or guidance on the multiple bioinformatic and taxonomic identification pipelines to prepare and assess high-throughput sequencing data (Gweon et al., 2015; Nguyen et al., 2016, Rognes et al., 2016; Somervuo et al. 2016; Bjørnsgaard et al., 2017; Anslan et al., 2017; Pauvert et al., 2019). Despite few attempts to optimize soil sampling protocols in high-throughput sequencing studies (Dickie et al., 2018), we still lack optimal soil sampling protocols to study fungal diversity and composition in Mediterranean soils. In addition, it is crucial to understand how these communities are structured in Mediterranean forests because of their potential important role in tree resistance against drought (See Mohan et al., 2014). For instance, Castaño et al. (2018) studied seasonal dynamics of these communities and how they respond to changing moisture and temperature, however lacked optimal sampling scheme to properly capture soil fungal diversity. Therefore, assessing the optimal sample pooling size in Mediterranean ecosystems is fundamental since it could affect the observed

diversity and community composition, which can potentially be detrimental to understand nutrient cycling and resistance against drought in these ecosystems (Mohan et al., 2014).

Most of the methodological studies have been performed in boreal and temperate ecosystems, but soil fungal communities in Mediterranean forest ecosystems seem to differ compared to boreal or temperate ecosystems (Castaño et al., 2018; Pérez-Izsquierdo et al., 2019). For instance, Mediterranean communities described in Castaño et al., (2018) were highly dominated by ectomycorrhizal species, and these were mainly species with short/contact exploration types (i.e. *Inocybe spp*; Castaño et al., 2018), which contrasts with many other boreal ecosystems, where medium-fringe or long exploration types may be more dominant (Sterkenburg et al., 2015). Since differences in exploration types determine how fungi explore spatially the soil (Agerer, 2001), distinct sampling approaches may be used depending on the dominating community. Therefore, sampling effort may be distinct for each community or habitat type, since fungal community members can have distinct growth, morphologies and trophic strategies, and mycelia can grow from few cm. to up to several meters (Agerer, 2001; Smith et al., 1992).

Correct assessment of soil fungal diversity or community composition using HTS methods requires an efficient soil sampling strategy, due to the species soil-area relationships and the complexity of the soil matrix (Grundmann and Debouzie, 2000; Ranjard et al., 2003). For instance, the heterogeneous distribution of fungi in the soil matrix has been recently highlighted (Ranjard et al. 2003), with fungal communities often distributed in forest soils in a patchy manner (Cairney, 2005). Fungal communities also operate in a distinct scale than other microbes such as bacteria, with a single genet

often occupying distances between <1 m and >5 m. (Dunham et al., 2003; Murata et al., 2005), up to 20 m, (Bonello et al., 1998; Sawyer et al., 1999). In addition, the amount of soil used to profile these communities employing molecular methods is typically limited to few grams or even < 1 g. Therefore, subsampling large amounts of soil to few grams is a common practice in fungal ecology studies dealing with soils (Kang and Mills, 2006). Moreover, the patchy distribution of fungi require that several samples are taken in a given site/plot, which are then often typically pooled before DNA analyses (Kang and Mills, 2006) or after DNA extraction (Dickie et al., 2018). If distinct soil samples are taken in a given area, it is crucial that samples are freeze-dried and grind to fine powder to facilitate homogenization (Lindahl et al., 2013). However, how the different number of pools (i.e. sampled volumes) and the number of samples taken in a given area may affect soil fungal diversity and community composition in samples with distinct ecological traits inhabiting distinct host species has not been tested yet in Mediterranean forest ecosystems.

It is well known that the observed number of plant and animal species increases with sampling area and volume (Arrhenius, 1921; McArthur, 1965; MacArthur and Wilson, 1968). For instance, Duarte et al. (2017), assessed the diversity of aquatic fungi across graded size of alder leaves and found that alpha diversity was positively influenced by increasing leaf area. Likewise, for microbes, Song et al. (2015) detected an increase in fungal OTU richness with increasing soil sample size from 0.25 g to 10 g in both prairie and forest soils. Therefore, increasing the number of soil sample pools may lead to a positive species/area relationship, and insufficient sampling may result in incorrect diversity estimations (Grey et al., 2018). The optimization of sample pooling size is a fundamental aspect for ecological studies as it may strongly affect results and their interpretations (Dickie et al., 2018). For example, insufficient number of samples may

lead to higher stochasticity in sampled communities, increasing sampling error and unexplained variation, which should be reflected in beta diversity values. Therefore, it is important to explore whether it is possible to establish a minimum optimal sampling size to reduce stochasticity and infer diversity estimates.

In this study, we aim to identify the minimum number of pooled samples needed to reach diversity plateau, i.e. optimal sample pooling size, for a set of distinct forest types in Mediterranean area. This might help us to detect reliable diversity and compositional values for a given area in order to answer subsequent ecological questions in forest ecosystems using appropriate sampling effort. It is well known that fungal diversity and community structure in forests is influenced by dominant tree species (Urbanová et al., 2015; Nagati et al., 2018; Geml, 2019). Therefore, we performed our study over three contrasting forest types, dominated by i) a widely distributed evergreen pine species (*P. sylvestris*), ii) a common broadleaf oak (*Quercus robur*) and ii) a mixed pine-oak forest of both species (*P. sylvestris*-*Quercus robur*). Here, *Quercus* and *Pinus* species possess different root systems occupying different soil layers (Sardans and Peñuelas, 2013) and different leaf traits, i.e. broadleaf vs. evergreen (Ishida et al., 2007), thus harbouring different fungal communities (Ishida et al., 2007; Cavard et al., 2011; Suz et al., 2017). Therefore, we expect different optimal sample pools sizes for each forest type. In line with these premises, we hypothesized that:

i) Considering the species-area theory (MacArthur and Wilson, 1968; Hill, 1973; Whittaker and Fernández- Palacios, 2007) fungal diversity will increase in pools with more soil samples until an optimal pooling size when the asymptotic plateau is reached.

ii) When we increase the number of sample pools, we expect to characterize the most dominating communities at plot level, reducing β-diversity. Similarly, when

pooling few samples, the probability to capture patchier communities increases, thus those species distributed in a patchier manner will cause an increase in soil fungal β-diversity in smaller sample pool sizes.

iii) Within each forest type, increasing the number of sample pools will produce a better characterization of the fungal community, because we will expect to sample the most abundant species as well as some species/communities distributed in a patchy manner. However, we hypothesize that these patchy distributed species will not have a great contribution to compositional differences but great effect over diversity.

2. Materials and Methods

2.1. Study sites and design

The study area was located in Northern-Eastern Spain (2º,4',18.61''E, 42º,15',46.42''N) at an altitude of 1149 m a.s.l., where three independent sites were selected. We choose three forest stands (100 m^2) in each site: a monospecific stand of *Pinus sylvestris*: named P, a monospecific stand of *Quercus robur*: named Q, and a mixed stand of *P. sylvestris* and *Q. robur* named M (total n=9). To avoid pseudo-replication, the forest stands at each site were randomly selected and the plots were more than 100 m distant from one another. Finally, to avoid tree proximity and represent under/out canopy, 20 samplings were considered in an area of 100 m^2, at least > 1 m from the nearby trees.

2.2. Soil sampling

In this study, 20 soil samples were randomly collected in November 2017 in each forest stand with a drillable cylinder corer (diameter: 5 cm; depth: 12 cm, 60 soil samples per forest type/site, 180 soil cores in total). In all cores, needles and oak leaves were eliminated, whereas humus and mineral soil were sampled together. Samples were sieved using 3 mm mesh and stored at 4 °C for less than 24 h until freeze-dried. Each sample was ground to fine powder using mortar and pestle to homogenize the soil core. The soil samples were manually pooled in order to obtain five composite independent samples representing an increasing gradient of mixing samples: pools of 3 samples, 6 samples, 10 samples, 15 samples and 20 samples. For this, the same volume (1 cm^3) from each soil sample that was used in the pooling was taken. This procedure was repeated for each plot in each site. From each of the 5 composite samples per stand we subsampled 500 mg of fine homogenized soil powder to extract the fungal DNA. The samples were coded with the corresponding forest type (P: *Pinus*, Q: *Quercus* and M: for mixed stands) followed by the number of soil samples pooled in each case, i.e. one sample pool: P1, Q1 and M1; for three sample pools: P3, Q3, M3; six sample pools: P6, Q6, M6; ten sample pools: P10, Q10, M10; fifteen sample pools: P15, Q15, M15; twenty sample pools: P20, Q20, M20. The resulting pooled samples were stored at −20 °C before DNA extraction.

2.3. Fungal community analyses

Fungal DNA was extracted from 500 mg aliquots using the NucleoSpin® NSP soil kit (Macherey-Nagel, Duren, Germany) following the manufacturer's protocol. Each sample was amplified using the gITS7 (Ihrmark et al., 2012) and ITS4 (White et al., 1990) primers to amplify the fungal ITS2 region, both fitted with unique 8- bp tags differing in at least three positions. The number of PCR cycles was optimised for each sample, with most of the samples amplifying at 23–26 cycles. The final concentrations

192 in the PCRs were: 1× Buffer, 200 µM of each nucleotide, 2.75 mM $MgCl_2$, primers at
193 500 nM (gITS7) and 300 nM (ITS4) and 0.025 U μl^{-1} polymerase (DreamTaq Green,
194 Thermo Scientific, Waltham, MA, USA). PCR cycling conditions were as follows: 5
195 min at 95°C, followed by 23-26 cycles of 30 s at 95°C, 30 s at 56°C, 30 s at 72°C and
196 final extension at 72°C for 7 min. Samples were amplified by triplicate together with
197 negative extraction and PCR controls. Amplicons were purified using the NucleoMag®
198 NGS Clean-up and Size Select (MACHEREY-NAGEL GmbH and Co) and quantified
199 using a Qubit fluorometer (Life Technologies, Carlsbad, CA, USA). Equal amounts of
200 DNA from each sample were pooled. Samples were sequenced at Stab Vida, Caparica,
201 Portugal on an Illumina MiSeq 2×300 bp.

202 *2.4. Bioinformatic analysis*

203 Sequences were quality filtered and clustered using the SCATA pipeline
204 (https://scata.mykopat.slu.se/). We first removed DNA sequences with length <200 bp
205 and were screened for sample tags and primers defining a primer match of at least 90%.
206 Sequences were pair-wise compared using 'usearch' (Edgar, 2010) after collapsing
207 homopolymers to 3 bp. Sequences were quality filtered removing data with amplicon
208 quality score of <20 (averaged per sequence) and with a score of <10 at any position.
209 Pairwise alignments were scored as follows: mismatch penalty of 1, gap open penalty of
210 0 and a gap extension penalty of 1. Putative chimera sequences were removed, and the
211 quality-filtered sequences were clustered into species hypotheses (Kõljalg et al., 2013)
212 using single linkage clustering, with a maximum distance of 1.5% to the closest
213 neighbour required to enter clusters. Global singletons were excluded from further
214 analyses. Switched tags were detected when the two primers from the same sequence
215 were found to have two distinct DNA tags and therefore these sequences were further
216 excluded from the data. Finally, the LULU (Frøsler et al., 2017) algorithm was applied

(minimum_ratio_type = “min”, minimum_match = 98.5, co_occ =0.8) to merge consistently co-occurring ‘daughter’ OTUs.Sequence data are archived at NCBI’s Sequence Read Archive under accession number PRJNA613458. (www.ncbi.nlm.nih.gov/sra).

2.5. Taxonomic identification

We taxonomically identified the 1000 most abundant OTUs. We selected the most abundant sequence from each OTU for taxonomic identification, using PROTAX software (Somervuo et al. 2016) implemented in PlutoF, using a 50% probability of correct classification (called by Somervuo et al. (2017) as “plausible identifications”). These identifications were confirmed and some of them improved using massBLASTer in PlutoF against the UNITE (Abarenkov et al. 2010). Taxonomic identities at species level were assigned based on >98.5% similarity with database references, or to other lower levels using the next criteria: genus on >97%, family on >95%, order on >92% and phylum on >90% similarity.

2.6. Statistical analyses

Statistical analyses were implemented in R software environment (version 3.6.0, R Development Core Team 2019), using the iNEXT (Hiesh et al., 2016) package for fungal diversity analyses, the *vegan* package (Oksanen et al., 2019) for the multivariate analyses, and *adespatial* package (Dray et al., 2018) was used for beta diversity analyses.

We used Hill’s diversity indices (Hills, 1973) to describe the differences in fungal diversity values between number of soil sample pools within each forest type. These analyses were performed on the overall fungal communities using the abundance-based matrices. Hill's diversity consists of three numbers: N0 is species richness; N1 is the

antilogarithm of Shannon's diversity index; and N2 is the inverse of Simpson's diversity index. Therefore, to test the effect of sample pooling on fungal diversity, the iNEXT function was used to build rarefactions curves pooling together the individual samples. The extrapolated confidence intervals were used to visualize the differences between the number of sample pools. Moreover, the number of sequences also rarefied to 4000 to assess interpolated richness with increasing number of sequences. For all compositional analyses, the species abundance matrix was Hellinger transformed (square root of relative abundance data) to account for taxa with low counts numbers (Legendre and Gallagher 2001) and then the dissimilarity matrices were calculated based on Bray-Curtis index. Also, compositional matrix was transformed to presence-absence and Jaccard dissimilarity was evaluated to test qualitative compositional changes. Differences in fungal overall community composition between number of sample pools were tested using permutational multivariate analyses of variance (PERMANOVA, function *"adonis"*). Then, the variance of Bray-Curtis matrix between the number of sample pools for each forest type was compared through using the *betadisper* function which is analogue to a Levene's test. Moreover, we expected species gains with increasing sample pools therefore, to assess ß-diversity patterns and whether the core of most abundant fungal species is maintained between sites, we evaluated for each pool the species (or abundances-per-species) losses (B) and species gains (C) using the beta-indices (tbi function, Legendre, 2019). Here, we used the one sample pool per each forest (sample 1) as a reference, and we compared pools with increasing number of samples (sample 3, 6, 10, 15 and 20) to identify species losses and gains. The statistical analyses' codes and some simulated data are freely accessible from the GitHub repository (Adamo et al. 2021, doi: 10.5281/zenodo.4434407).

3. Results

3.1. Sample pooling effect on fungal diversity

Species rarefaction curves showed significant differences in fungal richness across sample pools and between forest types. However, no clear differences in Shannon or Simpson fungal diversity indexes were detected across sample pools, since the extrapolated confidence intervals values overlapped. These two diversity variables ranged from 65.72-113.46/N1 and 52.11-125.26/N2 in *P. sylvestris*, from 52.11-136.21/N1 and 12.70-36.62/N2 in *Q. robur* and from 131.20-105.58/N1 and 52.11-125.26/N2 in mixed stands (Table S1). Considering species richness, there were significant differences between sample pools in *P. sylvestris* stands (Fig. 1a). The main difference was detected between P1, which had the lowest richness (= 428), and the other pools (> 650). The highest fungal richness was detected in P20 (= 916), followed by P15 (= 732), P10 (= 725) and P6 (= 704). In all cases, P3 observed richness values (= 657) were similar to observed values of higher number of sample pools (Fig.1a). Conversely, in *Q. robur* stands there were also significant differences in diversity across sample pools (Fig. 1b). Here, the extrapolated confidence intervals values of Q1 (714), Q6 and Q10 were significantly lower from Q15 (1019), and Q20 (868). On the other hand, no significant differences were detected between Q1, Q3, Q6 and Q10. Interestingly, Q3 richness values observed in *Q. robur* stands (857) were close to Q20 and Q15 (Fig.1b). Finally, in mixed pine-oak stands there were also significant differences in diversity across sample pools (Fig. 1c). The highest significant differences were detected between M1 or M3 (793) and the other sample pools. M6 showed the highest richness (1137) although it was not significantly different from M15 (1105) and M20 (1104). Moreover, no significant differences were detected between M3 and M6, therefore pooling from 3 to 6 samples will produce similar richness values

(Fig.1c). Finally, when the number of sequences were rarefied to 4000, differences in interpolated richness increased with increasing number of sequences (Fig. S1) similarly as previously described for interpolated and extrapolated Hill's N0 (Fig.1). The lower richness was detected in *P. sylvestris*, followed by *Q. robur* and mixed stands. For instance, 657 species were detected in P3, 857 in Q3, while 793 in M3. Conversely, mixed stands showed overall the highest richness values showing 30% more species than P and 10% more than Q stands (X^2 = 35.82, p <0.01).

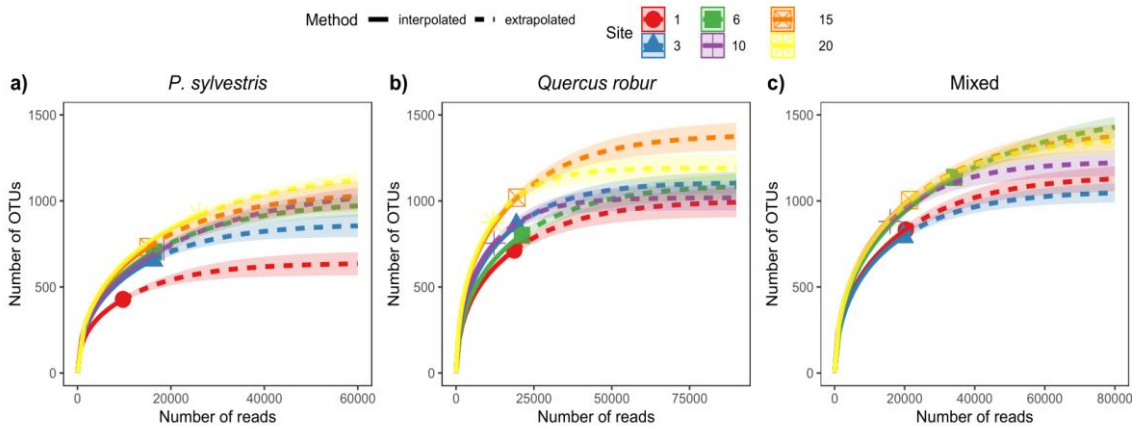


Fig. 1 Hill's N0 interpolated and extrapolated values across different sample pools in *P. sylvestris*, *Q. robur* and mixed pine-oak forest stand types. The values were obtained using the iNEXT fuction (iNEXT package, Hiesh et al., 2016). Hill's diversity consists of three numbers: N0 is species richness; N1 is the antilogarithm of Shannon's diversity index; and N2 is the inverse of Simpson's diversity index. Unbroken and dashed parts of the curve denote interpolated and extrapolated values respectively, and the shaded zone around each curve denotes the 95% confidence intervals. Significant differences appear where confidence interval do not overlap.

3.2. Sample pooling effect on fungal ß- diversity and species composition

ß-diversity values changed across sample pools in *P. sylvestris* and *Q. robur* stands ($F_{[5,12]}$=6.32, p-value < 0.01, $F_{[5,12]}$=13.12, p-value < 0.01) but not in mixed forest stands ($F_{[5,12]}$=0.67, p-value = 0.65; Fig.2). In contrast, no composition differences were

observed across soil sample pools in any of the three forest stands ($F_{[5,12]}$=0.61, p-value = 0.98, $F_{[5,12]}$=0.63, p-value = 0.99, $F_{[5,12]}$=0.47, p-value = 0.98) since SD-ellipses of the six groups were clearly superposed in the centre of the ordination (a) NMDS stress=0.07, b) NMDS stress=0.09, c) NMDS stress=0.10 Fig. S2). ß-diversity was highest in sample pools P1 and P3 of *P. sylvestris* stands ($F_{[5,12]}$=6.32, p-value < 0.01), while the ß-diversity steadily decreased with increasing number of sample pools (>P6), with no significant differences (Fig. 2a). Conversely, in *Q. robur* stands, ß-diversity was significantly higher in Q1 pools ($F_{[5,12]}$=13.12, p-value < 0.01) as compared to the other sample pools (>Q3, Fig. 2b). Here, ß-diversity values between larger pools other than Q1 were not significantly different (p>0.05). Finally, in mixed forest stands no significant differences in ß-diversity were detected between sample pools ($F_{[5,12]}$=0,67, p-value = 0.65, Fig. 2c), however, ß-diversity values increased but not significantly, from M1 to M20, with exception of M3. Similar results were obtained when the same analyses were performed over the presence-absence data using the dissimilarity matrices based on Jaccard index (data not shown).

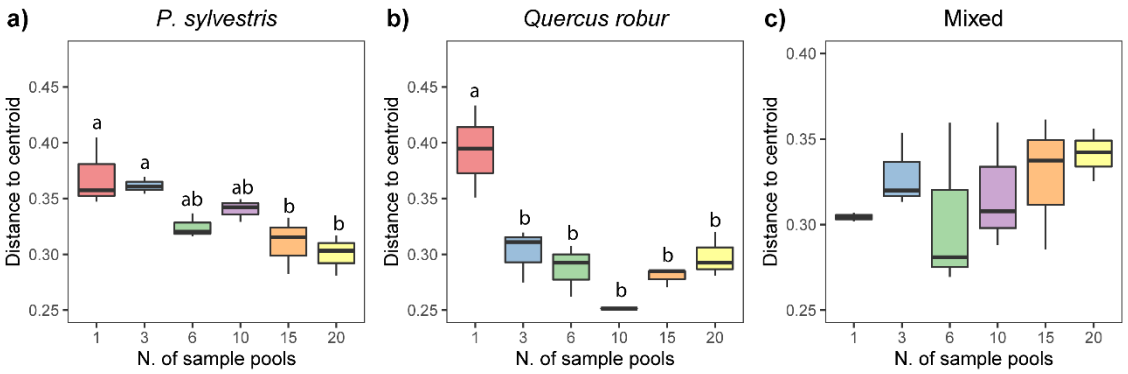


Fig. 2. Boxplots showing multivariate variance (Y-axis, ß-diversity values), sampled as distance to centroids, of each forest type in relation with the sample pools (X-axis). The species abundance matrix was Hellinger transformed and then the dissimilarity matrices were calculated based on Bray-Curtis index. Mean distance to centroids were compared with ANOVA and Tukey'HSD tests with letters denoting significant differences between number of sample pools.

There were differences in species loss and species gains between forest stands, however we did not find any significant p-values because of the low number of samples used in the permutations. In *P. sylvestris* stands, species loss values were not different between P3 and P10, while they slightly decreased between P15 (0.18) and P20 (0.16) (Table 1). Similarly, species gains values increased between P3 (0.27) and P20 (0.34). In *Q. robur* stands, species loss values were higher in Q3 (0.33) and Q10 (0.27), while they did not change across Q6, Q15 and Q20 (0.18). Conversely, no real changes in species gains were detected across Q3 and Q20, with the exception of Q6 (0.48) (Table 1). When mixed stands were analysed, species loss values decreased across M3 (0.39) and M6 (0.24) and did not change when they were compared with M1 and M20. On the other hand, species gains increased from M3 (0.36) to M6 (0.48), while there was a decrease in M10 (0.40). Yet, species gains values from M6 to M20 (0.44) decreased slightly (Table 1).

P. sylvestris	***Species loss***	***Species gains***	*p-value*
1-3	0.21 (± 0.05)	0.27 (± 0.05)	0.491
1-6	0.21 (± 0.05)	0.31 (± 0.06)	0.753
1-10	0.21 (± 0.01)	0.22 (± 0.03)	0.252
1-15	0.18 (± 0.04)	0.33 (± 0.02)	0.247
1-20	0.16 (± 0.02)	0.34 (± 0.02)	0.253
Q. robur	*Species loss*	*Species gains*	*p-value*
1-3	0.33 (± 0.11)	0.31 (± 0.12)	0.951
1-6	0.18 (± 0.07)	0.48 (± 0.04)	0.152
1-10	0.27 (± 0.11)	0.30 (± 0.08)	0.734
1-15	0.18 (± 0.03)	0.39 (± 0.04)	0.752
1-20	0.18 (± 0.04)	0.36 (± 0.04)	0.521
Mixed	*Species loss*	*Species gains*	*p-value*
1-3	0.39 (± 0.03)	0.36 (± 0.04)	0.953

1-6	0.24 (± 0.08)	0.48 (± 0.02)	0.502
1-10	0.37 (± 0.11)	0.40 (± 0.09)	0.814
1-15	0.31 (± 0.07)	0.45 (± 0.08)	0.712
1-20	0.32 (± 0.08)	0.44 (± 0.08)	0.758

Table 1. Mean (SE) ß-diversity components (loss and gain) across number of sample pools in *P. sylvestris, Q. robur* and mixed stand types. Temporal beta diversity was computed using the percentage difference index (Bray-Curtis) applied to the Hellinger transformed matrix. Total beta is the sum of 'species loss' and 'species gain' (Legendre, 2019). *P-values* were obtained using the *t.test.perm* option in the TBI function

4. Discussion

This study underlines the importance of sample pool size for accurate soil fungal diversity estimation in Mediterranean pure and mixed pine-oak forests, as increasing the number of soil sample pools, i.e. sampled volume, more reliable diversity predictions can be made with a positive species/area relationship (Whittaker and Fernández-Palacios, 2007). However, it seems not possible to standardise sampling pool protocols across distinct forest types, as our richness results showed that optimal soil sample pool size depended on forest type (e.g. pure or mixed forests). Moreover, increasing number of soil sample pools led to an increase in community similarity in pure forests, but not in mixed forests. Consequently, pools that represented less than three soil samples led to significant increases in ß-diversity values in pure forests, while values did not change in mixed forests. Finally, increasing the number of sample pools had no significant effect over species composition for any forest type, as we increased the sample pools while repeatedly sampling the same sites.

4.1. Sample pooling effect on fungal diversity

Our results demonstrate that increasing the number of soil sample pools leads to a positive species/area relationship regardless of the forest type investigated. Thus, the hypothesis 1 is accepted. These richness patterns are consistent with those reported in previous studies in agricultural fields and temperate forest sites, in which a positive relationship was detected between fungal diversity and increasing soil sample size (Ranjard et al., 2003; Song et al., 2015; Penton et al., 2016). Consequently, the number of samples pooled has important effects on the ecological interpretations also for fungal communities in soils, because insufficient sampling caused deviated richness values (Magurran, 2011). This implies that richness comparison between studies may be unreliable if distinct sampling strategies have been used, even comparing studies using

384 the same lab protocols. These results are very important for studies in which the total
385 diversity is targeted (i.e. biodiversity monitoring), but also when rare species are
386 targeted (Taberlet et al., 2018). The DNA extraction step also represents an important
387 source of bias in community composition (Plassart et al. 2012), however, here DNA
388 was carefully extracted following the same protocol for all the samples. In addition,
389 PCR step is also known to be a source of bias and may affect final community
390 composition. Nevertheless, we tried to keep biases as low as possible by reducing the
391 number of PCR cycles and using an optimized protocol for fungal metabarcoding
392 (Clemmensen et al., 2016). Finally, sequencing depth may also have an impact on the
393 perceived diversity (Smith and Peay, 2014), however based on the rarefaction curves
394 (Fig.1) our sequencing depth was able to capture similar coverage of the fungal
395 diversities of the community.
396 Surprisingly, neither Shannon nor Simpson fungal diversity indexes were affected by
397 sampling pooling, although they slightly increased but not significantly. Thus, for
398 Shannon and Simpson indexes the first hypothesis is not accepted. It is well known that
399 diversity is dependent on richness and evenness, then it seems that richness increases
400 are compensated in our case by evenness values (i.e. maintain or decrease slightly with
401 sample pools). Finally, although not tested here we argue that future studies should
402 consider both species-area and species-time relationship as it would lead to a deeper
403 understanding of fungal diversity patterns (Ladau et al., 2019).
404 In forest ecosystems, differences in dominant tree species identity can lead to
405 diversity and compositional changes (Ishida et al., 2007; Urbanová et al., 2015; Nagati
406 et al., 2018). Simultaneously, mixed forests are expected to harbour higher taxonomical
407 richness in all ecosystem compartments than pure stands (Ishida et al., 2007, Cavard et
408 al., 2011). For instance, Suz et al. (2017) reported higher ectomycorrhizal richness in

409 mixed pine-oak stands compared to pure pine stands. Our results follow these trends,
410 with greater richness in mixed stands compared to pure ones (Fig 1). Consequently, the
411 minimum number of sample pool size was different between pure and mixed stands. For
412 example, pooling at least three soil samples already provide consistent richness
413 estimations for *P. sylvestris and Q. robur* forests (same sampling effort), whereas for
414 mixed stands pools should include almost six soil samples.
415
416 *4.2. Sample pooling effect on fungal ß-diversity and species composition*
417 In this study, we observed a steady decrease of ß-diversity values with increasing
418 number of soil sample pools in both *P. sylvestris* and *Q. robur* stands, while there were
419 no significant changes in mixed forest stands (Fig.2). Thus, hypothesis 2 is partially
420 accepted. In pure *Pinus* and *Quercus* forest, the results followed the predicted trends,
421 with a decrease of dispersion values when increasing the number of sample pools. This
422 result indicates that pooling many samples reduces the ß-diversity estimation between
423 sites, which means a higher compositional similarity between different sites. This is
424 important, since by increasing the number of samples in each pool we may be able to
425 reduce the type II error and therefore reduce the error variance or unexplained variation.
426 The higher ß-diversity values observed in pools represented by low number of samples
427 in *Pinus* or *Quercus* is likely attributed to insufficient sampling effort that failed in
428 capturing the whole community in the site, with individual samples picking a different
429 subset of the community due to the patchiness distribution of each fungal species
430 (Cairney, 2012). Thus, it seems that smaller sample pools, i.e. lower than three, will
431 capture distinct subsets of the community, which would explain why there was much
432 higher heterogeneous communities between sites with lower pools than with larger soil
433 sample pools (Manter et al., 2010) since each new pool increased the species gains. Our

results agree with Ranjard et al. (2003), who found higher replicate variation in small sample sizes. It seems that in pure pine or oak forest, soil sample pools lower than three are prone to profile the community in a more biased manner. Conversely, our results showed that second hypothesis was not applicable for mixed stands, since increasing the number of sample pools does not significantly affect soil fungal ß-diversity. It is possible that the higher taxonomical richness and greater species coexistence present in mixed forests (Cavard et al., 2011) could explain why ß-diversity is not higher when pooling low number of samples. Further studies of mixed forest are needed to identify if increasing the number of sample pools over more than 20 cores causes a reduction of ß-diversity values.

Interestingly, our ß-diversity findings were supported by species loss and gain values between sample pools (Table 1). In pure *Pinus* and *Quercus* forest, while species gains values slightly increased or decreased, we detected almost constant species loss values across sample pools. Thus, the core of most abundant fungal species is maintained between sites, with low increases of less abundant species causing a reduction of ß-diversity. In contrast, species loss and gain values did not change in mixed forest, thus there are different ß-diversity patterns between forest types, being more heterogeneous the communities found in mixed forest, since interquartile ranges were higher than in pure stands (Fig. 2). In any case, it seems that we are not collecting enough number of samples to pool to characterize ß-diversity patterns and species gains and loses properly in mixed forest.

Finally, increasing the number of sample pools had no significant effect on species composition for any forest type. These results are consistent with our last hypothesis, as we expected to not detect any influence of sample pools on community composition in each forest type. Since each low sample pool reflect a subset of the higher pools

increasing the number of sample pools will not influence the species composition, qualitatively or quantitatively. Thus, it is possible that the main species are maintained, and the incorporation of new species is then reduced when increasing new sampling pools (see Fig. S1) (Magurran, 2011). Therefore, it seems that when profiling the core community (more abundant species) low sampling effort might be enough. However, an increase in the number of sampling cores may be desirable when targeting for rare or less abundant species since many important processes may be driven by specific, low abundant species (Red list fungal species, Quarantine pathogens).

4.3. Conclusions

In this study, increasing number of sample pools had a significant effect on fungal richness in all the three forest types, indicating a positive positive species/area relationship. Moreover, our results indicate that the minimum number of sample pools to adequately estimate fungal richness and species composition will be lower in monospecific stands, three in our case, than for more diverse mixed forest where the optimal pooling will be almost six samples. Our results shed light on best soil sample monitoring implementations to be applied for characterizing pure and mixed forests ecosystems. However, further research is needed to test if these results can be extrapolated to different ecosystems in the area or in similar areas.

Acknowledgments

This work was partially supported by the Spanish Ministry of Science, Innovation and Universities, grants AGL2015-66001-C3, RTI2018-093907-B-C21 and RTI2018-099315-A-I00. I.A. was supported by a Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie Cofund grant agreement No. 801596,

J.G.A. was supported by Ramon y Cajal fellowship (RYC-2016-20528) and J.A.B. benefitted from a Serra-Húnter Fellowship provided by the Generalitat of Catalunya.

Conflicts of interests

The authors declare they have no conflict of interest.

Authors' contribution

All authors contributed to the study conception and design. Material preparation, data collection and analysis were performed by Yasmin Piñuela, Carles Castaño, José Antonio Bonet, Irene Adamo and Josu G. Alday. The first draft of the manuscript was written by Irene Adamo and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

References

Adamo I, Pinuela Y, Bonet JA, Castaño C, Martinez de Aragón J, Parladé J, Pera J, Alday JG. 2021.Statistical-analyses-optimal-sampling-size-in-Mediterranean-forest-ecosystems:first release GitHub repository http://github.com/ire1990/Statistical-analyses-optimal-sampling-size-in-Mediterranean-forest-ecosystems. doi: 10.5281/zenodo.4434407)

Anslan S, Bahram M, Hiiesalu I, Tedersoo L, 2017. PipeCraft: Flexible open-source toolkit for bioinformatics analysis of custom high-throughput amplicon sequencing data. Mol. Ecol. Resour 17: e234–e240 https://doi.org/10.1111/1755-0998.12692

Arrhenius O, 1921. Species and Area. J. Ecol: 9: 95-99 . https://doi.org/10.2307/2255763

Bardgett RD, Wardle DA, 2010. Above-belowground linkages: biotic interaction, ecosystem processes and global chhange. Oxford University Press.

Bardgett RD, Van Der Putten WH, 2014. Belowground biodiversity and ecosystem

functioning. Nature 515: 505–511. https://doi.org/10.1038/nature13855

Bjørnsgaard Aas A, Davey, ML, Kauserud H, 2017. ITS all right mama: investigating the formation of chimeric sequences in the ITS2 region by DNA metabarcoding analyses of fungal mock communities of different complexities. Mol. Ecol. Resour 17:4, 730-741. https://doi.org/10.1111/1755-0998.12622

Bonello P, Bruns T D, Gardes M, 1998. Genetic structure of a natural population of the ectomycorrhizal fungus Suillus pungens. New Phytol. 138: 533-543263–264. https://doi.org/10.1046/j.1469-8137.1998.00122.x

Cairney JWG, 2005. Basidiomycete mycelia in forest soils: Dimensions, dynamics and roles in nutrient distribution. Mycol. Res 109: 7-20. https://doi.org/10.1017/S0953756204001753

Cairney JWG, 2012. Extramatrical mycelia of ectomycorrhizal fungi as moderators of carbon dynamics in forest soil. Soil Biol. Biochem.48: 198-208 https://doi.org/10.1016/j.soilbio.2011.12.029

Cavard X, Macdonald SE, Bergeron Y, Chen HYH, 2011. Importance of mixedwoods for biodiversity conservation: Evidence for understory plants, songbirds, soil fauna, and ectomycorrhizae in northern forests. Environ. Rev. 19: 142-161 https://doi.org/10.1139/A11-004

Clemmensen KE, Ihrmark K, Durling MB, Lindahl BD, 2016. Sample preparation for fungal community analysis by high-throughput sequencing of barcode amplicons, in: Methods in Environmental Genomics 1399: 61-88. https://doi.org/10.1007/978-1-4939-3369-3_4

De Filippis F, Laiola M, Blaiotta G, Ercolini D, 2017. Different amplicon targets for sequencing-based studies of fungal diversity. Appl. Environ. Microbiol 83: 1-9. https://doi.org/10.1128/AEM.00905-17

533 Dickie IA, Boyer S, Buckley HL, Duncan RP, Gardner PP, Hogg ID, Holdaway RJ,
534 Lear G, Makiola A, Morales SE, Powell JR, Weaver L, 2018. Towards robust and
535 repeatable sampling methods in eDNA-based studies. Mol. Ecol. Resour 18: 940-
536 952. https://doi.org/10.1111/1755-0998.12907
537 Dopheide A, Xie D, Buckley TR, Drummond AJ, Newcomb RD, 2019. Impacts of
538 DNA extraction and PCR on DNA metabarcoding estimates of soil biodiversity.
539 Methods Ecol. Evol 10: 120-133. https://doi.org/10.1111/2041-210X.13086
540 Dray S, Bauman D, Guillaume B, Borcard D, Clappe S, Guenard G, Jombart T, Laroque
541 G, Legendre P, Madi N, Wagner HH, 2018. Package "adespatial" 1–97.
542 Duarte S, Cássio F, Pascoal C, Bärlocher F, 2017. Taxa-area relationship of aquatic
543 fungi on deciduous leaves. PLoS One 12: 1–19.
544 https://doi.org/10.1371/journal.pone.0181545
545 Dunham SM, Kretzer A, Pfrender ME, 2003. Characterization of Pacific golden
546 chanterelle (Cantharellus formosus) genet size using co-dominant microsatellite
547 markers. Mol. Ecol. 12: 1607-1618. https://doi.org/10.1046/j.1365-
548 294X.2003.01837.x
549 Edgar CRC, 2010. Search and clustering orders of magnitude faster then BLAST.
550 Bionformatics 26:19, 2460-2461. https://doi.org/10.1093/bioinformatics/btq461
551 Frøslev TG, Kjøller R, Bruun HH, Ejrnæs R, Brunbjerg AK, Pietroni C, Hansen AJ,
552 2017. Algorithm for post-clustering curation of DNA amplicon data yields reliable
553 biodiversity estimates. Nat. Commun 8: 1188. https://doi.org/10.1038/s41467-017-
554 01312-x
555 Geml J, 2019. Soil fungal communities reflect aspect-driven environmental structuring
556 and vegetation types in a Pannonian forest landscape. Fungal Ecol. 39: 63–79.
557 https://doi.org/10.1016/j.funeco.2018.12.005

558 Grey EK, Bernatchez L, Cassey P, Deiner K, Deveney M, Howland KL, Lacoursière-
559 Roussel A, Leong S CY, Li Y, Olds B, Pfrender ME, Prowse TAA, Renshaw MA,
560 Lodge DM, 2018. Effects of sampling effort on biodiversity patterns estimated
561 from environmental DNA metabarcoding surveys. Scientific Reports 8, 8843.
562 https://doi.org/10.1038/s41598-018-27048-2
563 Grundmann GL, Debouzie DJ, 2000. Geostatistical analysis of the distribution of $NH4^-$
564 and $NO2^-$ oxidizing bacteria and serat the millimeter scale along a soil transect.
565 FEMS Micro Ecol. 34: 57-62. https://doi.org/10.1111/j.1574-6941.2000.tb00754.x
566 Gweon HS, Oliver A, Taylor J, Booth T, Gibbs M, Read DS, Griffiths RI, Schonrogge
567 K, 2015. PIPITS: An automated pipeline for analyses of fungal internal transcribed
568 spacer sequences from the Illumina sequencing platform. Methods Ecol. Evol 6(8):
569 973-980. https://doi.org/10.1111/2041-210X.12399
570 Hibbett D, AbarenkovK, Kõljalg U, Öpik M, Chai B, Cole J, Wang Q, Crous P, Robert
571 V, Helgason T, Herr JR, Kirk P, Lueschow S, O'Donnell K, Nilsson RH, Oono R,
572 Schoch C, Smyth C, Walker DM, Porras-Alfaro A, Taylor JW, Geiser DM, 2016.
573 Sequence-based classification and identification of Fungi. Mycologia 108: 1049-
574 1068. https://doi.org/10.3852/16-130
575 Hibbett DS, Ohman A, Kirk PM, 2009. Fungal ecology catches fire. New Phytol 184:
576 277-279. https://doi.org/10.1111/j.1469-8137.2009.03042.x
577 Hill MO, 1973. Diversity and evenness: a unifying notation and its consequences.
578 Ecology 54: 427-432.
579 Hsieh TC, Ma KH, Chao A, 2016. iNEXT: an R package for rarefaction and
580 extrapolation of species diversity (Hill numbers). Methods Ecol. Evol 7: 1451-
581 1456. https://doi.org/10.1111/2041-210X.12613
582 Ihrmark K, Bödeker ITM, Cruz-Martinez K, Friberg H, Kubartova A, Schenck J, Strid

Y, Stenlid J, Brandström-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. https://doi.org/10.1111/j.1574-6941.2012.01437.x

Ishida TA, Nara K, Hogetsu T, 2007. Host effects on ectomycorrhizal fungal communities: Insight from eight host species in mixed conifer-broadleaf forests. New Phytol 174: 430-440. https://doi.org/10.1111/j.1469-8137.2007.02016.x

Kang S, Mills AL, 2006. The effect of sample size in studies of soil microbial community structure. J. Microbiol. Methods 66: 242-250. https://doi.org/10.1016/j.mimet.2005.11.013

Koljalg U, Nilsson RH, Abarenkov K, Tedersoo L, Taylor AFS, Bahram M, 2013. Towards a unified paradigm for sequence-based identification of fungi. Mol. Ecol. 22: 5271–5277. https://doi.org/10.1111/mec.12481

Ladau J., Eloe-Fadrosh EA, 2019. Spatial, temporal, and phylogenetic scales of microbial ecology. *Trends Microbiol***27**:662–669. https://doi.org/10.1016/j.tim.2019.03.003

Legendre P, 2019. A temporal beta-diversity index to identify sites that have changed in exceptional ways in space–time surveys. Ecol. Evol 9: 3500-3514. https://doi.org/10.1002/ece3.4984

Legendre P, Gallagher ED, 2001. Ecologically meaningful transformations for ordination of species data. Oecologia 129: 271-280. https://doi.org/10.1007/s004420100716

Lindahl BD, Nilsson RH, Tedersoo L, Abarenkov K, Carlsen T, Kjøller R, Kõljalg U, Pennanen T, Rosendahl S, Stenlid J, Kauserud H, 2013. Fungal community analysis by high-throughput sequencing of amplified markers - a user's guide. New

608 Phytol. 199: 288-299. https://doi.org/10.1111/nph.12243
609 MacArthur RH, 1965. Patterns of biological diversity. Biol. Rev 40: 510-533.
610 https://doi.org/10.1111/j.1469-185x.1965.tb00815.x
611 MacArthur RH, Wilson EO, 1968. The Theory of Island Biogeography. Geogr. J 134:
612 592. https://doi.org/10.2307/1796430
613 Magurran AE, McGill BJ, 2011. Biological diversity: frontiers in measurement and
614 assessment. Oxford University Press.
615 Manter DK, Weir TL, Vivanco JM, 2010. Negative effects of sample pooling on PCR-
616 Based estimates of soil microbial richness and community structure. Appl.
617 Environ. Microbiol. 76: 2086–2090. https://doi.org/10.1128/AEM.03017-09
618 Mueller GM, Schmit JP, 2007. Fungal biodiversity: What do we know? What can we
619 predict? Biodivers. Conserv. 16: 1–5. https://doi.org/10.1007/s10531-006-9117-7
620 Murata H, Ohta A, Yamada A, Narimatsu M, Futamura N, 2005. Genetic mosaics in the
621 massive persisting rhizosphere colony “shiro” of the ectomycorrhizal
622 basidiomycete *Tricholoma matsutake*. Mycorrhiza 15: 505-512.
623 https://doi.org/10.1007/s00572-005-0358-1
624 Nagati M, Roy M, Manzi S, Richard F, Desrochers A, Gardes M, Bergeron Y, 2018.
625 Impact of local forest composition on soil fungal communities in a mixed boreal
626 forest. Plant Soil 432: 345–357. https://doi.org/10.1007/s11104-018-3806-3
627 Nilsson RH, Wurzbacher C, Bahram M, Coimbra VRM, Larsson E, Tedersoo L,
628 Eriksson J, Ritter CD, Svantesson, S, Sánchez-García M, Ryberg M, Kristiansson
629 E, Abarenkov K, 2016. Top 50 most wanted fungi. MycoKeys 12: 29-40.
630 https://doi.org/10.3897/mycokeys.12.7553
631 Nguyen NH, Song Z, Bates ST, Branco S, Tedersoo L, Menkel J, Schilling JS, Kennedy
632 PJ, 2016. FUNGuild: an open annotation tool for parsing fungal communty

633 datasetes by ecological guild. Fungal Ecol 20: 241-248
634 Oksanen J, Blanchet FG, Friendly M, Kindt R, Legendre P, McGlinn D, Minchin PR,
635 O'Hara RB, Simpson GL, Solymos P, Stevens MHH, Szoecs E, Wagner H, 2019.
636 vegan: Community Ecology Package. R package version 2,5-2. Cran R.
637 Pauvert C, Buèe M, Laval V, Edel-Hermann V, Fauchery L, Gautier A, Lesur I,
638 Valance J, Vacher C, 2019. Bionformatics matters: The accurancy of plant and soil
639 fungal communities data is highly dependent on the metabarcoding pipeline.
640 Fungal ecol. 41: 23-33. https://doi.org/10.1016/j.funeco.2019.03.005
641 Penton CR, Gupta VVSR, Yu J, Tiedje JM, 2016. Size matters: Assessing optimum soil
642 sample size for fungal and bacterial community structure analyses using high
643 throughput sequencing of rRNA gene amplicons. Front. Microbiol 7,824.
644 https://doi.org/10.3389/fmicb.2016.00824
645 Pérez-Izquierdo L, Zabal-Aguirre M , González-Martínez SC, Buée M, Verdú M,
646 Rincón A, Goberna M, 2019. Plant intraspecific variation modulates nutrient
647 cycling through its below ground rhizospheric microbiome. Journal of Ecology
648 107(4): 1594-1605. doi:10.1111/1365-2745.13202
649 Plassart P, Terrat S, Thomson B, Griffiths R, Dequiedt S, Lelievre M, Regnier T,Nowak
650 V, Bailey M, Lemanceau P, Bispo A, Chabbi A, Maron PA, Mougel C, Ranjard L,
651 2012. Evaluation of the ISO Standard 11063 DNA extraction for assessing soil
652 microbial abundance and community structure. Plos One 7 1-8.
653 https://doi.org/:10.1371/journal.pone.0044279
654 Ranjard L, Lejon DPH, Mougel C, Schehrer L, Merdinoglu D, Chaussod R, 2003.
655 Sampling strategy in molecular microbial ecology: Influence of soil sample size on
656 DNA fingerprinting analysis of fungal and bacterial communities. Environ.
657 Microbiol 5: 1111-1120. https://doi.org/10.1046/j.1462-2920.2003.00521.x

Rognes T, Flouri T, Nichols B, Quince C, Mahé F, 2016. VSEARCH: a versatile open source tool for metagenomics. PeerJ 4,: e2584. https://doi.org/10.7717/peerj.2584

Sardans J, Peñuelas J, 2013. Plant-soil interactions in Mediterranean forest and shrublands: Impacts of climatic change. Plant Soil 365: 1-33. https://doi.org/10.1007/s11104-013-1591-6

Sawyer NA, Chambers SM, Cairney JWG, 1999. Molecular investigation of genet distribution and genetic variation of Corrinarius rotundisporus in eastern Australian sclerophyll forests. New Phytol. 42: 561-568. https://doi.org/10.1046/j.1469-8137.1999.00417.x

Smith ML, Bruhn JN, Anderson JB. 1992. The fungus Armillaria bulbosa is among the largest and oldest living organisms. Nature 356: 428–431

Smith S, Read D, 2008. Mycorrhizal Symbiosis, Mycorrhizal Symbiosis 637-768. https://doi.org/10.1016/B978-0-12-370526-6.X5001-6

Somervuo P, Koskela S, Pennanen J, Nilsson RH, Ovaskainen O, 2016. Unbiased probabilistic taxonomic classification for DNA barcoding. Bioinformatics 32:2920-2927. https://doi.org/10.1093/bioinformatics/btw346

Song Z, Schlatter D, Kennedy P, Kinkel LL, Kistler HC, Nguyen N, Bates ST, 2015. Effort versus reward Preparing samples for fungal community characterization in high-throughput sequencing surveys of soils. PLoS One 1-13. https://doi.org/10.1371/journal.pone.0127234

Suz LM, Kallow S, Reed K, Bidartondo MI, Barsoum N, 2017. Pine mycorrhizal communities in pure and mixed pine-oak forests: Abiotic environment trumps neighboring oak host effects. For. Ecol. Manage. 40: 370–380. https://doi.org/10.1016/j.foreco.2017.09.030

Taberlet P, Bonin A, Zinger L, Coissac E, 2018. Environmental DNA: for biodiversity

research and monitoring, https://doi.org/10.1093/oso/9780198767220.001.0001

Tedersoo L, Bahram M, Põlme S, Kõljalg U, Yorou NS, Wijesundera R, Abarenkov K, 2014. Global diversity and geography of soil fungi. Science 346: 6213. https://doi.org/10.1126/science.1256688

Urbanová M, Šnajdr J, Baldrian P, 2015. Composition of fungal and bacterial communities in forest litter and soil is largely determined by dominant trees. Soil Biol. Biochem 84: 53-64. https://doi.org/10.1016/j.soilbio.2015.02.011

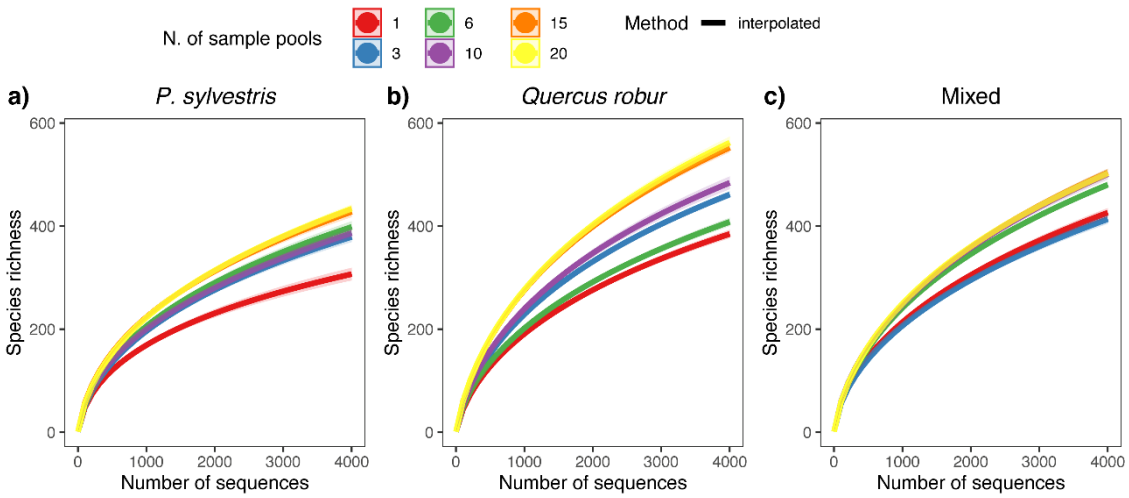
White TJ, Bruns T, Lee S, Taylor J, 1990. Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics in PCR protocols. 315–322. https://doi.org/10.1016/B978-0-12-372180-8.50042-1

Whittaker RJ, Fernández-Palacios JM, 2007. Island biogeography: ecology, evolution and conservation. 2nd edition, Oxford University Press.

Supplementary material

Table S1. Hill's N1 and Hill's N2 observed values across different sample pools in *P. sylvestris, Q. robur* and mixed pine-oak forest stand types. The values were obtained using the iNEXT fuction (iNEXT package, Hiesh et al., 2016). Hill's diversity consists of three numbers: N0 is species richness; N1 is the antilogarithm of Shannon's diversity index; and N2 is the inverse of Simpson's diversity index.

P. sylvestris	*Hill's N1*	*Hill's N2*
P1	65.72 (±1.29)	26.10 (±0.51)
P3	80.83 (±1.32)	25.88 (±0.46)
P6	84.80 (±1.27)	27.39 (±0.50)
P10	97.58 (±1.44)	39.83 (±0.59)
P15	113.47 (±1.83)	44.46 (±0.73)
P20	107.79 (±1.17)	38.03 (±0.54)
Q. robur	*Hill's N1*	*Hill's N2*
Q1	52.11 (±0.89)	12.70 (±0.24)
Q3	97.72 (±1.43)	30.74 (±0.51)
Q6	65.38 (±0.96)	16.18 (±0.25)
Q10	90.50 (±1.92)	22.40 (±0.52)
Q15	136.21 (±2.45)	36.62 (±0.64)
Q20	125.26 (±2.52)	27.52 (±0.76)
Mixed	*Hill's N1*	*Hill's N2*
M1	109.26 (±1.46)	46.41 (±0.60)
M3	105.57 (±1.36)	46.40 (±0.59)
M6	126.57 (±1.20)	44.23 (±0.55)
M10	131.12 (±1.90)	48.38 (±0.88)
M15	120.74 (±1.70)	33.74 (±0.6)
M20	126.77 (±1.62)	41.48 (±0.68)



711

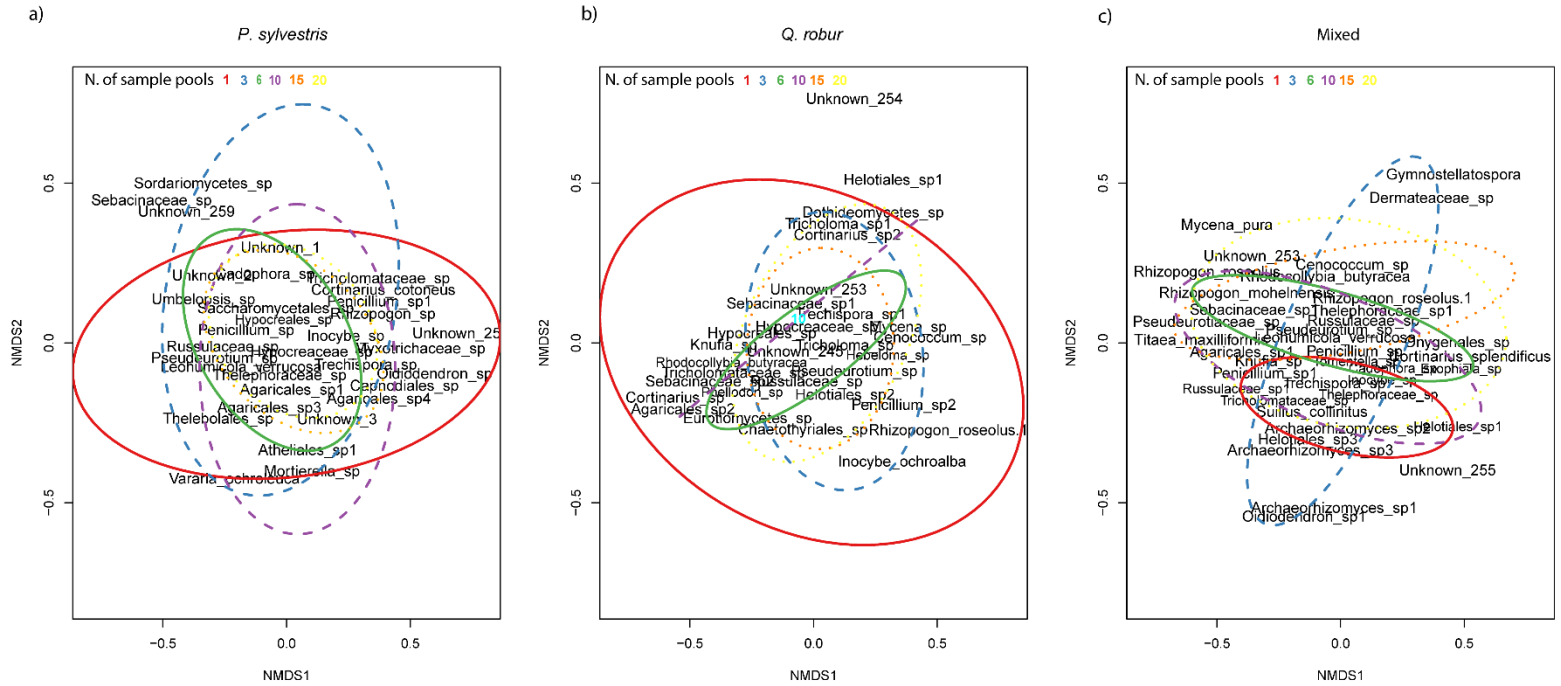
712 **Fig. S1.** Hill's N0 interpolated values across different sample pools in *P. sylvestris*, *Q.*

713 *robur* and mixed pine-oak forest stand types. The values were obtained using the iNEXT

714 function (iNEXT package, Hiesh et al., 2016). Hill's diversity consists of three numbers:

715 N0 is species richness; N1 is the antilogarithm of Shannon's diversity index; and N2 is

716 the inverse of Simpson's diversity index.



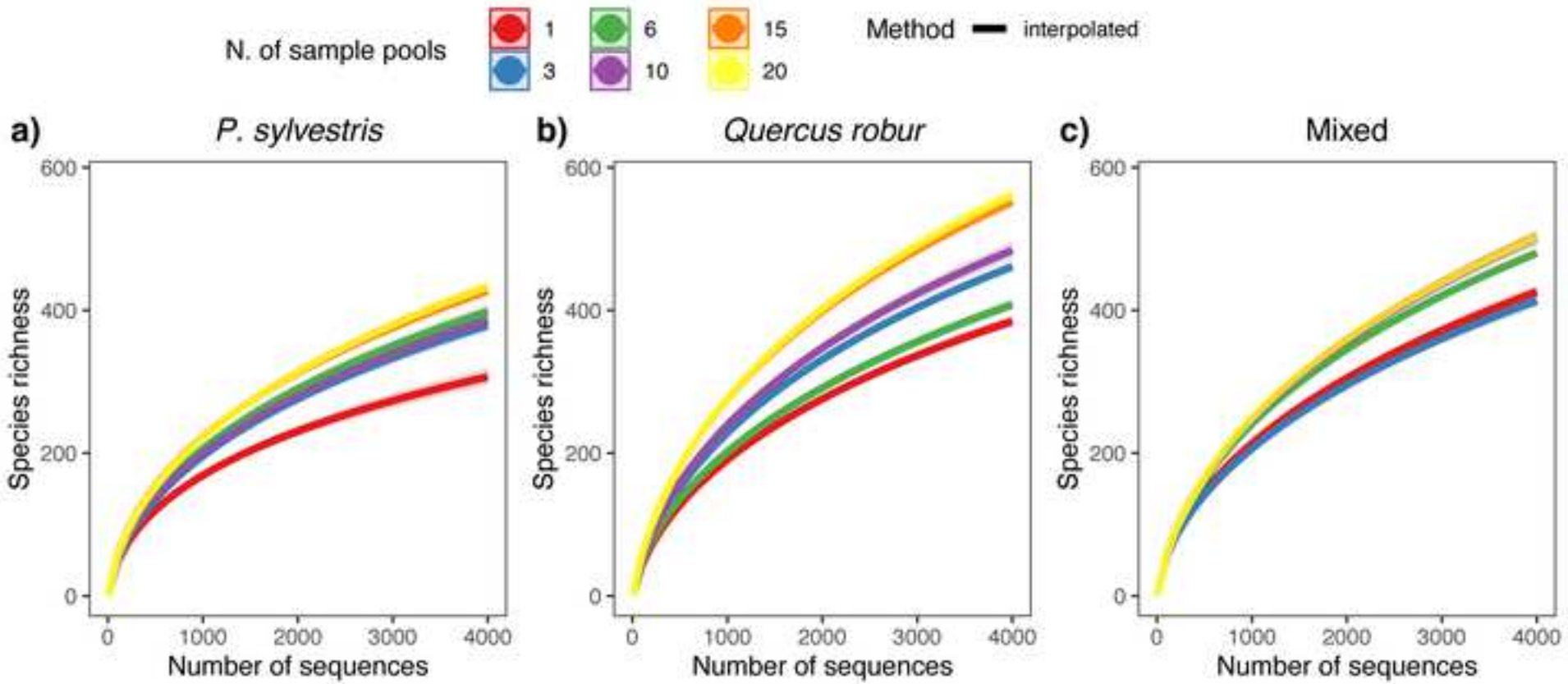
717 **Fig. S2**. Non-metric multidimensional scaling (NMDS) representing compositional

718 differences in the overall communities between number of soil sample pools (1, 3, 6, 10,

719 15, 20) in a) *P. sylvestris* b) *Q. robur* c) mixed forests.

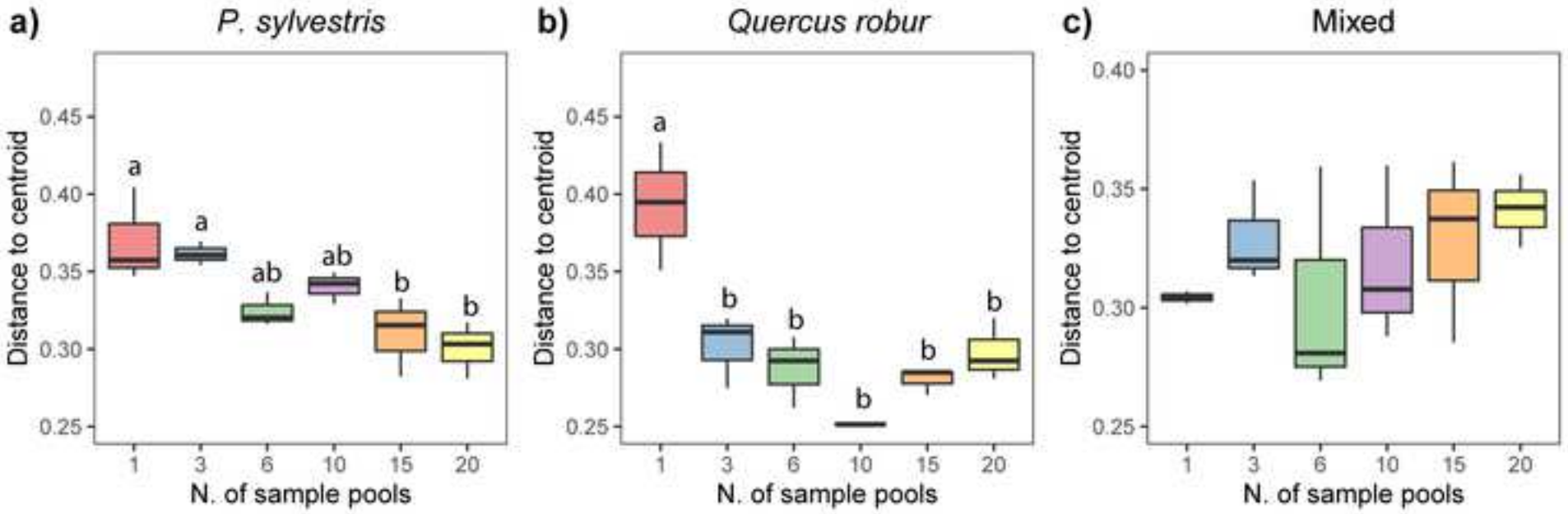
720

Fig S1



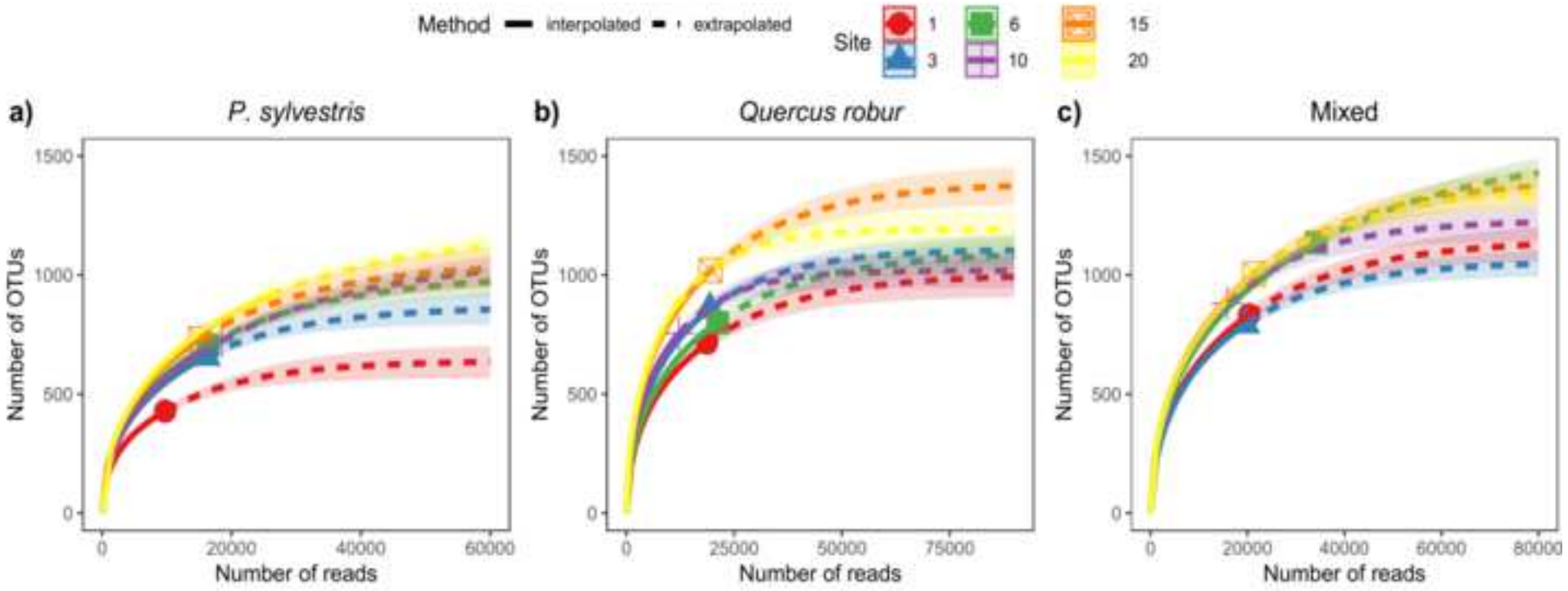
Figure(s)

Click here to access/download;Figure(s);fig2_tukey.tif



Figure(s)

Click here to access/download;Figure(s);Fig.1_fin.tiff



Figure(s)

Click here to access/download;Figure(s);Fig.S2New.tif