Author(s): Abrego, Nerea; Norros, Veera; Halme, Panu; Somervuo, Panu; Ali-Kovero, Heini; Ovaskainen, Otso

Title: Give me a sample of air and I will tell which species are found from your region : molecular identification of fungi from airborne spore samples

Year: 2018

Version:

Please cite the original version:

Abrego, N., Norros, V., Halme, P., Somervuo, P., Ali-Kovero, H., & Ovaskainen, O. (2018). Give me a sample of air and I will tell which species are found from your region : molecular identification of fungi from airborne spore samples. Molecular Ecology Resources, 18(3), 511-524. https://doi.org/10.1111/1755-0998.12755

DR. NEREA ABREGO (Orcid ID : 0000-0001-6347-6127)

Article type : Resource Article

Give me a sample of air and I will tell which species are found from your region – molecular identification of fungi from airborne spore samples

Nerea Abrego[1], Veera Norros[2,3], Panu Halme[4], Panu Somervuo[2], Heini Ali-Kovero[2], and Otso Ovaskainen[2,5]

[1]Department of Agricultural Sciences, University of Helsinki. P.O. Box 27, FI-00014, Finland.

[2]Department of Biosciences, University of Helsinki. P.O. Box 65, FI-00014, Finland.

[3]Marine Research Centre, Finnish Environment Institute, P.O. Box 140, FI-00251 Helsinki, Finland.

[4]Department of Biological and Environmental Science, University of Jyväskylä, P.O. Box 35, FI-40014 University of Jyväskylä, Finland.

[5]Centre for Biodiversity Dynamics, Department of Biology, Norwegian University of Science and Technology. N-7491 Trondheim, Norway.

Email addresses: nerea.abrego@helsinki.fi (NA), veera.norros@ymparisto.fi (VN), panu.halme@jyu.fi (PH), panu.somervuo@helsinki.fi (PS), heini.ali-kovero@helsinki.fi (HA-K) and otso.ovaskainen@helsinki.fi (OO).

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1111/1755-0998.12755

Corresponding author: NA. Mailing address: Department of Agricultural Sciences, University of Helsinki. P.O. Box 27, FI-00014, Finland. Email: nerea.abrego@helsinki.fi. Telephone: +358 400 151684.

Running title: Aerial fungal sampling

Abstract

Fungi are a megadiverse group of organisms, they play major roles in ecosystem functioning, and are important for human health, food production, and nature conservation. Our knowledge on fungal diversity and fungal ecology is however still very limited, in part because surveying and identifying fungi is time demanding and requires expert knowledge. We present a method that allows anyone to generate a list of fungal species likely to occur in a region of interest, with minimal effort and without requiring taxonomical expertise. The method consists of using a cyclone sampler to acquire fungal spores directly from the air to an Eppendorf tube, and applying DNA barcoding with probabilistic species identification to generate a list of species from the sample. We tested the feasibility of the method by acquiring replicate air samples from different geographical regions within Finland. Our results show that air sampling is adequate for regional-level surveys, with samples collected >100 km apart varying but samples collected <10 km apart not varying in their species composition. The data show marked phenology, and thus that obtaining a representative species list requires aerial sampling that covers the entire fruiting season. In sum, aerial sampling combined with probabilistic molecular species identification offers a highly effective method for generating a species list of air dispersing fungi. The method presented here has the potential to revolutionize fungal surveys, as it provides a highly cost-efficient

way to include fungi as a part of large-scale biodiversity assessments and monitoring programs.

Key-words: Atmospheric diversity, Fungal diversity, Fungal sampling method, Fungal survey, Molecular identification, Spore.

Introduction

There is increasing evidence that the atmosphere is a highly biodiverse environment (Womack, Bohannan, & Green, 2010; Gandolfi et al., 2013). The dispersal propagules of most sessile organisms such as plants, bacteria and fungi depend upon air for moving, and some microorganisms can use suspended air particles as their primary habitat (Côté et al., 2008; Klein et al., 2016). Recent studies have shown that the atmospheric biodiversity is especially high for fungal and bacterial microorganisms and that the aerial communities are not geographically or temporally homogeneous (Fierer et al., 2007; Fröhlich-Nowoisky et al., 2009; Bowers et al., 2010; Barberán et al., 2015). Yet, compared to aquatic and terrestrial environments, little is known about the composition and geographical variation of such biodiversity. Importantly, the knowledge on the potential of the atmospheric biodiversity in reflecting the biodiversity from the other environments is still very limited.

Most research on atmospheric fungi has been motivated by the need of detecting species affecting human health. Air sampling is used routinely in hospitals and human-populated areas to inform the public about the presence of allergenic fungal spores (e.g. Eames et al., 2009; Eduard et al., 2012; Polymenakou, 2012). Therefore, commonly used sampling

equipment and typically culture-dependent survey methods have been designed specifically to find pathogenic fungi (see Weissfeld et al., 2013). More recently, cultivation-independent survey methods and DNA identification techniques have been developed, giving the opportunity to record the presence of also many other fungal groups (e.g. Peccia & Hernandez, 2006; Hoisington et al., 2014; Be et al., 2015). Most widely used cultivation-independent spore sampling techniques do not capture spores directly from the air but rely on the spore deposition. Such sampling techniques include filter traps (e.g. Kivlin et al., 2014; Be et al., 2015; Castaño et al., 2017), collecting surfaces such as petri dishes (e.g. Adams et al., 2013), or filtering spores from rain water (e.g. Peay et al., 2012; Peay & Bruns, 2014). More recently developed "air samplers" (or "cyclone samplers") sample spores directly from air and provide thus a more direct and efficient possibility to explore atmospheric biodiversity (West & Kimber, 2015). Spore deposition increases greatly with spore size (Norros et al., 2014), and consequently air samplers are have a higher collection capability than deposition-based sampling techniques, in particular for small sized spores (West & Kimber, 2015). Thus far, the potential of air samplers combined with molecular species identification as a method to screen fungal communities has however remained only partially evaluated, the applications having been restricted to a single site (e.g. Fierer et al., 2007) or to a pre-defined set of focal species (e.g. Nicolaisen et al., 2017).

Fungi are difficult to survey, because most of them produce microscopic vegetative or reproductive structures (Cannon & Kirk, 2007) or because the reproductive structures (i.e. fruit-bodies) are often ephemeral (Halme & Kotiaho, 2012). Due to the difficulty of detecting a large proportion of fungal species by fruit-body based surveys, environmental DNA (eDNA) surveys are becoming increasingly used (Lindahl et al., 2013). Sampling fungi with the help of eDNA is highly advantageous because one can find many more species than from fruit-body surveys with less effort and expertise (Allmér et al., 2006; Ovaskainen et al.,

2013). However, eDNA surveys show that different substrates (e.g. soil, wood, litter) hold very different fungal communities, and that fungal communities have high spatial variation even at very small spatial scales of same substrates (Kubartová et al., 2012; Hazard et al., 2012). Thus obtaining a complete species list from a large area using substrate-specific eDNA samples may require an exhaustive sampling effort.

Assuming that the majority of the dispersal propagules of fungi are not dispersed too far from the substrates where they have been released, atmospheric fungal communities might be representative of their terrestrial diversity. If the fungal diversity in the air is representative of the terrestrial fungal diversity, this might greatly reduce the effort required to obtain fungal data and ultimately revolutionize our knowledge on global fungal diversity and biogeography. Yet, there are a number of issues which might compromise the usefulness of atmospheric fungal diversity as surrogate of the terrestrial diversity. First, it is not clear which spatial and temporal scales of the terrestrial fungal diversity the atmospheric communities represent. Fungal spores can be rapidly transported long distances from their source by wind (Rieux et al., 2014; Norros et al., 2014), which might cause a fast homogenization of the local atmospheric fungal communities. However, bacterial aerial communities have been shown to resemble closely their terrestrial source communities (Bowers et al., 2010), so this could be the case for fungi as well. The recent study by Womack et al (2015) suggests that aerial fungal communities may represent community variation at biome level, but the spatial resolution of such analogy is unknown. Second, the occurrence of fungi in the air greatly depends on the reproductive phenology of the species, and thus on the season of the year or time of the day (Kramer, 1982; Elbert et al., 2007; Pashley et al., 2012). Finally, although most fungi use air for dispersing, some species use primarily or entirely animal vectors (e.g.

hypogeous fungi) and thus are unlikely to be detected in the air (e.g. Lilleskov & Bruns, 2005; Nuñez et al., 2013).

The aim of this study is to test whether air samplers combined with DNA barcoding can be used to generate a list of air-dispersing fungal species likely to occur in a region of interest, with minimal sampling effort and without requiring taxonomical expertise. The method consists of using a cyclone sampler which acquires fungal spores directly from the air to an Eppendorf tube, and applying DNA barcoding with probabilistic species identification to generate a list of species from the sample. We tested the feasibility of the method by acquiring air samples from different natural environments in Finland. We assessed the effectiveness of the method by comparing the fungal composition identified from the air to previous inventories based on substrate-specific fruit-body and eDNA surveys. To determine the optimal season and sampling time required to obtain a representative species list, we also investigated the seasonal and diurnal variation of fungal communities in the air.

Materials and Methods

Study sites and collection of empirical data

We acquired spore samples from four study sites representing different geographical regions within Finland (Fig. 1A). Kuusimäki (called henceforth Site 1) is a 108 ha conservation site located in the municipality of Muurame in Central Finland (N 62.22º, E 25.48º). Hoikanpuro (called henceforth Site 2) is a 4 ha forest site located in the municipality of Pihtipudas in northern Central Finland (N 63.37º, E 25.55º). Rörstrand (called henceforth Site 3) is a 85 ha conservation site located in the municipality of Uusimaa in southern Finland (N 60.46º, E 25.20º). Bengtskär (called henceforth Site 4) is a small rocky island (2 ha) in the Baltic Sea,

located in the outer archipelago of the southwest coast of Finland (N 59.79º, E 22.49º). Sites 1 and 3 consist of large natural spruce-dominated forests (Norway spruce, *Picea abies*) with high amount of dead wood. Site 2 is also a natural spruce-dominated forest, but is a small area surrounded by clear-cuts. In Site 1 and Site 4 birch (*Betula* spp.) is the second most abundant tree species, followed by Scots pine (*Pinus sylvestris*) and aspen (*Populus tremula),* whereas in Site 2 only some birches are present in addition to spruce. Site 4 is a small and highly isolated island with no trees, located 25 km from the mainland and 9 km from the nearest island with forest. The distance to the nearest site is on average 165 km (±SD 21 km), and the maximum distance between sites is 433 km.

We used a cyclone sampler (Burkard Cyclone Sampler for Field Operation, Burkard Manufacturing Co Ltd; Emberlin & Baboonian 1995) to acquire spore samples from each study site (Fig. 1A). The cyclone sampler generates a circular airflow which captures particles with diameter at least 0.5 µm and deposits them directly into an Eppendorf tube. As summarized below and detailed in the Table S1, we acquired multiple samples from each study site to examine spatial, temporal and methodological variation in the samples.

In Site 1 we focused on examining the seasonal variation of aerial fungal communities. For this, we obtained samples of approximately one week duration (4-14 d) during the entire fruiting period from early spring (week 20, mid-May) to late autumn (week 42, mid-October). The snow cover had melted from this study site 2-3 weeks before the sampling started, and the first snowfalls happened already during the last days of sampling in the autumn. Additionally, in each month (May-October), we collected two samples of shorter duration between the one-week duration samples, one during the day (ca. 8-16 o'clock) and one

spanning over one night (ca. 16-8 o'clock). The day/night sampling coincided with monthly fruit-body inventories of wood-inhabiting fungi conducted at the study site (see Abrego et al., 2016). Altogether, we collected 29 samples in Site 1.

In Site 2 our focus was in examining the small-scale spatial and temporal variation of aerial fungal communities. To examine this, we acquired samples within and at different distances from Site 2. Inside the forest patch, we took eight spore samples at ground level and eight at the canopy level (at 10 m height). Outside the forest patch, we took spore samples at distances of 18 m, 50 m, 135 m, 368 m, 1000 m and > 5 km. To collect the samples at 18-1000 m distances, 8 sampling transects were established, corresponding to the cardinal and half-cardinal directions (Appendix 1, Fig. S7). The sampling was then always conducted downwind from the forest patch, i.e. at the transect that was closest to the wind direction at the start of the sampling period. The > 5 km samples were collected at different locations chosen randomly from the 5-20 km ring surrounding the study site, regardless of the wind direction. For each of the sampling distances, we acquired seven to nine 4-12-hour samples during a 16-day measuring period (21.9.-6.10.2009), each collected either during the day (defined as 6-18 o'clock) or during the night (18-6 o'clock). The exact sampling times and locations are provided in Table S1. 35 of the samples were collected on two separate occasions (e.g. 5 hours on day 1 at 50 m NE and 5 hours on day 7 at 50 m E, corresponding to the wind direction at each sampling occasion). This was done to reduce confounding effects of the precise sampling location and time on the samples collected at a given distance, while limiting the total number of samples. In addition to the distance gradient samples, we acquired 32 methodological samples at Site 2 aimed at examining how long a sampling time is required for obtaining a representative sample of the aerial fungal diversity. These samples were collected at the focal patch, with sampling time being either 1, 5, 15, 60 or 240 minutes,

with 10, 8, 6, 4 or 4 replicates for each of these, respectively. Altogether we collected 97 samples in Site 2.

In Site 3 and Site 4 our aim was to obtain a representative sample of the local species community, without any more specific study questions. To do so, we acquired a few replicate samples with some variation in sampling times and locations. We obtained 2 samples in Site 3, and 6 samples in Site 4.

To be able to compare the aerial fungal communities to the local communities observable as fruit-bodies, we compiled a list of wood-inhabiting fungi recorded as fruit-bodies for Sites 1-3. For Site 1 and Site 3 we supplemented information from previous fruit-body inventories (e.g. Ovaskainen et al., 2013; Abrego et al., 2016) by observations made during this study, whereas for Site 2 prior information was not available and thus we compiled the species list solely based on our surveys during this study. The target species included Basidiomycota species with polyporoid and prominent corticioid and hydnoid fruit-bodies. In the fruit-body inventory in Site 2, all dead wood objects at the focal patch with a diameter of 5 cm or greater were inspected for fruit-bodies of target species. As Site 4 has no trees and thus no dead wood, we assumed that in this study site wood-inhabiting fungi were absent. To further facilitate the comparison, we classified the molecularly identified species to wood-inhabiting fungi and to other fungi. These included not only the species recorded in fruit-body surveys but also agarics, large ascomycetes, heterobasidiomycetes, ramarioid fungi and corticioid fungi known to be wood-inhabiting.

DNA extraction, amplification and sequencing

As the first step in the laboratory, coarse material deposited in the Eppendorf tubes was removed using sterile tweezers and sample volumes were normalized to 300 µl with ultrapure Milli-Q water (EMD Millipore, MA USA). No attempt was made to remove finer material such as pollen. DNA was then extracted from the remaining sample following the Chelex procedure by incubating 1,5 h at 56°C with 75 µl of 24% Chelex® 100 resin solution (Bio-Rad, Hercules, CA USA) and 20 µl of 20 mg/ml Proteinase K (ThermoFisher Scientific, Waltham, MA USA).

The internal transcribed spacer (ITS) is one of the most widely sequenced DNA regions in fungi and has been selected as the universal genetic barcode for fungi (Schoch et al., 2012). Although for some particular species groups other regions are known to have superior resolution to classify species, in this study we used the ITS region because of the high availability of reference sequence data (Schoch et al., 2012). We amplified the internal transcribed spacer region 2 (ITS2) by using the primers ITS3 and ITS4 (White et al., 1990).

PCR was performed with Phusion HF polymerase (ThermoFisher Scientific, Waltham, MA) in two stages: first with untagged primers and then with primers including a sample specific identification tag and adaptors (adaptor A in the 5' end of ITS4, and B in the 5' end of ITS3). Two stages were needed as pilot runs showed that direct amplification with tagged primers resulted in drastic biases in the relative proportions of different species. Extraction blanks and PCR negative controls without DNA template were included.

The first stage of PCR reactions was run in a total volume of 20 µl for 25-28 cycles with an annealing temperature of 55°C and a primer concentration of 0,5 pmol/µl. Some DNA samples needed diluting up to 1:10 and the number of cycles was adjusted on a sample-by-sample basis to yield weak to moderately strong bands on agarose gel with approximately the same strength for all samples. 5 µl of each PCR product was run on 1% agarose gel and only those that were weak enough were chosen for further processing. The products were purified with 0,9x volume of AMPure magnetic beads (Beckman Coulter, Brea, CA USA) and diluted based on gel band intensity in a range from 1:3 to 1:12 for the second PCR. Amplification with tags and adaptors was performed in 10-15 cycles, followed by an agarose gel run and an AMPure cleaning as for the first PCR. The amplified DNA samples were further processed in the DNA Sequencing and Genomics lab of the Institute of Biotechnology at the University of Helsinki, and sequenced on the GS FLX Titanium platform (454 Life Sciences, Branford, CT USA).

The procedure described above was applied to all samples except the two samples from Site 3, which were collected at an earlier stage in the method development. For these samples, only one PCR of 25 cycles was performed with primers ITS1F (Gardes & Bruns, 1993) including a sample specific tag and the adaptor A, and ITS2 (White et al., 1990) containing the adaptor B. Also another platform of sequencing, 454 GS FLX, was used for the samples at the same institute.

Molecular species identification

Reliable molecular identification of fungi is challenging because of several reasons. First, the sequence to be classified may belong to a species that has not been described and that is thus unknown to science. Second, the sequence to be classified may belong to a species that has

been described but for which no reference sequences are available. Third, even if the sequence to be classified belongs to a species for which reference sequences are available, the resolution of the marker (here ITS2) may not be sufficient for species-level identification. Additionally, taxonomy and reference databases contain errors that generate false classifications. For these reasons, we performed a probabilistic taxonomic placement of the ITS reads using the PROTAX software (Somervuo et al., 2016), parameterized by the Index Fungorum taxonomy database and the UNITE reference sequence database (Somervuo et al., 2017). For each environmental sequence, PROTAX decomposes the total probability of one among all possible outcomes generated by the taxonomy database, including species present in the taxonomy database (whether or not they have reference sequences), including the possibility of unknown species (or higher taxonomic levels) that are not part of the taxonomy database, and accounting for the possibility of mislabelling of the reference sequences.

We extracted from these taxonomic placements all species level identifications that achieved at least 50% probability of correct classification (which Somervuo et al. 2017 called "plausible identifications"). The reason for this choice is that 50% can be considered as the most natural threshold to disentangle presences from absences: If a sequence is classified to a particular species with probability greater than 50%, it means that the particular species is the most likely true classification, whereas if the probability is lower than 50%, then some other species is the most likely true classification. We note that 50% probability threshold for true classification is a more strict criteria than using e.g. a 97% sequence similarity threshold, as the PROTAX approach accounts not only for sequence similarity, but also to uncertainties generated by e.g. incompleteness of reference databases (Somervuo et al. 2016, 2017). To test the robustness of the results to these choices, we repeated all analyses for species-level identifications that were assigned at least 90% probability (which Somervuo et al. 2017 called

"reliable identifications"), as well as for genus-level identifications that exceeded either 50% or 90% probability thresholds.

Statistical analyses

The starting point of all statistical analyses was a taxonomical units times samples matrix, where we counted the number of sequences assigned to each taxonomical unit (species or genus) on each sample, based on whether each sequence was assigned to that taxonomical unit with probability exceeding the chosen threshold (50% or 90%). We performed species-level analyses in two ways: either including all fungi, or including only wood-inhabiting fungi. The motivation for performing the analyses separately for wood-inhabiting fungi is that Sites 1-3 represented forests that were abundant in dead wood, that the fruit-body samples were collected solely from dead wood pieces, and that for Site 3 we also had eDNA data originating from dead wood samples.

To illustrate the taxonomical fungal composition in the study, we constructed Krona wheels (Ondov, Bergman, & Phillippy, 2011). As a measure of relative species abundance, we used (here and in all analyses below) relative sequence counts (number of sequences divided by sequencing depth). We raised the relative sequence counts to the power of 0.25 to downscale the variation among sequence reads, some of which is not biological but technical variation, caused e.g. by amplification biases among the species that are hard to control for (Baldrian et al., 2013). To assess how the samples collected from different locations or at different times differ from each other, we analysed the relative species abundance data by non-metric multidimensional scaling analysis (NMDS) with Bray-Curtis dissimilarity distance. In this analysis we excluded those samples with less than one hour of sampling duration. In NMDS

ordination space, distance among samples represents community dissimilarity, shorter distances indicating more similar communities. We examined if the samples collected from the four study sites separated from each other, and explored seasonal variation in the ordination space by adding a surface indicating the sampling week.

We computed species richness for each sample as the number of distinct taxonomical units identified. To examine variation in observed species richness among the study sites and its dependency on the sampling effort, we fitted a Poisson regression where the explanatory variables were the study site (categorical variable with four levels), the log-transformed duration of the sampling (continuous covariate used to control for sampling effort), and the log-transformed number of sequences (continuous covariate used to control for sequencing depth). Controlling for sequence depth allowed us to account for the possible bias due to the differences in PCR steps among sampling sites (Smith & Peay, 2014).

To evaluate whether the method gives signal on seasonal variation, we restricted the analysis to samples from Site 1. We modelled observed species richness with Poisson regression, with log-transformed sequencing depth, week and log-transformed duration of the sampling as continuous explanatory variables. We then plotted the expected species richness (corrected for sampling duration and sequencing depth) as a function of week. To visually examine seasonal variation in community composition, we performed NMDS analysis to the relative abundance data, to which we fitted a surface indicating the sampling week, and plotted the proportions of sequences belonging to major taxonomic groups (orders) over time.

To evaluate whether the method gives signal on small-scale spatial or temporal variation, we restricted the analysis to the samples from Site 2 (excluding the short-term methodological samples). We performed NMDS analysis to the relative abundance data on these samples to examine if samples taken at different distances from the forest patch, or during day versus night, separated in the ordination space. We modelled observed species richness with Poisson regression (including the short-term methodological samples), with log-transformed duration of the sampling, log-transformed distance to the forest patch, and log-transformed sequencing depth as explanatory variables. To assess how long sampling needs to be carried out to obtain a representative sample, we predicted the expected species richness (corrected for sequencing depth and distance) as a function of sampling duration.

To compare the results from the aerial sampling to results obtained by fruit-body inventories, we ranked the genera based on their mean (over samples) relative abundance, and checked whether genera belonging to the wood-inhabiting guild (all polypores and certain prominent corticioid and hydnoid species) had been recorded as fruit-bodies in Sites 1-3.

To examine how many times the sampling should be carried out over the season to obtain a representative sample of the species diversity, we used the weekly samples from Site 1 to construct species accumulation curves as a function of number of samples. We constructed four kinds of species accumulation curves, counting either (1) all species, (2) species occurring in at least two samples, (3) species observed with at least 10 sequence reads per sample, and (4) species with at least 100 sequence reads per sample. In the species accumulation curves, we accounted for variation in sequencing depth by resampling the data to the smallest number of sequences per sample.

Finally, to compare the results from the aerial sampling to results obtained by substrate-specific eDNA sampling, we re-analysed sequence data published by Mäkipää et al (2017). These data were collected in soil and dead wood substrates in Site 3. We conducted three kinds of analyses. First, we constructed Venn diagrams to pinpoint the numbers of taxonomical units shared among the air samples obtained in the present study and the wood and soil samples from Mäkipää et al (2017). Second, we applied a NMDS ordination to pooled relative abundance data to examine if and how the air, soil and wood samples separate from each other. Third, we constructed species accumulation curves also for the wood and soil samples. Unfortunately, wood and soil samples were not available for Site 1 for which we constructed the species accumulation curves based on the aerial samples, and thus the species accumulation curves are not fully comparable. To make the wood and soil accumulation curves of Site 3 as comparable as possible to the air accumulation curves of Site 1, we kept the sampling effort identical, thus having the same total amount of samples and resampling the sequences to the smallest amount of sequences per spore sample.

All analyses were carried out in the R environment (R Development Core Team, 2016). The NMDS analyses and species accumulation curves were carried out using the vegan package (Oksanen et al., 2015). In the NMDS analyses, the surfaces corresponding to continuous environmental variables were fitted using the 'ordisurf' function from the vegan package, and the Poisson regression models were fitted using the 'glm' basic R function.

Results

We report here the results for species-level identifications with reliability threshold 50%. The results based on data for genus-level identifications and analyses performed for 90% reliability thresholds were qualitatively similar and are reported in Appendix 1.

Altogether, we recorded 1021 distinct species in the 134 samples, out of which 399 were classified as wood-inhabiting fungi (Fig. 2, Appendices 2 and 3). The numbers of species from different phyla were the following: 572 Basidiomycota, 431 Ascomycota, 17 Zygomycota and 1 Chytridiomycota (Table S2). The number of species (species richness) varied among the samples from 14 to 206, with mean of 69 species. Wood-inhabiting fungal richness varied from 3 to 117, with mean of 33 species. In the Poisson regression model (Table S3), species richness was expectedly found to increase with sequencing depth (estimate = 0.33; $p < 0.001$) and sampling duration (estimate = 0.14; $p < 0.001$). It also varied among sites ($p < 0.001$), Site 1 holding significantly more species per sample than the rest of sites (119 species, out of which 61 were wood inhabiting). The species composition of the study sites separated well in the ordination space (Fig. 1B).

The ordination analysis of the samples from Site 1 showed a strong signal of seasonal variation in community composition, the early and late season samples separating very clearly (Fig. 3A, $p < 0.001$). Species richness decreased through the season (estimate = -0.01, $p = 0.0015$), a trend which was explained by the decrease in the number of Ascomycota species (Fig. 3B, Table S4, estimate = -0.02, $p < 0.001$). Polyporales and Agaricales were the dominating Basidiomycota orders throughout the season, whereas Helotiales and Lecanorales were the dominating Ascomycota orders (Fig. 3C).

The ordination analysis of the samples from Site 2 (excluding the short-term methodological samples) did not separate well the different sampling distances, but showed a clear signal of day versus night (Fig. 4A). The species richness (corrected for sequencing depth and duration) did not change with increasing distance from the focal forest centre (Table S5, Fig. S7, estimate = 0.0125, $p = 0.07$). The species richness decreased during the night compared to day in the analyses including all species (Fig. 4B, Table S5, estimate = -0.1052, $p < 0.001$), whereas an opposite trend was observed for wood-inhabiting species (Fig. 4C, Table S5, estimate = 0.0561, $p = 0.175$). The expected species richness (corrected for sequencing depth and distance) increased with increasing sampling time (Fig. 4D, Table S5, estimate = 0.7316, $p < 0.001$), with one minute sampling time yielding on average 24 species and one hour sampling yielding on average 46 species.

Among the most abundant genera identified with the highest identification probability threshold (0.9), 7 were classified as wood-inhabiting fungi in each of the Sites 1-3 (Table 1). Out of these, all were known from fruit-body inventories in Site 1 and Site 3, and five out of seven in Site 2. In terms of species richness, in the pooled data over Sites 1-3, in total 501 species were known as fruit-bodies and 1003 (out of which 684 were well-identified to the species level) from the aerial DNA data, out of which 177 species were recorded in both types of data. At the site level, we observed in the fruit-body inventories 441 species in Site 1, 81 species in Site 2 and 182 species in Site 3, whereas the numbers of species observed from the aerial DNA data were 709 (out of which 467 were identified to the species level) in Site 1, 572 (362) in Site 2, and 121 (75) in Site 3. We note that the great variation among the sites numbers represents variation in sampling intensity rather than biological variation for both the fruit-body and the DNA data. Site 1 was the target of intensive fruit-body surveys over many years (Abrego et al., 2016), and our aerial sampling there covers the entire fruiting

season. In contrast, the fruit-body surveys for Site 2 were conducted only for the purpose of this study, and the aerial DNA data for Site 3 included only two samples.

In the data subsampled to the smallest sequencing depth (2487 sequences) among the samples from Site 1, the species accumulation curve did not asymptote after acquiring 29 samples (Fig. 5A, Fig. S9). The species accumulation curves did not asymptote for soil samples from Site 3 or wood eDNA samples from Site 1 and Site 3 neither, and in all three cases the number of species was much lower than in the aerial eDNA samples (Fig. 5BC, Fig. S9). For all sampling substrates, the proportion of rare species was very high, as illustrated by the differences between the four types of species accumulation curves (Fig. 5ABC). The number of species found exclusively in air was tenfold the number of species found exclusively on wood and fivefold the number of species found exclusively on soil (Fig. 5D, Fig. S10).

Discussion

The method introduced in this paper has potentially game-changing implications for large-scale studies of fungal communities. Most importantly, our results show that aerial sampling combined with probabilistic molecular species identification offers a highly effective method for generating a species list of air dispersing fungi. Fungal biogeography is an emerging field with growing interest, as shown by the very recent pioneering studies (Sato et al., 2012; Tedersoo et al., 2014; Davison et al., 2015). Partly due to the high volume of samples that would be required for analysing all kinds of substrates globally, biogeographical studies in fungi have thus far been mainly focused on soil communities (but see Fröhlich-Nowoisky et al., 2012; Barberán et al., 2015). We argue that the methodological pipeline described in this

paper has great potential to gain data on fungal distributions from regional to global scales, at volumes as yet unseen.

Because of the cosmopolitan distributions of many fungal species and small size of the dispersing propagules, it has long been considered that biogeographical contingencies such as dispersal limitation do not determine fungal distributions (e.g. Moncalvo & Buchanan, 2008; Sato et al., 2012). However, recent studies have demonstrated that fungi can be dispersal limited at small spatial scales (e.g. Galante, Horton, & Swaney, 2011; Norros et al., 2012; Peay et al., 2012; Adams et al., 2013). Our results support dispersal limitation at intermediate to large scales: aerial fungal composition was clearly different between sites (> 100 km apart) but not within sites (< 10 km apart). These results concur with the US-wide study carried out by Barberán et al (2015), showing that the airborne fungal communities associated with dust accumulated on house surfaces are geographically structured among distances of hundreds of km. Conversely, our results also support the view that some particular species are able to travel long distances. In particular, supporting the findings of Peay et al. (2012), our results show that *Suillus* spores can travel long distances, since the spores of this ectomycorrhizal genus were abundant in Bengtskär, in spite of the site being a treeless isolated island.

Compared to terrestrial eDNA based surveys and to morphological identification of fungi by expert taxonomists, aerial sampling of spores is likely to provide a very cost-efficient method for profiling fungal communities comprehensively at large scales. As taxonomists are usually specialized in a single or few taxonomical groups, and/or spatial areas, comprehensive global sampling of all fungal groups based on fruit-body identification would require a huge effort. Furthermore, most fungi produce ephemeral fruit-bodies or fruit rarely and their detection

requires multiple visits, and/or intensive sampling in the region of interest (Halme & Kotiaho, 2012; van der Linde et al., 2012; Abrego et al., 2016). Substrate-specific eDNA surveys partly overcome detectability issues, because fungal species can be identified either as reproductive structures (fruit-bodies) or in vegetative stage (mycelia). Yet, as the results from studies based on substrate-specific eDNA surveys show (e.g. Kubartová et al., 2012; Hazard et al., 2012), fungal communities are highly variable at small scales and between substrate types. Thus, obtaining species lists from large areas such as entire forests using such techniques also requires high sampling effort (Runnel, Tamm, & Lõhmus, 2015). As we have demonstrated here, eDNA surveys based on aerial samples provide an attractive alternative for characterizing fungal composition simultaneously for fungi growing on many kinds of substrates. In our results, a high number of airborne species were not found from wood or soil, implying that the air contains also spores of species from other guilds, such as litter saprotrophs and lichenized fungi. We note, however, that by spore sampling we did not detect all species found in soil and wood, supporting the fact that not all fungi use primarily air for dispersing.

As expected, the longer the sampling time or the more samples we obtain along the season, the more comprehensive list of fungal species we get. The species accumulation curve did not flatten for the samples taken in our study, which is a common result when assessing fungal species richness either by eDNA (e.g. O'Brien et al., 2005) or fruit-body surveys (e.g. Unterseher et al., 2008). Thus, we recommend that comprehensive species surveys should combine information e.g. from 24 hr samples acquired weekly during the entire fruiting season.

In line with some previous studies (Fröhlich-Nowoisky et al., 2009; Pashley et al., 2012; Womack et al., 2015), we found that aerial fungal communities are comprised of more Basidiomycota than Ascomycota species. Studies showing a higher portion of Ascomycota have been carried out in urban environments (e.g. Fierer et al., 2007; Lee et al., 2010; Oh et al., 2014), where the elevated pollution and lack of plant debris may favour some Ascomycota species (Sterflinger & Prillinger, 2001), or during dry periods (e.g. Fierer et al., 2007; Lee et al., 2010), which conditions may favour Ascomycota dispersal (Elbert et al., 2007). Consistently with this, we found decreased Ascomycota diversity in autumn weeks, when precipitation is higher than in spring. Our result is also in line with an earlier study showing that in the studied area, Ascomycota fruit-body production peaks in spring (Purhonen et al., 2017). As suggested by Elbert et al (2007), we found that samples taken during the night hold different communities with generally less species than those taken during the day hours. The higher diversity of wood-inhabiting fungi in the night samples is consistent with observations of higher spore release rates during the night in saprotrophic basidiomycetes (Kramer, 1982).

We applied probabilistic species identification (Somervuo et al., 2016; 2017) to quantify the reliability of the generated species list. DNA sequences contain errors, the natural sequence variation within a species sometimes overlaps with that among species, and reference databases are incomplete and can have false annotations (e.g. Carlsen et al., 2012; Nilsson et al., 2012; Nguyen et al., 2015). This is especially problematic when trying to identify fungi from environmental samples, as a high portion of species are still to be sequenced (Taylor et al., 2014) or even to be described, and misidentification and mislabelling characterize the available databases (Nilsson et al., 2012). In frequently used software such as QIIME (Caporaso et al., 2010) and mothur (Schloss et al., 2009), taxonomic assignment is based on

e.g. the best BLAST hit or the naïve Bayes classifier of Wang et al. (2007), which methods do not account for the possibility of missing reference sequences or missing taxa. For this reason, we have applied here PROTAX, which provides calibrated probabilities of taxon membership also for the case of incomplete taxonomy and reference databases (Somervuo et al., 2017), making it possible to assess the reliability of the identifications.

The pipeline proposed in this paper offers exciting possibilities for large-scale fungal studies, such as generating comprehensive species distribution maps for predicting how species respond to e.g. climate warming, habitat loss or pollution. Other exciting applications include the early detection of pathogenic species causing important economic losses in agriculture, forestry or house construction already upon their arrival as spores. In our study for example, we detected *Heterobasidion annosum* and *Claviceps purpurea*, out of which the first is the most important forest pathogen of conifer trees in northern Europe, and the latter species is a common seed pathogen of cereal plants. We believe that the method presented here can also be used for gaining basic knowledge of the dispersal capabilities of fungi. For instance, one could design controlled experiments in which sporulation sources are generated and spores are captured at different distances.

Acknowledgements

We are grateful to the three anonymous reviewers for their constructive comments. We thank Lars Paulin and Petri Auvinen for performing the 454-sequencing. We are also grateful to Anni Rintoo, Katja Juutilainen, Jenna Purhonen, Asta Vaso, Lotta Sundström, Henrik Murdoch, Juho Kytömäki, Noora Vartija, Matti Kulju, Jonna Kuha, Jenni Nordén, Dmitry Schigel, Sami Haapanala as well as Paula and Per Wilson for field assistance and other help

with the data collection, and to Reijo Penttilä for providing the fruit-body species list of Rörstrand. The research was funded by the Academy of Finland (grants 1273253 and 284601 to OO), the Research Council of Norway (CoE grant 223257), LUOVA graduate school and Ella and Georg Ehrnrooth Foundation (grants to VN) and Finnish Ministry of Environment (PUTTE grant to PH).

References

Abrego, N., Halme, P., Purhonen, J., & Ovaskainen, O. (2016). Fruit body based inventories in wood-inhabiting fungi: Should we replicate in space or time? *Fungal Ecology, 20*, 225-232.

Adams, R. I., Miletto, M., Taylor, J. W., & Bruns, T. D. (2013). Dispersal in microbes: Fungi in indoor air are dominated by outdoor air and show dispersal limitation at short distances. *The Isme Journal, 7*, 1262.

Allmér, J., Vasiliauskas, R., Ihrmark, K., Stenlid, J., & Dahlberg, A. (2006). Wood-inhabiting fungal communities in woody debris of Norway spruce (*Picea abies* (L.) karst.), as reflected by sporocarps, mycelial isolations and T-RFLP identification. *FEMS Microbiology Ecology, 55*(1), 57-67.

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

Barberán, A., Ladau, J., Leff, J. W., Pollard, K. S., Menninger, H. L., Dunn, R. R., & Fierer, N. (2015). Continental-scale distributions of dust-associated bacteria and fungi. *Proceedings of the National Academy of Sciences, 112*(18), 5756-5761.

Be, N. A., Thissen, J. B., Fofanov, V. Y., Allen, J. E., Rojas, M., Golovko, G., . . . Jaing, C. J. (2015). Metagenomic analysis of the airborne environment in urban spaces. *Microbial Ecology, 69*(2), 346-355. Bowers, R. M., McLetchie, S., Knight, R., & Fierer, N. (2010). Spatial variability in airborne bacterial communities across land-use types and their relationship to the bacterial communities of potential source environments. *The ISME Journal, 5*(4), 601-612.

Cannon, P. F., & Kirk, P. M. (2007). *Fungal families of the world* (1st ed.). UK: CABI.

Caporaso, J. G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F. D., Costello, E. K., . . . Knight, R. (2010). QIIME allows analysis of high-throughput community sequencing data. *Nature Methods, 7*(5), 335-336.

Carlsen, T., Aas, A. B., Lindner, D., Vrålstad, T., Schumacher, T., & Kauserud, H. (2012). Don't make a mista(g)ke: Is tag switching an overlooked source of error in amplicon pyrosequencing studies? *Fungal Ecology, 5*(6), 747-749.

Castaño, C., Oliva, J., Martínez, d. A., Alday, J. G., Parladé, J., Pera, J., & Bonet, J. A. (2017). Mushroom emergence detected by combining spore trapping with molecular techniques. *Applied and Environmental Microbiology, 83*(13), e00600-17.

Côté, V., Kos, G., Mortazavi, R., & Ariya, P. A. (2008). Microbial and "de novo" transformation of dicarboxylic acids by three airborne fungi. *Science of the Total Environment, 390*(2–3), 530-537.

Davison, J., Moora, M., Öpik, M., Adholeya, A., Ainsaar, L., Bâ, A., . . . Zobel, M. (2015). Global assessment of arbuscular mycorrhizal fungus diversity reveals very low endemism. *Science, 349*(6251), 970-973.

Eames, I., Tang, J. W., Li, Y., & Wilson, P. (2009). Airborne transmission of disease in hospitals. *Journal of the Royal Society Interface, 6*, S697-S702.

Eduard, W., Heederik, D., Duchaine, C., & Green, B. J. (2012). Bioaerosol exposure assessment in the workplace: The past, present and recent advances. *Journal of Environmental Monitoring, 14*(2), 334-339.

Elbert, W., Taylor, P. E., Andreae, M. O., & Pöschl, U. (2007). Contribution of fungi to primary biogenic aerosols in the atmosphere: Wet and dry discharged spores, carbohydrates, and inorganic ions. *Atmospheric Chemistry and Physics, 7*(17), 4569-4588.

Emberlin, J. C., & Baboonian, C. (1995). The development of a new method of sampling airborne particles for immunological analysis. In A. Basomba, & F. Hernandez de Rojas (Eds.), *16th European congress of allergology and clinical immunology* (pp. 39-43). Bologna, Italy: Monduzzi Editore S.p.A.

Fierer, N., Liu, Z., Rodríguez-Hernández, M., Knight, R., Henn, M., & Hernandez, M. T. (2007). Short-term temporal variability in airborne bacterial and fungal populations. *Applied and Environmental Microbiology, 74*(1), 200-207.

Fröhlich-Nowoisky, J., Burrows, S. M., Xie, Z., Engling, G., Solomon, P. A., Fraser, M. P., . . . Pöschl, U. (2012). Biogeography in the air: Fungal diversity over land and oceans. *Biogeosciences, 9*(3), 1125-1136.

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

Galante, T. E., Horton, T. R., & Swaney, D. P. (2011). 95% of basidiospores fall within 1 m of the cap: A field-and modeling-based study. *Mycologia, 103*(6), 1175-1183.

Gandolfi, I., Bertolini, V., Ambrosini, R., Bestetti, G., & Franzetti, A. (2013). Unravelling the bacterial diversity in the atmosphere. *Applied Microbiology and Biotechnology, 97*(11), 4727-4736.

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

Halme, P., & Kotiaho, J. S. (2012). The importance of timing and number of surveys in fungal biodiversity research. *Biodiversity and Conservation, 21*(1), 205-219.

Hazard, C., Gosling, P., van der Gast, C.,J., Mitchell, D. T., Doohan, F. M., & Bending, G. D. (2012). The role of local environment and geographical distance in determining community composition of arbuscular mycorrhizal fungi at the landscape scale. *The ISME Journal, 7*(3), 498-508.

Hoisington, A. J., Maestre, J. P., King, M. D., Siegel, J. A., & Kinney, K. A. (2014). Impact of sampler selection on the characterization of the indoor microbiome via high-throughput sequencing. *Building and Environment, 80*, 274-282.

Kivlin, S. N., Winston, G. C., Goulden, M. L., & Treseder, K. K. (2014). Environmental filtering affects soil fungal community composition more than dispersal limitation at regional scales. *Fungal Ecology, 12*(Supplement C), 14-25.

Klein, A. M., Bohannan, B. J. M., Jaffe, D. A., Levin, D. A., & Green, J. L. (2016). Molecular evidence for metabolically active bacteria in the atmosphere. *Frontiers in Microbiology, 7*, 772.

Kramer, C. L. (1982). Production, release and dispersal of basidiospores. In J. C. Frankland, J. N. Hedger & M. J. Swift (Eds.), *Decomposer basidiomycetes: Their biology and ecology* (1st ed., pp. 33-49). UK: Cambridge University Press.

Kubartová, A., Ottosson, E., Dahlberg, A., & Stenlid, J. (2012). Patterns of fungal communities among and within decaying logs, revealed by 454 sequencing. *Molecular Ecology, 21*(18), 4514-4532.

Lee, S., Lee, H., Kim, S., Lee, H. M., Kang, H., & Kim, Y. P. (2010). Identification of airborne bacterial and fungal community structures in an urban area by T-RFLP analysis and quantitative real-time PCR. *Science of the Total Environment, 408*(6), 1349-1357.

Lilleskov, E. A., & Bruns, T. D. (2005). Spore dispersal of a resupinate ectomycorrhizal fungus, tomentella sublilacina, via soil food webs. *Mycologia, 97*(4), 762-769.

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

Mäkipää, R., Rajala, T., Schigel, D., Rinne, K.,T., Pennanen, T., Abrego, N., & Ovaskainen, O. (2017). Interactions between soil- and dead wood-inhabiting fungal communities during the decay of Norway spruce logs. *The Isme Journal, 11*(9), 1964-1974.

Moncalvo, J., & Buchanan, P. K. (2008). Molecular evidence for long distance dispersal across the southern hemisphere in the *Ganoderma applanatum-australe* species complex (basidiomycota). *Mycological Research, 112*(4), 425-436.

Nguyen, N. H., Smith, D., Peay, K., & Kennedy, P. (2015). Parsing ecological signal from noise in next generation amplicon sequencing. *New Phytologist, 205*(4), 1389-1393.

Nicolaisen, M., West, J. S., Sapkota, R., Canning, G. G. M., Schoen, C., & Justesen, A. F. (2017). Fungal communities including plant pathogens in near surface air are similar across northwestern europe. *Frontiers in Microbiology, 8*, 1729.

Nilsson, R. H., Tedersoo, L., Abarenkov, K., Ryberg, M., Kristiansson, E., Hartmann, M., . . . Kõljalg, U. (2012). Five simple guidelines for establishing basic authenticity and reliability of newly generated fungal ITS sequences. *MycoKeys, 4*, 37-63.

Norros, V., Rannik, Ü, Hussein, T., Petäjä, T., V., T., & Ovaskainen, O. (2014). Do small spores disperse further than large spores? *Ecology, 95*(6), 1612-1621.

Norros, V., Penttilä, R., Suominen, M., & Ovaskainen, O. (2012). Dispersal may limit the occurrence of specialist wood decay fungi already at small spatial scales. *Oikos, 121*(6), 961-974.

Nuñez, M. A., Hayward, J., Horton, T. R., Amico, G. C., Dimarco, R. D., Barrios-Garcia, M., & Simberloff, D. (2013). Exotic mammals disperse exotic fungi that promote invasion by exotic trees. *Plos One, 8*(6), e66832.

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

Oh, S., Fong, J. J., Park, M. S., Chang, L., & Lim, Y. W. (2014). Identifying airborne fungi in Seoul, Korea using metagenomics. *Journal of Microbiology, 52*(6), 465-472.

Oksanen, J., Blanchet, F. G., Kindt, R., Legendre, P., Minchin, P. R., O'Hara, B., . . . Wagner, H. (2015). *Vegan: Community ecology package* R package version 2.2-1.

Ondov, B. D., Bergman, N. H., & Phillippy, A. M. (2011). Interactive metagenomic visualization in a web browser. *BMC Bioinformatics, 12*(1), 385.

Ovaskainen, O., Schigel, D., Ali-Kovero, H., Auvinen, P., Paulin, L., Norden, B., & Norden, J. (2013). Combining high-throughput sequencing with fruit body surveys reveals contrasting life-history strategies in fungi. *ISME Journal, 7*(9), 1696-1709.

Pashley, C. H., Fairs, A., Free, R. C., & Wardlaw, A. J. (2012). DNA analysis of outdoor air reveals a high degree of fungal diversity, temporal variability, and genera not seen by spore morphology. *Fungal Biology, 116*(2), 214-224.

Peay, K. G., Schubert, M. G., Nguyen, N. H., & Bruns, T. D. (2012). Measuring ectomycorrhizal fungal dispersal: Macroecological patterns driven by microscopic propagules. *Molecular Ecology, 21*(16), 4122-4136.

Peay, K. G., & Bruns, T. D. (2014). Spore dispersal of basidiomycete fungi at the landscape scale is driven by stochastic and deterministic processes and generates variability in plant-fungal interactions. *New Phytologist, 204*(1), 180-191.

Peccia, J., & Hernandez, M. (2006). Incorporating polymerase chain reaction-based identification, population characterization, and quantification of microorganisms into aerosol science: A review. *Atmospheric Environment, 40*(21), 3941-3961.

Polymenakou, P. N. (2012). Atmosphere: A source of pathogenic or beneficial microbes? *Atmosphere, 3*(1), 87-102.

Purhonen, J., Huhtinen, S., Kotiranta, H., Kotiaho, J. S., & Halme, P. (2017). Detailed information on fruiting phenology provides new insights on wood-inhabiting fungal detection. *Fungal Ecology, 27*(Part B), 175-177.

R Development Core Team. (2016). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.

Rieux, A., Soubeyrand, S., Bonnot, F., Klein, E. K., Ngando, J. E., Mehl, A., . . . de Lapeyre de Bellaire, L. (2014). Long-distance wind-dispersal of spores in a fungal plant pathogen: Estimation of anisotropic dispersal kernels from an extensive field experiment. *Plos One, 9*(8), 1-13.

Runnel, K., Tamm, H., & Lõhmus, A. (2015). Surveying wood-inhabiting fungi: Most molecularly detected polypore species form fruit-bodies within short distances. *Fungal Ecology, 18*, 93-99.

Sato, H., Tsujino, R., Kurita, K., Yokoyama, K., & Agata, K. (2012). Modelling the global distribution of fungal species: New insights into microbial cosmopolitanism. *Molecular Ecology, 21*(22), 5599-5612.

Schloss, P. D., Westcott, S. L., Ryabin, T., Hall, J. R., Hartmann, M., Hollister, E. B., . . . Weber, C. F. (2009). Introducing mothur: Open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Applied and Environmental Microbiology, 75*(23), 7537-7541.

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

Smith, D. P., & Peay, K. G. (2014). Sequence depth, not PCR replication, improves ecological inference from next generation DNA sequencing. *Plos One, 9*(2), 1-12.

Somervuo, P., Yu, D. W., Xu, C. C. Y., Ji, Y., Hultman, J., Wirta, H., & Ovaskainen, O. (2017). Quantifying uncertainty of taxonomic placement in DNA barcoding and metabarcoding. *Methods in Ecology and Evolution, 8*(4), 398–407.

Somervuo, P., Koskela, S., Pennanen, J., Henrik Nilsson, R., & Ovaskainen, O. (2016). Unbiased probabilistic taxonomic classification for DNA barcoding. *Bioinformatics, 32:*(9), 2920-7.

Sterflinger, K., & Prillinger, H. (2001). Molecular taxonomy and biodiversity of rock fungal communities in an urban environment (Vienna, Austria). *Antonie Van Leeuwenhoek, 80*(3), 275-286.

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

Tedersoo, L., Bahram, M., Põlme, S., Kõljalg, U., Yorou, N. S., Wijesundera, R., . . . Abarenkov, K. (2014). Global diversity and geography of soil fungi. *Science, 346*(6213)

Unterseher, M., Schnittler, M., Dormann, C., & Sickert, A. (2008). Application of species richness estimators for the assessment of fungal diversity. *FEMS Microbiology Letters, 282*(2), 205-213.

van der Linde, S., Holden, E., Parkin, P. I., Alexander, I. J., & Anderson, I. C. (2012). Now you see it, now you don't: The challenge of detecting, monitoring and conserving ectomycorrhizal fungi. *Fungal Ecology, 5*(5), 633-640.

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

Weissfeld, A. S., Joseph, R. A., Le, T. V., Trevino, E. A., Schaeffer, M. F., & Vance, P. H. (2013). Optimal media for use in air sampling to detect cultivable bacteria and fungi in the pharmacy. *Journal of Clinical Microbiology, 51*(10), 3172-3175.

West, J. S., & Kimber, R. B. E. (2015). Innovations in air sampling to detect plant pathogens. *Annals of Applied Biology, 166*(1), 4-17.

White, T. J., Bruns, T. D., Lee, S., & Taylor, J. (1990). Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. In M. A. Innis, D. H. Gelfaud, J. J. Sninsky & T. J. White (Eds.), *PCR protocols: A guide to methods and applications* (pp. 315-322)

Womack, A. M., Artaxo, P. E., Ishida, F. Y., Mueller, R. C., Saleska, S. R., Wiedemann, K. T., . . . Green, J. L. (2015). Characterization of active and total fungal communities in the atmosphere over the amazon rainforest. *Biogeosciences, 12*(21), 6337-6349.

Womack, A. M., Bohannan, B. J. M., & Green, J. L. (2010). Biodiversity and biogeography of the atmosphere. *Philosophical Transactions of the Royal Society B: Biological Sciences, 365*(1558), 3645-3653.

Author contributions: NA, OO and VN wrote the first draft of the paper, VN and PH collected the data, HA-K performed the molecular analyses, and OO, NA and PS performed the statistical analyses. All authors contributed to the writing of the final version of the paper.

Data accessibility: The sequence data, matrices of identified species and measured environmental variables are available in Dryad data repository (doi:10.5061/dryad.9121q).

Supplemental information for online publication

Appendix 1. Supplementary results containing Figs. A1-A10 and Tables A1-A5.
Appendix 2. Piechart showing the taxonomic composition of all species in the dataset, derived using the probabilistic taxonomical placement method.
Appendix 3. Piechart showing the taxonomic composition of wood-inhabiting fungal species in the dataset, derived using the probabilistic taxonomical placement method.

Figure legends

Figure 1. An overview of the sampling design and variation in community composition among all aerial DNA samples acquired for the present study. Panel A shows the spatial location of the four study sites and illustrates the device (cyclone sampler) used for the aerial fungal sampling. Panel B shows the distribution of the samples in the ordination space of non-metric multidimensional scaling analysis (NMDS), with the four study sites indicated by different symbols and the sampling weeks by green isoclines. In panel B, the analyses are based on species-level identifications with reliability threshold 50%. For the same ordination figure based on genus-level identifications, wood-inhabiting fungal species, and analyses performed for other identification thresholds, see Fig. S1 in Appendix 1.

Figure 2. The taxonomic distribution of the samples for the pooled data, either for Ascomycota (A) or Basidiomycota (B). The pie charts are snapshots from interactive web page (provided as Appendix 2), in which the taxonomic compositions can be studied in detail. The colors corresponding to confidence level 1-2 highlight >50% reliable classifications to well-identified taxa (see Somervuo et al. 2017 for explanations of the other confidence levels).

Figure 3. Seasonal variation among weekly samples acquired in Site 1. Panel A shows the distribution of the samples in the ordination space of non-metric multidimensional scaling analysis (NMDS), day/night/full week samples indicated by different symbols and the sampling weeks by green isoclines. Panel B shows expected species richness (corrected for sequencing depth and duration) as a function of week of sampling, for all species and for Basidiomycota and Ascomycota species separately. Panel C shows the proportion of the species belonging to the five most abundant orders during the sampling season. In all panels, the data have been restricted to species-level identifications with reliability threshold 50%. For the same figures based on genus-level identifications, and analyses performed for other identification thresholds, see the Figs. S2-S4 in Appendix 1.

Figure 4. Spatial, temporal and methodological variation among samples acquired for Site 2. Panel A shows the distribution of the samples in the ordination space of non-metric multidimensional scaling analysis (NMDS), with the distance to the focal forest patch indicated by different symbols and samples taken during the day hours and night hours indicated by red and blue colors, respectively. Panel B and C show the species richness for samples taken during the day hours (red) and night hours (blue), for all species (panel B) and for wood inhabiting fungal species (panel C). The boxes and whiskers show the median, upper and lower quartiles, and maximum and minimum values of species richness. Panel D shows species richness as a function of sampling duration. In panel D, the line shows the prediction of Poisson regression, in which sequencing depth has been standardized to its mean value over the data. In all panels, the analyses are based on species-level identifications with reliability threshold 50%. For the same figures based on genus-level identifications, and analyses performed for 90% identification thresholds, see Figs. S5-S8 in Appendix 1.

Figure 5. Comparison in species richness and community composition among samples taken from different substrates. Panels ABC show species accumulation along increasing number of samples for aerial samples (A), wood samples (B), and soil samples (C). The non-metric multidimensional scaling ordination (NMDS) of panel D shows variation in community composition among samples taken in the three substrates, and the Venn diagram of panel E illustrates the numbers of species shared or not shared among the three substrates. The different species accumulation curves within each panel are based on i) all species (black line showing mean, grey polygon 95% confidence interval), ii) species occurring in at least two samples (black line, cyan polygon), iii) species with at least 10 sequence reads per sample (dashed line, grey polygon), and iv) (species with at least 100 sequence reads per sample (dotted line, grey polygon). The aerial samples originate from all sites (S1-S4) in panels DE but are restricted to S1 in panel A, whereas the wood and soil samples (panels BC) originate from S3. For the same figures based on genus-level identifications, and analyses performed for 90% identification thresholds, see Figs. S9-S10 in Appendix 1.

Table 1. Twenty most abundant genera in each sampling site, based on 90% identification threshold. The genera belonging to wood-inhabiting fungal species (all polypores and prominent corticioid and hydnoid species) are indicated with †. The numbers show the abundance from aerial surveys measured as relative sequence number raised to the power of 0.25, and the number of species observed as fruit-bodies from sites S1-S3. Note that due to variation in effort in fruit-body surveys, the numbers are not directly comparable among the sites.

Most abundant genera in S1	Abundance	Fruit-body species	Most abundant genera in S2	Abundance	Fruit-body species	Most abundant genera in S3	Abundance	Fruit-body species	Most abundant genera in S4	Abundance
Rhodotorula	0.21	-	*Baeospora*	0.21	-	*Penicillium*	0.32	-	*Rhodotorula*	0.33
Lophodermium	0.19	-	*Taphrina*	0.17	-	*Antrodia*†	0.25	2	*Filobasidium*	0.26
Penicillium	0.19	-	*Melampsoridium*	0.17	-	*Phellinus*†	0.24	4	*Suillus*	0.20
Trametes†	0.18	2	*Exobasidium*	0.16	-	*Russula*	0.23	-	*Laetiporus*†	0.19
Mycena	0.17	-	*Chrysomyxa*	0.15	-	*Taphrina*	0.23	-	*Baeospora*	0.19
Taphrina	0.17	-	*Phlebia*†	0.14	2	*Phlebia*†	0.23	3	*Curvibasidium*	0.17
Ramularia	0.16	-	*Antrodiella*†	0.14	1	*Fomes*†	0.20	1	*Taphrina*	0.16
Phlebia†	0.16	8	*Phenoliferia*	0.13	-	*Pyrenophora*	0.19	-	*Sporobolomyces*	0.14
Antrodia†	0.15	3	*Cortinarius*	0.11	-	*Fomitopsis*†	0.18	2	*Cryptococcus*	0.14
Botryobasidium†	0.14	7	*Trechispora*†	0.10	2	*Cortinarius*	0.17	-	*Phlebia*†	0.13
Fomes†	0.13	1	*Phlebia*†	0.10	2	*Mrakia*	0.16	-	*Pluteus*	0.12
Trechispora†	0.11	7	*Penicillium*	0.09	-	*Neoerysiphe*	0.16	-	*Physcia*	0.12
Cladonia	0.09	-	*Trametes*†	0.09	2	*Cladosporium*	0.16	-	*Antrodiella*†	0.12
Stereum†	0.09	3	*Pseudeurotium*	0.08	-	*Postia*†	0.14	12	*Peniophora*	0.11
Exobasidium	0.09	-	*Mycena*	0.08	-	*Cytospora*	0.14	-	*Naganishia*	0.11
Lepraria	0.09	-	*Melampsora*	0.07	-	*Coniophora*†	0.14	3	*Phaeophyscia*	0.10
Filobasidium	0.09	-	*Bryoria*	0.07	-	*Butyrea*	0.14	-	*Malassezia*	0.10
Pluteus	0.09	-	*Lophodermium*	0.07	-	*Chaetosphaeria*	0.11	-	*Stereum*†	0.10
Bryoria	0.08	-	*Stereum*†	0.05	0	*Mortierella*	0.10	-	*Cortinarius*	0.10
Xeromphalina	0.08	-	*Chondrostereum*†	0.05	0	*Alternaria*	0.09	-	*Melampsoridium*	0.09