The diversity and seasonality of the indoor mycobiome

Eva Lena F. Estensmo

Dissertation presented for the degree of

Philosophiae Doctor (PhD)

2021

Department of Biosciences

Faculty of Mathematics and Natural Sciences

University of Oslo

Series of dissertations submitted to the
Faculty of Mathematics and Natural Sciences, University of Oslo
No. 2411

ISSN 1501-7710

Cover: Hanne Baadsgaard Utigard.
Print production: Reprosentralen, University of Oslo.

Acknowledgements

During my PhD, I have been able to be a part of a scientific journey that resulted in this thesis. First, I would like to thank my main supervisor Håvard Kauserud for believing in me and for the opportunity to do a PhD in your research group. Thank you for the encouragements, contributions and excellent advices. I would also like to thank my co-supervisors Inger Skrede and Sundy Maurice. I have really enjoyed our professional relationships, but also our everyday talks and your support. I have learned incredible much from all of you, and I could not have asked for better supervisors. I would also like to thank the rest of the IndoorMyco team at the University of Oslo, Pedro M. Martin-Sanchez, Luis Morgado and Synnøve S. Botnen, for your contributions to the research of the indoor mycobiome. It has been a pleasure to work with you. I am also grateful for being a part of Oslo Mycology Group, and I will like to thank you all for making me feel welcome, being helpful and for all the social and professional activities.

I would like to thank Digital Life Norway for the opportunity to do an "innovation" PhD and collaborate with an industry. It has been a pleasure to do my duty work at Mycoteam, and I would like to thank Kolbjørn M. Jenssen and my co-supervisors Ingeborg B. Engh and Johan Mattsson for including me in your firm and for your contributions to the research projects. I will also thank Ingrid Grønli for contributing to the sampling, Maria Nuñes for fruitful discussions, and Malin Siljedal for great collaborations in the laboratory.

Furthermore, I would like to thank all the citizen scientists who provided samples and data of their houses and daycare centers. Your contribution have been essential for this thesis. I would also like to acknowledge Peter Horvath et al. from the Natural History Museum for providing environmental metadata in our large-scale studies of houses and daycares. In addition, I would like to thank the Norwegian Asthma and Allergy Association (NAAF) for funding.

At last, but not at least, I would like to thank my family and friends for your support and for all the good times we had together. Thank you for all the amazing trips, great dinners, sparkling wine and hard workouts. I am so grateful to have you in my life, especially you Dennis. Your support and love means a lot to me, and I could not have done this without you. In addition, a special thanks to you Dad. It breaks my heart to write this thesis knowing that you are not going to be with us for much longer. I am so grateful for your endless support, and I have always looked up to you. You are my hero, and I will always remember our time together.

Table of Contents

List of papers

Paper I

The influence of intraspecific sequence variation during DNA metabarcoding: A case study of eleven fungal species. Eva Lena Estensmo, Sundy Maurice, Luis Morgado, Pedro M. Martin-Sanchez, Inger Skrede and Håvard Kauserud (2021). *Molecular Ecology Resources.* https://doi.org/10.1111/1755-0998.13329

Paper II

Analyzing indoor mycobiomes through a large-scale citizen science study in Norway. Pedro M. Martin-Sanchez, Eva Lena F. Estensmo, Luis N. Morgado, Sundy Maurice, Ingeborg B. Engh, Inger Skrede and Håvard Kauserud (2021). *Molecular Ecology.* https://doi.org/10.1111/mec.15916

Paper III

The indoor mycobiome of daycare centers is affected by occupancy and climate. Eva Lena F. Estensmo, Synnøve Smebye Botnen, Sundy Maurice, Pedro M. Martin-Sanchez, Luis Morgado, Ingeborg Bjorvand Engh, Klaus Høiland, Inger Skrede and Håvard Kauserud. *Manuscript intended for Applied and Environmental Microbiology.*

Paper IV

Spatiotemporal variation of the indoor mycobiome in daycare centers. Eva Lena Estensmo, Luis Morgado, Sundy Maurice, Pedro M. Martin-Sanchez, Ingeborg B. Engh, Johan Mattsson, Håvard Kauserud and Inger Skrede (2021). *Under review in Microbiome.*

Summary

Buildings harbor unique and complex microbial communities, including fungi. From earlier work, we know that indoor fungal communities, the indoor mycobiome, can vary significantly in different geographic regions and in different seasons. However, more baseline information about the indoor mycobiome is necessary for improved indoor air quality and for identification of tentative health risks. This thesis aims to improve our understanding about the indoor mycobiome in Norway.

One approach to investigate the indoor mycobiome composition is DNA metabarcoding, which is based on high throughput sequencing (HTS) of amplified markers. Ideally, the amplified marker will discriminate among species. The rDNA ITS region is the recognized barcoding region for fungi, and the ITS1 or ITS2 regions are the most commonly used markers for fungal DNA metabarcoding studies. In this thesis, ITS2 was used during the DNA metabarcoding analyses of all four papers. However, the ITS region may include considerable intraspecific variation. This variation can lead to over-splitting of species during DNA metabarcoding analyses. In Paper I, we assessed the effects of intraspecific sequence variation in DNA metabarcoding by analyzing local populations of eleven fungal species. All the eleven species, except one, included some level of intraspecific variation in the ITS2 region. The presence of this intraspecific variation in ITS2 suggest that clustering is needed to approach species-level resolution in metabarcoding studies of fungal communities.

In order to improve the knowledge of the indoor mycobiome, we analysed fungal communities in indoor environments associated with private homes (Paper II) and daycare centers (Paper III) at large geographic scales in Norway using a citizen science approach. Dust samples were collected from doorframes from 125 daycares and 271 private homes in three different house compartments: outside the building (main entrance), living room and bathroom. The fungal community composition and diversity were determined by DNA metabarcoding. The fungal community composition was clearly different between indoor and outdoor samples in both daycares and private homes, but there were no marked differences between the two indoor compartments in either of the studies. The fungal richness and compositional variation could be ascribed to numerous indoor and outdoor variables, and there was a clear geographic signal in the indoor mycobiome composition that mirrored the outdoor climate. In both studies, the indoor mycobiomes represent a mixture of fungi from both indoor and outdoor sources. In the

daycares, the indoor mycobiomes included considerably more yeasts and molds compared to the outdoor, with Saccharomycetales as the dominant fungal order. In the private homes, the mycobiomes were mainly dominated by molds from the fungal orders Capnodiales and Eurotiales. The observed differences between the daycares and private homes may be due to the large number of occupants, and children in particular, in the daycares.

Finally, we investigated the spatiotemporal dynamics of the indoor mycobiomes in two daycare centers (Paper IV). Dust samples were collected throughout a year in order to evaluate indoor air quality, and the effect of occupancy and seasonality. We collected dust samples from different rooms and analyzed their mycobiomes using DNA metabarcoding. The fungal community composition in rooms with limited occupancy was different from rooms with high occupancy and more similar to the outdoor samples. A strong seasonal pattern was observed in the mycobiome composition, mainly structured by the outdoor weather conditions. Therefore, the temporal variability should be accounted for in indoor mycobiome studies and in evaluations of indoor air quality.

Introduction

The fungal kingdom

The Fungal kingdom is believed to have originated between 760 million and 1 billion years ago (Lücking *et al.* 2009). Through time, a high diversity of fungal species have originated. About 148 000 fungal species have been described so far (Antonelli *et al.* 2020), but the total number is highly uncertain and estimates spans between 0.8 to 5.1 million fungal species (Antonelli *et al.* 2020; Blackwell 2011; Hawksworth & Lücking 2017; Tedersoo *et al.* 2014). Currently, the fungal kingdom is divided into nine phyla: Opisthosporidia, Chytridiomycota, Neocallimastigomycota, Blastocladiomycota, Zoopagomycota, Mucoromycota, Glomeromycota, Ascomycota and Basidiomycota (Naranjo-Ortiz & Gabaldón 2019). The most species rich groups are Basidiomycota and Ascomycota, which I will mainly focus on in this thesis.

The uncertainty of the number of fungal species is mainly based on their microscopic size and that they generally live hidden within their substrate, except during the fruiting stage. Fungi can be multicellular (filamentous), growing with hyphae that may branch and create large mycelial networks. Filamentous fungi may form large macroscopic structures, such as fruit bodies commonly found in nature (Figure 1). Filamentous fungi are mainly spread with microscopic spores through air, which can be of both sexual and asexual origin (see asexual spores of *Aspergillus* in Figure 1). Fungi can also be unicellular organisms (yeast). Some fungal species can also be dimorphic and shift between growth forms during different life stages (producing either yeasts or hyphae). Both yeasts and filamentous growth are found in species of both Basidiomycota and Ascomycota.

Figure 1. Illustration of different fungal growth forms. From the left: yeasts growth in *Candida*; microscopic asexual spore production of *Aspergillus,* a filamentous fungi; the macroscopic fruit bodies of *Craterellus*. Photo: Eva Lena Estensmo.

Fungi are heterotrophic organisms that absorb nutrients from their surroundings. They are essential components in most ecosystems and exhibit varied ecology and life strategies. They can be decomposers of dead organic material (saprotrophs), have various forms of mutualistic relationships (e.g. lichens and mycorrhiza) or they can be plant pathogens or parasitize animals. Some fungi can switch from one life strategy or growth mode to another to adapt to environmental change. For example, *Candida* species are common commensal fungi associated with the human body, but the fungus can impact our immune system by switching between yeast and hyphal growth forms and turn pathogenic (Limon *et al.* 2017; Underhill & Iliev 2014).

In boreal and temperate climatic regions, there are significant temporal changes in temperature and precipitation throughout a year. Fungi adapts their life stages to these environmental conditions. Many mycorrhizal and saprotrophic fungi produce fruit bodies and spread their spores during the fall, whereas plant pathogens may reproduce asexually or sexually in various time periods during the entire plant growth season (Chen *et al.* 2018). Thus, the fungal spore diversity and composition in the air can vary significantly throughout the year (Moore *et al.* 2008). The fungal response to the environment may be subtle, but currently a large shift in fruiting season of fungi can be observed due to climate changes. Analyses of collections of fruit bodies through time show that spring-fruiting fungi are now fruiting earlier, and some fungi have extended their fruiting season to both earlier in the spring and later in the fall (Andrew *et al.* 2018; Boddy *et al.* 2014; Gange *et al.* 2007).

Fungi in the built environment

Many fungi have expanded their niche into the built environment (Balasundaram *et al.* 2018; Gilbert & Stephens 2018; Nevalainen *et al.* 2015; Schmidt 2007). The built environment includes all manufactured structures such as buildings, transportation systems and other constructions. Humans spend a majority of their lifetime in buildings, such as private homes, workplaces, daycares and schools. These constructions are extreme environments in many ways, designed to be inhospitable for microbial life with many processed materials, which are limiting available nutrients and water. Nevertheless, buildings harbor unique and complex microbial communities, including invertebrates, fungi, bacteria, archaea and viruses, that are able to survive in the extreme environments (Gilbert & Stephens 2018; Martin *et al.* 2015; Nevalainen *et al.* 2015).

The overall assembly of the fungal communities in buildings is referred to as the indoor mycobiome. The indoor mycobiome includes fungi that originate from both indoor and outdoor sources and are structured by numerous factors, including geography, climate, building features, building usage, the number and type of occupants and our behavior (Adams *et al.* 2016; Horve *et al.* 2020; Nevalainen *et al.* 2015). Previous studies of indoor environments suggest that the indoor mycobiome are highly affected by outdoor air (Adams *et al.* 2013; Barberán *et al.* 2015b; Frankel *et al.* 2012; Pitkäranta *et al.* 2008). Common fungi in the outdoor air include species from the genera *Cladosporium*, *Penicillium*, *Aspergillus* and *Alternaria*, and these are also ubiquitous in the indoor environment (Adams *et al.* 2013). The most important indoor sources of fungi and indoor factors that structures the mycobiome was reviewed by Prussin and Marr (2015), which suggested that they include occupants, pets, food, waste, plants, plumbing systems, mold damages, heating, ventilation and air conditioning (Prussin & Marr 2015).

Fungi in the indoor environment mainly includes saprotrophs that can degrade organic substrates. Their capability to grow is often restricted by the availability of water, but where excess moisture is available, fungi can start to grow. Many fungi degrade the actual building materials and cause damages to wood constructions and other building materials (Singh 1999; WHO 2009). The most infamous wood decaying fungus of the built environment in temperate and boreal regions is the dry rot fungus *Serpula lacrymans*, which is able to quickly decay construction material under optimal conditions (Kauserud *et al.* 2012; Schmidt 2007). Luckily, most indoor saprotrophic fungi that cause damages to building materials are mainly restricted to surface growth. The most common fungal species that frequently colonize moist building materials includes the genera *Cladosporium*, *Penicillium*, *Aspergillus*, *Alternaria*, *Epicoccum*, *Wallemia*, *Ulocladium*, *Stachybotrys*, *Chaetomium* and *Acremonium* (Horner 2003) – often recognized as molds on moist surfaces. Nevertheless, most of the fungi in the indoor environment are not causing any damage to buildings (Adams *et al.* 2013; Amend *et al.* 2010; Barberán *et al.* 2015a). Saprotrophic fungi in the buildings may instead live on ephemeral substrates as food and food waste, or other organic substances.

In addition to the many sources of fungal growth in the indoor and outdoor environment, the human body itself is a significant source of fungi in the indoor environment. Most of the fungal species associated with the human body is harmless and a part of our natural microbiota. Several fungal genera can be found in the gastro intestinal tract, including *Candida*, *Cladosporium*,

Aureobasidium, *Aspergillus*, *Malassezia*, *Epicoccum* and *Saccharomyces* (Dupuy *et al.* 2014; Ghannoum *et al.* 2010). Many fungi are also associated with human skin. The basidiomycete yeast *Malassezia* is the dominant fungus on human skin, and is particularly prevalent on adults (White *et al.* 2014). In children, the fungal skin community is more diverse, including genera like *Aspergillus*, *Epicoccum*, *Cladosporium*, *Cryptococcus* and *Phoma*, in addition to *Malassezia* (Jo *et al.* 2016). Although not very harmful, particularly fungi growing on human skin are associated with common disorders, including dandruff, atopic dermatitis, ringworm, and nail infections. These disorders are caused by *Malassezia* and various ascomycete dermatophytes, including species in the genera *Trichophyton*, *Microsporum*, and *Epidermophyton* (White *et al.* 2014). *Candida* can cause oral infections in children, elderly and sick humans, commonly known as thrush (Scully *et al.* 1994). However, some commensal species in the genera *Aspergillus*, *Aureobasidium*, *Malassezia*, *Candida*, and *Cryptococcus* are also opportunistic pathogens that can cause harmful diseases in humans under certain conditions (Limon *et al.* 2017). Especially people with immune-compromised immune systems can be attacked by a high number of fungi that otherwise appear as commensals. *Aspergillus* can cause serious invasive aspergillosis in immuno-compromised patients, unfortunately, often with a deadly outcome.

Even though we are surrounded with a variety of fungi, we rarely notice their presence. However, some fungi can produce compounds affecting and reducing the indoor air quality in our buildings. These compounds includes volatile organic compounds and mycotoxins (Nevalainen *et al.* 2015), and they can easily get in contact with our skin or enter the human body through eyes, nose or mouth. In most buildings, these compounds exist in low concentrations and does not cause any harm. However, in buildings with dampness- and mold-related indoor air quality problems, these compounds, together with fungal spores, can reach high concentrations due to fungal growth. The direct effect of human health from volatiles and mycotoxins in the air is not well known, but it has been shown that high levels of fungal volatile organic compounds cause developmental defects in *Drosophila* (Inamdar & Bennett 2015). Further, high levels of fungal spores can cause adverse health effects, such as allergies, asthma and other respiratory symptoms (Bornehag *et al.* 2001; Mendell *et al.* 2011). It is estimated that 5% of humans will have some allergic airway symptoms from molds during their lifetime (Hardin *et al.* 2003). Nevertheless, the causative agents and mechanisms of such health effects are insufficiently understood.

In Norway, and many other parts of the world, children spend significantly amount of time in daycares. The air quality in the daycares is probably important for the development and health of these children. It has been shown that the skin of children can host a higher fungal diversity than adult skin, and it may therefore be expected that the fungal diversity is higher in daycares than in most private homes. Moreover, daycares usually have higher occupancy and activity level than private homes.

Methods to study the indoor mycobiome

Assessments of indoor mycobiomes are mainly based on air and dust samples, as well as inspection and sampling from building materials if fungal growth is suspected. The presence of fungi in the indoor air can be monitored by collecting airborne fungal particles passively or actively. Passive methods include sedimentation of fungal particles on collection plates or petri dishes with media for culturing (Napoli *et al.* 2012). The fungal particles, or dust, can then be sampled using swabs or tape. Active methods include air sampling with instruments extracting a certain amount of air, from which particles are extracted or deposited on growth media, or dust collection by vacuum cleaners (Napoli *et al.* 2012). By active air sampling, one can collect an exact volume of air during a specific period of time, and record which fungi are circulating in the air in that moment. By passive or active dust sampling, one can collect fungi that has been circulating in the air and settled on surfaces over long time. For how long the dust has been settled would be dependent on the experimental setup or the frequency of cleaning of the surfaces.

Usually, the scientists themselves perform the sampling, but by selecting passive sampling methods that require limited equipment, one can recruit volunteers to perform the sampling. This can be referred to as citizen science, where networks of non-scientists help to collect data as part of a research project (Cohn 2008; Dickinson *et al.* 2010; Gura 2013). This can introduce some bias, such as variation in sampling performance, the volunteers' ability to follow instructions, and the quality of the material they sample (Cohn 2008; Dickinson *et al.* 2010). However, citizen science is a powerful approach where sample equipment can be sent out along with detailed instructions by post, and hundreds of samples, covering large geographic areas, are returned.

The fungal content of the collected samples can be analysed using different approaches. Traditional methods include culturing, microscopy or the use of biomarkers and chemical

analyses by estimating e.g. fungal biomass from ergosterol (Nevalainen *et al.* 2015). Culture-based methods have well-known limitations; only a fraction of fungi can grow on a chosen culture medium at the selected incubation conditions since they have specific growth requirements (Macher 2001). In addition, taxonomic assignment can be difficult, as some fungi are hard to morphologically distinguish under laboratory conditions. Therefore, a shift toward DNA-based methods has taken place in the recent years. Real-time quantitative PCR (qPCR) has been widely used to detect and quantify indoor fungi based on DNA (Vesper *et al.* 2007), but this approach only monitors fungi that you *a priori* suspect are present, and it does not provide new knowledge about other taxa. This means that you need to have deep knowledge of what to expect in your sample.

High throughput sequencing (HTS) of amplified markers (DNA metabarcoding) has become a powerful tool for specie-level identification of fungal communities (Goodwin *et al.* 2016; Lindahl *et al.* 2013; Taberlet *et al.* 2018; Taberlet *et al.* 2012). DNA metabarcoding of fungi normally relies on the analysis of the Internal Transcribed Spacer (ITS) region of ribosomal RNA (rRNA) gene (Gardes & Bruns 1993; Koljalg *et al.* 2005; Nilsson *et al.* 2008; Schoch *et al.* 2012). The ITS region lies between the conserved ribosomal subunits: the small subunit 18S and the large subunit 28S (Nilsson *et al.* 2019). The ITS region consists of the variable regions ITS1 and ITS2, which is separated by the conserved region 5.8S. This allows us to amplify the whole ITS region or the two regions separately. Reference sequences of the full ITS region is included in the UNITE database, which is commonly used for taxonomic assignments (Kõljalg *et al.* 2013). Due to length limitations of the sequencing technologies, either the ITS1 or the ITS2 region is typically analyzed. The ITS2 region has been suggested as the preferred barcode because of less amplification bias due to length differences and, additionally, the development of less biased primers (Tedersoo *et al.* 2015; Tedersoo & Lindahl 2016). However, there is still no consensus about which region that should be used (Blaalid *et al.* 2013; Mbareche *et al.* 2020; Tedersoo & Lindahl 2016)

A major obstacle during HTS-based analyses are the bioinformatics analyses, where the massive amount of DNA sequences needs to be processed (Mahe *et al.* 2015). The sequences need to be demultiplexed, quality filtered and error corrected, merged and grouped into biological entities that can be used downstream in community analyses. One approach is to cluster sequences into operational taxonomic units (OTUs) based on a fixed sequence similarity threshold (Caporaso *et al.* 2010; Edgar 2013; Schloss *et al.* 2009; Westcott & Schloss 2015).

Another approach aims to identify the underlying haplotypes present in the original biological samples that gave rise to all the sequence variability. The software DADA2 aims to identify these haplotypes (Callahan *et al.* 2019), and the term amplicon sequence variants (ASVs) has been coined for the output of DADA2 analyses. These ASVs can serve as input for downstream analyses (Callahan *et al.* 2017). However, for the ITS marker it can be problematic to use an ASV as an estimate of a species, as the ITS may have high levels of intraspecific variation (Nilsson *et al.* 2008). We have little knowledge on how the intraspecific sequence variation translates into OTU delineation in DNA metabarcoding studies of fungal communities.

In North America and other parts of the world, the indoor mycobiome has been assessed using high throughput sequence analyses of fungal DNA (Amend *et al.* 2010; Barberán *et al.* 2015a; Barberán *et al.* 2015b; Weikl *et al.* 2016). However, so far, there have been few studies in Europe and Scandinavia implementing HTS to study the indoor mycobiome, despite that indoor fungal communities in different geographic regions can show significant differences and vary through space and time.

Objectives

The overarching aim of my PhD was to improve our understanding about the indoor mycobiome in Norway: which fungi are present and which factors drive the composition and diversity of the indoor mycobiome. Other overarching aims were to evaluate whether high throughput sequencing of DNA obtained from dust samples can be used to survey the indoor mycobiome and whether citizens can be engaged to obtain samples.

More specifically, in Paper I we wanted to assess how intraspecific sequence variation in the ITS2 region affects DNA metabarcoding, and whether this variation leads to over-splitting of species. This is an important topic to address, since over-splitting of species may lead to an overestimation of the fungal diversity in environmental samples.

In paper II and III, we aimed to establish baseline information about which fungi occur within private homes (Paper II) and daycares (Paper III) throughout Norway. In these studies, we also aimed to assess whether outdoor environmental conditions, building features or inhabitant characteristics were most important in structuring the indoor mycobiomes.

Finally, in Paper IV, our aim was to investigate the spatiotemporal dynamics of the indoor mycobiome in two daycare centers in Norway, by collecting and analyzing dust samples

throughout a year. Also in this study, we wanted to evaluate the importance of the outdoor environment versus building and inhabitant characteristics in structuring the indoor mycobiome.

Results

The main results from each of the four studies are summarized in the following section.

Paper I: The influence of intraspecific sequence variation during DNA metabarcoding: A case study of eleven fungal species

The fungal ITS region is the main DNA barcode region for fungi, and is widely used in DNA metabarcoding studies of fungal biodiversity. However, this region may include considerable intraspecific sequence variation, which can lead to over-splitting of species during DNA metabarcoding and, hence, and overestimation of the diversity in environmental samples. To address this topic, we performed DNA metabarcoding on 177 fungal specimens of 11 basidiomycete species and compared the obtained amplicon sequence variants (ASVs), an approximating for haplotypes, to the corresponding Sanger sequences. By denoising the sequence data using DADA2, we tested whether the same ITS2 haplotypes were identified by DNA metabarcoding and Sanger sequencing, and analyzed the allelic diversity of ITS2. We identified between 1 and 11 haplotypes for each species, resulting in 65 haplotypes. This means that all species, except one, had intraspecific variation in the ITS2 region. There was a high correspondence between haplotypes generated by Sanger sequencing and HTS, but a few additional haplotypes were detected in low frequencies using either approach. These additional haplotypes were likely due to PCR and sequencing errors or intragenomic variation in the rDNA region. After clustering the sequences at 97% identity, we obtained 13 sequence clusters (OTUs) for the 11 species. Because of the presence of intraspecific variation in ITS2 region, we suggest that haplotypes (or ASVs) should not be used as basic units in ITS-based fungal community analyses. An extra clustering step is needed to approach species-level resolution.

Paper II: Analyzing indoor mycobiomes through a large-scale citizen science study in Norway

In the second study, we investigate which fungi that are present in private houses throughout Norway and which factors, such as climate, building features and occupant characteristics that structure the indoor mycobiomes. Through a citizen science sampling campaign, we obtained 807 dust samples from 271 houses, collected from door frames from three different locations: outside, living room and bathroom. The dust mycobiomes were analyzed by DNA

metabarcoding of the ITS2 region. The community composition was clearly different between indoor and outdoor samples, but there were no significant differences between the different indoor rooms in composition or diversity. The selected variables, related to climate, building features and occupant characteristics, accounted for 15% of the variation in community composition. The sampling location (indoor versus outdoor) was the most important factor (7.6%), followed by regional-scale climate (4.2%), building features (1.4%) and occupant characteristics (1.1%). The indoor mycobiomes showed higher species richness compared to the outdoor mycobiome, which is probably due to the accumulation of fungi from both outdoor and indoor sources. The indoor mycobiomes were mainly dominated by ascomycetes, with indicator fungi belonging to two ecological groups with allergenic potential: xerophilic molds and skin-associated yeasts. The xerophilic molds included mainly *Penicillium* and *Aspergillus*, whereas the skin-associated yeasts included mainly *Malassezia*, *Debaryomyces*, *Candida* and *Rhodotorula*. These results show that the indoor mycobiomes includes a mixture of fungi from both indoor and outdoor sources, and is structured by a multitude of indoor and outdoor variables.

Paper III: The indoor mycobiome of daycare centers is affected by occupancy and climate

In the third study, we investigate the indoor mycobiomes of 125 daycare centers throughout Norway, covering major gradients in environmental conditions. Dust samples were collected using citizen science, where the staff at the daycare centers sampled from specific locations inside and outside the daycare centers. The samples were analyzed using DNA metabarcoding of the ITS2 region. We observed a clear separation between the indoor and outdoor mycobiomes throughout the entire region, with no difference in the mycobiomes of the two indoor rooms. The richness and compositional variation could be ascribed to numerous factors, both outdoor climatic conditions such as temperature and insolation, geographic variables like proximity to water, as well as indoor variables related to the buildings. There was a clear geographic signal in the mycobiome composition that mirrored outdoor climate, stretching from humid areas in western Norway to drier and colder areas in the east and north of Norway. The mycobiomes were mainly made up of ascomycetes and basidiomycetes, with marked differences in the outdoor and indoor mycobiomes. The indoor mycobiomes included considerably more yeast fungi and molds compared to the outdoors, with Saccharomycetales as the dominant indoor fungal group. The number of children in the daycare centers and

building features also affected the indoor mycobiome composition, and numerous fungal genera associated with the human body were detected, such as *Malassezia*, *Candida*, and *Saccharomyces*.

Paper IV: Spatiotemporal variation of the indoor mycobiome in daycare centers

In study II and III we analyzed the indoor mycobiome present at a certain time point. However, how temporally stable the indoor mycobiome is throughout the year is unclear. We therefore investigated the spatiotemporal variation in indoor mycobiome in two daycare centers in Norway. We did this by collecting dust samples from identical glass plates placed out in different rooms throughout a year. The mycobiome were analyzed using DNA metabarcoding of the ITS2 region in order to evaluate the indoor air quality and the effect of occupancy and seasonality. The community composition of the mycobiome in rooms with limited occupancy (auxiliary rooms), such as the basement and the loft, was similar to the outdoor samples. These rooms had a higher abundance of fungi from Basidiomycota. The rooms with higher occupancy (main rooms), such as the central room and bathroom, were clearly different in community composition from the auxiliary rooms. There were no significant difference in community composition between the different main rooms, and they all contained a higher abundance of Ascomycota compared to the auxiliary rooms. In addition, we observed a strong seasonal pattern in the mycobiome composition, mainly structured by the outdoor climate and especially moisture and temperature. Typical outdoor basidiomycetes in the orders Agaricales and Polyporales were more abundant during summer and fall, whereas ascomycetes of the orders Saccharomycetales and Capnodiales were dominant during winter and spring. These results shows that the indoor mycobiome in daycare centers are clearly structured by occupancy and seasonality.

Discussion

In the following, I will first briefly discuss methodological aspects encountered in different studies in this thesis. Then I will discuss the composition and structure of the indoor mycobiome, and compare the mycobiome from daycares and private houses.

Methodological considerations

In the large-scale studies of the mycobiome in private homes (Paper II) and daycares (Paper III), we choose a citizen science sampling strategy to obtain a high number of samples from a wide geographical distribution. Citizen science sampling can introduce biases due to variation in e.g. sampling performance, the volunteers' ability to follow instructions, and the quality of the material they sample (Cohn 2008; Dickinson *et al.* 2010). In addition, the recruitment process can be biased towards volunteers that are concerned about their indoor air quality. Nevertheless, very few outlier samples appeared, a clear separation between outdoor and indoor samples were observed, and significant correlations to metadata were found. This suggests that the samples were not too biased. Our results regarding the mycobiome composition, are in line with previous studies of dust sampling of the built environment (Adams *et al.* 2013; Weikl *et al.* 2016; Yamamoto *et al.* 2015), and I argue that the citizen science approach turned out to be a successful strategy.

When we designed our studies, we wanted to investigate which fungi that was common in the indoor air circulating around the house. We therefore choose sampling locations located about 2 meter above ground, which allows dust deposition without a direct human influence. Dust sampling is a common approach to investigate the indoor mycobiome in the built environment (Adams *et al.* 2013; Weikl *et al.* 2016; Yamamoto *et al.* 2015). However, there are methodological considerations that should be assessed. Dust samples obtained with swabs and filters provides limited sample material resulting in relatively small amount of DNA. In the two daycares were temporal sampling were done (Paper IV), the rooms with limited occupancy had considerably lower amount of visible dust on the swabs and potentially lower amount of DNA. Likewise, the outdoor air samples collected during the winter months contained lower amount of DNA, most likely due to considerably lower number of fungal spores in the air compared to the other seasons. Another important factor for dust sampling is the dust deposition-time, which is dependent on the frequency of cleaning. The dust deposition-time was controlled in the

temporal sampling in Paper IV, but may vary significantly in the citizen science sampling in Paper II and III. Anyways, the dust sampling of doorframes and glass frames conducted in our studies provided sufficient sample material for most samples for the metabarcoding analyses.

The DNA metabarcoding workflow includes numerous steps in the laboratory, as outlined by Lindahl and colleagues (Lindahl *et al.* 2013). Even though all samples were treated equally, some steps may introduce biases, such as variability in DNA extraction efficiency among organisms, contamination, primer bias to different fungal taxonomic group, PCR bias and sequencing errors. In addition, in the DNA metabarcoding workflow, numerous samples are tagged, pooled and sequenced together. The unique tags are used to link the sequences to the original samples after sequencing, but can in some cases switch during PCR or sequencing. These biases might affect the revealed fungal community composition. Therefore, controls and technical replicate samples were included in all studies. The controls included DNA extraction controls (using clean swabs or filters as starting material), PCR negatives and positive controls (mock communities). Most of the DNA extraction controls and PCR negatives were filtered out during the bioinformatics due to too low number of reads. A few of these controls remained with a low number of reads, and their OTUs were assessed according to their abundance and frequency in the dataset. OTUs that were regarded as contaminants were removed. The positive controls consisted of a mock community of known species, and these samples showed a similar pattern where the reads from these species were almost exclusively detected in the positive controls in all of the studies, suggesting that the tag-switching rate was low. The similarity of the community composition in the technical replicates, which included duplicates of dust samples in different PCR pools and sequencing libraries, was assessed by NMDS and the ordinations confirmed the reproducibility of the DNA metabarcoding workflow.

It is also important to consider the effects of the bioinformatics workflow that we use to analyze our metabarcoding data. We mainly used the software DADA2 for analyzing our data, which is known for single-nucleotide resolution and improved error correction (Callahan *et al.* 2019). However, in Paper I we detected a few more additional ASVs (haplotypes) using HTS compared to Sanger sequencing. Some of these ASVs occurred in very low frequencies, which might be due to PCR errors and DADA2 failing to identify these as artifacts. A similar pattern was observed in a study by Callahan and colleagues (2019), where they did full 16S sequencing of known bacteria and ended up with a few additional ASVs. In addition, although the DADA2 algorithm has a chimeric sequence filter implemented, five obvious chimeric ASVs occurred in

the filtered DNA metabarcoding dataset. This exemplifies that a few ASVs can be erroneous even after DADA2 processing.

It has been suggested to use ASVs as the basic units in microbial community analyses (Callahan *et al.* 2017). As Paper I shows, this is problematic when analyzing fungal ITS data that may contain intraspecific sequence variation. Our results show that a clustering step is needed after error correction to approach species-level resolution. The commonly used similarity threshold is 97%, and this similarity threshold is thought to retain a balance between intraspecific sequence variation and sequencing errors (Blaalid *et al.* 2013; Nilsson *et al.* 2019). However, if you are investigating community composition, clustering levels ranging from 87-99% has little influence on the overall structure if strong underlying gradients are present in the data (Botnen *et al.* 2018).

The indoor mycobiome

Spatial distribution and community composition

Norway possesses marked climatic and environmental gradients, spanning from the warm and wet west coast to the cold and dry inland (Figure 2). These gradients can be used to assess to what degree the outdoor environment influences the indoor mycobiomes, compared to characteristics of the buildings and the occupants. For the large-scale spatial pattern, we observed clear geographic signals in the indoor mycobiome community composition that largely mirrored large-scale environmental gradients in Norway. One main gradient is the continentality-oceanity gradient, which also corresponds with plant phenology. Hence, sampling at the same time along this gradient will to some extent mirror also plants and fungi being active at different stages.

In our studies of indoor mycobiomes, climatic variables linked to temperature, moisture and seasonality were among the most important drivers of the mycobiome composition in both daycares and private houses. These findings are in agreement with previous mycobiome studies in the built environment (Amend *et al.* 2010; Barberán *et al.* 2015a; Barberán *et al.* 2015b). Amend et al. (2010) performed a global survey of fungi from 72 indoor environments and found that the local environmental outside was the strongest determinant of indoor fungal composition. In the study by Barberán et al. (2015), they analyzed dust microbiomes collected

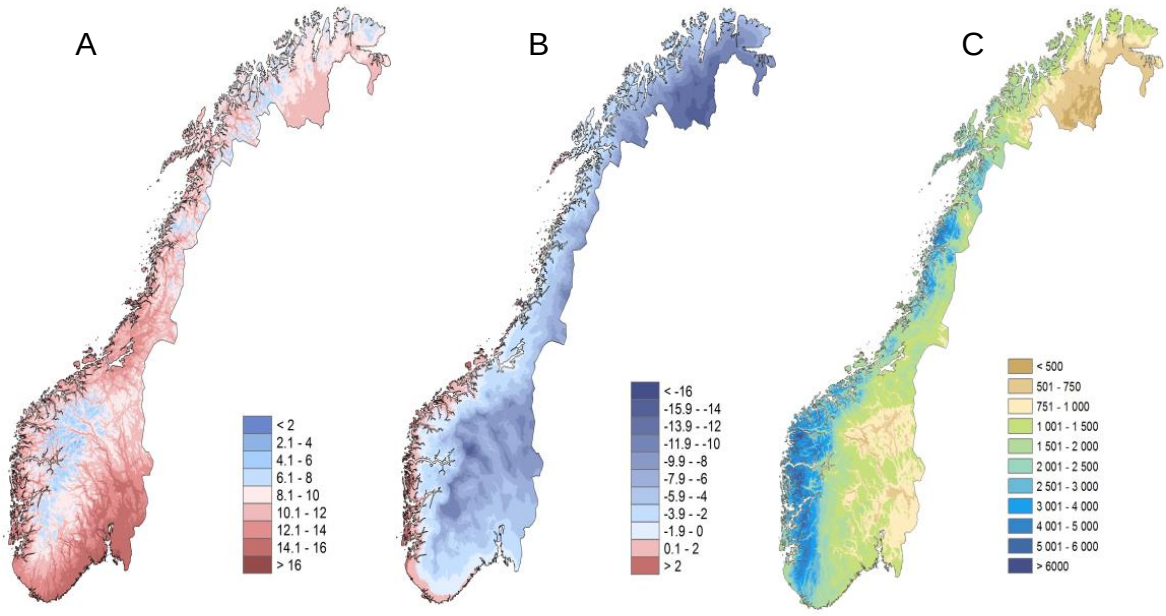


Figure 2. The average temperature (°C) and precipitation (mm) in Norway from 1985-2014. (A) Summer temperature, (B) winter temperature and (C) annual precipitation. Retrieved from the Norwegian climate service center: https://klimaservicesenter.no/

inside 1,200 houses across the United States and identified geographic patterns in the indoor mycobiomes that could be explained by various climate and soil variables (Barberán *et al.* 2015b).

In addition to climate, building and occupants characteristics significantly influenced the mycobiome composition in both daycares and private houses. Interestingly, the presence of pests explained some of the variation for indoor samples in both Paper II and Paper III. The volunteers reported in particular three kind of pests: mice, rats and long-tailed silverfish. Presence of rodents turned out as a significant variable in the daycares. Rodents could act as carriers for fungal spores from the outside, from other parts of the building or from themselves as they carry their own mycobiome associated with their skin, fur, gut or feces (Mims *et al.* 2021; Sanjar *et al.* 2020; Stejskal *et al.* 2005).

Fungal richness and evenness were consistently higher in indoor than outdoor samples in both daycares and private houses. This trend has also been reported in previous studies (Barberán *et al.* 2015a; Yamamoto *et al.* 2015). Barberán et al. (2015) suggested that this tendency might be partially due to the dominance of a few taxa in the outdoor communities, masking some of the

infrequent species. Moreover, a higher indoor diversity may also be due to a mixture of fungi from both outdoor and indoor sources in indoor environments.

We also observed a strong difference in community composition between outdoor and indoor samples. Previous studies of indoor environments suggest that the indoor mycobiomes are highly affected by outdoor air (Adams *et al.* 2013; Barberán *et al.* 2015b; Frankel *et al.* 2012; Pitkäranta *et al.* 2008). However, in these studies they did not study indoor environments with different levels of activity. In Paper II-IV we demonstrated that the number of occupants affects the indoor mycobiome composition. In particular, in Paper IV, the community composition of the mycobiome in rooms with limited occupancy (auxiliary rooms) was more connected to the outdoor samples, while the rooms with higher occupancy (main rooms) were clearly different in community composition. The separation in mycobiome composition of the main rooms and auxiliary rooms in Paper IV can be explained by the number of people accessing and using the rooms, suggesting that occupancy is an important factor shaping the indoor mycobiome in addition to the outdoor air.

In Paper II-IV, we chose to sample the bathroom and the central room, as these rooms are frequently used. We expected that the bathrooms were more prone to moisture problems, hence fungal growth may occur more often in the bathrooms, which may affect the indoor mycobiome. In Paper III, we observed that the fungal richness was somewhat higher in the bathrooms, in line with our expectations. However, there was only a slight difference in community composition between the two room types in the daycares, accounting for 2% of the overall variation.

In Paper II and Paper III, we observed more ascomycetes in the main rooms compared to the outdoors. Several indoor ascomycetes are known to cause allergies and disease in humans, and a previous study of the indoor air in school environments showed that emissions from occupants contributed more to the allergenic fungal populations than fungi entering from the outdoor air (Yamamoto *et al.* 2015). Emissions of fungi from humans might be a reason for why the indoor mycobiome composition is highly affected by occupancy in Paper II-IV. Therefore, it is important to understand this spatial variation of the indoor mycobiome, as this will reflect how the occupants are affected by these fungal species.

Daycares vs. private houses

We observed a clear separation of the indoor samples from the daycares and private houses (Figure 3). The two datasets were sampled in the months of April and May. More daycares were sampled in May than April, while more private houses were sampled in April than May, which might lead to a sampling bias. However, the same pattern was observed when analyzing a subset of the samples from a two-week period in April/May separately. Thus, the results seem to reflect real differences in community compositions in daycares and private houses rather than temporal sampling bias.

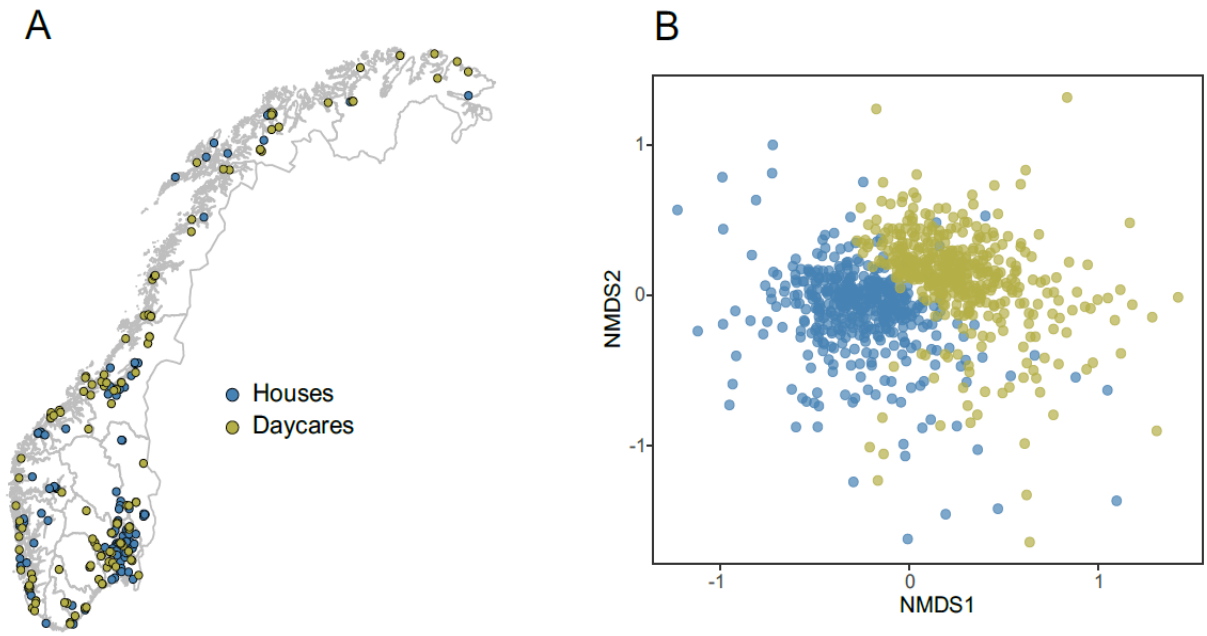


Figure 3. Overview of the buildings and samples from private houses (blue) and daycares (green). (A) A map of the location and distribution of the houses and daycares throughout Norway. (B) NMDS ordination plot displaying compositional variation in the indoor dust mycobiomes in private homes and daycares. Each point represents one dust sample. We combined the sequences from the large-scale citizen study of houses and daycares, ran the bioinformatics and balanced the dataset to include 428 indoor samples from 214 houses and 411 indoor samples from 124 daycares. The NMDS ordination was performed in a similar manner as described in Paper II.

In the private houses (Paper II), the mycobiome was clearly dominated by the order Capnodiales, but also xerophilic molds as Eurotiales and Wallemiales. Xerophilic molds are

fungi that can grow and reproduce in conditions with low availability of water. Although less abundant than in the daycares, Malasseziales and Saccharomycetales were also abundant in private houses.

In the daycares (Paper III), the orders Saccharomycetales and Capnodiales were the most abundant. Saccharomycetales are yeasts including e.g. the well-known genus *Saccharomyces* associated largely with foods, and the human associated fungi in the genus *Candida*. Capnodiales, with the widespread genus *Cladosporium*, includes both plant and human pathogens (Crous *et al.* 2009). In addition, the orders Malasseziales and Mucorales were abundant in the daycares. Malassesziales are basidiomycete yeasts that are associated with the human skin (White *et al.* 2014), whereas Mucorales includes mainly soil saprotrophs that is commonly associated with food production and food spoilage (Hoffmann *et al.* 2013).

By comparing the abundance of different growth forms of fungi in the daycares and the private houses, we observed a higher abundance of yeasts in the daycares (Figure 4). Daycares are characterized by high occupancy and high levels of activity compared to private homes. Previously, higher fungal concentrations have been detected in daycares (Madureira *et al.* 2015). We know that the human body is a significant source of fungi in the indoor environment, and therefore, higher occupancy could contribute to higher levels of human associated fungi. Several fungal genera can be found in the gastro intestinal tract, including different yeasts like *Candida*, *Malassezia* and *Saccharomyces* (Dupuy *et al.* 2014; Ghannoum *et al.* 2010). *Malassezia* are also associated with human skin. The high occupancy in daycares, in addition to the children's natural mycobiome, could explain the high abundance of yeasts in the daycares.

Seasonal variation

In the temporal study (Paper IV), we observed a strong seasonal pattern in the mycobiome composition throughout the year of sampling. The mycobiome composition was mainly structured by the outdoor climate and especially moisture and temperature. Our observed patterns mirror those found in seasonal studies on outdoor mycobiomes (Karlsson *et al.* 2020; Reponen *et al.* 1992). Since the outdoor fungal community has a strong impact on indoor mycobiomes, it is expected that seasonal changes in the outdoor environment also affect which fungi occur indoor. Typical outdoor basidiomycetes in the orders Agaricales and Polyporales were more abundant during summer and fall, whereas ascomycetes of the orders

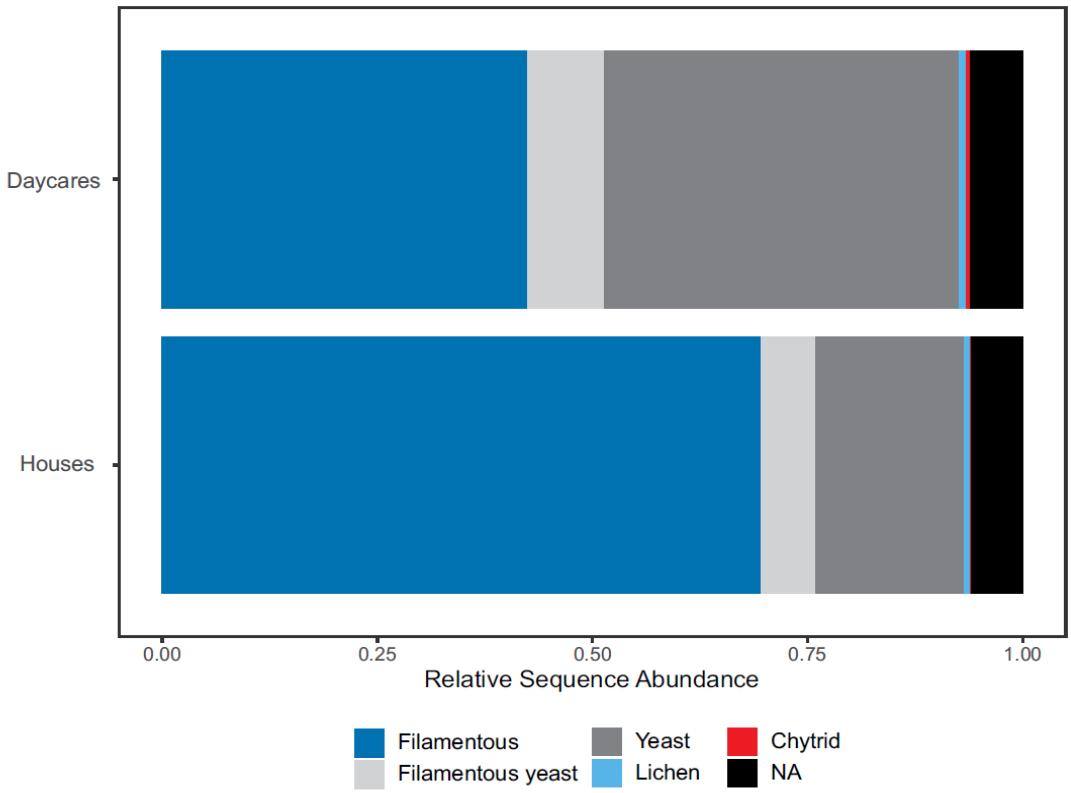


Figure 4. The relative sequence abundance of different groups of fungi in daycares and houses. OTUs were annotated into different growth forms of fungi: filamentous, filamentous yeast, yeast, lichen, chytrid and others (NA), and their relative sequence abundance were compared between daycares and houses.

Saccharomycetales and Capnodiales were dominant during winter and spring. Similar findings were reported in a seasonal study of indoor mycobiomes of four office complexes, where ascomycete molds as Capnodiales and Eurotiales, and basidiomycete yeasts such as *Malassezia* were more common in the winter and spring (Pitkäranta *et al.* 2008).

In Paper II and Paper III, all samples were collected during spring. Seen in light of the results from Paper IV, the fungal communities in Paper II and III show a typical winter/spring pattern with low levels of outdoor basidiomycetes. Thus, these studies are heavily affected by sampling time. However, we do expect that a sampling in the summer or fall would have been even more affected by outdoor fungi and the surrounding environment. Seasonal changes are highly important for indoor air quality, as it affects which fungi that occurs indoor.

Conclusions and future perspectives

I conclude that dust sampling through citizen science coupled to DNA metabarcoding was a successful strategy for characterizing the indoor mycobiome of a large set of houses and daycares throughout Norway. Future large-scale studies on indoor mycobiome should preferably target other countries and other institutions, to reveal whether similar trends are present in different buildings and under different seasons and climates. In the studies included in this thesis, the health effect of the mycobiome on the occupants could not be evaluated. Such analyses are difficult, due to the complexity of mycobiomes and health aspects of the occupants. In the future, more studies addressing which effect the mycobiome have on the occupant's health are needed. In this regard, it will be good to reduce the number of variables and rather focus on the ones with highest explanatory power, such as number of inhabitants.

The studies in this thesis demonstrates that DNA metabarcoding, based on HTS and error correction with DADA2 followed by clustering, is a powerful approach to investigate the indoor mycobiome. However, there are limitations associated with DNA metabarcoding analyses using a single marker, as multiple independent DNA markers are often required for proper species delineation. We are still not in a position to generate datasets with multiple unlinked markers from most environmental samples, although technical advancements, such as single cell technologies, may enable this in the future. Third generation sequencing technologies, such as PacBio and Oxford Nanopore, are promising to generate longer barcodes and improve the taxonomic resolution (Kennedy *et al.* 2018; Tedersoo *et al.* 2018). Another basic difficulty associated with fungal ITS metabarcoding is the different levels of intraspecific sequence variation across species. Our analyses, as well as previous literature, demonstrate that most fungal species include some levels of intraspecific sequence variation and that sequence clustering therefore is needed to approach species level resolution. More complete reference sequence databases and the implementation of reference-based delineation of species instead of de novo clustering, may be one future solution to separate species in a more dynamic way (Cline *et al.* 2017). However, this requires improvement of taxonomic coverage in current reference databases.

Our studies demonstrate clear differences in the dust mycobiome composition between indoor and outdoor environments, between rooms with different occupancy and between daycares and private homes. The more occupants and human activity, the more the indoor mycobiome differs

from the outdoor mycobiome composition. In addition, our findings are in line with previous indoor mycobiome studies, identifying climatic variables as the key determinants of indoor mycobiome. Our results demonstrate how the mycobiome composition follows a strong seasonal trend, mirroring outdoor weather conditions. Knowledge about the seasonal trends will have important implications for monitoring and evaluation of indoor air quality.

Acknowledgements

I would like to acknowledge Håvard Kauserud and Inger Skrede for valuable feedback on this introduction.

References

Adams RI, Bhangar S, Dannemiller KC, *et al.* (2016) Ten questions concerning the microbiomes of buildings. *Building and Environment* **109**, 224-234.

Adams RI, Miletto M, Taylor JW, Bruns TD (2013) Dispersal in microbes: fungi in indoor air are dominated by outdoor air and show dispersal limitation at short distances. *The ISME Journal* **7**, 1262-1273.

Amend AS, Seifert KA, Samson R, Bruns TD (2010) Indoor fungal composition is geographically patterned and more diverse in temperate zones than in the tropics. *Proceedings of the National Academy of Sciences* **107**, 13748-13753.

Andrew C, Heegaard E, Høiland K, *et al.* (2018) Explaining European fungal fruiting phenology with climate variability. *Ecology* **99**, 1306-1315.

Antonelli A, Fry C, Smith R, *et al.* (2020) State of the world's plants and fungi 2020. *Royal botanic gardens, Kew*.

Balasundaram SV, Hess J, Durling MB, *et al.* (2018) The fungus that came in from the cold: dry rot's pre-adapted ability to invade buildings. *The ISME journal* **12**, 791-801.

Barberán A, Dunn RR, Reich BJ, *et al.* (2015a) The ecology of microscopic life in household dust. *Proceedings of the Royal Society B: Biological Sciences* **282**.

Barberán A, Ladau J, Leff JW, *et al.* (2015b) Continental-scale distributions of dust-associated bacteria and fungi. *Proceedings of the National Academy of Sciences* **112**, 5756-5761.

Blaalid R, Kumar S, Nilsson RH, *et al.* (2013) ITS1 versus ITS2 as DNA metabarcodes for fungi. *Molecular Ecology Resources* **13**, 218-224.

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

Boddy L, Büntgen U, Egli S, *et al.* (2014) Climate variation effects on fungal fruiting. *Fungal Ecology* **10**, 20-33.

Bornehag CG, Blomquist G, Gyntelberg F, *et al.* (2001) Dampness in buildings and health. Nordic interdisciplinary review of the scientific evidence on associations between exposure to "dampness" in buildings and health effects (NORDDAMP). *Indoor Air* **11**, 72-86.

Botnen SS, Davey ML, Halvorsen R, Kauserud H (2018) Sequence clustering threshold has little effect on the recovery of microbial community structure. *Molecular Ecology Resources* **18**, 1064-1076.

Callahan BJ, McMurdie PJ, Holmes SP (2017) Exact sequence variants should replace operational taxonomic units in marker-gene data analysis. *The ISME Journal* **11**, 2639-2643.

Callahan BJ, Wong J, Heiner C, *et al.* (2019) High-throughput amplicon sequencing of the full-length 16S rRNA gene with single-nucleotide resolution. *Nucleic Acids Research* **47**, e103.

Caporaso JG, Kuczynski J, Stombaugh J, *et al.* (2010) QIIME allows analysis of high-throughput community sequencing data. *Nature Methods* **7**, 335-336.

Chen W, Hambleton S, Seifert KA, *et al.* (2018) Assessing performance of spore samplers in monitoring seromycobiota and fungal plant pathogen diversity in canada. *Applied and Environmental Microbiology* **84**, e02601-02617.

Cline LC, Song Z, Al-Ghalith GA, Knights D, Kennedy PG (2017) Moving beyond de novo clustering in fungal community ecology. *New Phytologist* **216**, 629-634.

Cohn JP (2008) Citizen science: can volunteers do real research? *BioScience* **58**, 192-197.

Crous PW, Schoch CL, Hyde KD, *et al.* (2009) Phylogenetic lineages in the Capnodiales. *Studies in Mycology* **64**, 17-47S17.

Dickinson JL, Zuckerberg B, Bonter DN (2010) Citizen science as an ecological research tool: challenges and benefits. *Annual Review of Ecology, Evolution, and Systematics* **41**, 149-172.

Dupuy AK, David MS, Li L, *et al.* (2014) Redefining the human oral mycobiome with improved practices in amplicon-based taxonomy: discovery of *Malassezia* as a prominent commensal. *PLOS ONE* **9**, e90899.

Edgar RC (2013) UPARSE: highly accurate OTU sequences from microbial amplicon reads. *Nature Methods* **10**, 996.

Frankel M, Bekö G, Timm M, *et al.* (2012) Seasonal variations of indoor microbial exposures and their relation to temperature, relative umidity, and air exchange rate. *Applied and Environmental Microbiology* **78**, 8289.

Gange AC, Gange EG, Sparks TH, Boddy L (2007) Rapid and recent changes in fungal fruiting patterns. *Science* **316**, 71.

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

Ghannoum MA, Jurevic RJ, Mukherjee PK, *et al.* (2010) Characterization of the oral fungal microbiome (mycobiome) in healthy individuals. *PLOS Pathogens* **6**, e1000713.

Gilbert JA, Stephens B (2018) Microbiology of the built environment. *Nature Reviews Microbiology* **16**, 661-670.

Goodwin S, McPherson JD, McCombie WR (2016) Coming of age: ten years of next-generation sequencing technologies. *Nature Reviews Genetics* **17**, 333-351.

Gura T (2013) Citizen science: amateur experts. *Nature* **496**, 259-261.

Hardin BD, Kelman BJ, Saxon A (2003) Adverse human health effects associated with molds in the indoor environment. *Journal of Occupational and Environmental Medicine* **45**, 470-478.

Hawksworth DL, Lücking R (2017) Fungal diversity revisited: 2.2 to 3.8 million species. *Microbiology Spectrum* **5**.

Hoffmann K, Pawłowska J, Walther G, *et al.* (2013) The family structure of the Mucorales: a synoptic revision based on comprehensive multigene-genealogies. *Persoonia* **30**, 57-76.

Horner WE (2003) Assessment of the indoor environment: evaluation of mold growth indoors. *Immunology and Allergy Clinics* **23**, 519-531.

Horve PF, Lloyd S, Mhuireach GA, *et al.* (2020) Building upon current knowledge and techniques of indoor microbiology to construct the next era of theory into microorganisms, health, and the built environment. *Journal of Exposure Science & Environmental Epidemiology* **30**, 219-235.

Inamdar AA, Bennett JW (2015) Volatile organic compounds from fungi isolated after hurricane katrina induce developmental defects and apoptosis in a *Drosophila melanogaster* model. *Environmental Toxicology* **30**, 614-620.

Jo J-H, Deming C, Kennedy EA, *et al.* (2016) Diverse human skin fungal communities in children converge in adulthood. *Journal of Investigative Dermatology* **136**, 2356-2363.

Karlsson E, Johansson A-M, Ahlinder J, *et al.* (2020) Airborne microbial biodiversity and seasonality in Northern and Southern Sweden. *PeerJ* **8**, e8424-e8424.

Kauserud H, Knudsen H, Högberg N, Skrede I (2012) Evolutionary origin, worldwide dispersal, and population genetics of the dry rot fungus *Serpula lacrymans*. *Fungal Biology Reviews* **26**, 84-93.

Kennedy PG, Cline LC, Song Z (2018) Probing promise versus performance in longer read fungal metabarcoding. *New Phytologist* **217**, 973-976.

Koljalg U, Larsson KH, Abarenkov K, *et al.* (2005) UNITE: a database providing web-based methods for the molecular identification of ectomycorrhizal fungi. *New Phytologist* **166**, 1063-1068.

Kõljalg U, Nilsson RH, Abarenkov K, *et al.* (2013) Towards a unified paradigm for sequence-based identification of fungi. *Molecular Ecology* **22**, 5271-5277.

Limon JJ, Skalski JH, Underhill DM (2017) Commensal fungi in health and disease. *Cell Host & Microbe* **22**, 156-165.

Lindahl BD, Nilsson RH, Tedersoo L, *et al.* (2013) Fungal community analysis by high-throughput sequencing of amplified markers--a user's guide. *New Phytologist* **199**, 288-299.

Lücking R, Huhndorf S, Pfister DH, Plata ER, Lumbsch HT (2009) Fungi evolved right on track. *Mycologia* **101**, 810-822.

Macher JM (2001) Review of methods to collect settled dust and isolate culturable microorganisms. *Indoor Air* **11**, 99-110.

Madureira J, Paciência I, Rufo JC, *et al.* (2015) Assessment and determinants of airborne bacterial and fungal concentrations in different indoor environments: Homes, child day-care centres, primary schools and elderly care centres. *Atmospheric Environment* **109**, 139-146.

Mahe F, Rognes T, Quince C, de Vargas C, Dunthorn M (2015) Swarm v2: highly-scalable and high-resolution amplicon clustering. *PeerJ* **3**, e1420.

Martin LJ, Adams RI, Bateman A, *et al.* (2015) Evolution of the indoor biome. *Trends in Ecology & Evolution* **30**, 223-232.

Mbareche H, Veillette M, Bilodeau G, Duchaine C (2020) Comparison of the performance of ITS1 and ITS2 as barcodes in amplicon-based sequencing of bioaerosols. *PeerJ* **8**, e8523.

Mendell MJ, Mirer AG, Cheung K, Tong M, Douwes J (2011) Respiratory and allergic health effects of dampness, mold, and dampness-related agents: A review of the epidemiologic evidence. *Environmental Health Perspectives* **119**, 748-756.

Mims TS, Abdallah QA, Stewart JD, *et al.* (2021) The gut mycobiome of healthy mice is shaped by the environment and correlates with metabolic outcomes in response to diet. *Communications Biology* **4**, 281.

Moore D, Gange AC, Gange EG, Boddy L (2008) Chapter 5 Fruit bodies: Their production and development in relation to environment. In: *British Mycological Society Symposia Series* (eds. Boddy L, Frankland JC, van West P), pp. 79-103. Academic Press.

Napoli C, Marcotrigiano V, Montagna MT (2012) Air sampling procedures to evaluate microbial contamination: a comparison between active and passive methods in operating theatres. *BMC Public Health* **12**, 594-594.

Naranjo-Ortiz MA, Gabaldón T (2019) Fungal evolution: diversity, taxonomy and phylogeny of the Fungi. *Biological reviews of the Cambridge Philosophical Society* **94**, 2101-2137.

Nevalainen A, Täubel M, Hyvärinen A (2015) Indoor fungi: companions and contaminants. *Indoor Air* **25**, 125-156.

Nilsson RH, Anslan S, Bahram M, *et al.* (2019) Mycobiome diversity: high-throughput sequencing and identification of fungi. *Nature Reviews Microbiology* **17**, 95-109.

Nilsson RH, Kristiansson E, Ryberg M, Hallenberg N, Larsson K-H (2008) Intraspecific ITS variability in the kingdom fungi as expressed in the international sequence databases and its implications for molecular species identification. *Evolutionary Bioinformatics* **4**, EBO.S653.

Pitkäranta M, Meklin T, Hyvärinen A, *et al.* (2008) Analysis of fungal flora in indoor dust by ribosomal DNA sequence analysis, quantitative PCR, and culture. *Applied and Environmental Microbiology* **74**, 233-244.

Prussin AJ, Marr LC (2015) Sources of airborne microorganisms in the built environment. *Microbiome* **3**, 78.

Reponen T, Nevalainen A, Jantunen M, Pellikka M, Kalliokoski P (1992) Normal range criteria for indoor air bacteria and fungal spores in a subarctic climate. *Indoor Air* **2**, 26-31.

Sanjar F, Weaver AJ, Peacock TJ, *et al.* (2020) Temporal shifts in the mycobiome structure and network architecture associated with a rat (Rattus norvegicus) deep partial-thickness cutaneous burn. *Medical Mycology* **58**, 107-117.

Schloss PD, Westcott SL, Ryabin T, *et al.* (2009) Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Applied and Environmental Microbiology* **75**, 7537-7541.

Schmidt O (2007) Indoor wood-decay basidiomycetes: damage, causal fungi, physiology, identification and characterization, prevention and control. *Mycological Progress* **6**, 281-281.

Schoch CL, Seifert KA, Huhndorf S, *et al.* (2012) Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for fungi. *Proceedings of the National Academy of Sciences* **109**, 6241-6246.

Scully C, el-Kabir M, Samaranayake LP (1994) Candida and oral candidosis: a review. *Critical Reviews in Oral Biology & Medicine* **5**, 125-157.

Singh J (1999) Dry rot and other wood-destroying fungi: their occurrence, biology, pathology and control. *Indoor and Built Environment* **8**, 3-20.

Stejskal V, Hubert J, Kubátová A, Váňová M (2005) Fungi associated with rodent feces in stored grain environment in the Czech Republic. *Journal of Plant Diseases and Protection* **112**, 98-102.

Taberlet P, Bonin A, Coissac E, Zinger L (2018) *Environmental DNA: for biodiversity research and monitoring.* Oxford University Press.

Taberlet P, Coissac E, Pompanon F, Brochmann C, Willerslev E (2012) Towards next-generation biodiversity assessment using DNA metabarcoding. *Molecular Ecology* **21**, 2045-2050.

Tedersoo L, Anslan S, Bahram M, *et al.* (2015) Shotgun metagenomes and multiple primer pair-barcode combinations of amplicons reveal biases in metabarcoding analyses of fungi. *MycoKeys* **10**, 1.

Tedersoo L, Bahram M, Põlme S, *et al.* (2014) Global diversity and geography of soil fungi. *Science* **346**, 1256688.

Tedersoo L, Lindahl B (2016) Fungal identification biases in microbiome projects. *Environmental Microbiology Reports* **8**, 774-779.

Tedersoo L, Tooming-Klunderud A, Anslan S (2018) PacBio metabarcoding of fungi and other eukaryotes: errors, biases and perspectives. *New Phytologist* **217**, 1370-1385.

Underhill DM, Iliev ID (2014) The mycobiota: interactions between commensal fungi and the host immune system. *Nature Reviews Immunology* **14**, 405-416.

Vesper S, McKinstry C, Haugland R, *et al.* (2007) Development of an environmental relative moldiness index for US homes. *Journal of Occupational and Environmental Medicine* **49**, 829-833.

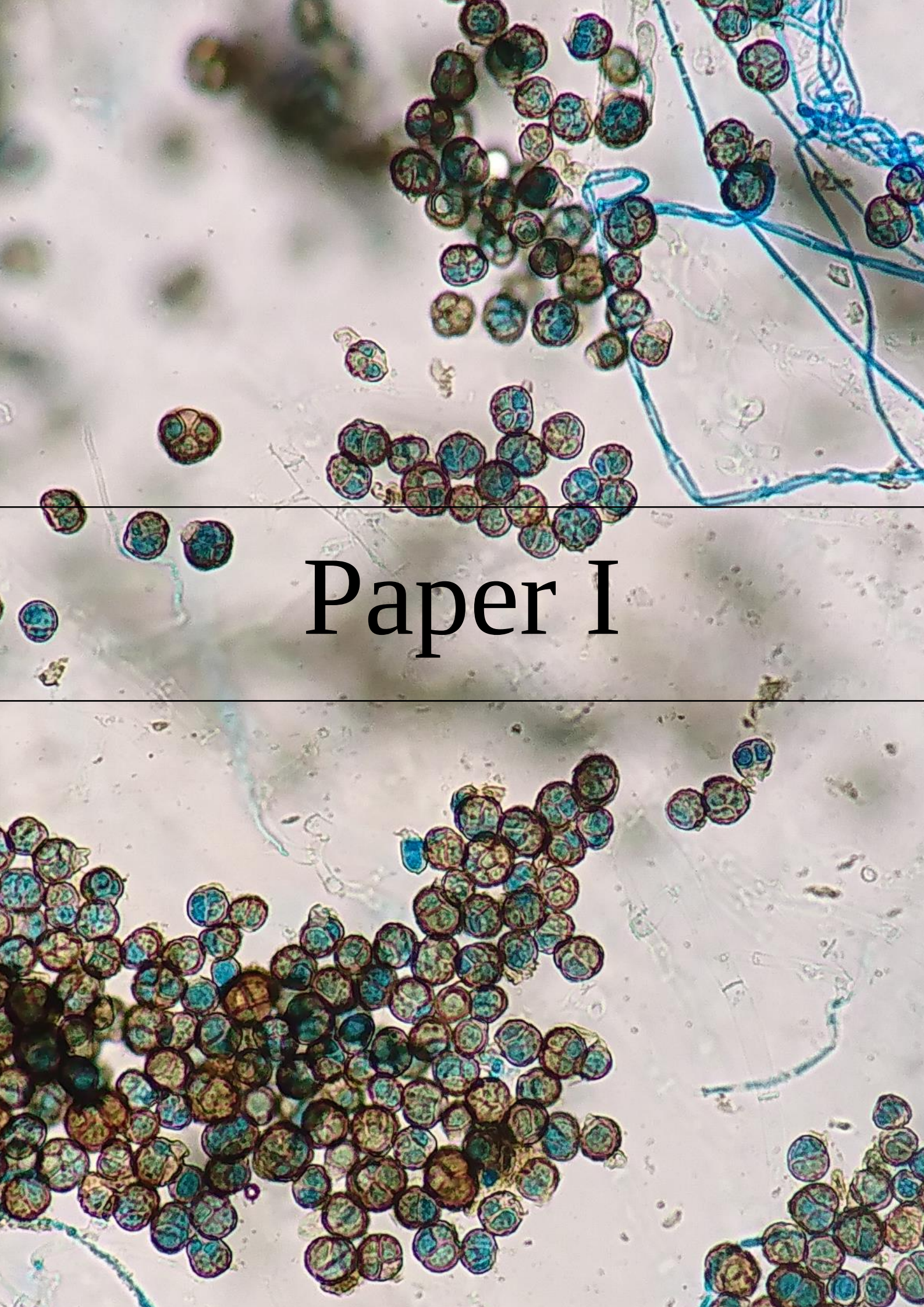
Weikl F, Tischer C, Probst AJ, *et al.* (2016) Fungal and bacterial communities in indoor dust follow different environmental determinants. *PLOS ONE* **11**, e0154131-e0154131.

Westcott SL, Schloss PD (2015) De novo clustering methods outperform reference-based methods for assigning 16S rRNA gene sequences to operational taxonomic units. *PeerJ* **3**, e1487.

White TC, Findley K, Dawson TL, *et al.* (2014) Fungi on the skin: dermatophytes and *Malassezia*. *Cold spring harbor perspectives in medicine* **4**, a019802.

WHO (2009) *Guidelines for Indoor Air Quality: Dampness and Mould* WHO, Geneva: World Health Organization.

Yamamoto N, Hospodsky D, Dannemiller KC, Nazaroff WW, Peccia J (2015) Indoor emissions as a primary source of airborne allergenic fungal particles in classrooms. *Environmental Science & Technology* **49**, 5098-5106.



Paper I

Received: 14 August 2020 | Revised: 20 November 2020 | Accepted: 6 January 2021

DOI: 10.1111/1755-0998.13329

RESOURCE ARTICLE

MOLECULAR ECOLOGY RESOURCES WILEY

The influence of intraspecific sequence variation during DNA metabarcoding: A case study of eleven fungal species

Eva Lena F. Estensmo[1] | Sundy Maurice[1] | Luis Morgado[1,2] | Pedro M. Martin-Sanchez[1] | Inger Skrede[1] | Håvard Kauserud[1]

[1]Section for Genetics and Evolutionary Biology (Evogene, Department of Biosciences, University of Oslo, Oslo, Norway

[2]Naturalis Biodiversity Center, RA Leiden, The Netherlands

Correspondence
Eva Lena F. Estensmo and Håvard Kauserud, Section for Genetics and Evolutionary Biology (Evogene), Department of Biosciences, University of Oslo, P. O. Box 1066 Blindern, 0316 Oslo, Norway
Email: eva.estensmo@outlook.com (ELFS); havard.kauserud@ibv.uio.no (HK)

Funding information
Norges Forskningsråd, Grant/Award Number: 254746

Abstract

DNA metabarcoding has become a powerful approach for analysing complex communities from environmental samples, but there are still methodological challenges limiting its full potential. While conserved DNA markers, like 16S and 18S, often are not able to discriminate among closely related species, other more variable markers – like the fungal ITS region, may include considerable intraspecific variation, which can lead to oversplitting of species during DNA metabarcoding analyses. Here we assessed the effects of intraspecific sequence variation in DNA metabarcoding by analysing local populations of eleven fungal species. We investigated the allelic diversity of ITS2 haplotypes using both Sanger sequencing and high throughput sequencing (HTS) coupled with error correction with the software DADA2. All the eleven species, except one, included some level of intraspecific variation in the ITS2 region. Overall, we observed a high correspondence between haplotypes generated by Sanger sequencing and HTS, with the exception of a few additional haplotypes detected using either approach. These extra haplotypes, typically occurring in low frequencies, were probably due to PCR and sequencing errors or intragenomic variation in the rDNA region. The presence of intraspecific (and possibly intragenomic) variation in ITS2 suggest that haplotypes (or ASVs) should not be used as basic units in ITS-based fungal community analyses, but an extra clustering step is needed to approach species-level resolution.

KEYWORDS
community ecology, DNA metabarcoding, fungi, haplotypes, ITS

1 | INTRODUCTION

High throughput sequencing (HTS) of amplified markers, i.e., DNA metabarcoding, has become a powerful tool to study microbial communities (Goodwin et al., 2016; Lindahl et al., 2013; Taberlet et al., 2012, 2018). DNA metabarcoding has considerably improved our understanding of the structure and function of microbial communities in different habitats (Bahram et al., 2018; Tedersoo et al., 2014), and is also a well-established approach for surveying the biodiversity (Barsoum et al., 2019) and ecosystem biomonitoring (Douglas et al., 2012; Stat et al., 2017).

The commonly used DNA barcoding region for microorganisms lie within the nuclear ribosomal DNA (rDNA). Parts of this region offer conserved primer sites that can be used to amplify broad taxonomic groups, combined with areas of high inter- and low intraspecific variation in-between, which can provide some degree of

taxonomic resolution. The most used rDNA barcoding markers for microorganisms include the internal transcribed spacer (ITS) region for fungi (Nilsson et al., 2008; Schoch et al., 2012), the 16S region for bacteria and archaea (Stackebrandt & Goebel, 1994) and the 18S region for microeukaryotes (Hadziavdic et al., 2014). Due to different evolutionary rates, these markers include contrasting levels of sequence variability and, thus, provide various levels of resolution. In general, the fungal ITS marker includes considerably more sequence variability compared to 18S, and consequently provides higher interspecific resolution, but also some degree of intraspecific variability (Nilsson et al., 2008; Schoch et al., 2012).

Although often ignored, the peculiarities of these taxonomic markers imply that the sequences should be processed differently during DNA metabarcoding analyses. For example, in the case of more conserved markers, like 16S and 18S, merging of taxa is a common problem as it underestimates the species diversity, while for the variable ITS marker, splitting of taxa based on intraspecific sequence variation is also a concern in community analyses. In addition, PCR and sequencing errors introduce artificial sequence variation that can be hard to disentangle from naturally occurring intraspecific sequence variability.

A wide array of different bioinformatics approaches has been developed to group and delineate the HTS data into biological entities that are used downstream in community analyses. One early approach was to cluster sequences into operational taxonomic units (OTUs; approximations for biological taxonomic entities) based on a fixed sequence similarity threshold, for example 97% (Caporaso et al., 2010; Edgar, 2013; Schloss et al., 2009; Westcott & Schloss, 2015). Later, more elaborate approaches were developed in order to better distinguish between PCR and sequencing artefacts and biological sequence variation (Boyer et al., 2016; Callahan et al., 2016; Mahe et al., 2015), and thus, return OTUs better at approximating the biological entities.

Although somewhat different solutions have been developed in various software (Pauvert et al., 2019), a common basic aim in the more recent methods is to identify the underlying haplotypes present in the template DNA, giving rise to all the sequence variability generated during PCR and sequencing. In a recent study (Callahan et al., 2019), it was shown that the software DADA2 is able to provide single-nucleotide resolution when analysing the entire bacterial 16S rDNA region. The term amplicon sequence variants (ASVs) has been coined for the output of DADA2 analyses, which are approximations for the underlying haplotypes. For conserved markers like 16S and 18S, where one single base pair (bp) difference can reflect, at least, differences between species and genera, ASVs can serve as input for downstream analyses (Callahan et al., 2017). However, for markers with high level of intraspecific variation, like the ITS marker used for fungi (Nilsson et al., 2008), this can be highly problematic since the diversity will be tremendously overestimated by treating each ITS haplotype as a biological entity in downstream statistical analyses. Hence, ASVs will (at best) represent different allelic variants of ITS region, while community ecology is typically based on species-level analyses. To correct for the intraspecific ITS variation, an extra clustering step may be needed to group haplotypes (or ASVs) into species-level OTUs. For fungi and the ITS region, it has been debated at which similarity level sequences should be clustered to approximate the species-level (Caporaso et al., 2010; Edgar, 2013; Westcott & Schloss, 2015). Several studies have indicated that 97% represents a reasonable approximation (Blaalid et al., 2013; Nilsson et al., 2008). However, such a general threshold might lead to splitting of some taxa and lumping of others (Blaalid et al., 2013).

Despite the high level of intraspecific ITS sequence variation in fungi, we have little knowledge on how this variation translates into OTU delineation in DNA metabarcoding studies of fungal communities. Here, we assessed how DNA metabarcoding, using the fungal ITS2 marker, is able to deal with intraspecific sequence variation, and to what degree this variation leads to oversplitting of taxa. To address this topic, we performed DNA metabarcoding on 177 fungal specimens of 11 basidiomycetes species and compared their ASVs to the corresponding Sanger sequences. By denoising the sequence data using DADA2 (Callahan et al., 2016), we tested whether the same ITS2 haplotypes were identified by DNA metabarcoding and Sanger sequencing, and to what degree further sequence clustering is needed to approach species-level resolution.

2 | MATERIALS AND METHODS

Eleven wood-decay fungal species (Table 1) were sampled in an old-growth spruce forest in southeastern Finland (Issakka, Kuhmo). For each species, 16 individual fruit bodies were collected on distinct spruce logs. Given that these fungi typically spread by sexual basidiospores, no clonal dispersal between spruce logs is expected. The fruit body tissue of these fungi is made up of dikaryotic hyphae, and heterozygous genotypes are therefore expected if intraspecific ITS2 variation is present (see Figure S1 for example).

Between 10–15 small pieces of approximately 5 mm^2 of tissue were cut out from each fruit body and grinded in 800 µl of 2% CTAB and 1% beta-mercaptoethanol using a Retsch MM200 mixer (4 × 45 s at 25 oscillations). DNA was extracted using a modified CTAB extraction protocol (Gardes & Bruns, 1993; Murray & Thompson, 1980) and cleaned with the E. Z. N. A Soil DNA kit (Omega Biotek) by adding the HTR reagent and then following the manufacturer´s guidelines. DNA was eluted in 100 µl elution buffer, quantified with Qubit ds DNA BR Assay kit (Life Technologies) and standardized with 10 mM Tris to a concentration range of 5–10 ng/µl.

The 177 DNA samples extracted from individual fruit bodies were distributed into 2 x 96 PCR plates. Eleven samples were included as replicates, with at least one sample for each species, which were equally distributed between the plates. Each PCR plate also contained a negative PCR sample, together with a mock community composed of six fungal basidiomycetes species (*Trametes versicolor*, *Climacocystis borealis*, *Cystostereum murrayi*, *Serpula lacrymans*, *Heterobasidion annosum*, *Heterobasidion occidentale*) with low expected prevalence in our data set. The resulting 192 samples were processed into two libraries using a combination of 96 uniquely

tagged primers with tags (*x*) ranging from 6–9 bp. The fungal ITS2 region was targeted with the gITS7 (5′-*x*GTGA*R*TCATCGA*R*TCTTTG) (Ihrmark et al., 2012) and ITS4 (5′-*x*CTCCGCTTATTGATATG) (White et al., 1990) primers. The PCR mixture in 25 µl final volume consisted of 14.6 µl Milli-Q water, 2.5 µl 10x Gold buffer, 0.2 µl dNTP's (25 nM), 1.5 µl reverse and forward primers (10 µM), 2.5 µl $MgCl_2$ (50 mM), 1.0 µl BSA (20 mg/ml), 0.2 µl AmpliTaq Gold polymerase (5 U/µl) and 5–10 ng/µl of DNA template. The following cycling parameters were used for amplification: enzyme activation at 95°C for 5 min, followed by 25 cycles of denaturation at 95°C for 30 s, annealing at 55°C for 30 s, extension at 72°C for 1 min, and a final extension step at 72°C for 10 min.

The quality of PCR products was controlled by electrophoresis on a 2% agarose gel prior to normalization using the SequalPrep Normalization Plate Kit (Invitrogen) and eluted in 20 µl elution buffer. The 96 PCR products within each library were pooled, concentrated and purified using Agencourt AMPure XP magnetic beads (Nerliens Meszansky AS) and the DNA concentration was measured with Qubit ds DNA BR Assay kit (Life Technologies). The two libraries were barcoded with Illumina adapters, spiked with 20% PhiX and sequenced in one Illumina MiSeq (Illumina) lane with 2 × 300 bp paired-end reads at StarSEQ (StarSEQ GmbH).

For comparison, we generated Sanger sequences of ITS2 for the 177 fruiting bodies. Amplification was performed with ITS3 (5′-GCATCGATGAAGAACGCAGC) and ITS4 (5′-TCCTCCGCTTATTGATATGC) primers (White et al., 1990), with the same PCR mix and program as above. The resulting amplicons were cleaned with ExoProStar (Sigma Aldrich) and sequenced in both directions by Eurofins Genomics (Ebersberg).

The resulting metabarcoding data set comprised 25,953,804 reads. The sequences were demultiplexed with CUTADAPT v. 2.7 (Martin, 2011) and low quality reads were removed (at least 26 bp overlap between query and target, no indel and minimum length of 100 bp). DADA2 v. 12 (Callahan et al., 2016) was used to filter low quality sequences, with a maximum expected error of 2.5 and without the TruncLen option in order to keep the sequence length. The samples were dereplicated, error corrected and denoised (dada: err = NULL, pool = "pseudo", selfConsist = TRUE). We then merged the error corrected forward and reverse sequences using a minimum overlap of 5 bp. Chimeras were filtered out using the bimera algorithm (method = "pooled"), and the remaining sequences were used to create the ASVs table. Otherwise, we used default options for unspecified parameters. Taxonomy was assigned using BLAST v. 2.8.1 to the raw ASVs by the UNITE database v. 8.0 (Koljalg et al., 2005) (evalue 1×10^{-4}, max_target_seqs 1). The resulting ASV table consisted of 2,965,749 reads accounting for 3,647 ASVs. For downstream analyses, we retained only 57 ASVs which were assigned to the 11 target species, and excluded the numerous others appearing in the HTS data, which belonged to fungicolous fungi growing inside the fruit bodies. The technical replicates provided largely the same ASVs. However, there were some indications of minor levels of tag-switching, leading to the presence of some ASVs in other samples in very low sequence abundance (see Table S2). The technical replicates were excluded from further analyses. The replicated mock community provided exactly the same ASVs.

Both the ASVs and the Sanger sequences were further processed in GENEIOUS PRIME v. 2020.0.5 (https://www.geneious.com). The Sanger sequences were manually curated and poor-quality sequences were excluded from the data set. Heterozygous sites were characterized according to the IUPAC nucleotide code, and the forward and reverse reads were merged when possible (depending on quality). Separate sequence alignments were generated from ASVs and Sanger sequences, which were then concatenated to a joint alignment for each species.

TABLE 1 Comparison of Sanger and HTS sequences. Only specimens for which sequences were obtained from both approaches are shown. Sequence length (base pair) is the overlap between Sanger and HTS sequence alignments, number of dephased sequences correspond to the ITS2 sequence from each dikaryotic (n + n) individual and ASVs stands for amplicon sequence variants. Total haplotypes (Hap.) include common haplotypes from both Sanger and HTS and additional haplotypes identified by either approach

			Sanger sequences			ASVs			
Species	Specimen	Sequence length (bp)	Dephased sequences	Polymorphic sites	Hap.	Reads	Polymorphic sites	Hap.	Total haplotypes
Amylocystis lapponica	9	308	18	0	1	66,364	0	1	1
Antrodia serialis	16	188	32	3	4	178,787	3	4	4
Fomitopsis pinicola	16	229	32	5	5	238,143	7	8	9
Fomitopsis rosea	15	277	30	5	5	168,991	6	8	8
Gloeophyllum separium	16	271	32	5	5	429,990	5	5	6
Phlebia centrifuga	16	276	32	1	2	355,067	10	4	4
Phellinus ferrugineofuscus	16	282	32	5	4	178,295	5	5	5
Phellopilus nigrolimitatus	6	296	12	9	6	46,677	8	7	11
Phellinus viticola	16	281	32	4	5	120,306	4	5	5
Postia caesia	10	232	20	2	3	191,483	10	4	4
Trichaptum abietinum	15	268	30	7	5	148,066	7	6	8

The Sanger sequences, many with heterozygous sites due to allelic variability in the dikaryotic tissue, were dephased i.e., the consensus sequence of each sample was split into two homozygous sequence strands, and analysed for DNA polymorphisms in DNASP v. 6 (Rozas et al., 2017). Hence, for each species, we obtained one haplotype data set from the Sanger sequences and another from the HTS and compared their relative abundance in R (v. 3.6.2; R Core Team, 2018). Haplotype networks for the 11 species were generated with POPART v. 1.7 (Leigh & Bryant, 2015), displaying the level of intraspecific variation in ITS2. For the calculation of haplotype networks, indels were included as characters, where multiposition gaps were scored as one mutational event. A biplot showing the correspondence in relative abundance of each haplotype across the two data sets was made in R (v. 3.6.2; R Core Team, 2019). At last, the sequences from the two haplotype data sets were clustered with 97% identity by VSEARCH v. 2.13 (Rognes et al., 2016). This step included clustering (--sizein --sizeout, --minseqlength 10, --qmask none, --centroids), sorting and filtering and mapping of reads against OTU representatives (--usearch_global, --minseqlength 10, --strand plus, --maxaccepts 0, --qmask none).

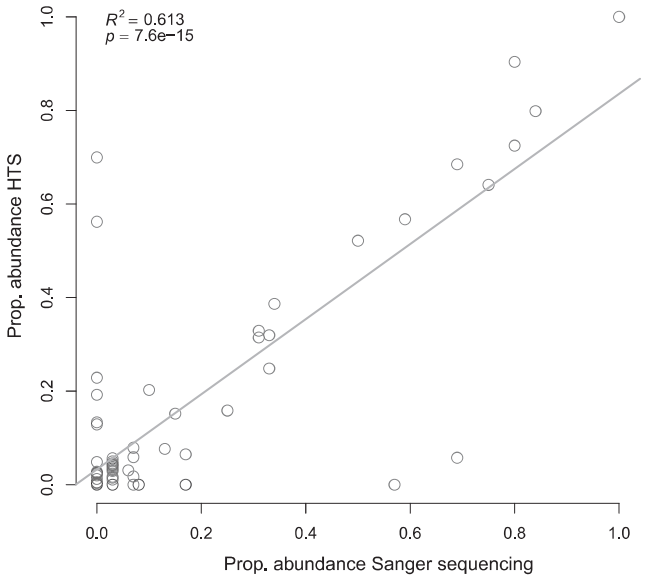


FIGURE 1 Biplot showing the correspondence of haplotypes in relative abundance across the two datasets (Sanger sequencing on x-axis and HTS on y-axis)

3 | RESULTS

We obtained high quality ITS2 Sanger sequences for 151 out of 177 fruit bodies, ranging from 6 to 16 fruit bodies per species. The remaining fruit bodies either did not amplify or resulted in low-quality sequences, due to fungicolous fungi growing inside the fruit bodies (generating multiple templates) or high level of heterozygosity of indels, leading to chromatograms that were hard to interpret. The ITS2 sequences were dephased into one to six ITS2 haplotypes per species (Table 1), identifying a total 45 haplotypes from the Sanger data set. For all species, except *Amylocystis lapponica* represented by a single haplotype, some level of intraspecific ITS2 sequence variation were present in the local population.

Although we obtained HTS data for 163 out of 177 fruit bodies distributed across the eleven species, for comparative purposes we only focused on the specimens for which Sanger sequences were available. After removing all ITS2 sequences corresponding to fungicolous fungi, a total of 2,316,395 ITS2 sequences were attributed to the 11 target species. After denoising the sequences using DADA2 and removing five additional chimeric sequences, we identified between one and eight haplotypes (ASVs) for each species (Table 1), totaling 57 haplotypes. The technical replicates (one sample from each species) provided largely the same results, except for the presence of some low abundance ASVs that were probably caused by tag-switching from other samples (i.e., these ASV were not unique but occurred with high abundance in other samples; Table S2).

Overall, we detected 65 different haplotypes from the combined data set, of which 37 (57%) were shared between the two approaches (Table S1), eight (12.3%) only from Sanger sequencing, while 20 (30.8%) were specific to the HTS data. With some exceptions, a high correspondence was found in the relative abundance of haplotypes across the two data sets (Figure 1). The haplotype networks (Figure 2) illustrate the relationship between the haplotypes identified from the two data sets and demonstrate the level of intraspecific variation across species, varying from one haplotype (in *Amylocystis lapponica*) to 11 (in *Phellopilus nigrolimitatus*). The networks also indicate that most haplotypes were closely related, separated by a few mutational steps. Five haplotypes were present in very low abundances in the HTS data set, 10-fold lower than what would be expected from a single allele being present in the population (i.e., total read number divided by number of alleles, Table S1). These rare haplotypes probably represent PCR and sequencing errors, or alternatively, intragenomic variation.

After clustering the sequences at 97% identity, we obtained 13 clusters or OTUs for the 11 species. Each species was represented by one OTU, except for two, *Phellopilus nigrolimitatus* and *Phlebia centrifuga*, which were represented by two OTUs.

4 | DISCUSSION

In general, we observed a good correspondence between the two methods, Sanger sequencing versus DNA metabarcoding, in assessing allelic variation in the ITS2 marker across the 11 fungal species, with 57% of the detected haplotypes shared across the two data sets. We also observed a high correlation in relative abundances of haplotypes across the data sets, where the most striking mismatches were caused by single bp indels. The additional haplotypes detected by one of the approaches can either be due to methodological errors introduced at various steps, or they may represent de facto sequence variation that one of the methods failed to detect.

In some fungal species, intragenomic variation in ITS occurs due to lack of concerted evolution homogenizing the paralogs (Lindner

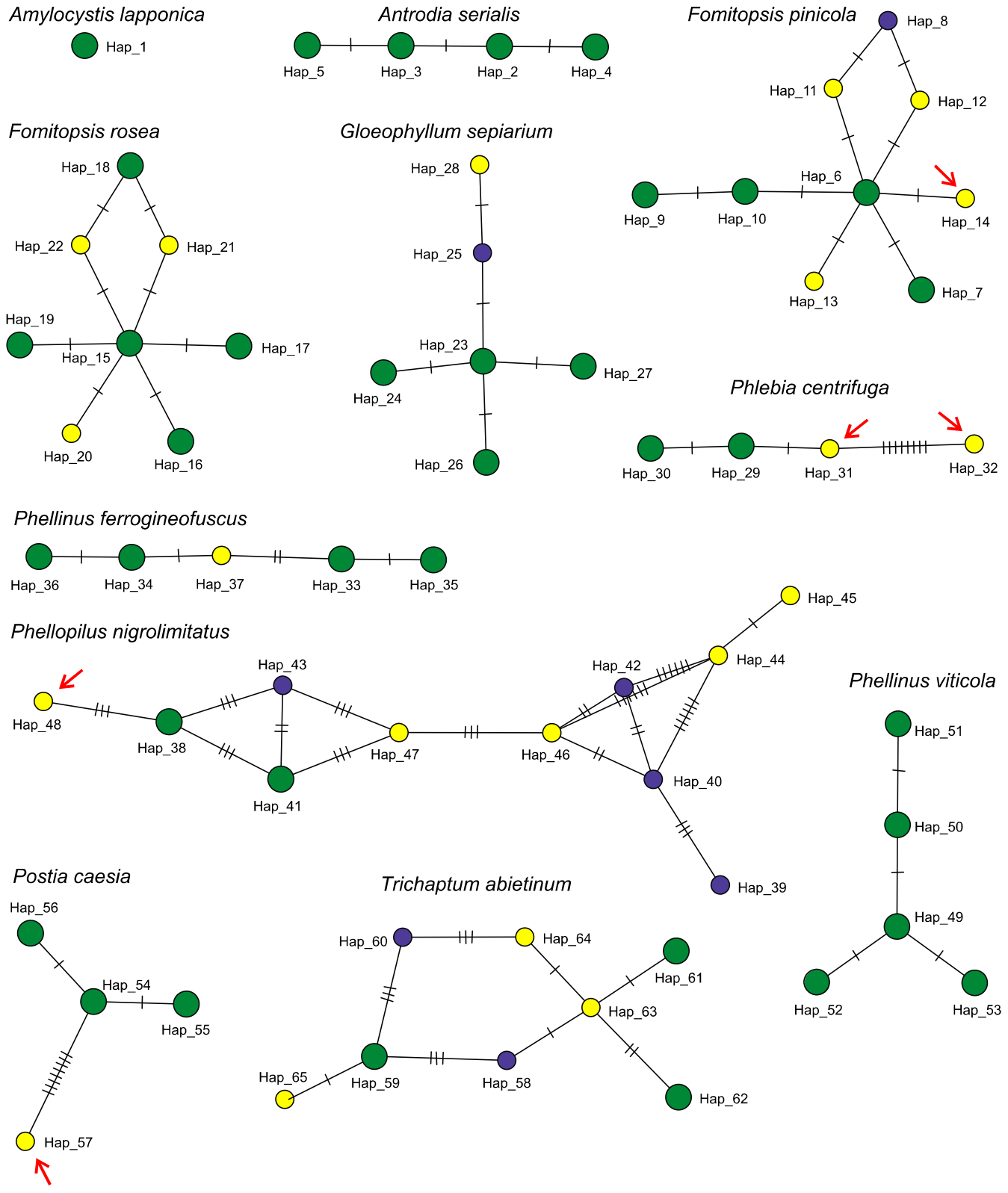


FIGURE 2 Haplotype networks displaying the level of intraspecific variation in ITS2 for the 11 fungal species. Each circle represents one haplotype and each dash represent one mutational step. Green colour indicates haplotypes detected both by Sanger sequencing and HTS, yellow haplotypes from HTS data set, and blue haplotypes were only detected by Sanger sequencing. Red arrows indicate haplotypes occurring with very low sequence abundance in the HTS data i.e., 10 times below what would be expected from a single allele in the population. The naming of haplotypes (Hap_1 to Hap_65) follows Table S1

& Banik, 2011). Such variation is hard to detect with direct Sanger sequencing, since a consensus sequence is derived from the multiple DNA templates. Although intragenomic ITS paralogs are rare (Lindner et al., 2013), we cannot rule out the possibility that some of the extra haplotypes detected by HTS represent ITS paralogues.

Alternatively, some of the unique haplotypes appearing in low abundance in the DNA metabarcoding data set might be due to PCR errors introduced during the initial PCR cycles and that DADA2 failed to identify these as artifacts. Although the DADA2 algorithm has a chimeric sequence filter implemented, five obvious chimeric haplotypes occurred in the filtered DNA metabarcoding data set, with chimeric breakpoints towards either the beginning or the end of the sequences. This exemplifies that a few haplotypes (ASVs), can be erroneous even after DADA2 processing. By analysing full-length 16S rRNA of mock communities of bacteria sequenced with PacBio SCC, a high correspondence was detected between the original templates and the obtained ASVs (Callahan et al., 2016). However, also in this case, some additional ASVs detected were either due to PCR or sequencing errors, or alternatively, intragenomic 16S variation (Větrovský & Baldrian, 2013). It is important to keep in mind that ASVs are probabilistic sequence reconstruction based on error models and thus have an associated uncertainty. Analyses of the technical replicates, as well as the replicated mock communities, demonstrated a high level of consistency, indicating that low levels of errors are introduced during PCR and sequencing, except for what seems to be a low level of tag-switching, the latter having no consequences on our results. When it comes to the additional haplotypes in the Sanger data set, this could result from erroneous dephasing of the original Sanger sequences or wrong basecalls due to ambiguous peaks.

For all the target species, except one, some level of intraspecific variation in the ITS2 region was detected, even at the fine geographic scale (i.e., a single forest). This corresponds well with the previous literature on intraspecific ITS variability in the fungal kingdom (Nilsson et al., 2008; Smith et al., 2007). Nilsson et al. (2008) reported an intraspecific sequence variability of 3.33% (± standard deviation of 5.62) for Basidiomycota. For some of the target species, sequence variation in the ITS region has been previously reported across regional spatial scales (Kauserud & Schumacher, 2002, 2003), and are in line with our results. The level of intraspecific ITS sequence variation in fungi varies widely, from some extreme cases approaching 8% sequence divergence in ITS (Kauserud & Schumacher, 2002), to some species showing identical haplotypes across broad areas (Kauserud et al., 2007). These varying levels reflects the species natural histories, where evolutionary old species with high population sizes may include higher sequence variability in ITS compared to more recent species with smaller populations, that might have experienced recent genetic bottlenecks.

It has recently been advocated to use the term ASVs (in our study largely referred to as haplotypes) as the basic units in microbial community analyses (Callahan et al., 2017). Indeed, this is a reasonable approach for conserved markers, like 16S and 18S, when a single bp mutation may separate between species or even genera. This is however not the case for variable markers with intraspecific variation. Our results show that in the variable ITS marker, a clustering step is needed after error correction to approach species-level resolution. After clustering our haplotypes, we obtained 13 OTUs representing the 11 species, with two species represented by two haplotypes. The importance of clustering depends on the study aims. In studies emphasizing beta diversity (community turnover), it has previously been shown that comparable results can be obtained using ASVs or OTUs representing sequence clusters (Glassman & Martiny, 2018). In line with this, Botnen et al. (2018) demonstrated that beta diversity patterns are highly robust against different clustering levels, ranging from 85% sequence similarity to 99%, both for ITS and 16S data. The most abundant OTUs (or ASVs) drive the community pattern and they largely show the same distributions across different data treatments (Botnen et al., 2018).

According to our results, we conclude that DNA metabarcoding, based on HTS and error-correction with DADA2, to a large extent reflects the allelic variation in natural populations and is a powerful approach to resolve complex communities. However, there are inherent limitations associated with single-locus DNA metabarcoding analyses. Multiple independent DNA markers are often required for proper species delineation. Yet, we are still not in a position to generate multilocus data sets from most environmental samples, although technical advancements in for example single-cell genomics may enable this in the future. Third generation sequencing technologies (e.g., PacBio, Oxford Nanopore) are promising to generate longer barcodes (e.g., 500–1,500 bp for 16S, >700 bp for ITS and 650 bp for COI) and improve taxonomic resolution (Kennedy et al., 2018; Tedersoo et al., 2018). Another basic difficulty associated with fungal ITS metabarcoding is the different levels of intraspecific sequence variation across species. Our analyses, as well as previous literature, demonstrate that most fungal species include some levels of intraspecific sequence variation and that sequence clustering therefore is needed to approach species level resolution. More complete reference sequence databases and the implementation of reference-based delineation of species instead of de novo clustering, may be one future solution (Cline et al., 2017). However, the taxonomic coverage in current reference databases are still generally too shallow for this approach.

ACKNOWLEDGEMENTS

We acknowledge Teppo Helo, Ari Meriruoko from Metsähallitus and Gergely Várkonyi from the Friendship Park Research Center in Kuhmo for help with sampling organization. The research was financially supported by the Research Council of Norway (grant No 254746) and the Department of Biosciences at the University of Oslo.

AUTHOR CONTRIBUTIONS

H.K., S.M. and E.L.F.E. designed and conceptualized the research. S.M. and H.K. performed sampling. S.M. processed samples and extracted DNA. S.M. and E.L.F.E. performed HTS library preparation

and Sanger sequencing. E.L.F.E. analysed data with contributions from H. K., S. M., and L. M. E.L.F.E., S.M. and H.K. wrote the manuscript. All authors edited and approved the manuscript.

DATA AVAILABILITY STATEMENT

The MiSeq raw sequence data is available on the NCBI short read archive (SUB8582638) under Bioproject PRJNA680258. Consensus Sanger sequences are deposited in GenBank (Submission SUB8713788) with corresponding accession numbers provided in Table S3. The bioinformatics codes, together with the OTU table, the metadata, intermediate files for generating the haplotype network and the raw Sanger sequences are deposited on Dryad (https://doi.org/10.5061/dryad.h18931zjq). Supporting Information is available for download on the *Molecular Ecology Resources* website.

ORCID

Eva Lena F. Estensmo https://orcid.org/0000-0001-6215-3166
Sundy Maurice https://orcid.org/0000-0002-5376-0981
Pedro M. Martin-Sanchez https://orcid.org/0000-0003-2674-8650
Inger Skrede https://orcid.org/0000-0002-1113-7403
Håvard Kauserud https://orcid.org/0000-0003-2780-6090

REFERENCES

Bahram, M., Hildebrand, F., Forslund, S. K., Anderson, J. L., Soudzilovskaia, N. A., Bodegom, P. M., Bengtsson-Palme, J., Anslan, S., Coelho, L. P., Harend, H., Huerta-Cepas, J., Medema, M. H., Maltz, M. R., Mundra, S., Olsson, P. A., Pent, M., Põlme, S., Sunagawa, S., Ryberg, M., … Bork, P. (2018). Structure and function of the global topsoil microbiome. *Nature*, *560*, 233–237.

Barsoum, N., Bruce, C., Forster, J., Ji, Y.-Q., & Yu, D. W. (2019). The devil is in the detail: Metabarcoding of arthropods provides a sensitive measure of biodiversity response to forest stand composition compared with surrogate measures of biodiversity. *Ecological Indicators*, *101*, 313–323.

Blaalid, R., Kumar, S., Nilsson, R. H., Abarenkov, K., Kirk, P. M., & Kauserud, H. (2013). ITS1 versus ITS2 as DNA metabarcodes for fungi. *Molecular Ecology Resources*, *13*, 218–224.

Botnen, S. S., Davey, M. L., Halvorsen, R., & Kauserud, H. (2018). Sequence clustering threshold has little effect on the recovery of microbial community structure. *Molecular Ecology Resources*, *18*, 1064–1076.

Boyer, F., Mercier, C., Bonin, A., Le Bras, Y., Taberlet, P., & Coissac, E. (2016). Obitools: A unix-inspired software package for DNA metabarcoding. *Molecular Ecology Resources*, *16*, 176–182.

Callahan, B. J., McMurdie, P. J., & Holmes, S. P. (2017). Exact sequence variants should replace operational taxonomic units in marker-gene data analysis. *ISME Journal*, *11*, 2639–2643.

Callahan, B. J., McMurdie, P. J., Rosen, M. J., Han, A. W., Johnson, A. J. A., & Holmes, S. P. (2016). DADA2: High-resolution sample inference from Illumina amplicon data. *Nature Methods*, *13*, 581.

Callahan, B. J., Wong, J., Heiner, C., Oh, S., Theriot, C. M., Gulati, A. S., McGill, S. K., & Dougherty, M. K. (2019). High-throughput amplicon sequencing of the full-length 16S rRNA gene with single-nucleotide resolution. *Nucleic Acids Research*, *47*, e103.

Caporaso, J. G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F. D., Costello, E. K., Fierer, N., Peña, A. G., Goodrich, J. K., Gordon, J. I., Huttley, G. A., Kelley, S. T., Knights, D., Koenig, J. E., Ley, R. E., Lozupone, C. A., McDonald, D., Muegge, B. D., Pirrung, M., … Knight, R. (2010). QIIME allows analysis of high-throughput community sequencing data. *Nature Methods*, *7*, 335–336.

Cline, L. C., Song, Z., Al-Ghalith, G. A., Knights, D., & Kennedy, P. G. (2017). Moving beyond de novo clustering in fungal community ecology. *New Phytologist*, *216*, 629–634.

Douglas, W. Y., Ji, Y., Emerson, B. C., Wang, X., Ye, C., Yang, C., & Ding, Z. (2012). Biodiversity soup: Metabarcoding of arthropods for rapid biodiversity assessment and biomonitoring. *Methods in Ecology and Evolution*, *3*, 613–623.

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

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

Glassman, S. I., & Martiny, J. B. (2018). Broadscale ecological patterns are robust to use of exact sequence variants versus operational taxonomic units. *mSphere*, *3*(4), https://doi.org/10.1128/mSphere.00148-18

Goodwin, S., McPherson, J. D., & McCombie, W. R. (2016). Coming of age: Ten years of next-generation sequencing technologies. *Nature Reviews Genetics*, *17*, 333–351.

Hadziavdic, K., Lekang, K., Lanzen, A., Jonassen, I., Thompson, E. M., & Troedsson, C. (2014). Characterization of the 18S rRNA gene for designing universal eukaryote specific primers. *PLoS One*, *9*, e87624.

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

Kauserud, H., & Schumacher, T. (2002). Population structure of the endangered wood decay fungus Phellinus nigrolimitatus (Basidiomycota). Canadian journal of botany. *Journal Canadien De Botanique*, *80*, 59–606.

Kauserud, H., & Schumacher, T. (2003). Regional and local population structure of the pioneer wood-decay fungus Trichaptum abietinum. *Mycologia*, *95*, 416–425.

Kauserud, H., Svegården, I. B., Saetre, G. P., Knudsen, H., Stensrud, Ø., Schmidt, O., Doi, S., Sugiyama, T., & Högberg, N. (2007). Asian origin and rapid global spread of the destructive dry rot fungus Serpula lacrymans. *Molecular Ecology*, *16*, 3350–3360.

Kennedy, P. G., Cline, L. C., & Song, Z. (2018). Probing promise versus performance in longer read fungal metabarcoding. *New Phytologist*, *217*, 973–976.

Koljalg, U., Larsson, K. H., Abarenkov, K., Nilsson, R. H., Alexander, I. J., Eberhardt, U., Erland, S., Hoiland, K., Kjoller, R., Larsson, E., Pennanen, T., Sen, R., Taylor, A. F. S., Tedersoo, L., Vralstad, T., & Ursing, B. M. (2005). UNITE: A database providing web-based methods for the molecular identification of ectomycorrhizal fungi. *New Phytologist*, *166*, 1063–1068.

Leigh, J. W., & Bryant, D. (2015). popart: Full-feature software for haplotype network construction. *Methods in Ecology and Evolution*, *6*, 1110–1116.

Lindahl, B. D., Nilsson, R. H., Tedersoo, L., Abarenkov, K., Carlsen, T., Kjoller, R., Koljalg, 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 Phytologist*, *199*, 288–299.

Lindner, D. L., & Banik, M. T. (2011). Intragenomic variation in the ITS rDNA region obscures phylogenetic relationships and inflates estimates of operational taxonomic units in genus Laetiporus. *Mycologia*, *103*, 731–740.

Lindner, D. L., Carlsen, T., Henrik Nilsson, R., Davey, M., Schumacher, T., & Kauserud, H. (2013). Employing 454 amplicon pyrosequencing to reveal intragenomic divergence in the internal transcribed spacer rDNA region in fungi. *Ecology and Evolution*, *3*, 1751–1764.

Mahe, F., Rognes, T., Quince, C., de Vargas, C., & Dunthorn, M. (2015). Swarm v2: Highly-scalable and high-resolution amplicon clustering. *PeerJ*, *3*, e1420.

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal*, *17*, 10–12.

Murray, M. G., & Thompson, W. F. (1980). Rapid isolation of high molecular weight plant DNA. *Nucleic Acids Research*, *8*, 4321–4325.

Nilsson, R. H., Kristiansson, E., Ryberg, M., Hallenberg, N., & Larsson, K.-H. (2008). Intraspecific ITS variability in the kingdom fungi as expressed in the international sequence databases and its implications for molecular species identification. *Evolutionary Bioinformatics*, *4*, EBO.S653.

Pauvert, C., Buée, M., Laval, V., Edel-Hermann, V., Fauchery, L., Gautier, A., Lesur, I., Vallance, J., & Vacher, C. (2019). Bioinformatics matters: The accuracy of plant and soil fungal community data is highly dependent on the metabarcoding pipeline. *Fungal Ecology*, *41*, 23–33.

R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing. Austria: Vienna. https://www.R-project.org/;2018

Rognes, T., Flouri, T., Nichols, B., Quince, C., & Mahé, F. (2016). VSEARCH: A versatile open source tool for metagenomics. *PeerJ Preprints*, *4*, e2409v2401.

Rozas, J., Ferrer-Mata, A., Sanchez-DelBarrio, J. C., Guirao-Rico, S., Librado, P., Ramos-Onsins, S. E., & Sanchez-Gracia, A. (2017). DnaSP 6: DNA sequence polymorphism analysis of large data sets. *Molecular Biology Evolution*, *34*, 3299–3302.

Schloss, P. D., Westcott, S. L., Ryabin, T., Hall, J. R., Hartmann, M., Hollister, E. B., Lesniewski, R. A., Oakley, B. B., Parks, D. H., Robinson, C. J., Sahl, J. W., Stres, B., Thallinger, G. G., Van Horn, D. J., & Weber, C. F. (2009). Introducing mothur: Open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Applied and Environmental Microbiology*, *75*, 7537–7541.

Schoch, C. L., Seifert, K. A., Huhndorf, S., Robert, V., Spouge, J. L., Levesque, C. A., Chen, W., Bolchacova, E., Voigt, K., Crous, P. W., Miller, A. N., Wingfield, M. J., Aime, M. C., An, K.-D., Bai, F.-Y., Barreto, R. W., Begerow, D., Bergeron, M.-J., Blackwell, M., … Schindel, D. (2012). Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for fungi. *Proceedings of the National Academy of Sciences USA*, *109*, 6241–6246.

Smith, M. E., Douhan, G. W., & Rizzo, D. M. (2007). Intra-specific and intra-sporocarp ITS variation of ectomycorrhizal fungi as assessed by rDNA sequencing of sporocarps and pooled ectomycorrhizal roots from a quercus woodland. *Mycorrhiza*, *18*, 15–22.

Stackebrandt, E., & Goebel, B. M. (1994). Taxonomic note: A place for DNA-DNA reassociation and 16S rRNA sequence analysis in the present species definition in bacteriology. *International Journal of Systematic and Evolutionary Microbiology*, *44*, 846–849.

Stat, M., Huggett, M. J., Bernasconi, R., DiBattista, J. D., Berry, T. E., Newman, S. J., Harvey, E. S., & Bunce, M. (2017). Ecosystem biomonitoring with eDNA: Metabarcoding across the tree of life in a tropical marine environment. *Scientific Reports*, *7*, 1–11.

Taberlet, P., Bonin, A., Coissac, E., & Zinger, L. (2018). *Environmental DNA: For biodiversity research and monitoring*. Oxford University Press.

Taberlet, P., Coissac, E., Pompanon, F., Brochmann, C., & Willerslev, E. (2012). Towards next-generation biodiversity assessment using DNA metabarcoding. *Molecular Ecology*, *21*, 2045–2050.

Tedersoo, L., Bahram, M., Põlme, S., Kõljalg, U., Yorou, N. S., Wijesundera, R., Ruiz, L. V., Vasco-Palacios, A. M., Thu, P. Q., Suija, A., Smith, M. E., Sharp, C., Saluveer, E., Saitta, A., Rosas, M., Riit, T., Ratkowsky, D., Pritsch, K., Põldmaa, K., … Abarenkov, K. (2014). Global diversity and geography of soil fungi. *Science*, *346*, 1256688.

Tedersoo, L., Tooming-Klunderud, A., & Anslan, S. (2018). PacBio metabarcoding of fungi and other eukaryotes: Errors, biases and perspectives. *New Phytologist*, *217*, 1370–1385.

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

Westcott, S. L., & Schloss, P. D. (2015). De novo clustering methods outperform reference-based methods for assigning 16S rRNA gene sequences to operational taxonomic units. *PeerJ*, *3*, e1487.

White, T. J., Bruns, T., Lee, S., & Taylor, J. (1990). Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. *PCR Protocols: A Guide to Methods and Applications*, *18*, 315–322.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

How to cite this article: Estensmo EL, Maurice S, Morgado L, Martin-Sanchez PM, Skrede I, Kauserud H. The influence of intraspecific sequence variation during DNA metabarcoding: A case study of eleven fungal species. *Mol Ecol Resour.* 2021;00:1–8. https://doi.org/10.1111/1755-0998.13329

Supplementary information

Table S1. Overview of the 65 haplotypes identified across 177 fruit bodies of 11 species and their distribution in the Sanger and HTS datasets. Number in brackets indicates the minimum number of reads expected from each allele present in the population (i.e. #reads/2 x sample size). For each species, Sanger sequences were dephased into haplotypes (Dephased Hap.) and their relative abundance (RA) were estimated, empty cells in the column Dephased Hap. refer to haplotypes identified from HTS only (yellow), haplotypes identified by HTS only are in blue and common haplotypes to the two approaches are in green. #Reads correspond to the total number of sequences per haplotype in the HTS dataset and estimate of their relative proportions (Prop.).

Haplotype	Dephased Hap.	RA	#Reads	Prop. (R)
Amylocystis lapponica			(3,687)	
Hap_1	18	1	66,364	1
Antrodia serialis			(5,587)	
Hap_2	19	0.59	101,440	0.5674
Hap_3	11	0.34	69,103	0.3865
Hap_4	1	0.03	5,217	0.0292
Hap_5	1	0.03	3,027	0.0169
Fomitopsis pinicola			(7,442)	
Hap_6	27	0.84	190,157	0.7985
Hap_7	2	0.06	7,303	0.0307
Hap_8	1	0.03		
Hap_9	1	0.03	13,441	0.0564
Hap_10	1	0.03	11,874	0.0499
Hap_11		0.00	6,537	0.0274
Hap_12		0.00	5,830	0.0245
Hap_13		0.00	2,931	0.0123
Hap_14		0.00	70	0.0003
Fomitopsis rosea			(5,633)	
Hap_15	24	0.80	122,473	0.7247
Hap_16	2	0.07	9,932	0.0588
Hap_17	2	0.07	13,353	0.0790
Hap_18	1	0.03	1,831	0.0108
Hap_19	1	0.03	5,855	0.0346
Hap_20		0.00	8,215	0.0486
Hap_21		0.00	3,950	0.0234
Hap_22		0.00	3,382	0.0200
Gloeophyllum sepiarium			(13,434)	
Hap_23	24	0.80	388,544	0.9038
Hap_24	2	0.07	7,513	0.0175
Hap_25	2	0.07		
Hap_26	1	0.03	16,206	0.0377
Hap_27	1	0.03	6,520	0.0152
Hap_28		0.00	11,117	0.0259
Phlebia centrifuga				
Hap_29	22	0.69	243,136	0.6848
Hap_30	10	0.31	111,691	0.3146
Hap_31		0.00	222	0.0006
Hap_32		0.00	12	0.0000

Haplotype	Dephased Hap.	RA	#Reads	Prop. (R)
Phellinus ferrogineofuscus			(5,572)	
Hap_33	22	0.69	10,265	0.0576
Hap_34	8	0.25	28,248	0.1584
Hap_35	1	0.03	7,820	0.0439
Hap_36	1	0.03	7,232	0.0406
Hap_37		0.00	124,730	0.6996
Phellopilus nigrolimitatus			(3,890)	
Hap_38	4	0.33	11,591	0.2483
Hap_39	2	0.17		
Hap_40	2	0.17		
Hap_41	2	0.17	3,027	0.0648
Hap_42	1	0.08		
Hap_43	1	0.08		
Hap_44		0.00	10,674	0.2287
Hap_45		0.00	6,228	0.1334
Hap_46		0.00	8,972	0.1922
Hap_47		0.00	6,016	0.1289
Hap_48		0.00	169	0.0036
Phellinus viticola			(3,760)	
Hap_49	16	0.50	62,741	0.5215
Hap_50	10	0.31	39,566	0.3289
Hap_51	4	0.13	9,190	0.0764
Hap_52	1	0.03	5,036	0.0419
Hap_53	1	0.03	3,773	0.0314
Postia caesia			(9,574)	
Hap_54	15	0.75	122,730	0.6409
Hap_55	3	0.15	29,089	0.1519
Hap_56	2	0.10	38,778	0.2025
Hap_57		0.00	886	0.0046
Trichaptum abietinum			(11,268)	
Hap_58	17	0.57		
Hap_59	10	0.33	108,017	0.3195
Hap_60	1	0.03		
Hap_61	1	0.03	14,404	0.0426
Hap_62	1	0.03	15,497	0.0458
Hap_63		0.00	189,983	0.5620
Hap_64		0.00	8,437	0.0250
Hap_65		0.00	1,711	0.0051

Table S2. The table summarizes the results from the analyses of technical replicates of the 11 species. The two technical replicates for each species are shown pairwise in the columns. The rows (OTUid) represent the detected ASVs, where the taxonomic matches are shown in the last column. Dark green indicates sequences belonging to the target species, demonstrating a high consistency across the two replicates. Light red indicates low-abundance presences that likely are a result of tag-switching. The light red indicates tentative tag-switches of ASVs present in other replicate samples, dark red indicates tentative tag-switches from other samples in the overall setup (i.e. not among the replicates) and light green reflects tentative tag-switches from other samples of the same species. Nevertheless, the tentative tag-switches do not generate new ASVs/haplotypes, but lead to some leakage among samples. This does not influence on our results, since we look at the overall distribution of ASVs within each species.

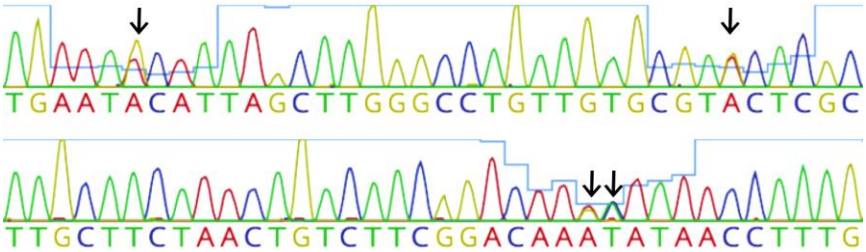
OTUid	Glosep		Phecen		Fompin		Triabi		Fomros		Antser		Phefer		Amylap		Phevit		Phenig		Poscae		Sequence match (UNITE)
	S065	S190	S138	S091	S033	S188	S170	S093	S049	S189	S017	S187	S102	S088	S001	S186	S122	S090	S106	S089	S154	S092	
seq1	21820	37743	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Gloeophyllum_sepiarium_SH1505693.08FU
seq41	0	11	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Fomitopsis_rosea_SH1519483.08FU
seq3	0	1	1603	2122	0	0	0	0	0	0	0	0	0	0	223	0	1	0	0	0	0	0	Phlebia_centrifuga_SH1550104.08FU
seq9	0	0	17213	22387	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Phlebia_centrifuga_SH1550104.08FU
seq3476	0	0	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Phlebia_centrifuga_SH1550104.08FU
seq15	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Phellinus_viticola_SH1554178.08FU
seq4	0	0	0	2	21309	49407	0	0	2	0	0	0	0	0	0	3	0	0	0	0	0	1	Fomitopsis_pinicola_SH1519477.08FU
seq5	0	0	1	0	0	0	12379	3931	0	0	0	0	1	0	0	0	0	0	0	1	0	0	Trichaptum_abietinum_SH1539679.08FU
seq8	1	1	0	0	0	0	6676	1862	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Trichaptum_abietinum_SH1539679.08FU
seq6	12	18	0	1	0	0	0	0	18115	44166	0	0	0	0	0	0	0	0	0	0	0	0	Fomitopsis_rosea_SH1519483.08FU
seq7	0	0	3	0	0	0	0	0	0	0	4751	29848	0	15	0	0	0	0	0	0	0	0	Antrodia_serialis_SH1879451.08FU
seq12	1	0	0	0	0	0	0	0	0	1	5320	32567	0	0	0	0	0	0	0	0	0	0	Antrodia_serialis_SH1879451.08FU
seq10	0	0	0	0	0	0	0	0	0	0	0	0	13534	19027	0	0	0	0	0	0	0	0	Hymenochaetaceae
seq55	0	0	0	0	0	0	0	0	0	0	0	0	0	12	0	0	0	0	0	0	0	0	Antrodia_serialis_SH1879451.08FU
seq11	0	0	0	0	0	0	0	0	0	0	0	0	0	1	4075	15946	0	0	0	0	0	0	Amylocystis_lapponica_SH1510121.08FU
seq13	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3774	3306	0	0	0	0	Phellinus_viticola_SH1554178.08FU
seq47	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3773	3517	0	0	0	0	Phellinus_viticola_SH1554178.08FU
seq22	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	Phellopilus_nigrolimitatus_SH1540507.08FU
seq20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5773	2029	0	0	Phellopilus_nigrolimitatus_SH1540506.08FU
seq33	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6228	4213	0	0	Phellopilus_nigrolimitatus_SH1540507.08FU
seq38	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	35	0	0	Phellopilus_nigrolimitatus_SH1540507.08FU
seq72	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	344	369	Postia_cyanescens_SH1518080.08FU
seq236	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	61	54	Postia_cyanescens_SH1518080.08FU

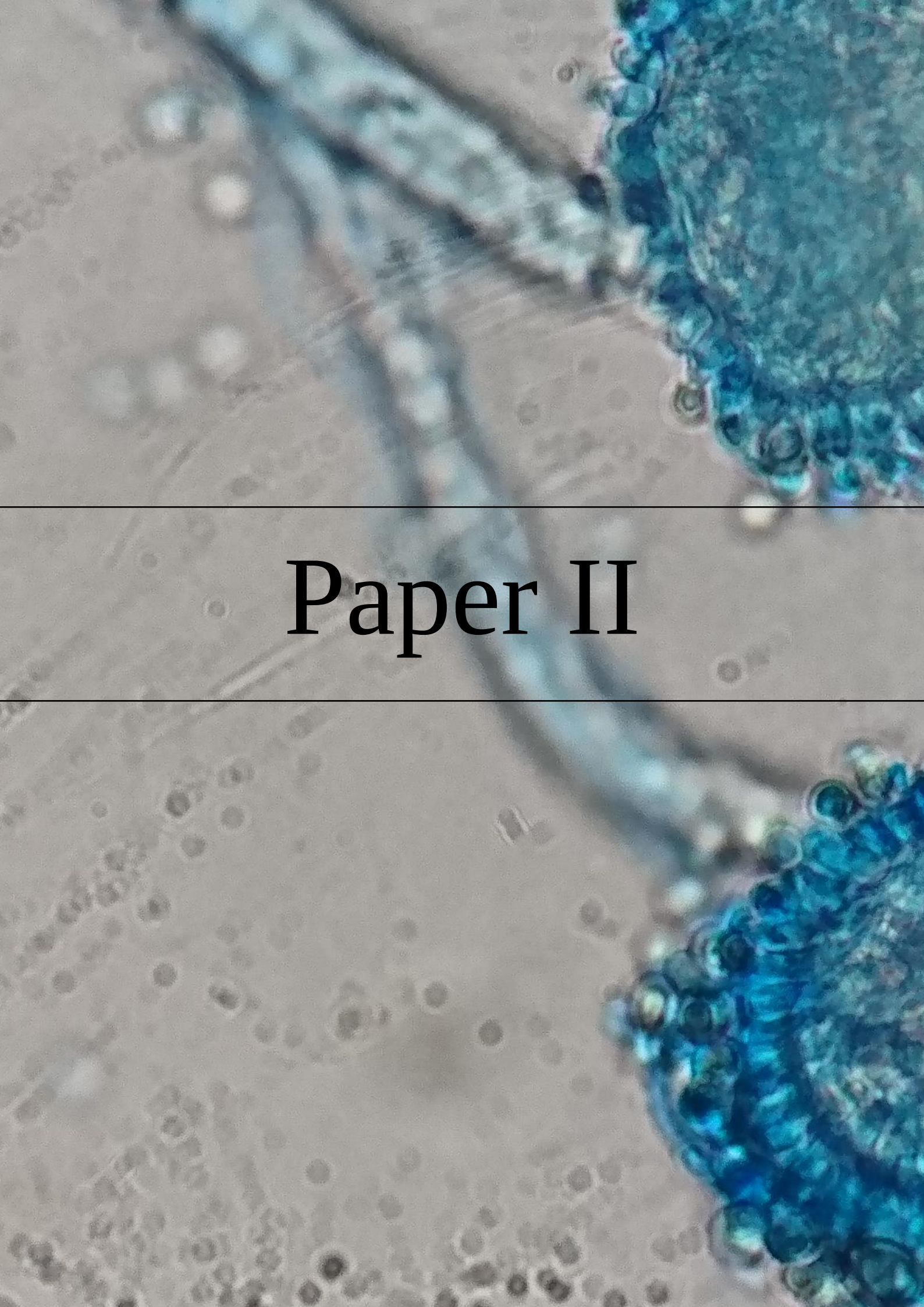
Table S3. ITS Sanger sequences from the fungal specimens (Sample ID) deposited in GenBank (Submission SUB8713788) with corresponding accession numbers (Sequence ID).

Sample ID	Sequence ID	Sample ID	Sequence ID
Amylap15	MW358287	Phecen16	MW358361
Amylap1	MW358288	Phecen23	MW358362
Amylap16	MW358289	Phecen15	MW358363
Amylap2	MW358290	Phecen22	MW358364
Amylap21	MW358291	Phecen21	MW358365
Amylap8	MW358292	Phecen13	MW358366
Amylap4	MW358293	Phecen2	MW358367
Amylap3	MW358294	Phecen11	MW358368
Amylap7	MW358295	Phecen19	MW358369
Antser11	MW358296	Phecen1	MW358370
Antser21	MW358297	Phecen6	MW358371
Antser1	MW358298	Phecen14	MW358372
Antser6	MW358299	Phefer19	MW358373
Antser3	MW358300	Phefer10	MW358374
Antser13	MW358301	Phefer5	MW358375
Antser22	MW358302	Phefer3	MW358376
Antser2	MW358303	Phefer21	MW358377
Antser5	MW358304	Phefer2	MW358378
Antser19	MW358305	Phefer6	MW358379
Antser4	MW358306	Phefer18	MW358380
Antser18	MW358307	Phefer16	MW358381
Antser17	MW358308	Phefer14	MW358382
Antser25	MW358309	Phefer13	MW358383
Antser10	MW358310	Phefer4	MW358384
Antser20	MW358311	Phefer20	MW358385
Fompin6	MW358312	Phefer17	MW358386
Fompin17	MW358313	Phefer1	MW358387
Fompin5	MW358314	Phenig2	MW358388
Fompin10	MW358315	Phenig3	MW358389
Fompin1	MW358316	Phenig20	MW358390
Fompin20	MW358317	Phenig12	MW358391
Fompin18	MW358318	Phenig17	MW358392
Fompin4	MW358319	Phenig1	MW358393
Fompin3	MW358320	Phevit17	MW358394
Fompin24	MW358321	Phevit7	MW358395
Fompin22	MW358322	Phevit16	MW358396
Fompin15	MW358323	Phevit6	MW358397
Fompin14	MW358324	Phevit15	MW358398
Fompin23	MW358325	Phevit5	MW358399
Fompin13	MW358326	Phevit14	MW358400
Fompin19	MW358327	Phevit4	MW358401
Fomros10	MW358328	Phevit13	MW358402
Fomros24	MW358329	Phevit22	MW358403
Fomros8	MW358330	Phevit12	MW358404
Fomros23	MW358331	Phevit19	MW358405
Fomros21	MW358332	Phevit11	MW358406
Fomros19	MW358333	Phevit18	MW358407
Fomros16	MW358334	Phevit1	MW358408
Fomros15	MW358335	Phevit9	MW358409
Fomros13	MW358336	Poscae23	MW358410
Fomros22	MW358337	Poscae18	MW358411
Fomros12	MW358338	Poscae3	MW358412
Fomros18	MW358339	Poscae16	MW358413
Fomros17	MW358340	Poscae22	MW358414
Glosep1	MW358341	Poscae13	MW358415
Glosep22	MW358342	Poscae20	MW358416
Glosep24	MW358343	Poscae2	MW358417

Glosep17	MW358344	Poscae19	MW358418
Glosep23	MW358345	Poscae5	MW358419
Glosep15	MW358346	Triabi24	MW358420
Glosep21	MW358347	Triabi13	MW358421
Glosep14	MW358348	Triabi10	MW358422
Glosep13	MW358349	Triabi4	MW358423
Glosep2	MW358350	Triabi18	MW358424
Glosep11	MW358351	Triabi11	MW358425
Glosep19	MW358352	Triabi20	MW358426
Glosep10	MW358353	Triabi3	MW358427
Glosep18	MW358354	Triabi1	MW358428
Glosep20	MW358355	Triabi23	MW358429
Glosep16	MW358356	Triabi21	MW358430
Phecen18	MW358357	Triabi22	MW358431
Phecen4	MW358358	Triabi12	MW358432
Phecen17	MW358359	Triabi19	MW358433
Phecen3	MW358360		

Figure S1. Section from a chromatogram of a *Trichaptum abietinum* ITS2 sequence showing four heterozygous sites (black arrows). Heterozygous sites were scored according to the IUPAC nucleotide code and heterozygous genotypes were split into two haplotypes when dephased.





Paper II

Received: 12 July 2020 | Revised: 30 March 2021 | Accepted: 1 April 2021

DOI: 10.1111/mec.15916

ORIGINAL ARTICLE

MOLECULAR ECOLOGY WILEY

Analysing indoor mycobiomes through a large-scale citizen science study in Norway

Pedro M. Martin-Sanchez[1] | Eva-Lena F. Estensmo[1] | Luis N. Morgado[1,2] | Sundy Maurice[1] | Ingeborg B. Engh[3] | Inger Skrede[1] | Håvard Kauserud[1]

[1]Section for Genetics and Evolutionary Biology (Evogene), Department of Biosciences, University of Oslo, Oslo, Norway

[2]Naturalis Biodiversity Center, Leiden, the Netherlands

[3]Mycoteam AS, Oslo, Norway

Correspondence
Pedro M. Martin-Sanchez, Section for Genetics and Evolutionary Biology, Department of Biosciences, University of Oslo, PO Box 1066 Blindern, 0316 Oslo, Norway.
Emails: p.m.martin-sanchez@ibv.uio.no; pedrom.martin@yahoo.es

Funding information
European Union's Horizon 2020, MSCA-IF, Grant/Award Number: 741332; University of Oslo; Centre for Digital Life Norway

Abstract

In the built environment, fungi can cause important deterioration of building materials and have adverse health effects on occupants. Increased knowledge about indoor mycobiomes from different regions of the world, and their main environmental determinants, will enable improved indoor air quality management and identification of health risks. This is the first citizen science study of indoor mycobiomes at a large geographical scale in Europe, including 271 houses from Norway and 807 dust samples from three house compartments: outside of the building, living room and bathroom. The fungal community composition determined by DNA metabarcoding was clearly different between indoor and outdoor samples, but there were no significant differences between the two indoor compartments. The 32 selected variables, related to the outdoor environment, building features and occupant characteristics, accounted for 15% of the overall variation in community composition, with the house compartment as the key factor (7.6%). Next, climate was the main driver of the dust mycobiomes (4.2%), while building and occupant variables had significant but minor influences (1.4% and 1.1%, respectively). The house-dust mycobiomes were dominated by ascomycetes (~70%) with Capnodiales and Eurotiales as the most abundant orders. Compared to the outdoor samples, the indoor mycobiomes showed higher species richness, which is probably due to the mixture of fungi from outdoor and indoor sources. The main indoor indicator fungi belonged to two ecological groups with allergenic potential: xerophilic moulds and skin-associated yeasts. Our results suggest that citizen science is a successful approach for unravelling the built microbiome at large geographical scales.

KEYWORDS
buildings, dust, fungi, indicator species, indoor and outdoor environments, ITS2 metabarcoding

1 | INTRODUCTION

Throughout the world, people spend a major part of their lifetime in the built environment, including houses, workplaces, kindergartens and schools. They harbour unique and complex microbial assemblages (fungi, bacteria, archaea and viruses), whose ecological roles and impact on human health remain largely unknown (Gilbert & Stephens, 2018).

Fungi, one of the most diverse kingdoms of life, with essential ecosystem functions (Willis, 2018), are also present in the built environment, where the extreme environmental conditions (dry and generally warm) favour certain species. The overall assembly of fungi in buildings can be termed the "indoor mycobiome" and is largely composed of saprotrophs that degrade available organic substrates and stress-tolerant ascomycetes, including ubiquitous airborne mould genera (e.g., *Cladosporium*, *Penicillium*, *Aspergillus* and *Alternaria*). Wherever enough moisture is present, fungi grow and subsequently emit spores, fragments of hyphae, volatile organic compounds and mycotoxins that act as sources of indoor pollutants (Flannigan & Miller, 2011; Nevalainen et al., 2015; Rintala et al., 2012). Dampness- and mould-related indoor air quality problems are a public health concern due to their association with adverse health effects, such as allergies, asthma and other respiratory symptoms (Fisk et al., 2007; Mendell et al., 2011).

Microbiological assessments in the built environment focus mainly on air and dust samples indicative of human exposure indoors. The fungal content of these samples can be analysed using different approaches: microscopy, culturing, chemical analyses and DNA-based methods (Nevalainen et al., 2015). Considering the well-known limitations of culture-based methods (Amann et al., 1995), a shift toward DNA-based methods has taken place in recent decades. High-throughput sequencing (HTS) of amplified markers (DNA metabarcoding) has recently become a key tool for surveying fungal communities in environmental samples (Lindahl et al., 2013; Nilsson et al., 2019). In the last decade, many studies have used DNA metabarcoding to reveal the microbiome of residential buildings in different parts of the world (Gilbert & Stephens, 2018), mainly focusing on bacteria (Adams et al., 2015; Lax et al., 2014), but also on fungi (Adams et al., 2013a, 2013b; Amend et al., 2010; Barberán et al., 2015a; Tong et al., 2017).

The indoor mycobiome is determined primarily by large-scale environmental gradients such as climate, but local environmental variation within individual buildings, including differences in construction features and building functions, can also contribute to shaping the fungal diversity and composition (Adams et al., 2016; Gilbert & Stephens, 2018; Stephens, 2016). A first global survey analysing 72 settled-dust samples from buildings in six continents revealed that the indoor fungal diversity is significantly higher in temperate zones than in the tropics, with latitude being the best predictor of the indoor mycobiome composition, while neither building design nor function had any significant effect (Amend et al., 2010).

Both culture- and DNA-based studies have demonstrated that outdoor air is the main source of indoor fungi. Adams et al. (2013a, 2013b) observed that indoor fungi are dominated by those spreading from outdoor air, and the mycobiome of indoor surfaces displayed similar patterns to outdoor air in the same locality. Barberán et al. (2015a, 2015b) analysed dust microbiomes collected inside and outside 1200 houses across the USA and confirmed that most indoor fungi were derived from outdoor sources. They further identified geographical patterns in the indoor mycobiomes that could be explained by climate, soil and vegetation variables.

However, a variety of internal secondary sources must be considered as well, such as organic materials (food, waste and potted plants), certain surfaces (drains and carpets) and occupants (humans and pets) (Adams et al., 2013b; Flannigan & Miller, 2011; Haines et al., 2019; Nevalainen et al., 2015; Rintala et al., 2012). DNA-based dust studies have indicated that various building features and occupant characteristics are also key determinants of the indoor mycobiome (Dannemiller et al., 2016; Kettleson et al., 2015). In this regard, Yamamoto et al. (2015) claimed that indoor emissions associated with occupant activities were the primary sources of airborne allergenic fungal particles. However, taken together, it is well accepted that the indoor mycobiome is determined largely by the outdoor environment, while bacteria are more strongly influenced by occupants and their activities (Adams et al., 2016; Barberán et al., 2015a; Gilbert & Stephens, 2018; Lax et al., 2014; Stephens, 2016).

Except for the pioneering global study by Amend et al. (2010), and the continental-scale study across the USA by Barberán et al. (2015a, 2015b), the majority of existing DNA-based mycobiome studies have focused on specific building units at a local scale. A few regional studies have also targeted some large cities, like Munich (Weikl et al., 2016) and Hong Kong (Tong et al., 2017). Given that the indoor mycobiome is highly influenced by the outdoor air, we can expect significant differences between houses inherent to their outdoor regional climate and environment. Revealing the indoor mycobiome and characterizing the variations across houses from different geographical regions of the world will provide basic knowledge for improved indoor air quality management and the identification of health risks.

Our study area, Norway, possesses marked climatic and environmental gradients, enabling us to assess to what degree the outdoor environment, vs. building features and occupant characteristics, influence the indoor mycobiomes. To represent a broad sample of buildings, we organized a citizen science dust sampling campaign in houses throughout Norway coupled with subsequent DNA metabarcoding analyses of the mycobiomes. Previous studies have demonstrated that citizen science, coupled with HTS approaches, is a promising avenue for conducting large-scale microbiome studies, including the built and human microbiomes (Barberán et al., 2015a; McDonald et al., 2018).

More specifically, we addressed and tested the following research questions and hypotheses: (i) which factors shape the indoor mycobiomes? In this regard, we investigate whether regional-scale variation in climate (and other regional-scale variables), building features or occupant characteristics are the main determinants. Here we hypothesize (H1) that all three categories influence the indoor mycobiomes, but regional-scale climate is the most important driver. Next, we ask (ii) which fungi dominate the house-dust mycobiomes in Norway. We hypothesize (H2) that ascomycetes, and especially

stress-tolerant ascomycetes, are the dominant groups in this environment. We also ask (iii) how much of the indoor mycobiome overlaps with the outdoor mycobiome. In relation to this question, we hypothesize (H3) that a major fraction of the indoor fungi derives from outdoor sources, while a relatively minor fraction originates from indoor sources.

2 | MATERIALS AND METHODS

2.1 | Citizen science dust sampling campaign

To increase the number of study houses and cover a broad geographical area, citizen scientists were recruited through scientific

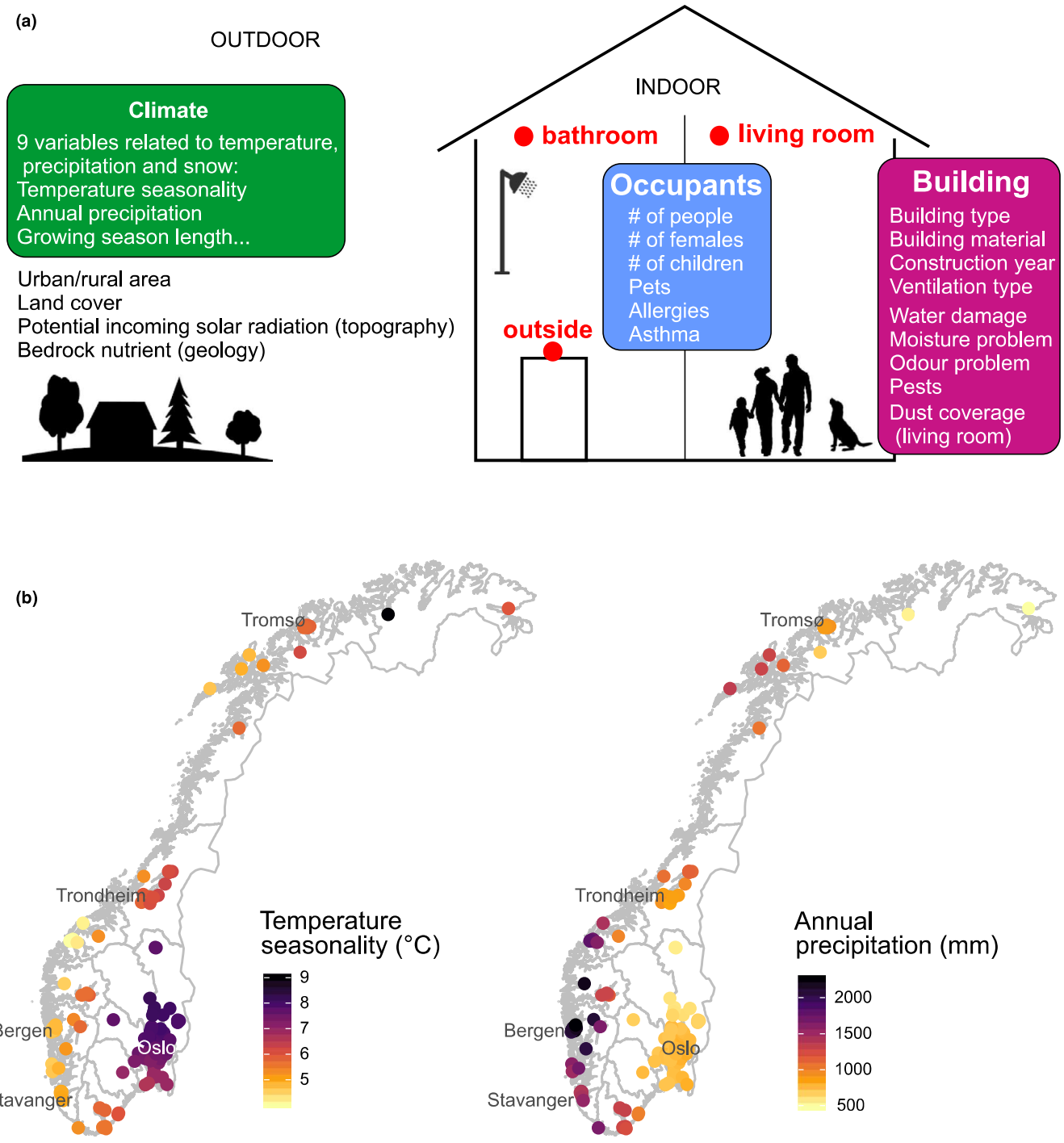


FIGURE 1 Overview of the citizen science dust sampling campaign in Norway. (a) Schematic overview of the metadata for each house: outdoor metadata that mainly include climatic variables (green), building features (violet) and occupant characteristics (blue). The sampling points (house compartments) are indicated with red dots. The building variable "Dust coverage" corresponds to the percentage of dust covering the study surface at the living room, as measured on the adhesive tape (Mycotape2). (b) Maps showing the location of the 269 houses (in mainland Norway) coloured according to their temperature seasonality (left; standard deviation of mean monthly temperatures = BIO4/100) and the annual precipitation (right; BIO12)

networks and diverse actions in social and public media. A total of 359 volunteers signed up in this study and provided relevant information (metadata) about their houses (Figure 1a). Sampling kits (Figure S1), including instructions, return envelope, three sterile FLOQSwabs in tubes (Code 552C; Copan Italia spa) and two adhesive tapes (Mycotape2; Mycoteam AS) were sent to volunteers by post. Following our instructions, volunteers swabbed dust samples from the upper doorframes located in three compartments of their houses: outside of the building (main entrance), living room and bathroom. No specific surface area was predetermined in the instructions. Doorframes act as passive collectors of dust deposited during an unknown amount of time. In addition, one adhesive tape was collected from other areas not frequently cleaned (e.g., shelves) in the living room to calculate the percentage of dust coverage, which was later included as an environmental variable in the study. The samples were sent back to the University of Oslo (UiO) by post, where they were registered and the swabs were stored at −80°C until DNA extractions. However, the adhesive tapes were immediately scanned using an Epson Perfection V850 Pro scanner (Seiko Epson Corporation) and the dust coverage was calculated on a surface area of 45 × 18 mm by image analysis using the OLYMPUS STREAM version 1.9 software.

To minimize the influence of seasonality effects, all samples were collected in a short time span during spring 2018, mainly in May (from April 27 to June 5). In total, 269 houses were sampled from mainland Norway, covering its major climatic gradients (Figure 1b; Figure S2). Two houses from Longyearbyen, in the Arctic Archipelago of Svalbard, were also included.

2.2 | Environmental data

Metadata about the study houses and their occupants were provided by the volunteers through an online questionnaire at the UiO website. In addition to the location of houses including their addresses and geographical coordinates (latitude and longitude), the following 15 variables (with categories for categorical variables, excluding those present in a very low number of houses) were extracted from the questionnaire: building type (detached house/semidetached house/block), area (urban/rural), construction year, building material (wood/brick and concrete), ventilation type (natural/mechanical/balanced), number of people, number of children, number of females, pets (no/dog/cat), allergies (no/pollen/food/skin), asthma (yes/no), moisture problem (yes/no), water damage (yes/no), odour problem (yes/no) and pests (no/mice/rats/grey silverfish) (Figure 1a). Data about the location of dust samples in the house were included as two categorical variables: house compartment (outside/living room/bathroom) and indoor vs. outdoor (indoor/outdoor).

Based on the geographical coordinates of study houses, data for six relevant WorldClim 2 bioclimatic variables (annual mean temperature BIO1, temperature seasonality BIO4, mean temperature of the driest quarter BIO9, mean temperature of the warmest quarter BIO10, mean temperature of the coldest quarter BIO11 and annual precipitation BIO12) were extracted at 30-seconds resolution (~1 km^2) using the R package DISMO (Fick & Hijmans, 2017). Moreover, data for 116 environmental variables related to geology, topography, climate and hydrology were also explored. They were kindly provided by the authors of a recent study modelling the vegetation types in Norway (Horvath et al., 2019). The contribution of all continuous variables, 46 of 116 from this data set plus the six previously extracted from WorldClim, were evaluated by principal component analysis (PCA) using the R package ADE4 (Dray & Dufour, 2007) (Figure S3). Based on this PCA, 10 continuous variables were selected for the statistical analyses: the six detailed WorldClim bioclimatic variables plus growing season length, snow-covered area in February, snow water equivalent in April and potential incoming solar radiation. Two additional categorical variables from the vegetation study (Horvath et al., 2019): land cover AR50 (developed area/agricultural area/forest/barren land/bog and fen/fresh water) and bedrock nutrient (poor/average/rich), as well as the dust coverage measured on the adhesive tapes, were included in the final selection (32 variables; Figure 1a).

2.3 | Fungal DNA metabarcoding: DNA extraction, amplification and sequencing

DNA was extracted from the swabs using chloroform and the EZNA Soil DNA Kit (Omega Bio-tek). Swab tips were transferred to the kit's disruptor tubes that contain glass beads and 800 μl SLX-Mlus buffer. After a first bead-beating cycle (1 min at 4.5 m s^{-1}) using the FastPrep-24 homogenizer (MP Biomedicals), the samples were frozen at −20°C for at least 30 min. Afterwards, samples were incubated at 70°C for 15 min and again shaken using the FastPrep-24 (two cycles of 30 s at 4.5 m s^{-1}). After adding 600 μl chloroform, samples were vortexed for 30 s and centrifuged at 15,000 *g* for 5 min. DNA from the aqueous top phase was further purified using the HiBind DNA Mini Column from the EZNA Soil DNA Kit following the manufacturer's instructions. Final DNA extracts were eluted in 30 μl EB buffer and quantified using the fluorometric Qubit dsDNA HS Assay Kit (Invitrogen, Thermo Fisher Scientific). Low DNA yield, ranging from 0.05 to 1 ng μl^{-1}, was recovered from the swabs, which was expected considering the small amount of dust collected with dry swabs. Nine blank controls (unused sterile swabs) from different extraction batches were included through the complete DNA metabarcoding protocol.

The internal transcribed spacer 2 (ITS2) region of the nuclear rDNA was amplified using the primers gITS7 5′-GTGARTCATCGARTCTTTG-3′ (Ihrmark et al., 2012) and ITS4 5′-TCCTCCGCTTATTGATATGC-3′ (White et al., 1990). Both forward and reverse primers were designed with 96 unique tags (barcodes) of 7–9 bases at the 5′-end, which differed in at least three positions from each other. To avoid tag switching errors (Carlsen et al., 2012), samples were combined in pools of 96 samples, each with a unique tag combination (Table S1). Nine pools (96 samples each) were analysed in this study, and each of them included an extraction blank,

a PCR (polymerase chain reaction) negative and a mock community that was used as a positive control (details in the Supporting Information). In total, 17 dust samples were duplicates and used as technical replicates across different PCR libraries.

PCRs in 25 µl contained 1 unit of AmpliTaq Gold DNA polymerase (Applied Biosystems, Thermo Fisher Scientific), 0.4 µM of each primer, 0.8 mg ml^{-1} bovine serum albumin (BSA; Thermo Scientific, Thermo Fisher Scientific), 1× Buffer II, 2.5 mM of $MgCl_2$, 0.2 mM each of dNTPs and 4 µl of DNA extract (~0.2–4 ng of DNA). Amplifications were carried out using the following cycling parameters: an initial denaturing step at 95°C for 5 min followed by 35 cycles consisting of 30 s at 95°C, 30 s at 55°C and 1 min at 72°C, and a final elongation step at 72°C for 10 min. PCR products of each library were initially purified and normalized using a SequalPrep Normalization Plate Kit (Applied Biosystems, Thermo Fisher Scientific) and subsequently pooled. After an additional purification using 0.8 volume of Agencourt AMPure XP magnetic beads (Beckman Coulter), DNA concentration and length of the final pooled amplicons were checked using a Qubit dsDNA HS Assay Kit and Bioanalyzer High Sensitivity DNA chip (Agilent Technologies), respectively. Sequencing was carried out at Fasteris SA using the Metafast protocol, which incorporated Illumina adapters using a PCR-free ligation procedure to minimize errors such as chimera formation and tag switching. Three full Illumina 250-bp paired-end MiSeq version 3 runs (Illumina) were used. Each run included three pooled libraries labelled with specific indexes in their Illumina barcodes. The complete resulting data set contained 55,568,124 paired reads and is available on the European Nucleotide Archive (ENA) at EMBL-EBI under accession no. PRJEB42161. Accession numbers for the nine libraries correspond to the BioSamples SAMEA7740226–SAMEA7740234.

2.4 | Bioinformatics pipeline

After an initial quality checking of sequencing results using FASTQC version 0.11.2 (Babraham Bioinformatics Team), samples were demultiplexed independently (R1 and R2) with CUTADAPT version 1.8 (Martin, 2011) allowing zero mismatches in tags and primers; these were simultaneously removed along with sequences shorter than 100 bases. The demultiplexed R1 and R2 reads were kept separate for the next analyses using DADA2version 1.12 (Callahan et al., 2016): (i) quality filtering and trimming, (ii) dereplication, (iii) generating error models and denoising, (iv) merging in contigs, (v) creating the table of amplicon sequence variants (ASVs) and (vi) removal of chimeras. Additional clustering of ASVs in operational taxonomic units (OTUs), as recommended in previous studies (Estensmo et al., 2021), was done using VSEARCH version 2.11.1 (Rognes et al., 2016) at 98% similarity. This clustering level is similar to the 98.5% level used to define the species hypotheses (SHs) in the UNITE database (Kõljalg et al., 2013). OTUs containing only one read (singletons) were removed after clustering. To correct for potential over-splitting of OTUs due to remaining sequencing errors, the OTU table was curated using LULU with default settings (Frøslev et al., 2017). Taxonomic assignment of the OTUs was carried out using VSEARCH against the eukaryotic ITS data set from UNITE version 8.0 (UNITE Community, 2019a). Two filters were subsequently applied on the resulting OTU table to select those OTUs that contained at least 10 reads and showed at least 70% identity in the taxonomic assignment. Finally, we selected the OTUs assigned to the kingdom Fungi on the quality-filtered table. To refine the taxonomic annotation of the top-100 most abundant fungi, a double-checking was done on those OTUs that initially failed at the species level. This was performed using BLAST+ version 2.8 against both UNITE and NCBI databases (UNITE Community, 2019b). Ecological trophic modes and guilds for the identified taxa were annotated using the FUNGUILD tool (Nguyen et al., 2016). More details on the bioinformatics analyses and the assessment of control and replicates samples are provided in the Supporting Information (Table S2, Figure S4 and supplementary methods).

2.5 | Statistical analyses

Statistical analyses were conducted in R version 3.5.2 (R Core Team, 2018) through RSTUDIO version 1.2.1335. TIDYVERSE version 1.2.1 (Wickham, 2017) and the VEGAN version 2.5-6 (Oksanen et al., 2019) R packages were used for data manipulation and plotting, and ecological analyses, respectively. Initially, the OTU table was rarefied (×10 times resampling with the median value taken per OTU) to 2000 reads per sample using the function *rrarefy*, and further adapted for the three data sets: all samples (full data set), indoor samples and outdoor samples.

Alpha diversity was assessed by calculating species richness (number of observed OTUs) and evenness (equitability between OTUs), as well as Shannon and inverse Simpson indices. Beta diversity was assessed using *betadisper* to test the homogeneity of variance in different groups of samples. Significant differences in the variance of these parameters were evaluated with the analysis of variance (ANOVA) test. Beta diversity was also assessed with non-metric multidimensional scaling (NMDS) ordination of both dust samples and OTUs using *metaMDS*, Bray–Curtis dissimilarity index and 200 random starts in search of stable solution. After an initial comparison of NMDS results obtained from four types of OTU tables—rarefied data with or without three further transformations: logarithmic, Hellinger (Legendre & Gallagher, 2001) and Cumulative Sum Scaling (CSS; Paulson et al., 2013)—we chose to conduct the final NMDS analyses on the Hellinger-transformed rarefied OTU tables. Continuous environmental variables and alpha diversity indices were regressed against NMDS ordination and added as vectors on the ordination plots using *gg_envfit* of the R package GGORDIPLOT version 0.3.0 (Quensen, 2018) to visualize their association with the dust mycobiomes. To evaluate the correlation between environmental variables and the observed variance in fungal community composition, permutational multivariate analysis of variance (PERMANOVA; 999 permutations) was performed individually on each variable using *adonis2* and the resulting *p*-values were adjusted

using the conservative Bonferroni correction method. The effects of three groups of variables ("building," "occupants" and "climate"), compared to the factor "house compartment," were assessed by variation partitioning analysis (VPA) based on the Bray–Curtis dissimilarities using *varpart* and *vegdist*.

To evaluate the overlap between outdoor and indoor mycobiomes, we compared the OTUs detected in the three house compartments using two different estimates: percentages of OTUs across overall data (before and after removing the OTUs with <10 reads per sample) and mean percentages of OTUs per house. To reveal the significant associations ($p < .05$) between OTUs and some relevant environmental variables, the indicator species analysis was performed using *multipatt* of the R package INDICSPECIES (De Cáceres & Legendre, 2009). Finally, to unravel the most relevant variables predicting (i) the species richness per sample and (ii) the percentage of shared OTUs between indoor and outdoor, we conducted generalized linear model (GLM) analyses using the *glm* function. A forward selection was performed using Akaike's information criterion (AIC) to assess model improvement in comparison with the null model.

3 | RESULTS

3.1 | Data features and overall fungal diversity

After quality filtering, denoising and sequence clustering, the final fungal data set contained 7110 OTUs (22,622,815 reads), distributed among 811 dust samples from 271 houses. The number of reads per sample varied widely, from 424 to 245,588 (mean = 27,929), and the number of OTUs per sample (richness) ranged from seven to 867 (mean = 270) (Figure S5). Likewise, the abundance of OTUs varied extensively, from 10 to 2,040,802 reads per OTU (mean = 3182), while their occurrences ranged from one to 807 dust samples (mean = 31). For further statistical analyses, we resampled the data to a relatively low sequencing depth (2000 reads per sample) in order to keep the majority of samples (four samples were excluded) and houses, representative of a wide geographical area. The rarefied data set contained 6632 OTUs distributed across 807 samples.

The average richness and diversity (Shannon) per sample were significantly higher in the indoor compared to outdoor samples (Figure 2a). The evenness as well as the inverse Simpson index followed a similar trend (Figure S6). The two indoor compartments, that is living room and bathroom, had similar levels of richness and diversity. Sample origin (indoor vs. outdoor) was the strongest predictor of fungal richness according to the GLM ($p = 7.48e-05$). Several other variables (mean temperatures of the warmest and driest quarters, temperature seasonality, annual precipitation, snow-covered area in February, latitude and number of children) significantly improved the AIC of the null model ($p < .05$). However, adding these variables to the strongest predictor (indoor vs. outdoor) had no significant effect on model outcome. Houses from higher latitudes (northern Norway) possessed on average higher fungal diversity (Shannon) compared to houses in the south (Figure S7). In contrast to richness and diversity, the compositional dissimilarity (beta diversity) was higher among the outdoor samples (Figure 2b).

3.2 | Determinants of the community composition

We observed a marked compositional difference between indoor and outdoor mycobiomes, as revealed by NMDS ordination of all dust samples (Figure 3a). House compartment (outside, living

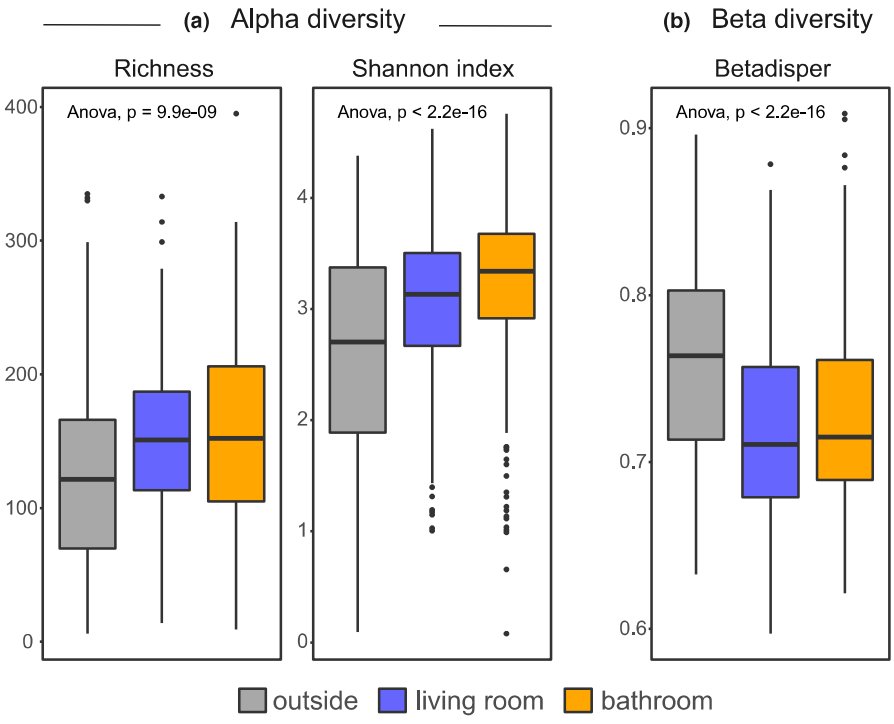
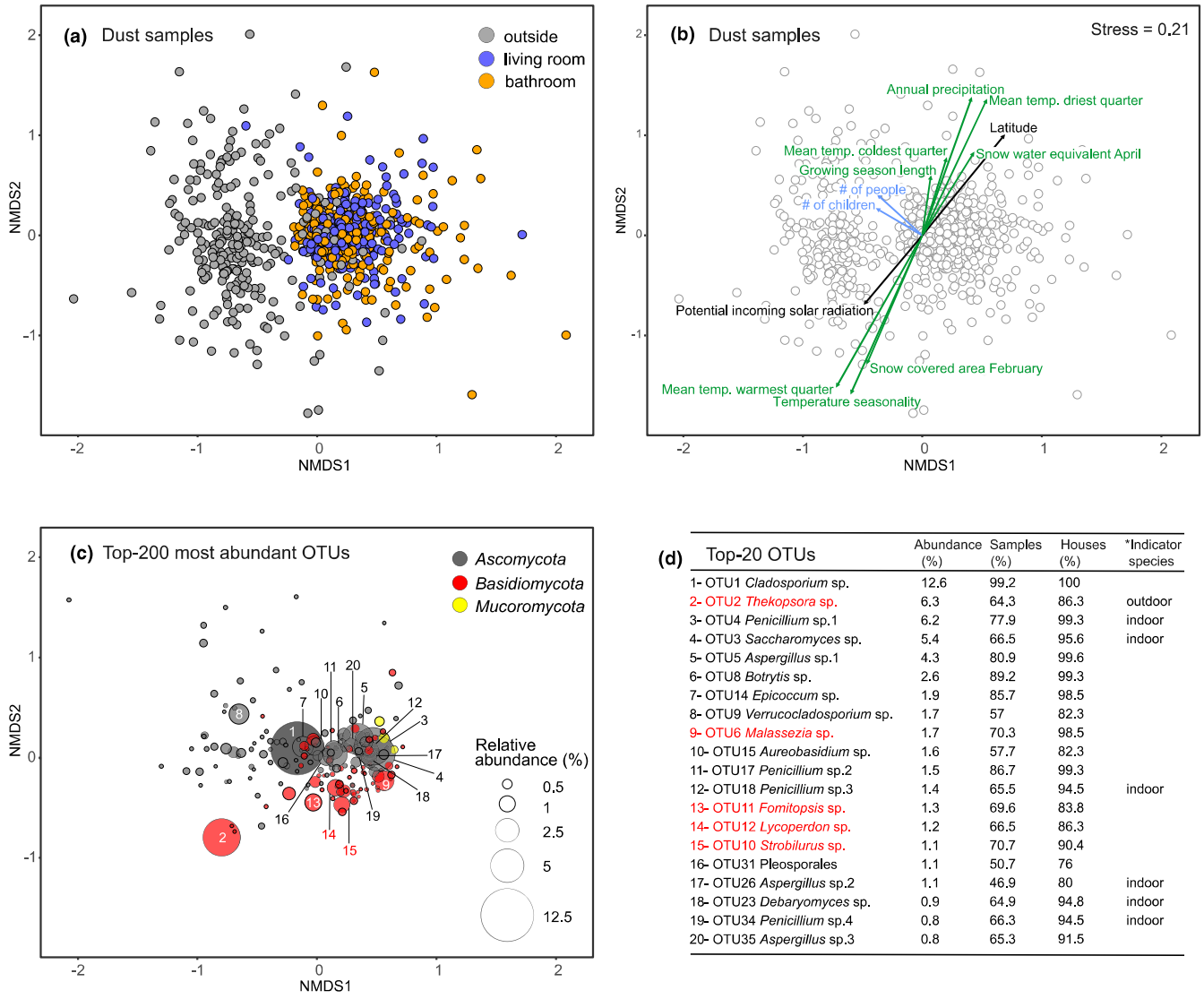


FIGURE 2 Box plots visualizing diversity patterns in the three studied house compartments. A total of 269 houses were assessed, including dust samples from the outside ($n = 266$), living room ($n = 270$) and bathroom ($n = 271$). (a) Alpha diversity (richness) and Shannon index; (b) beta diversity. All statistics were calculated from the rarefied matrix (6632 OTUs). All differences between outdoor and indoor compartments (outside vs. living room and outside vs. bathroom) were highly significant according to Tukey HSD test ($p < 1e-05$), while no significant difference was found between living room and bathroom ($p > .05$)



(d) Top-20 OTUs

	Abundance (%)	Samples (%)	Houses (%)	*Indicator species
1- OTU1 *Cladosporium* sp.	12.6	99.2	100	
2- OTU2 *Thekopsora* sp.	6.3	64.3	86.3	outdoor
3- OTU4 *Penicillium* sp.1	6.2	77.9	99.3	indoor
4- OTU3 *Saccharomyces* sp.	5.4	66.5	95.6	indoor
5- OTU5 *Aspergillus* sp.1	4.3	80.9	99.6	
6- OTU8 *Botrytis* sp.	2.6	89.2	99.3	
7- OTU14 *Epicoccum* sp.	1.9	85.7	98.5	
8- OTU9 *Verrucocladosporium* sp.	1.7	57	82.3	
9- OTU6 *Malassezia* sp.	1.7	70.3	98.5	
10- OTU15 *Aureobasidium* sp.	1.6	57.7	82.3	
11- OTU17 *Penicillium* sp.2	1.5	86.7	99.3	
12- OTU18 *Penicillium* sp.3	1.4	65.5	94.5	indoor
13- OTU11 *Fomitopsis* sp.	1.3	69.6	83.8	
14- OTU12 *Lycoperdon* sp.	1.2	66.5	86.3	
15- OTU10 *Strobilurus* sp.	1.1	70.7	90.4	
16- OTU31 Pleosporales	1.1	50.7	76	
17- OTU26 *Aspergillus* sp.2	1.1	46.9	80	indoor
18- OTU23 *Debaryomyces* sp.	0.9	64.9	94.8	indoor
19- OTU34 *Penicillium* sp.4	0.8	66.3	94.5	indoor
20- OTU35 *Aspergillus* sp.3	0.8	65.3	91.5	

FIGURE 3 Fungal community composition in house-dust samples. (a, b) NMDS ordination plot displaying compositional variation in the dust mycobiomes in the complete data set (n = 807). Each point indicates one dust sample. (a) Colour indicates the three compartments (outside, living room and bathroom). (b) Linear-regression of continuous variables with significant association (p < .05) with the NMDS. Climatic variables are shown as green arrows, occupant characteristics as light blue and geography/topography as black. (c) NMDS ordination of the species optima of the 200 most abundant OTUs. Bubble size indicates their relative abundance as a proportion of the total number of rarefied reads and colour indicates their phylum assignment. The numbers indicate the taxonomic assignment of the 20 most abundant OTUs, as provided in panel d. (d) OTU ID, taxonomic assignment, relative abundance, occurrence in dust samples and houses, as well as their classification as outdoor or indoor indicator species of the 20 most abundant OTUs (*; extracted from Table 2). Ascomycota are indicated in black and Basidiomycota in red

room or bathroom) was the key factor structuring the fungal community composition, accounting for 7.66% of the overall variation (variation partitioning analysis; Figure 4). However, there was no difference between the two indoor compartments: living room vs. bathroom (Figure 3a). A relatively low proportion of the variation in fungal community composition was explained by the assessed variables (Table 1), altogether accounting for about 15% of the variation (Figure 4). Climatic variables were also important for the fungal community composition in the dust samples, as seen in the ordination plot. Various climatic variables correlated with the second ordination axis (Figure 3b; Table S3). Together, climatic variables accounted for 4.18% of the variation among all dust samples, which increased to 6.79% for the outdoor samples when analysed separately (Figure 4). The four most important climatic variables were annual temperature variation (temperature seasonality BIO4), mean temperature of the warmest (BIO10) and the driest (BIO9) quarter, as well as annual precipitation (BIO12) (Table 1). There was a clear geographical signal in the fungal community composition. This was especially the case for the outdoor samples, but also, to a lesser extent, for the indoor samples (Figure S8a,c, Table S3), which again relate to the regional climate variability in the study area (Figure 1). Building features and occupant

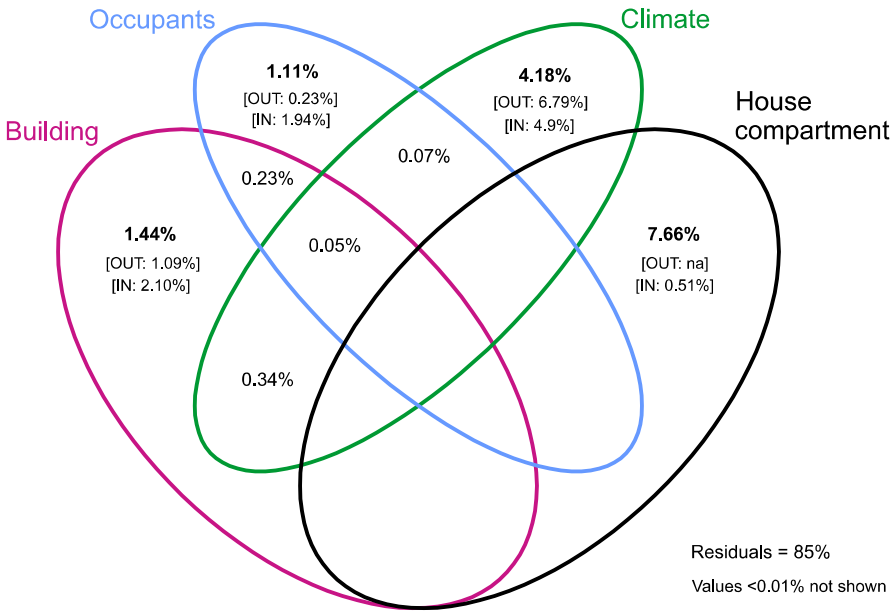


FIGURE 4 Venn diagram summarizing the variation partitioning analysis (VPA). The three groups of variables are indicated in colours ("Building," "Occupants" and "Climate") and compared to the factor "House compartment"; see Table 1 for selection. The percentage of variation explained by each group alone is in bold for the complete data set. VPA values in square brackets were obtained for the partial data sets when analysed separately: OUT, outdoor data set; IN, indoor data set. The unexplained variation (residual) was 85% and variables explaining <0.01% are not shown in the Venn diagram

characteristics accounted for only 1.44% and 1.11% of the overall variation in fungal community composition. Their contributions increased to 2.1% (building features) and 1.94% (occupant characteristics), respectively, when analysed on indoor samples exclusively (Figure 4, Table 1). According to the PERMANOVA results, the percentage of dust coverage measured on living room surfaces was also a significant explanatory variable with low R^2 value (0.4%) for the indoor data set (Table 1). The more occupants there were in houses, the more similar the indoor samples were to outdoor samples in fungal community composition (Figure 3b).

3.3 | Dominant fungi in house dust

The taxonomic assignment for the most abundant fungi is shown in Figure 5 and Figure S9 (FUNGuild annotation) and Table S4 (top-200 most abundant OTUs). High proportions of OTUs could not be identified at different taxonomic levels: 7.9%, 14.6%, 37.8% and 57.7% at the phylum, order, genus and species levels, respectively. From the nine phyla identified, Ascomycota dominated in both indoor and outdoor samples, including on average 70% of the sequences per sample, while Basidiomycota made up around 25% (Figure 5a). The third most abundant phylum was Mucoromycota, showing higher percentages of sequences in the indoor samples (2.1% living room and 1.5% bathroom) compared to outside (0.3%). Six other fungal phyla were detected in much smaller proportions (<0.1%) and with more limited distribution, sorted by decreasing abundance: Mortierellomycota, Olpidiomycota, Chytridiomycota, Rozellomycota, Entomophthoromycota and Entorrhizomycota.

As seen from the OTU ordination plot in Figure 3c, there was a broad-scale structuring of the major taxonomic groups. Both for the indoor and the outdoor samples, ascomycetes were in general more associated with areas with higher precipitation, lower mean temperature of the warmest quarter and low degree of seasonality in temperature, while the basidiomycetes showed the opposite pattern, being associated with more continental climates (Figure 3b,c).

At the order level, there were also marked differences between indoor and outdoor samples (Figure 5b). Eurotiales, the most common order, including 20.8% of the total sequence count, was far more abundant in the indoor compartments (30.5% in living rooms and 25.2% in bathrooms) than outside (6.5%). The same trend appeared for Saccharomycetales, Agaricales, Helotiales, Malasseziales and Mucorales. In contrast, Capnodiales, Pucciniales, Lecanorales and Chaetothyriales were clearly more abundant in the outdoor samples. Like for the order level, there were also clear trends for the most common genera (Figure 5c): *Penicillium*, *Aspergillus*, *Saccharomyces*, *Malassezia* and *Botrytis* were far more abundant indoors, while *Cladosporium*, *Thekopsora*, *Verrucocladosporium*, *Scoliciosporum* and *Hypogymnia* were more abundant outside. *Cladosporium* was the overall most abundant genus, representing 13.3% of all sequences, which mostly correspond to the most abundant OTU (OTU1 with 12.6% of the total sequences; Figure 3d). In addition, the yeast genera *Malassezia* and *Aureobasidium* were particularly abundant in bathrooms.

3.4 | Indoor vs. outdoor mycobiomes

A large proportion of the fungi (36.3% of the OTUs) were present in all three house compartments and 50.6% of the OTUs were shared between indoor and outdoor compartments (Figure 6a left). However, after excluding low-abundance OTUs (with <10 reads per sample), only 27.4% of the OTUs were shared between outdoor and indoor samples (Figure 6a right), indicating that the relatively high overlap was largely driven by rare fungi. In addition, comparing the overlap on a house-by-house basis revealed that only 15% of the OTUs on average were

TABLE 1 PERMANOVA results summarizing the variability explained by each variable on the compositional variation of mycobiomes

Variable (source)[a]	VPA group	All samples R^2 (%)[b]	Outdoor R^2 (%)[b]	Indoor R^2 (%)[b]
House compartment	House compartment	7.49*	n/a	0.64*
Indoor vs. outdoor		7.12*	n/a	n/a
Temperature seasonality (BIO4)	Climate	1.98*	2.84*	2.40*
Mean temperature of the warmest quarter (BIO10)	Climate	1.82*	2.73*	2.22*
Mean temperature of the driest quarter (BIO9)	Climate	1.70*	2.43*	2.06*
Annual precipitation (BIO12)	Climate	1.66*	2.45*	2.03*
Snow-covered area in February (MET)	Climate	1.58*	2.50*	1.82*
Latitude		1.51*	2.29*	1.86*
Mean temperature of the coldest quarter (BIO11)	Climate	1.35*	1.99*	1.60*
Pests	Building	1.35*	2.48	1.99*
Growing season length (MET)	Climate	1.26*	2.08*	1.41*
Land cover AR50 (NIBIO)		1.20*	2.19	1.76*
Allergies	Occupants	1.07	2.66	1.61*
Annual mean temperature (BIO1)	Climate	1.04*	1.65*	1.22*
Potential incoming solar radiation (Geodata)		1.04*	1.47*	1.30*
Snow water equivalent in April (MET)	Climate	0.99*	1.72*	1.17*
Longitude		0.96*	1.60*	1.11*
Pets	Occupants	0.87*	1.30	1.46*
Building type	Building	0.85*	1.54*	1.24*
Building material	Building	0.70*	1.21	1.03*
Bedrock nutrient (NGU)		0.63*	0.90	0.87*
Ventilation type	Building	0.61*	1.32	0.96*
No. of children	Occupants	0.56*	0.46	1.02*
No. of people	Occupants	0.55*	0.43	1.03*
Urban/rural area		0.50*	0.90*	0.64*
No. of females	Occupants	0.44*	0.41	0.75*
Construction year	Building	0.34*	0.53	0.51*
Moisture problem	Building	0.28*	0.40	0.44*
Dust coverage (Mycotape2)	Building	0.22	0.32	0.40*
Odour problem	Building	0.21	0.37	0.32
Water damage	Building	0.20	0.46	0.30*
Asthma	Occupants	0.17	0.42	0.26

Note: Asterisks indicate significant R^2 values according to their Bonferroni-corrected *p*-values ($p < .05$).

Abbreviation: n/a, not applicable.

[a]Data for the majority of studied variables were collected from the volunteers through an online survey. Sources for other variables: WorldClim for bioclimatic variables (BIO), The Norwegian Meteorological Institute (MET), Norwegian Institute of Bioeconomy Research (NIBIO) and Geodata AS (Geodata). Dust coverage data were generated from the analysis of adhesive tape samples (Mycotape2) collected from the living room.

[b]Three data sets: All samples ($n = 807$), Outdoor ($n = 266$) and Indoor ($n = 541$).

shared between outdoor and indoor samples, while 75% of the OTUs appeared uniquely in one of the house compartments (Figure 6b). Based on a GLM analysis, none of the assessed variables significantly explained the varying degree of overlap in community composition between indoor and outdoor compartments ($p > .05$).

The indicator species analysis revealed 241 OTUs (3.6% of the total number of OTUs) to be significantly ($p < .05$) associated with indoor environments, while 550 OTUs (8.3%) were associated with outdoor environments. In line with the taxonomic results (Figure 5), there were many indoor indicator OTUs in Eurotiales (16.9% of the

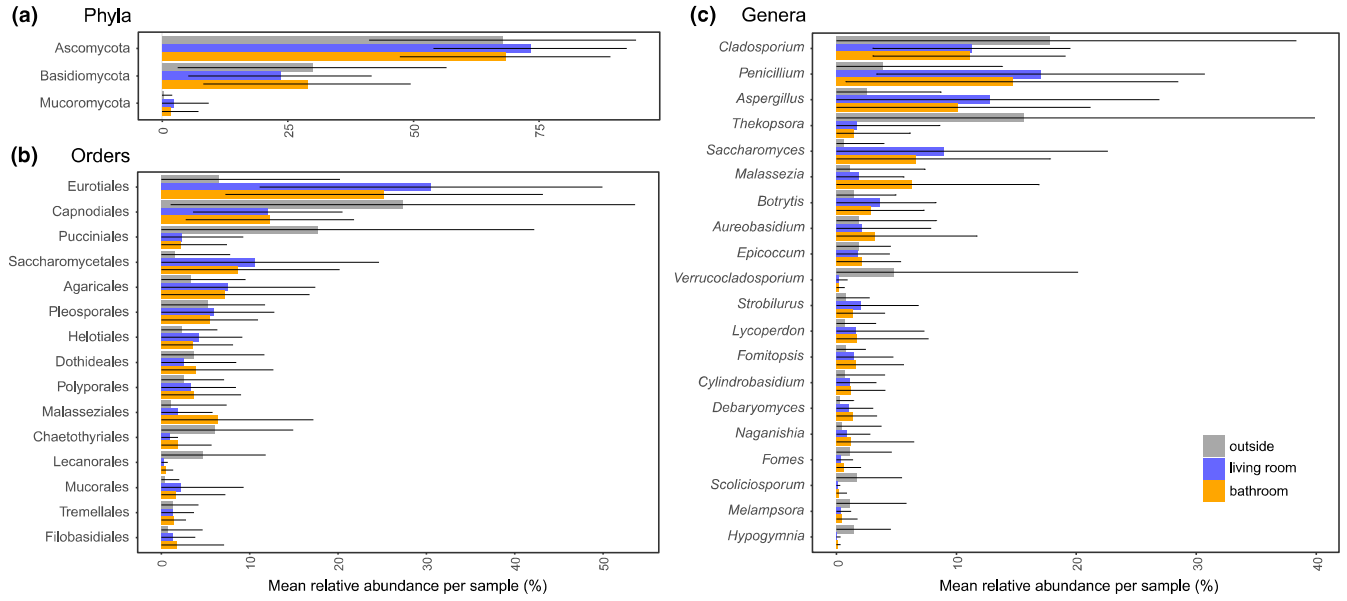


FIGURE 5 Taxonomic distribution of the most abundant fungi detected in the studied houses. Taxa are split according to the three compartments: outside, living room and bathroom, and shown at three taxonomic levels: (a) top-three phyla, (b) top-15 orders, and (c) top-20 genera. Relative abundance of taxa are mean values per sample calculated based on the complete rarefied data set. Unidentified OTUs at the corresponding taxonomic levels (7.9%, 14.6% and 37.8% of the OTUs at the phylum, order and genus levels, respectively) were excluded

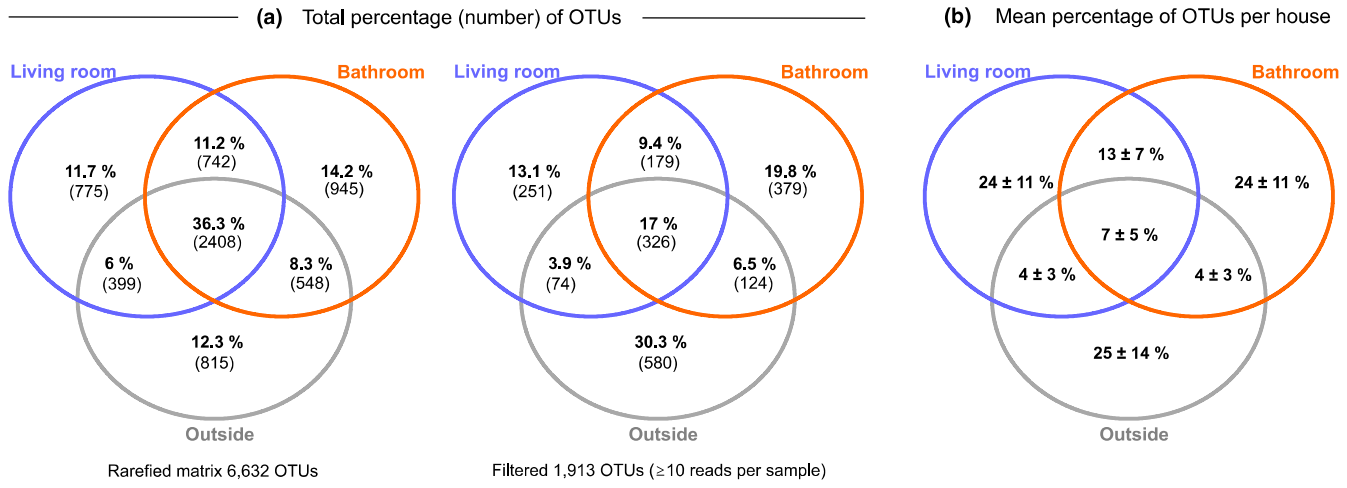


FIGURE 6 Venn diagrams showing the distribution of dust mycobiomes across the three house compartments. The three diagrams show the proportions of OTUs across overall data (a, left), after removing low-abundance OTUs (<10 reads per sample) (a, right), and when comparing at a house-by-house basis (b). Mean percentages of OTUs are shown together with standard deviations

indicator OTUs) and Agaricales (15.5%), and numerous outdoor indicator OTUs in Lecanorales (16.5%), Chaetothyriales (16.5%) and Capnodiales (13.4%) (Figure S10). OTUs with the highest indicator values (IndVal > 50%) for indoor and outdoor environments are detailed in Table 2. Overall, indoor indicator fungi were mostly characterized by their allergenic potential and association with human skin and material colonization, while outdoor indicator fungi were associated with rock-inhabiting fungal taxa.

4 | DISCUSSION

4.1 | Determinants for the indoor dust mycobiome

From previous studies, in other parts of the world (mainly USA), there is limited knowledge on intrinsic and extrinsic factors contributing to the indoor mycobiomes. To narrow this gap, we evaluated the importance of numerous factors related to outdoor

TABLE 2 Indicator species (IndVal > 50%; $p < .05$)

OTU ID	Phylum	Order	Genus	Traits[b]	RA[c] (%)	Occurrence[d] (% houses)	IndVal (%)
Indoor indicators							
OTU3[a]	Ascomycota	Saccharomycetales	*Saccharomyces*	O, A, S	5.42	95.6	91.2
OTU4[a]	Ascomycota	Eurotiales	*Penicillium*	O, A, S, M	6.23	99.3	89.5
OTU18[a]	Ascomycota	Eurotiales	*Penicillium*	O, A, S, M	1.40	94.5	86.0
OTU49	Ascomycota	Eurotiales	*Penicillium*	O, A, S, M	0.51	91.9	83.9
OTU23[a]	Ascomycota	Saccharomycetales	*Debaryomyces*	O, S	0.87	94.8	83.2
OTU34[a]	Ascomycota	Eurotiales	*Penicillium*	O, A, S, M	0.79	94.5	82.2
OTU44	Ascomycota	Eurotiales	*Aspergillus*	O, A, S, M	0.53	90.0	79.6
OTU51	Basidiomycota	Malasseziales	*Malassezia*	A, S	0.27	81.9	77.8
OTU65	Basidiomycota	Sporidiobolales	*Rhodotorula*	O, A, S, M	0.30	81.2	75.1
OTU47	Ascomycota	Eurotiales	*Aspergillus*	O, A, S, M	0.41	84.9	74.1
OTU32	Basidiomycota	Filobasidiales	*Naganishia*		0.51	79.3	73.8
OTU91	Ascomycota	Eurotiales	*Penicillium*	O, A, S, M	0.22	71.2	70.5
OTU26[a]	Ascomycota	Eurotiales	*Aspergillus*	O, A, S, M	1.07	80.1	70.1
OTU130	Ascomycota	Saccharomycetales	*Candida*	O, A, S	0.13	74.2	69.7
OTU63	Basidiomycota	Malasseziales	*Malassezia*	A, S	0.23	71.6	68.8
OTU58	Ascomycota	Eurotiales	*Penicillium*	O, A, S, M	0.43	60.9	65.9
OTU87	Ascomycota	Capnodiales	*Cladosporium*	A, M	0.17	64.2	61.6
OTU127	Ascomycota	Eurotiales	*Aspergillus*	O, A, S, M	0.13	60.1	61.5
OTU112	Basidiomycota	Polyporales	*Phlebia*		0.07	62.7	61.1
OTU80	Basidiomycota	Sporidiobolales	*Sporobolomyces*	O	0.12	60.5	60.3
OTU115	Basidiomycota	Agaricales	*Panellus*		0.07	61.2	57.6
OTU148	Ascomycota	Saccharomycetales	*Candida*	O, A, S	0.11	50.2	55.8
OTU164	Basidiomycota	Malasseziales	*Malassezia*	A, S	0.07	48.3	53.6
OTU41	Mucoromycota	Mucorales	*Mucor*	M	0.47	45.4	53.5
OTU250	Basidiomycota	Polyporales	*Rigidoporus*		0.04	52.4	52.5
OTU66	Basidiomycota	Wallemiales	*Wallemia*	O	0.14	46.5	52.3
OTU81	Basidiomycota	Wallemiales	*Wallemia*	O	0.11	45.0	51.5
OTU261	Basidiomycota	Agaricales	*Agaricus*		0.07	38.7	51.4
Outdoor indicators							
OTU2[a]	Basidiomycota	Pucciniales	*Thekopsora*	P	6.26	86.3	85.9
OTU74	Ascomycota	Chaetothyriales		R	0.48	87.8	84.4
OTU57	Ascomycota	Lecanorales	*Scoliciosporum*	R, L	0.55	84.1	80.5
OTU114	Ascomycota	Lecanorales	*Parmelia*	R, L	0.23	67.5	75.3
OTU122	Ascomycota	Lecanorales	*Hypogymnia*	R, L	0.23	65.7	73.4
OTU147	Ascomycota	Chaetothyriales		R	0.21	64.2	70.8
OTU152	Ascomycota	Lecanorales	*Hypogymnia*	R, L	0.20	55.7	69.9
OTU85	Ascomycota	Capnodiales	*Neocatenulostroma*	R	0.34	73.4	65.3
OTU319	Ascomycota	Lecanorales	*Scoliciosporum*	R, L	0.08	43.9	59.5
OTU227	Ascomycota	Chaetothyriales		R	0.11	55.7	58.1
OTU113	Ascomycota	Capnodiales	*Devriesia*	R	0.24	43.2	53.7
OTU150	Ascomycota				0.16	48.3	53.2
OTU349	Ascomycota	Dothideales	*Perusta*	R	0.06	40.9	52.7
OTU484	Ascomycota	Lecanorales	*Scoliciosporum*	R, L	0.05	39.5	51.8
OTU188	Ascomycota	Myriangiales		R	0.08	40.6	51.8

(Continues)

TABLE 2 (Continued)

OTU ID	Phylum	Order	Genus	Traits[b]	RA[c] (%)	Occurrence[d] (% houses)	IndVal (%)
OTU525	Ascomycota	Chaetothyriales		R	0.03	39.8	51.7
OTU244	Ascomycota	Chaetothyriales	*Cladophialophora*	R	0.11	43.9	51.4
OTU510	Ascomycota	Capnodiales		R	0.04	44.3	51.4
OTU265	Ascomycota				0.05	34.7	51.1
OTU369	Ascomycota	Chaetothyriales		R	0.04	38.7	50.3

[a]Top-20 most abundant fungi, also detailed in the ordination plot for OTUs (Figure 3c,d).

[b]Relevant traits: osmotolerant fungi (O) (Flannigan & Miller, 2011; Gostinčar et al., 2018), allergenic fungi (A) (Esch et al., 2001; Yamamoto et al., 2012), skin-associated fungi (S) (Findley et al., 2013), material-colonizing fungi (M) (Andersen et al., 2011; Flannigan & Miller, 2011), plant pathogenic fungi (P), rock-inhabiting fungi (R) and lichen-forming fungi (L) (Ametrano et al., 2019; Gostinčar et al., 2018).

[c]Relative abundance as proportion of the total number of rarefied reads.

[d]Percentage of study houses (n = 271) where the indicator species were detected, considering the three compartments.

conditions, building features and occupant characteristics, in order to identify tentative drivers of fungal diversity and community composition in Norwegian houses. This first study on indoor mycobiomes at a large geographical scale in Europe (Norway spans the latitudes 57–81°N) revealed that 15% of the overall community composition variation can be explained by the assessed variables. The fungal community composition in house dust was clearly different between indoor and outdoor samples. After accounting for the key effect of the house compartment (7.66% of the variation), our results corroborated the first hypothesis (H1), namely that regional-scale climate is the most important driver of the mycobiome (4.18%), while building and occupant factors have significant influence, but to a much lesser extent (1.44% and 1.11%, respectively). These findings are in agreement with previous mycobiome studies in the built environment (Adams et al., 2013a; Amend et al., 2010; Barberán et al., 2015a, 2015b; Stephens, 2016). Amend et al. (2010) first suggested that large-scale (extrinsic) factors are driving the fungal composition in buildings, rather than specific building features. Likewise, Barberán et al. (2015a) reported that climatic variables (particularly mean annual temperature and precipitation) were the best predictors for indoor mycobiomes across North America, explaining 14% of the variation in indoor mycobiomes, in contrast to the 5% explained by building features. In our study, the three climatic variables (i) temperature seasonality, and mean temperature of the (ii) warmest and (iii) driest quarters showed better explanatory power than annual precipitation and annual mean temperature.

Although having limited explanatory power, most of the building features were significantly related to the indoor mycobiome. The presence of pests was the most relevant building factor, accounting for 1.99% of the variation among indoor samples. The volunteers reported in particular three kind of pests: mice, rats and long-tailed silverfish. The prevalence of the long-tailed silverfish (*Ctenolepisma longicaudata*) has increased notably in Europe in recent years and is considered a major nuisance pest in modern buildings in Norway (Aak et al., 2019). Madden et al. (2016) reported that arthropod and microbial (fungi and bacteria) diversities follow parallel trends in settled-dust samples. Other building factors studied (type of building, material, ventilation, construction year and moisture-related problems) also explained smaller fractions of the indoor mycobiome variation (R^2 values between 1.24% and 0.44%).

The most relevant occupant-related variable was the presence of allergy cases (including pollen, food and skin reactions), associated with 1.61% of the indoor variation in fungal community composition. In indoor samples, we found a striking abundance of taxa with allergenic effects on humans, and such taxa were also identified as indoor indicator species. Furthermore, the number of occupants and the presence of pets were also significant explanatory variables, with R^2 values of 1.46% and 1.03%, respectively. Dannemiller et al. (2016) previously reported the influence of occupancy (people, children and pets) on the fungal community composition, with an increased richness associated with the presence of pets. Nevertheless, to the best of our knowledge, this is the first study that has revealed a positive correlation between the number of occupants and mycobiome composition, where an increased number of occupants drive the community towards the outdoor species composition. This trend may partially be explained by a higher exchange of particles (i.e., aerosols and dust) transported through clothing and shoes.

4.2 | Fungal diversity in Norwegian houses

Fungal richness, evenness and alpha diversity were consistently higher in indoor than outdoor samples. The same trend has been reported in previous studies as well (Barberán et al., 2015a; Yamamoto et al., 2015). As suggested by Barberán et al. (2015a), this tendency may be due to two inter-related phenomena: (i) the dominance of a few taxa in the outdoor communities and (ii) the higher richness/diversity indoors, including a mixture of outdoor and indoor fungi. Both phenomena were probably relevant in our study. Dominant outdoor taxa from the genera *Cladosporium*, *Thekopsora* and *Verrucocladosporium* are among the top-20 OTUs (1st, 2nd and 8th most abundant OTUs) and occurred in more than 80% of the houses, in both outdoor and indoor compartments.

In contrast, studies of specific building units reported the opposite trend, with higher fungal diversity and richness outdoors

(Adams et al., 2013a, 2014; Sylvain et al., 2019). This trend was reported for the fungal diversity and biomass in settled dust from water-damaged units of a housing complex in San Francisco, with the lowest diversity inside units with visible moulds (Sylvain et al., 2019). However, that finding was associated with the influence of a few dominant taxa, which were probably growing and spreading from mould colonies indoors. In this regard, Adams, Amend, Taylor, and Bruns (2013) demonstrated that local sources of abundantly sporulating fungi might distort the perception of species richness and community composition assessed by PCR-based HTS approaches, where a few abundant species can mask the presence of rarer fungi during the PCR.

In addition, several studies have reported a global trend for fungal diversity and richness that increase with latitude (Amend et al., 2010; Větrovský et al., 2019). Our study also supports this trend, as slightly higher alpha diversities were obtained for houses in northern Norway.

In agreement with previous studies in the built environment, which mainly described air- and dust-borne communities, the mycobiomes in studied houses were clearly dominated by ascomycetes (~70%) with Capnodiales and Eurotiales as major orders in abundance, corroborating our hypothesis H2. These orders are well known for their stress tolerance; Capnodiales (with *Cladosporioum* as the dominant genus in our data set) is particularly rich in extremotolerant species, including saprobes, plant pathogens, endophytes, epiphytes and rock-inhabiting fungi (Ametrano et al., 2019; Crous et al., 2009), while *Eurotiales* contains many xerophilic fungi (especially *Aspergillus* and *Penicillium* species) that are able to grow on substrates with low water activity (*aw* ≤ 0.85) like household dust (Flannigan & Miller, 2011; Pettersson & Leong, 2011).

Interestingly, we observed a distinct difference in the overall distribution of Ascomycota and Basidiomycota; the former was to a higher extent connected to areas with high annual precipitation and longer growing season, while basidiomycetes were more prevalent in continental areas with high degree of seasonality and high snow cover during winter. More than reflecting the actual biogeography of the two phyla, we speculate that this pattern may partly be due to temporal differences in the vegetation period across the study area. During the sampling campaign in May, plant growth had probably progressed more in areas with a longer growing season, meaning that a larger proportion of plant-associated ascomycetes (including, for example, pathogens, endophytes and saprotrophs) had become dominant in these areas. Furthermore, several of the most dominant basidiomycetes, including *Fomitopsis* sp. and *Strobilurus* sp., are known to be prevalent in coniferous forests that are more abundant in continental climates.

4.3 | Overlap between indoor and outdoor mycobiomes

In light of previous studies (Adams et al., 2013a; Barberán et al., 2015a), we expected that a major part of the indoor fungi originated from outdoor sources (H3). Barberán et al. (2015a) reported that 65% of the indoor fungal OTUs were also present outdoors. In our study, this overlap was 58% and to a considerable extent driven by low-abundance fungi (39% overlap after excluding OTUs with <10 reads per sample). However, on a house-by-house basis, only 15% of the OTUs were present in both outdoor and indoor environments, and only 13% of the OTUs in both indoor compartments (living room and bathroom). The low overlap between compartments in single houses may be due to the limited representativeness of the collected samples (one per house compartment) and/or the influence of indoor fungal sources nearby the sampled surfaces. Considering these results, our hypothesis H3 has been partly refuted, as we cannot conclude that the major fraction of indoor fungi was from outdoor sources. As stated by Yamamoto et al. (2015), the indoor emissions related to occupant activities may also act as primary sources for the indoor mycobiome. They reported that 70% of indoor fungal aerosol particles (80% for allergenic taxa), collected from seven classrooms of four different countries, were associated with indoor emissions. Diverse indoor fungal sources, including spoiled materials and food, waste, potted plants, drains and skin debris, have been recognized in previous studies (Adams et al., 2013b; Nevalainen et al., 2015; Tong et al., 2017). Presumably, the indoor mycobiome is assembled by a combination of outdoor and indoor sources and their exact contributions are hard to tease apart. More detailed experimental studies are needed for this purpose.

The indoor core mycobiome from Norwegian houses (i.e., those fungi significantly associated with their indoor environments) is similar to what has been reported in other countries. We detected two main groups of indoor fungi: (i) the well-known household xerophilic moulds belonging to Eurotiales (17% of indoor indicator OTUs; mostly to the genera *Penicillium* and *Aspergillus*) and the basidiomycete genus *Wallemia* (3%), and (ii) yeasts belonging to the orders Saccharomycetales (6%; genera *Saccharomyces*, *Debaryomyces* and *Candida*) and Sporidiobolales (2%; genera *Rhodotorula* and *Sporobolomyces*), as well as the basidiomycete genus *Malassezia* (4%).

Penicillium and *Aspergillus* species are ubiquitous fungi found in dust and air samples, both indoors and outdoors, during all seasons (Barberán et al., 2015b; Flannigan, 2011; Nevalainen et al., 2015; Rintala et al., 2012; Shelton et al., 2002). They are especially abundant indoors, as part of household dust or colonizing building materials and foodstuffs, which become relevant sources for further conidial dispersion (Andersen et al., 2011; Flannigan & Miller, 2011). *Wallemia* is an extreme xerophilic basidiomycete, commonly found in dust due to its ability to grow at low water potential, *aw* < 0.75 (Flannigan & Miller, 2011; Zajc & Gunde-Cimerman, 2018). The yeast genera *Malassezia*, *Debaryomyces*, *Candida* and *Rhodotorula* are commensal fungi associated with human skin, showing a prevalence in indoor environments (Dannemiller et al., 2014; Findley et al., 2013; Flannigan, 2011; Maestre et al., 2018; Rintala et al., 2012; Tong et al., 2017). The fourth most-abundant species (OTU3, 5.4% of the total reads, present in 96% of houses), with the highest indoor IndVal (91.2%), was identified as *Saccharomyces* sp., a relevant genus in food production that

includes *S. cerevisiae* (baker's and brewer's yeast) and has previously been reported in indoor environments (Barberán et al., 2015a; Flannigan, 2011). The majority of these indoor fungi have been described as potential allergenic taxa (Esch et al., 2001; Yamamoto et al., 2012). Lastly, there was a significantly higher occurrence (mean = 21% of study houses) of indoor indicator species compared to outdoor indicators (9%), supporting that there is a consistent indoor core mycobiome.

Outdoor dust mycobiomes, collected at the doorframe of the main entrance outside the buildings, also showed striking differences compared to the indoor mycobiomes. Besides the prevalence of *Cladosporium* and *Thekopsora* (18% and 16% mean relative abundance per samples, respectively), the indicator species analysis revealed that outdoor samples were distinctly enriched in so-called rock-inhabiting fungi, including lichen-forming fungi of the order Lecanorales (16% of outdoor indicator OTUs), as well as fungi affiliated to Chaetothyriales (16%) and Capnodiales (13%). They are well known for their multistress tolerance and prevalence in diverse outdoor environments including rocks and buildings, where they are exposed to solar radiation, desiccation and rehydration, temperature fluctuations, osmotic stress, pollutants and lack of nutrients (Ametrano et al., 2019; Gorbushina, 2007).

4.4 | Concluding remarks

In summary, we have shown that numerous factors are related to the composition of the indoor mycobiomes, but together only explain a small fraction of the community composition. This seems to be a general feature of fungal communities. Further observational or experimental studies should be addressed to assess the causal effect(s) of one or a few factors using a balanced and cross-factorial design. For example, regional environmental variation can be removed by focusing on a smaller geographical area where two factors, such as number of occupants and building types, can be systematically evaluated.

Our main findings are in line with previous indoor mycobiome studies, identifying climatic variables as the key determinants of the indoor mycobiome. Building features and occupant characteristics had a significant but smaller influence. The indoor dust mycobiome represents a mixture of fungi from outdoor and indoor sources, which could also be the reason why a higher fungal richness was observed indoors. The indoor core mycobiome is characterized by two ecological groups with allergenic potential, xerophilic moulds and skin-associated yeasts. In contrast, rock-inhabiting fungi, well known for their multistress tolerance and ability to form biofilms on buildings, were the main outdoor indicator fungi.

Despite methodological limitations related to the citizen science sampling (e.g., nonuniform means of collection, small amount of dust collected with subsequently low DNA yields and low number of samples per house), this approach turned out to be highly effective and we were able to obtain a large number of samples covering Norway in a relatively short time. The DNA analyses revealed that most samples could be used in statistical analyses, with no divergent outlier samples. Moreover, most indoor and outdoor samples fall into two separate clusters, supporting that the samples were collected according to our instructions. We believe the citizen science approach holds large opportunities for further broad-scale sampling within countries and continents, but also at a global scale. Not only can indoor environments be sampled this way, but also various outdoor environments such as soil and plants. In addition to democratization of science, citizen science is a way to reduce unnecessary travelling and related carbon emissions.

ACKNOWLEDGEMENTS

This study received funding from the European Union's Horizon 2020 research and innovation programme through a Marie Skłodowska-Curie Individual Fellowship to P.M.M.S., under the grant agreement MycoIndoor No. 741332. We acknowledge the University of Oslo and Centre for Digital Life Norway for funding to H.K. All volunteers who provided samples and data relating to their houses are acknowledged for their essential contributions. We acknowledge Peter Horvath and Anders Bryn, from the Natural History Museum at UiO, for providing the metadata of environmental variables. Maria Nuñez from Mycoteam AS is acknowledged for her valuable feedback on fungal taxonomy and building ecology.

AUTHOR CONTRIBUTIONS

P.M.M.S., E.L.F.E., I.S. and H.K. conceived and designed the study. E.L.F.E. organized the citizen science sampling with contribution from I.B.E. P.M.M.S. performed laboratory work. P.M.M.S. and L.N.M. analysed data (bioinformatics and statistics) and prepared figures. S.M. provided technical advice on laboratory work and contributed to statistical analyses. P.M.M.S. wrote the first draft of the manuscript. All authors edited and approved the final manuscript.

DATA AVAILABILITY STATEMENT

All supplementary figures and tables are supplied as supporting information. Raw sequences (fastq files) of this study are available on ENA at EMBL-EBI under accession no. PRJEB42161 (https://www.ebi.ac.uk/ena/browser/view/PRJEB42161). All data sets analysed during the current study are available on Dryad https://doi.org/10.5061/dryad.59zw3r24w (Martin-Sanchez et al., 2021), including map files, metadata, the original OTU table, the final fungal rarefied OTU table, the taxonomic assignment and representative ITS2 sequences of their OTUs, as well as the R scripts used for data analyses. Personal data of the citizen scientists (i.e., addresses and geographical coordinates of their houses) were omitted from the metadata.

ORCID

Pedro M. Martin-Sanchez https://orcid.org/0000-0003-2674-8650
Sundy Maurice https://orcid.org/0000-0002-5376-0981
Inger Skrede https://orcid.org/0000-0002-1113-7403
Håvard Kauserud https://orcid.org/0000-0003-2780-6090

REFERENCES

Aak, A., Rukke, B. A., Ottesen, P. S., & Hage, M. (2019). *Long-tailed silverfish* (Ctenolepisma longicaudata) – *Biology and control*. Norwegian Institute of Public Health Report (pp. 1–43). Norwegian Institute of Public Health.

Adams, R. I., Amend, A. S., Taylor, J. W., & Bruns, T. D. (2013). A unique signal distorts the perception of species richness and composition in high-throughput sequencing surveys of microbial communities: A case study of fungi in indoor dust. *Microbial Ecology*, *66*, 735–741. https://doi.org/10.1007/s00248-013-0266-4

Adams, R. I., Bateman, A. C., Bik, H. M., & Meadow, J. F. (2015). Microbiota of the indoor environment: A meta-analysis. *Microbiome*, *3*, 49. https://doi.org/10.1186/s40168-015-0108-3

Adams, R. I., Bhangar, S., Dannemiller, K. C., Eisen, J. A., Fierer, N., Gilbert, J. A., Green, J. L., Marr, L. C., Miller, S. L., Siegel, J. A., Stephens, B., Waring, M. S., & Bibby, K. (2016). Ten questions concerning the microbiomes of buildings. *Building and Environment*, *109*, 224–234.

Adams, R. I., Miletto, M., Lindow, S. E., Taylor, J. W., & Bruns, T. D. (2014). Airborne bacterial communities in residences: similarities and differences with fungi. *PLoS One*, *9*, e91283. https://doi.org/10.1371/journal.pone.0091283

Adams, R. I., Miletto, M., Taylor, J. W., & Bruns, T. D. (2013a). Dispersal in microbes: Fungi in indoor air are dominated by outdoor air and show dispersal limitation at short distances. *The ISME Journal*, *7*, 1262–1273. https://doi.org/10.1038/ismej.2013.28

Adams, R. I., Miletto, M., Taylor, J. W., & Bruns, T. D. (2013b). The diversity and distribution of fungi on residential surfaces. *PLoS One*, *8*, e78866. https://doi.org/10.1371/journal.pone.0078866

Amann, R. I., Ludwig, W., & Schleifer, K. H. (1995). Phylogenetic identification and in situ detection of individual microbial cells without cultivation. *Microbiological Reviews*, *59*, 143–169.

Amend, A. S., Seifert, K. A., Samson, R., & Bruns, T. D. (2010). Indoor fungal composition is geographically patterned and more diverse in temperate zones than in the tropics. *Proceedings of the National Academy of Sciences of the United States of America*, *107*, 13748–13753. https://doi.org/10.1073/pnas.1000454107

Ametrano, C. G., Muggia, L., & Grube, M. (2019). Extremotolerant black fungi from rocks and lichens. In S. Tiquia-Arashiro & M. Grube (Eds.), *Fungi in extreme environments: Ecological role and biotechnological significance* (pp. 119–143). Springer.

Andersen, B., Frisvad, J. C., Søndergaard, I., Rasmussen, I. S., & Larsen, L. S. (2011). Associations between fungal species and water-damaged building materials. *Applied and Environmental Microbiology*, *77*, 4180–4188. https://doi.org/10.1128/AEM.02513-10

Barberán, A., Dunn, R. R., Reich, B. J., Pacifici, K., Laber, E. B., Menninger, H. L., Morton, J. M., Henley, J. B., Leff, J. W., Miller, S. L., & Fierer, N. (2015a). The ecology of microscopic life in household dust. *Proceedings of the Royal Society B*, *282*, 20151139. https://doi.org/10.1098/rspb.2015.1139

Barberán, A., Ladau, J., Leff, J. W., Pollard, K. S., Menninger, H. L., Dunn, R. R., & Fierer, N. (2015b). Continental-scale distributions of dust-associated bacteria and fungi. *Proceedings of the National Academy of Sciences of the United States of America*, *112*, 5756–5761. https://doi.org/10.1073/pnas.1420815112

Callahan, B. J., McMurdie, P. J., Rosen, M. J., Han, A. W., Johnson, A. J., & Holmes, S. P. (2016). DADA2: High-resolution sample inference from Illumina amplicon data. *Nature Methods*, *13*, 581–583. https://doi.org/10.1038/nmeth.3869

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*, 747–749. https://doi.org/10.1016/j.funeco.2012.06.003

Crous, P. W., Schoch, C. L., Hyde, K. D., Wood, A. R., Gueidan, C., de Hoog, G. S., & Groenewald, J. Z. (2009). Phylogenetic lineages in the *Capnodiales*. *Studies in Mycology*, *64*, 17–47S7. https://doi.org/10.3114/sim.2009.64.02

Dannemiller, K. C., Gent, J. F., Leaderer, B. P., & Peccia, J. (2016). Influence of housing characteristics on bacterial and fungal communities in homes of asthmatic children. *Indoor Air*, *26*, 179–192. https://doi.org/10.1111/ina.12205

Dannemiller, K. C., Mendell, M. J., Macher, J. M., Kumagai, K., Bradman, A., Holland, N., Harley, K., Eskenazi, B., & Peccia, J. (2014). Next generation DNA Sequencing reveals that low fungal diversity in house dust is associated with childhood asthma development. *Indoor Air*, *24*, 236–247. https://doi.org/10.1111/ina.12072

De Cáceres, M., & Legendre, P. (2009). Associations between species and groups of sites: Indices and statistical inference. *Ecology*, *90*, 3566–3574. https://doi.org/10.1890/08-1823.1

Dray, S., & Dufour, A. B. (2007). The ade4 package: Implementing the duality diagram for ecologists. *Journal of Statistical Software*, *22*, 1–20. https://doi.org/10.18637/jss.v022.i04

Esch, R. E., Hartsell, C. J., Crenshaw, R., & Jacobson, R. S. (2001). Common allergenic pollens, fungi, animals, and arthropods. *Clinical Reviews in Allergy & Immunology*, *21*, 261–292. https://doi.org/10.1385/CRIAI:21:2-3:261

Estensmo, E. L. F., Maurice, S., Morgado, L. N., Martin-Sanchez, P. M., Skrede, I., & Kauserud, H. (2021). The influence of intraspecific sequence variation during DNA metabarcoding: A case study of eleven fungal species. *Molecular Ecology Resources*, *21*, 1141–1148. https://doi.org/10.1111/1755-0998.13329

Fick, S. E., & Hijmans, R. J. (2017). Worldclim 2: New 1-km spatial resolution climate surfaces for global land areas. *International Journal of Climatology*, *37*, 4302–4315. https://doi.org/10.1002/joc.5086

Findley, K., Oh, J., Yang, J., Conlan, S., Deming, C., Meyer, J. A., Schoenfeld, D., Nomicos, E., Park, M., Kong, H. H., & Segre, J. A. (2013). Topographic diversity of fungal and bacterial communities in human skin. *Nature*, *498*, 367–370. https://doi.org/10.1038/nature12171

Fisk, W. J., Lei-Gomez, Q., & Mendell, M. J. (2007). Meta-analyses of the associations of respiratory health effects with dampness and mold in homes. *Indoor Air*, *17*, 284–296. https://doi.org/10.1111/j.1600-0668.2007.00475.x

Flannigan, B. (2011). Microorganisms in indoor air. In B. Flannigan, R. A. Samson, & J. D. Miller (Eds.), *Microorganisms in home and indoor work environments: Diversity, health impacts, investigation and control* (pp. 25–43). CRC Press.

Flannigan, B., & Miller, J. D. (2011). Microbial growth in indoor environments. In B. Flannigan, R. A. Samson, & J. D. Miller (Eds.), *Microorganisms in home and indoor work environments: Diversity, health impacts, investigation and control* (pp. 57–108). CRC Press.

Frøslev, T. G., Kjøller, R., Bruun, H. H., Ejrnæs, R., Brunbjerg, A. K., Pietroni, C., & Johansen, A. J. (2017). Algorithm for post-clustering curation of DNA amplicon data yields reliable biodiversity estimates. *Nature Communications*, *8*, 1188. https://doi.org/10.1038/s41467-017-01312-x

Gilbert, J. A., & Stephens, B. (2018). Microbiology of the built environment. *Nature Reviews Microbiology*, *16*, 661–670. https://doi.org/10.1038/s41579-018-0065-5

Gorbushina, A. A. (2007). Life on the rocks. *Environmental Microbiology*, *9*, 1613–1631. https://doi.org/10.1111/j.1462-2920.2007.01301.x

Gostinčar, C., Zajc, J., Lenassi, M., Plemenitaš, A., de Hoog, S., Al-Hatmi, A. M. S., & Gunde-Cimerman, N. (2018). Fungi between extremotolerance and opportunistic pathogenicity on humans. *Fungal Diversity*, *93*, 195–213. https://doi.org/10.1007/s13225-018-0414-8

Haines, S. R., Adams, R. I., Boor, B. E., Bruton, T. A., Downey, J., Ferro, A. R., Gall, E., Green, B. J., Hegarty, B., Horner, E., Jacobs, D. E., Lemieux, P., Misztal, P. K., Morrison, G., Perzanowski, M., Reponen, T., Rush, R. E., Virgo, T., Alkhayri, C., … Dannemiller, K. C. (2019). Ten questions concerning the implications of carpet on indoor

chemistry and microbiology. *Building and Environment*, *170*, 1–16. https://doi.org/10.1016/j.buildenv.2019.106589

Horvath, P., Halvorsen, R., Stordal, F., Tallaksen, L. M., Tang, H., & Bryn, A. (2019). Distribution modelling of vegetation types based on area-frame survey data. *Applied Vegetation Science*, *22*, 547–560. https://doi.org/10.1111/avsc.12451

Ihrmark, K., Bödeker, I. T. M., Cruz-Martinez, K., Friberg, H., Kubartova, A., Schenck, J., Strid, Y., Stenlid, J., Brandström-Durling, M., Clemmensen, K. E., & Lindahl, B. D. (2012). New primers to amplify the fungal ITS2 region – Evaluation by 454-sequencing of artificial and natural communities. *FEMS Microbiology Ecology*, *82*, 666–677. https://doi.org/10.1111/j.1574-6941.2012.01437.x

Kettleson, E. M., Adhikari, A., Vesper, S., Coombs, K., Indugula, R., & Reponen, T. (2015). Key determinants of the fungal and bacterial microbiomes in homes. *Environmental Research*, *138*, 130–135. https://doi.org/10.1016/j.envres.2015.02.003

Kõljalg, U., Nilsson, R. H., Abarenkov, K., Tedersoo, L., Taylor, A. F. S., Bahram, M., Bates, S. T., Bruns, T. D., Bengtsson-Palme, J., Callaghan, T. M., Douglas, B., Drenkhan, T., Eberhardt, U., Dueñas, M., Grebenc, T., Griffith, G. W., Hartmann, M., Kirk, P. M., Kohout, P., … Larsson, K.-H. (2013). Towards a unified paradigm for sequence-based identification of fungi. *Molecular Ecology*, *22*, 5271–5277. https://doi.org/10.1111/mec.12481

Lax, S., Smith, D. P., Hampton-Marcell, J., Owens, S. M., Handley, K. M., Scott, N. M., Gibbons, S. M., Larsen, P., Shogan, B. D., Weiss, S., Metcalf, J. L., Ursell, L. K., Vazquez-Baeza, Y., Van Treuren, W., Hasan, N. A., Gibson, M. K., Colwell, R., Dantas, G., Knight, R., & Gilbert, J. A. (2014). Longitudinal analysis of microbial interaction between humans and the indoor environment. *Science*, *345*, 1048–1052. https://doi.org/10.1126/science.1254529

Legendre, P., & Gallagher, E. D. (2001). Ecologically meaningful transformations for ordination of species data. *Oecologia*, *129*, 271–280. https://doi.org/10.1007/s004420100716

Lindahl, B. D., Nilsson, R. H., 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 Phytologist*, *199*, 288–299. https://doi.org/10.1111/nph.12243

Madden, A. A., Barberán, A., Bertone, M. A., Menninger, H. L., Dunn, R. R., & Fierer, N. (2016). The diversity of arthropods in homes across the United States as determined by environmental DNA analyses. *Molecular Ecology*, *25*, 6214–6224. https://doi.org/10.1111/mec.13900

Maestre, J. P., Jennings, W., Wylie, D., Horner, S. D., Siegel, J., & Kinney, K. A. (2018). Filter forensics: Microbiota recovery from residential HVAC filters. *Microbiome*, *6*, 22. https://doi.org/10.1186/s40168-018-0407-6

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal*, *17*, 10–12. https://doi.org/10.14806/ej.17.1.200

Martin-Sanchez, P. M., Estensmo, E. L. F., Morgado, L. N., Maurice, S., Engh, I. B., Skrede, I., & Kauserud, H. (2021). Data from: Analyzing indoor mycobiomes through a large-scale citizen science study. *Dryad Digital Repository*; v2. https://doi.org/10.5061/dryad.59zw3r24w

McDonald, D., Hyde, E., Debelius, J. W., Morton, J. T., Gonzalez, A., Ackermann, G., Aksenov, A. A., Behsaz, B., Brennan, C., Chen, Y., DeRight Goldasich, L., Dorrestein, P. C., Dunn, R. R., Fahimipour, A. K., Gaffney, J., Gilbert, J. A., Gogul, G., Green, J. L., Hugenholtz, P., … Gunderson, B. (2018). American Gut: An open platform for citizen science microbiome research. *mSystems*, *3*, e00031-18. https://doi.org/10.1128/mSystems.00031-18

Mendell, M. J., Mirer, A. G., Cheung, K., Tong, M., & Douwes, J. (2011). Respiratory and allergic health effects of dampness, mold, and dampness-related agents: A review of the epidemiologic evidence. *Environmental Health Perspectives*, *119*, 748–756. https://doi.org/10.1289/ehp.1002410

Nevalainen, A., Täubel, M., & Hyvärinen, A. (2015). Indoor fungi: Companions and contaminants. *Indoor Air*, *25*, 125–156. https://doi.org/10.1111/ina.12182

Nguyen, N. H., Song, Z., Bates, S. T., Branco, S., Tedersoo, L., Menke, J., Schilling, J. S., & Kennedy, P. G. (2016). FUNGuild: An open annotation tool for parsing fungal community datasets by ecological guild. *Fungal Ecology*, *20*, 241–248. https://doi.org/10.1016/j.funeco.2015.06.006

Nilsson, R. H., Anslan, S., Bahram, M., Wurzbacher, C., Baldrian, P., & Tedersoo, L. (2019). Mycobiome diversity: High-throughput sequencing and identification of fungi. *Nature Reviews Microbiology*, *17*, 95–109. https://doi.org/10.1038/s41579-018-0116-y

Oksanen, J., Blanchet, F. G., Kindt, R., Legendre, P., Minchin, P. R., O'Hara, R. B., Simpson, G. L., Solymos, P., Stevens, M. H. H., & Wagner, H. (2019). *vegan: Community ecology package*. R package version 2.5-6. https://CRAN.R-project.org/package=vegan

Paulson, J. N., Stine, O. C., Bravo, H. C., & Pop, M. (2013). Differential abundance analysis for microbial marker-gene surveys. *Nature Methods*, *10*, 1200–1202. https://doi.org/10.1038/nmeth.2658

Pettersson, O. V., & Leong, S. L. (2011). Fungal xerophiles (osmophiles). In *Encyclopedia of Life Science, eLS (Microbiology)* (pp. 1–9). Chichester, UK: John Wiley and Sons Ltd.

Quensen, J. (2018). *ggordiplots: Make ggplot versions of vegan's ordiplots*. R package version 0.3.0. 2018. https://github.com/jfq3/ggordiplots

R Core Team (2018). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Rintala, H., Pitkäranta, M., & Täubel, M. (2012). Microbial communities associated with house dust. *Advances in Applied Microbiology*, *78*, 75–120. https://doi.org/10.1016/B978-0-12-394805-2.00004-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

Shelton, B. G., Kirkland, K. H., Flanders, W. D., & Morris, G. K. (2002). Profiles of airborne fungi in buildings and outdoor environments in the United States. *Applied and Environmental Microbiology*, *68*, 1743–1753. https://doi.org/10.1128/AEM.68.4.1743-1753.2002

Stephens, B. (2016). What have we learned about the microbiomes of indoor environments? *mSystems*, *1*, e00083-16. https://doi.org/10.1128/mSystems.00083-16

Sylvain, I. A., Adams, R. I., & Taylor, J. W. (2019). A different suite: The assemblage of distinct fungal communities in water-damaged units of a poorly-maintained public housing building. *PLoS One*, *14*, e0213355. https://doi.org/10.1371/journal.pone.0213355

Tong, X., Leung, M. H. Y., Wilkins, D., & Lee, P. K. H. (2017). City-scale distribution and dispersal routes of mycobiome in residences. *Microbiome*, *5*, 131. https://doi.org/10.1186/s40168-017-0346-7

UNITE Community (2019a). *UNITE USEARCH/UTAX release for eukaryotes*. Version 18.11.2018. https://doi.org/10.15156/BIO/786346

UNITE Community (2019b). Full UNITE+INSD dataset for Fungi. Version 18.11.2018. Retrieved from https://doi.org/10.15156/BIO/786347

Větrovský, T., Kohout, P., Kopecký, M., Machac, A., Man, M., Bahnmann, B. D., Brabcová, V., Choi, J., Meszárošová, L., Human, Z. R., Lepinay, C., Lladó, S., López-Mondéjar, R., Martinović, T., Mašínová, T., Morais, D., Navrátilová, D., Odriozola, I., Štursová, M., … Baldrian, P. (2019). A meta-analysis of global fungal distribution reveals climate-driven patterns. *Nature Communications*, *10*, 5142. https://doi.org/10.1038/s41467-019-13164-8

Weikl, F., Tischer, C., Probst, A. J., Heinrich, J., Markevych, I., Jochner, S., & Pritsch, K. (2016). Fungal and bacterial communities in indoor dust follow different environmental determinants. *PLoS One*, *11*, e0154131. https://doi.org/10.1371/journal.pone.0154131

White, T. J., Bruns, T. D., Lee, S. B., & Taylor, J. W. (1990). Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. In M. A. Innis, D. H. Gelfand, J. J. Sninsky, & T. J. White (Eds.), *PCR – Protocols and applications – A laboratory manual* (pp. 315–322). Academic Press.

Wickham, H. (2017). *tidyverse: Easily Install and Load the 'Tidyverse'*. R package version 1.2.1. https://CRAN.R-project.org/package=tidyverse

Willis, K. J. (2018). *State of the world's fungi 2018*. Report. : Royal Botanic Gardens, Kew.

Yamamoto, N., Bibby, K., Qian, J., Hospodsky, D., Rismani-Yazdi, H., Nazaroff, W. W., & Peccia, J. (2012). Particle-size distributions and seasonal diversity of allergenic and pathogenic fungi in outdoor air. *The ISME Journal*, *6*, 1801–1811. https://doi.org/10.1038/ismej.2012.30

Yamamoto, N., Hospodsky, D., Dannemiller, K. C., Nazaroff, W. W., & Peccia, J. (2015). Indoor emissions as a primary source of airborne allergenic fungal particles in classrooms. *Environmental Science & Technology*, *49*, 5098–5106. https://doi.org/10.1021/es506165z

Zajc, J., & Gunde-Cimerman, N. (2018). The genus Wallemia – From contamination of food to health threat. *Microorganisms*, *6*, 46. https://doi.org/10.3390/microorganisms6020046

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

How to cite this article: Martin-Sanchez PM, Estensmo E-L, Morgado LN, et al. Analysing indoor mycobiomes through a large-scale citizen science study in Norway. *Mol Ecol*. 2021;00:1–17. https://doi.org/10.1111/mec.15916

MOLECULAR ECOLOGY

Supplemental Information for:

Analyzing indoor mycobiomes through a large-scale citizen science study in Norway

Pedro M. Martin-Sanchez, Eva Lena F. Estensmo, Luis N. Morgado, Sundy Maurice, Ingeborg B. Engh, Inger Skrede, Håvard Kauserud

Table of Contents:

MOLECULAR ECOLOGY

Figure S1. (**a**) Sampling kit sent to the volunteers, including return envelope, instructions, three FLOQSwabs (**b**) and two adhesive tapes (**c**; Mycotape2).

MOLECULAR ECOLOGY

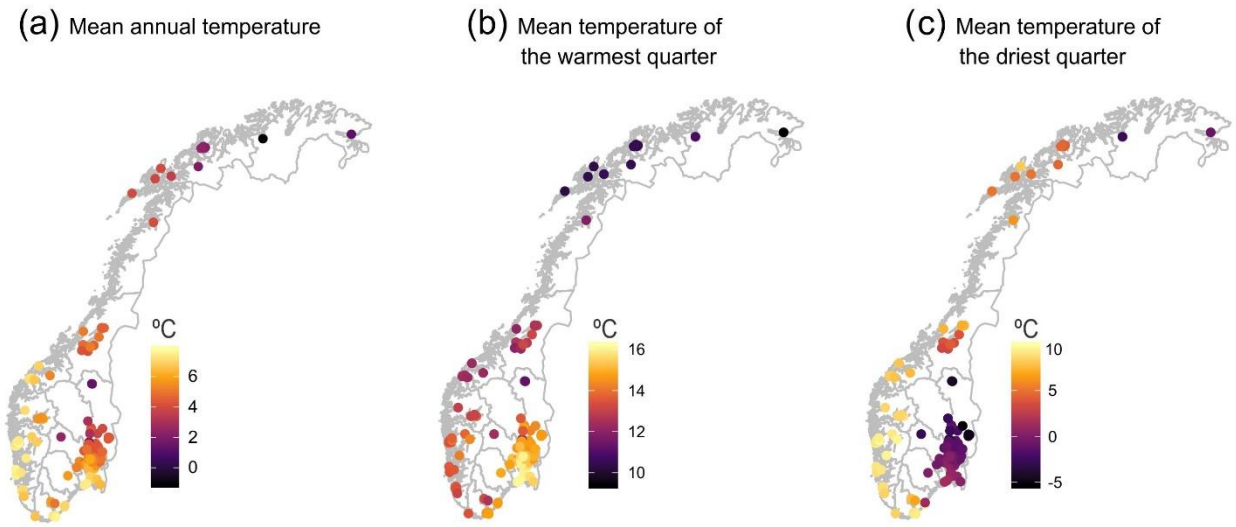


Figure S2. Maps showing the location of the 269 houses (in mainland Norway) colored according to their mean annual temperatures (**a**; BIO1), mean temperatures of the warmest quarter (**b**; BIO10) and mean temperatures of the driest quarter (**c**; BIO9).

MOLECULAR ECOLOGY

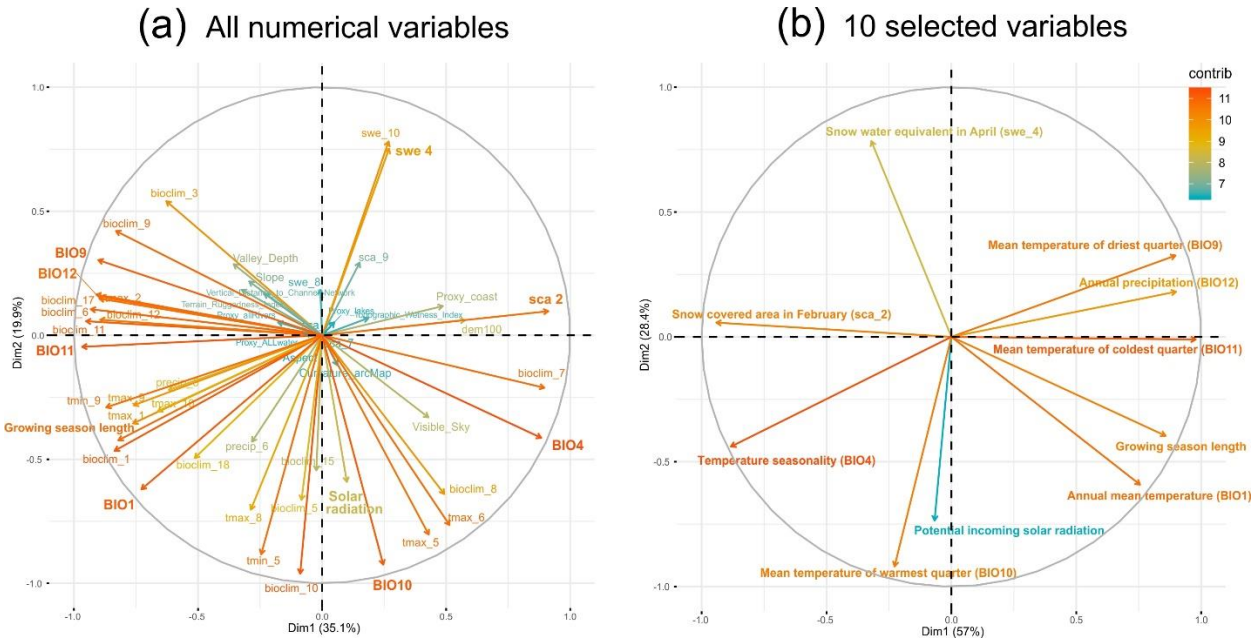


Figure S3. (**a**) PCA analysis for data from 52 environmental continuous variables explored in this study: 46 explanatory variables analyzed in a recent study modelling the vegetation types in Norway (Horvath et al. 2019; variables related to climate, snow, hydrology and topography) plus six relevant WorldClim 2 bioclimatic variables (BIO). Note that five of these WorldClim 2 variables were analyzed by duplicate ("BIO" and "bioclim") because the "bioclim" ones were slightly different (after kriging the spatial resolution from 1 km to 100 m; Horvath et al. 2019). The selected variables are in bold. (**b**) PCA analysis for data from the 10 selected environmental continuous variables.

MOLECULAR ECOLOGY

Table S1. Unique tag combinations used for the PCR libraries.

Barcodes – Forward primer (gITS7)												
	1	2	3	4	5	6	7	8	9	10	11	12
A	NAACAAC	NNNTCACTC	NNTTGAGT	NCTTGGT	NNNAAGGTC	NNGTAACA	NNTTCGGA	NNGCGGTT	NNTCGTTA	NNGAAGCT	NNGTCTTA	NNTATCTG
B	NNAACCGA	NGAACTA	NNNAAGCAG	NNTCCAGC	NGGCGCA	NNNAATCCT	NNNCGACGT	NNNACACAA	NNNTGTGGC	NNNGATATT	NNNTATACC	NNNCGGCCA
C	NNNCCGGAA	NNCCGTCC	NTTGCAA	NNNACTTCA	NNTCGACG	NAGACCG	NCTCATG	NGCTCCG	NCTCTGC	NAGCTGG	NACCTAT	NCCTAAT
D	NAGTGTT	NNNAAGACA	NNCACGTA	NGCGAGA	NNNCCTGTC	NNNTGGCGG	NNNTGTATA	NNNTACTTC	NNNATGGAT	NNNCGCGAT	NNNAGGTAA	NNNACGCGC
E	NNCCGCTG	NCGTGCG	NNNTAACAT	NNTGGAAC	NAGAAGA	NCTATAA	NACAACC	NGTTGCC	NACAGGT	NACATTG	NATTCTA	NGTGTAG
F	NNNAACGCG	NNGGTAAG	NTGCGTG	NNNGTACAC	NNAATAGG	NNAATGAA	NNTCAGAG	NNGTATGT	NNTCCGCT	NNCCAAGG	NNGTGATC	NNTTCCTT
G	NGGCTAC	NNNATAATT	NNGGTCGA	NAAGTGT	NNNGGTTCT	NNNCGAATC	NNNGTAGTG	NNNGTCAAT	NNNGTCCGG	NNNACCATA	NNNGACGGC	NNNCAGAGC
H	NNTTCTCG	NCGTCAC	NNNCACTCT	NNTCTTGG	NTAATGA	NAGAGAC	NAGCACT	NAGCCTC	NCATTAG	NAGGATG	NGTGCCT	NTGATCC
Barcodes – Reverse primer (ITS4)												
A	NNAACAAC	NTCACTC	NNNTTGAGT	NNCTTGGT	NAAGGTC	NNNGTAACA	NNNTTCGGA	NNNGCGGTT	NNNTCGTTA	NNNGAAGCT	NNNGTCTTA	NNNTATCTG
B	NNNAACCGA	NNGAACTA	NAAGCAG	NNNTCCAGC	NNGGCGCA	NAATCCT	NCGACGT	NACACAA	NTGTGGC	NGATATT	NTATACC	NCGGCCA
C	NCCGGAA	NNNCCGTCC	NNTTGCAA	NACTTCA	NNNTCGACG	NNNAGACCG	NNNCTCATG	NNNGCTCCG	NNNCTCTGC	NNNAGCTGG	NNNACCTAT	NNNCCTAAT
D	NNAGTGTT	NAAGACA	NNNCACGTA	NNGCGAGA	NCCTGTC	NTGGCGG	NTGTATA	NTACTTC	NATGGAT	NCGCGAT	NAGGTAA	NACGCGC
E	NNNCCGCTG	NNCGTGCG	NTAACAT	NNNTGGAAC	NNAGAAGA	NNCTATAA	NNACAACC	NNGTTGCC	NNACAGGT	NNACATTG	NNATTCTA	NNGTGTAG
F	NAACGCG	NNNGGTAAG	NNTGCGTG	NGTACAC	NNNAATAGG	NNNAATGAA	NNNTCAGAG	NNNGTATGT	NNNTCCGCT	NNNCCAAGG	NNNGTGATC	NNNTTCCTT
G	NNGGCTAC	NATAATT	NNNGGTCGA	NNAAGTGT	NGGTTCT	NCGAATC	NGTAGTG	NGTCAAT	NGTCCGG	NACCATA	NGACGGC	NCAGAGC
H	NNNTTCTCG	NNCGTCAC	NCACTCT	NNNTCTTGG	NNTAATGA	NNAGAGAC	NNAGCACT	NNAGCCTC	NNCATTAG	NNAGGATG	NNGTGCCT	NNTGATCC

MOLECULAR ECOLOGY

Table S2. Summary of the bioinformatics pipeline used in this study.

Main bioinformatics steps	Tool	# of reads, # of ASVs or OTUS
Initial quality checking of Illumina MiSeq sequences	FastQC	55,568,124 reads
Demultiplexing, removal of primers and short reads (<100 bp)	Cutadapt	48,200,134 (R1) 40,181,127 (R2)
Quality filtering and trimming	DADA2	31,218,183 (R1) 31,218,183 (R2)
Dereplication, denoising and merging	DADA2	30,177,465 contigs
Removal of chimeras (5.7 %)	DADA2	28,452,189 contigs 28,066 ASVs
Clustering at 98% and removal of singletons	VSEARCH	13,942 OTUs
Post-clustering curation	LULU	11,625 OTUs
Removal of OTUs: <10 reads, <70% identity in taxonomic assignment	R	8,059 OTUs
Assessing and removal of controls and replicate samples	R	8,033 OTUs
Filtering OTUs assigned to the kingdom Fungi (88.5 %)	R	7,110 OTUs (811 samples) 22,622,815 reads
Rarefied OTU table (2,000 reads per sample)	R	6,632 OTUs (807 samples)

MOLECULAR ECOLOGY

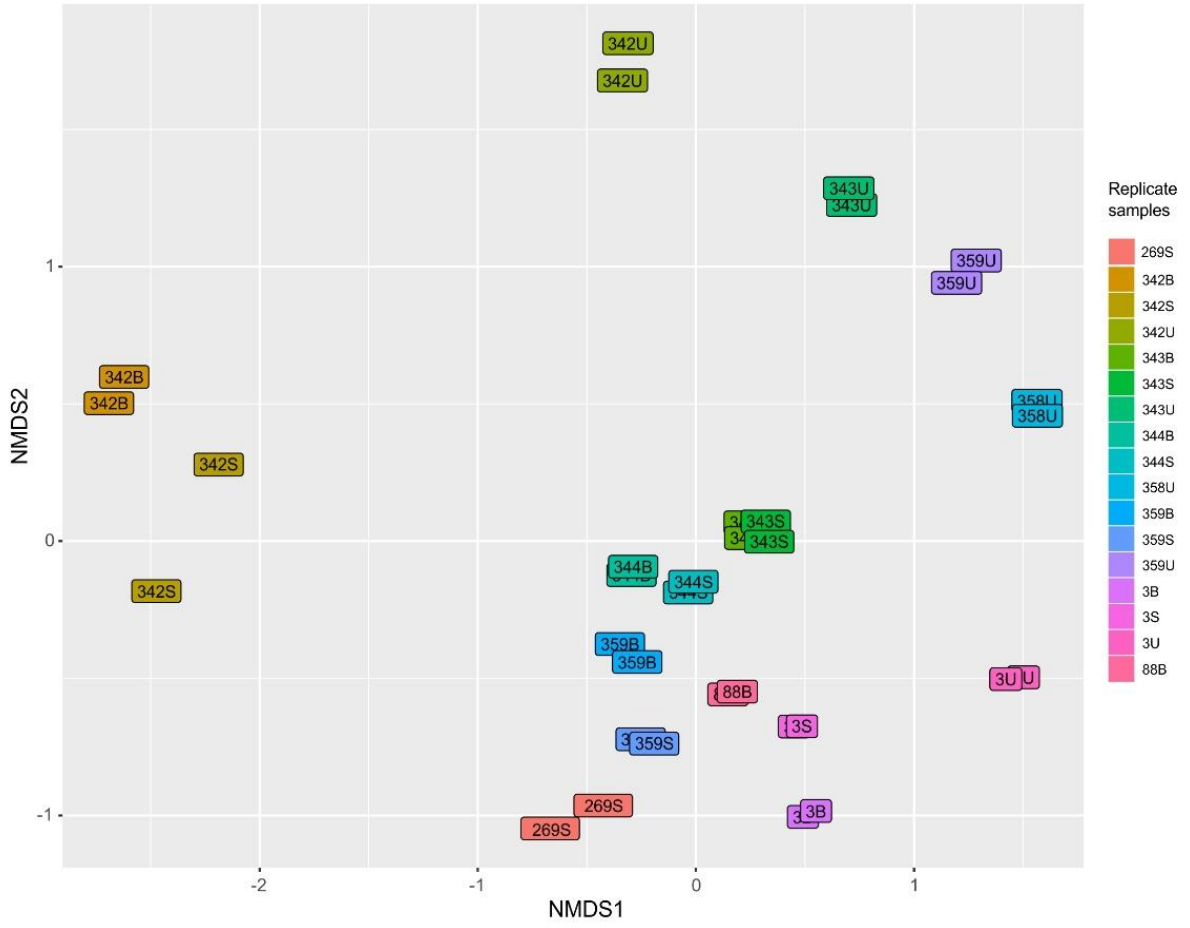


Figure S4. NMDS ordination plot for the 17 technical replicates included in this study, based on the quality filtered OTU table for all organisms before filtering fungi.

MOLECULAR ECOLOGY

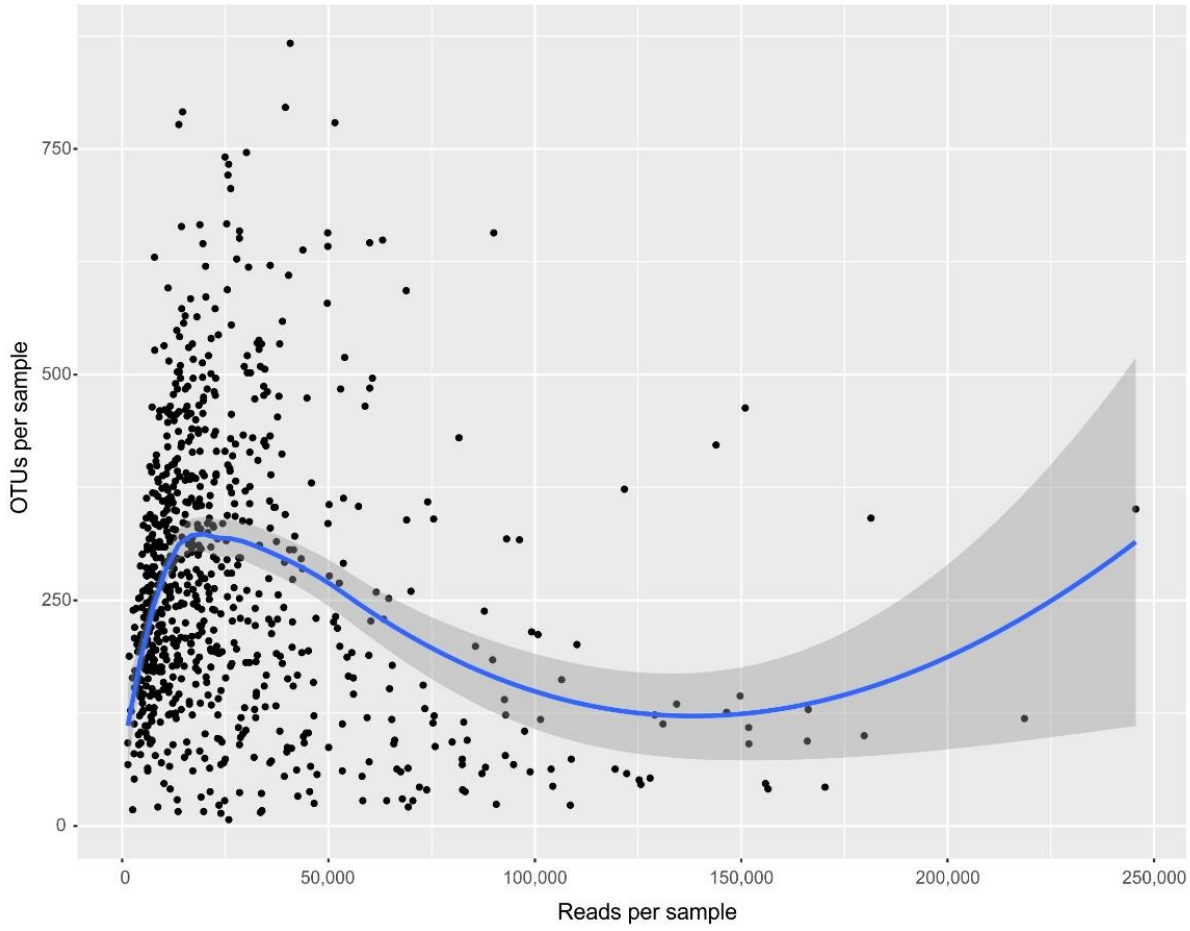


Figure S5. OTUs vs. reads plot for the 811 dust samples included in the final quality filtered fungal OTU table (7,110 OTUs and 22,622,815 reads).

MOLECULAR ECOLOGY

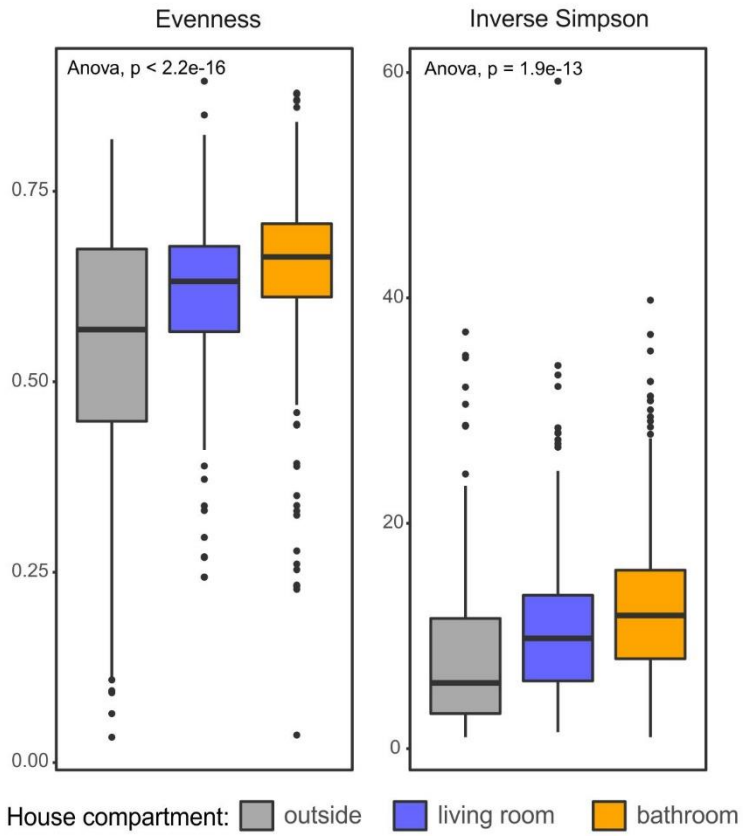


Figure S6. Diversity indices (evenness and inverse Simpson) in house dust samples (complete dataset, *n* = 807) for each house compartment: outside, living room and bathroom. All differences in pairwise comparisons were significant according to Tukey HSD test ($p < 0.05$).

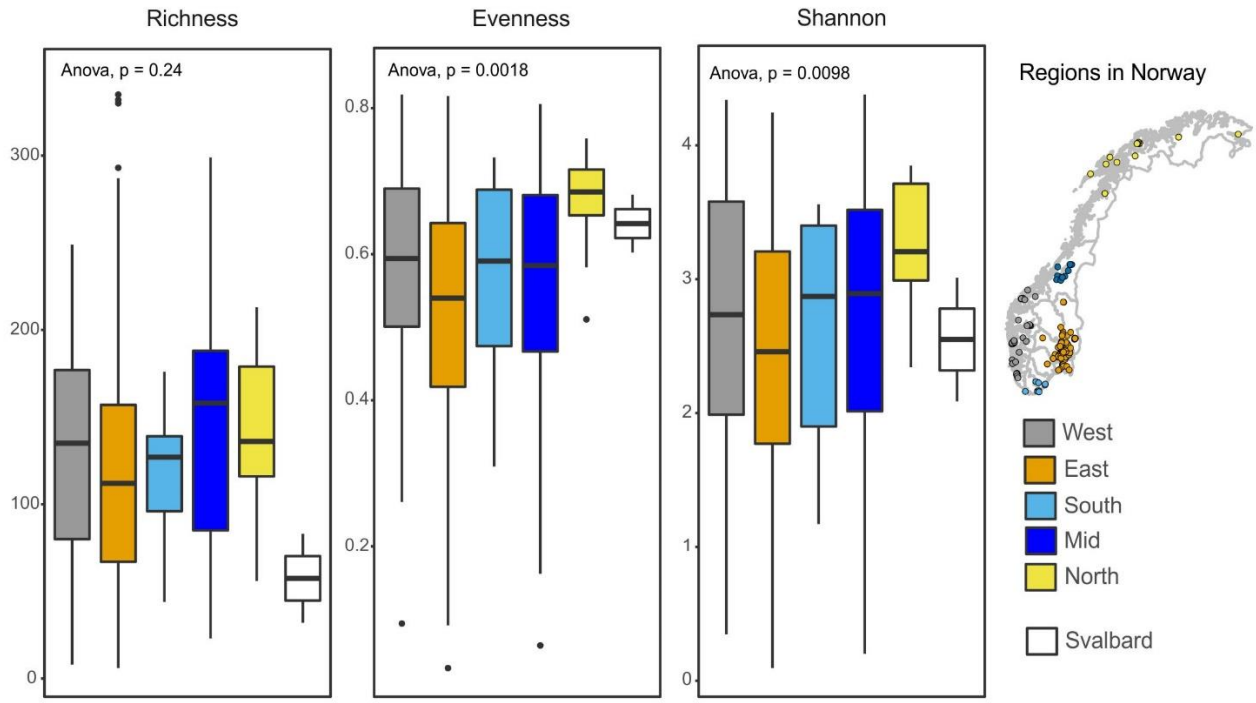


Figure S7. Diversity indices (richness, evenness and Shannon) in outdoor dust samples collected from different regions in Norway. The outdoor dataset (n = 266) includes the regions: west (45), east (164), south (9), mid (29), north (17) and Svalbard (2). In the Tukey HSD tests, the only significant differences ($p < 0.05$) in pairwise comparisons were those between east and north for both evenness and Shannon.

MOLECULAR ECOLOGY

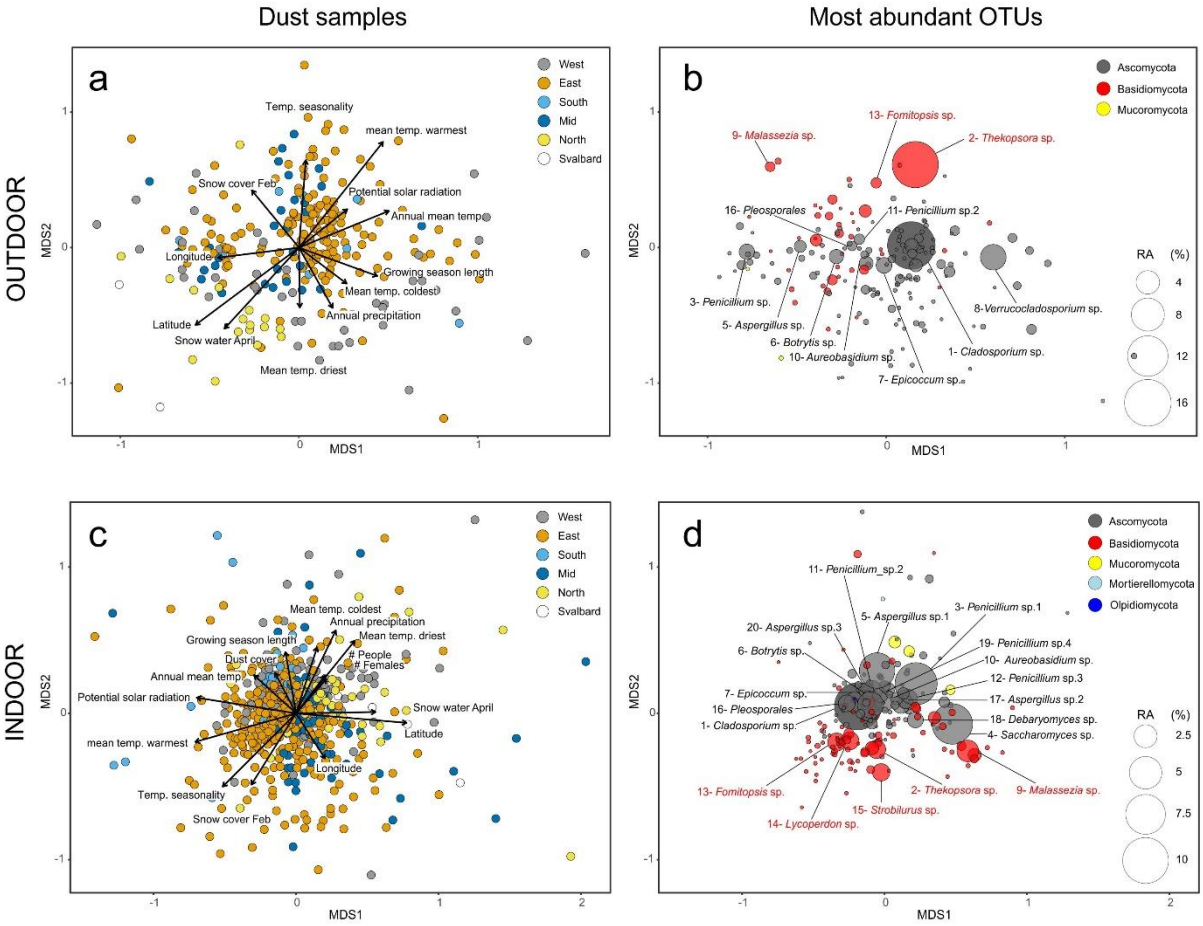


Figure S8. (**a**, **c**) NMDS ordination plots showing the clustering of the analyzed dust samples (**a**: outdoor dataset, *n* = 266; **c**: indoor dataset, *n* = 541), color indicates the five large geographical regions in Norway, samples from Svalbard (Artic islands) are empty circles. Linear regression of continuous variables with significant association ($p < 0.05$) with the NMDS ordinations, which reflect geography, climate, topography, occupants and fungal diversity, are plotted as vectors. (**b**, **d**) Top-200 most abundant OTUs plotted on the same NMDS ordinations (**b**: outdoor dataset; **d**: indoor dataset), bubble size indicates their relative abundance as proportion of the total number of rarefied reads, color indicates their phyla assignment, and labels highlight the top-20 OTUs considering the complete dataset (more details in Figure 3d).

MOLECULAR ECOLOGY

Table S3. Significance and correlation between the continuous variables fitted to the NMDS ordinations for the complete dataset (Fig. 3b), and the outdoor and indoor datasets (Fig. S7a, c). The significant correlations (p<0.05) are in bold.

	Complete dataset		Outdoor dataset		Indoor dataset	
Variables	R^2	*p*-val	R^2	*p*-val	R^2	*p*-val
Annual mean temperature (BIO1)	0.004	0.161	0.103	**<0.001**	0.029	**<0.001**
Temperature seasonality (BIO4)	0.107	**<0.001**	0.131	**<0.001**	0.097	**<0.001**
Mean temperature of driest quarter (BIO9)	0.080	**<0.001**	0.061	**<0.001**	0.078	**<0.001**
Mean temperature of warmest quarter (BIO10)	0.105	**<0.001**	0.261	**<0.001**	0.102	**<0.001**
Mean temperature of coldest quarter (BIO11)	0.030	**<0.001**	0.045	**0.005**	0.042	**<0.001**
Annual precipitation (BIO12)	0.077	**<0.001**	0.073	**<0.001**	0.075	**<0.001**
Growing season length	0.014	**0.005**	0.074	**<0.001**	0.034	**<0.001**
Snow water equivalent in April	0.036	**<0.001**	0.167	**<0.001**	0.059	**<0.001**
Snow covered area in February	0.058	**<0.001**	0.078	**<0.001**	0.063	**<0.001**
Potential incoming solar radiation	0.027	**<0.001**	0.048	**0.002**	0.094	**<0.001**
Longitude	0.003	0.342	0.067	**<0.001**	0.025	**0.003**
Latitude	0.055	**<0.001**	0.208	**<0.001**	0.114	**<0.001**
# of people	0.011	**0.012**	0.005	0.504	0.020	**0.005**
# of females	0.007	0.054	0.006	0.49	0.019	**0.009**
# of children	0.008	**0.043**	0.004	0.546	0.008	0.133
Dust coverage (*Mycotape2*)	0.008	0.051	<0.001	0.844	0.018	**0.004**

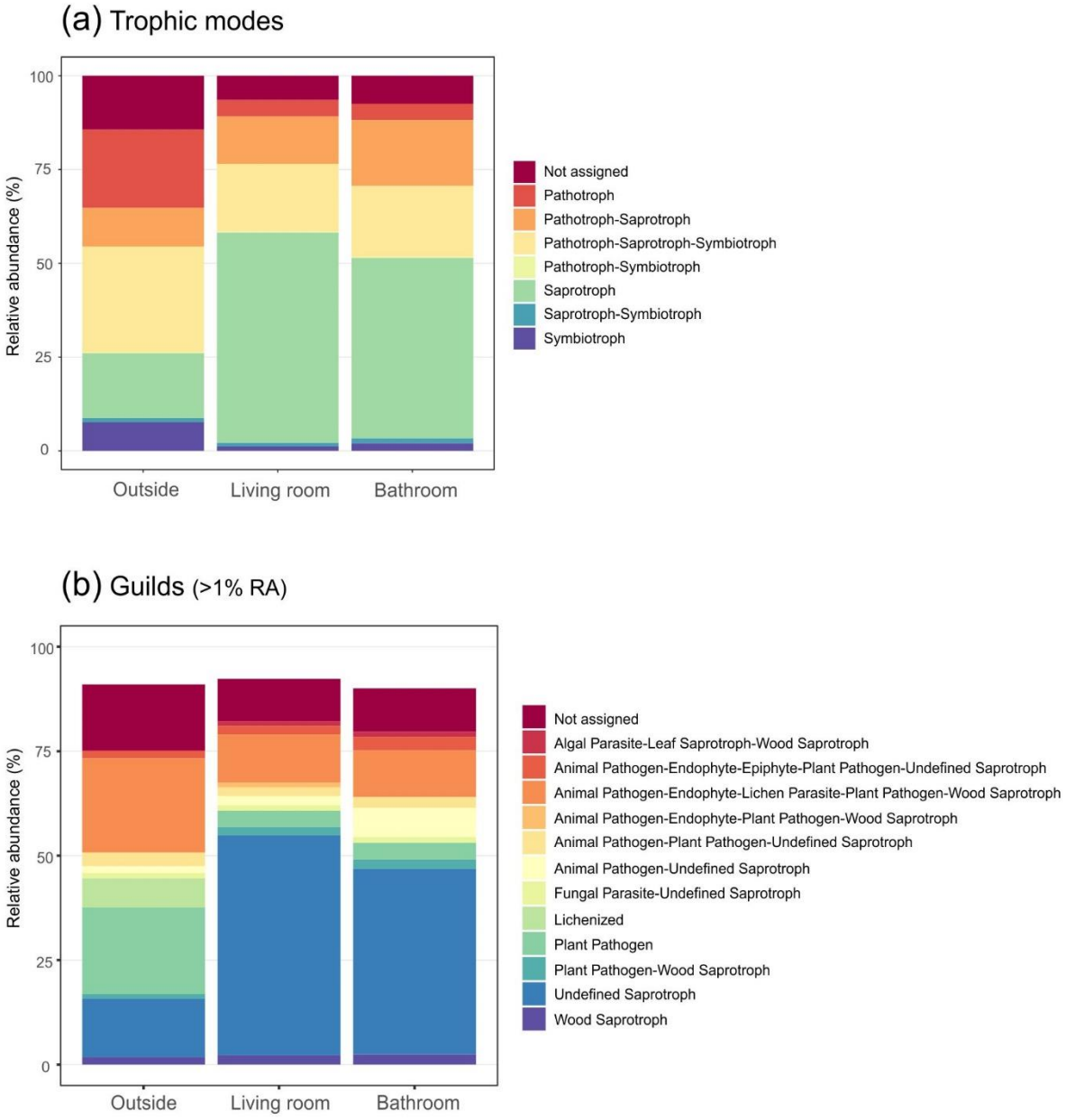


Figure S9. Relative abundances of the trophic modes (**a**) and the major guilds (**b**) assigned through the FUNGuild annotation.

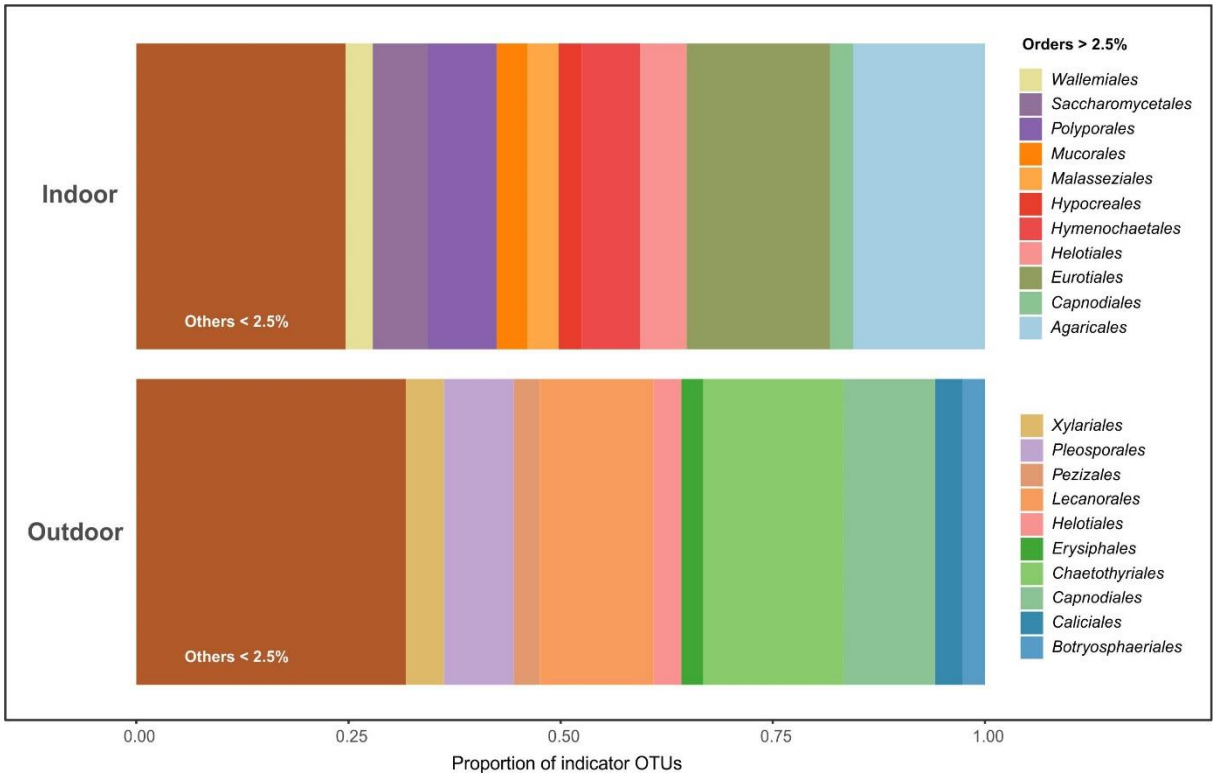


Figure S10. Taxonomic affiliation of indicator species ($p < 0.05$) for indoor (241 OTUs) and outdoor (550 OTUs) environments. Orders containing 2.5% or more of indicator OTUs are shown, while orders representing < 2.5% are clustered as 'Others'. Unidentified OTUs at the order level were excluded, which corresponded to 9.1% and 22.7% for indoor and outdoor indicators, respectively.

MOLECULAR ECOLOGY

Supplementary methods

Mock community

Positive controls contained 1 ng of an equimolar mixture of DNA from three fungal species that are not expected in the Norwegian built environment: *Mycena belliarum*, *Pycnoporellus fulgens* and *Inonotus dryadeus*. They were included to evaluate the efficiency of the DNA metabarcoding workflow, and more specifically, to assess potential tag switching errors (Carlsen et al., 2012).

Assessment of control and replicates samples

Prior to filtering the fungal OTUs, the quality of controls and replicates were assessed on the matrix that contained 8,033 OTUs, 88.5% attributed to Fungi, 11.2% to Viridiplantae (green plants mostly belonging to the phyla Streptophyta and Anthophyta), and the remaining 0.2% (19 OTUs) corresponded to other kingdoms. Previous studies have reported that gITS7/ITS4 primers can also amplify plant DNA (Ihrmark et al., 2012).

The number, identity and abundance of OTUs in the controls (extraction blanks, PCR negatives and positives) were checked and corrected considering their frequency in the study samples. All positive controls (mock community of three fungal species) included in the nine sequencing libraries showed an identical pattern composed of the same four OTUs. The three major OTUs corresponded to the mock-community members, identified as *Mycena belliarium*, *Pycnoporellus fulgens* and *Inonotus hispidus*, which represented ~99.96% of reads present in positive controls. The additional minor OTU (~ 0.04% of reads) detected in the positives corresponds to *Saccharomyces* sp. (OTU3), one of the most abundant and widely distributed OTU in the whole dataset. Remarkably, reads from mock species were exclusively detected in the positive controls, with the exception of a few reads (< 23) present in two dust samples, suggesting that the tag switching rate was insignificant in this study.

Regarding the negative controls, six extraction blanks (unused sterile swabs) and three PCR negatives contained a relatively low number of reads, representing an average of 4.1±2.6 OTUs per negative control. After checking the abundance and frequency of these OTUs in the study samples, two of them (< 10 reads in two samples) were deleted. The remaining 22 OTUs were kept because they were widely distributed in the dataset and correspond to ubiquitous fungi in the built environment.

The similarity of the community profiles for 17 technical replicates (duplicates in different PCR pools and sequencing libraries) was confirmed by NMDS (Figure S8), and the replicate with lower number of reads were discarded. Hence, confirming the reproducibility of the DNA metabarcoding workflow.

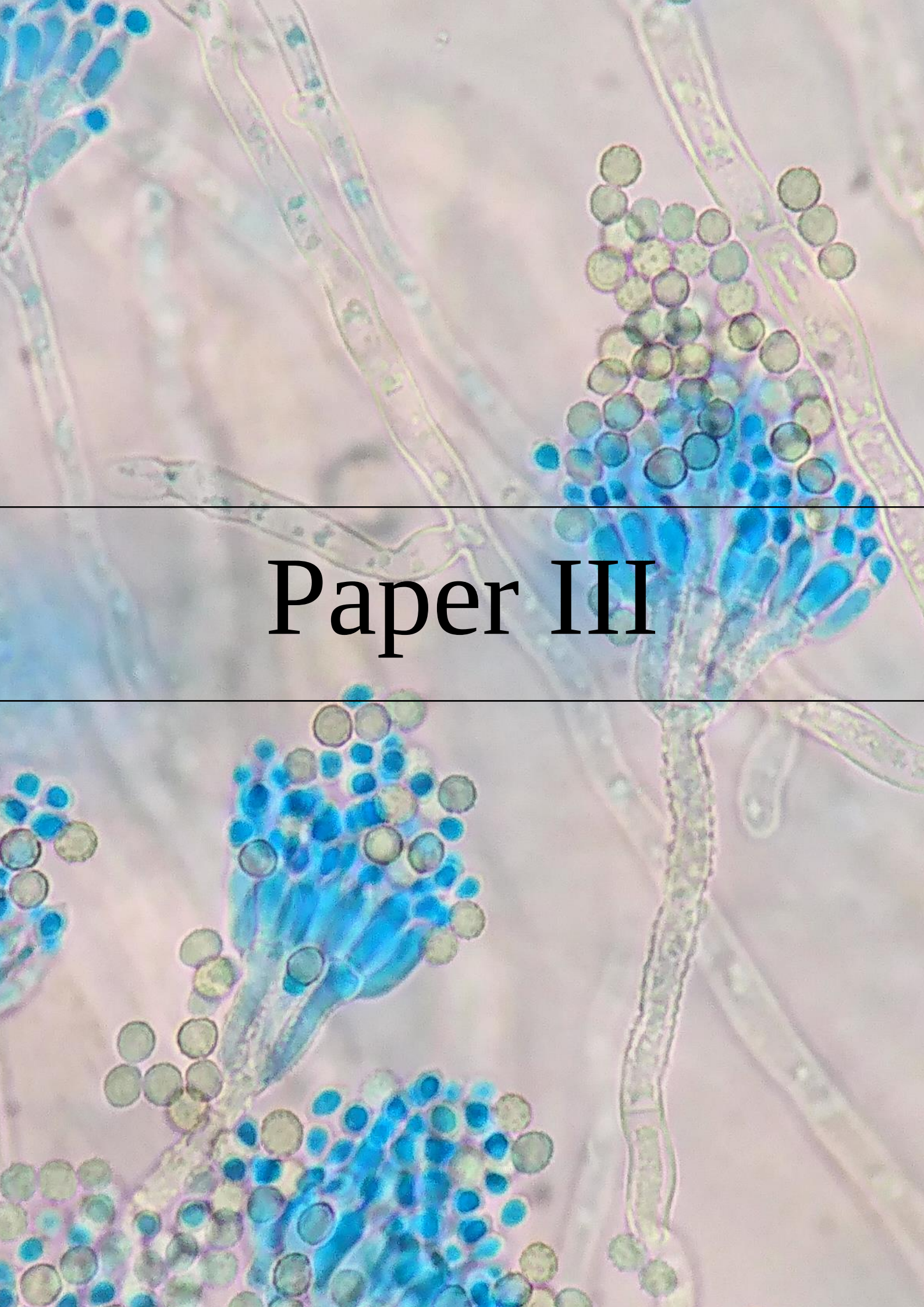
References:

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, 747–749. doi:10.1016/j.funeco.2012.06.003

Ihrmark, K., Bödeker, I. T. M., Cruz-Martinez, K., Friberg, H., Kubartova, A., Schenck, J., ... Lindahl, B. D. (2012). New primers to amplify the fungal ITS2 region - evaluation by 454-sequencing of artificial and natural communities. FEMS Microbiology Ecology, 82, 666–677. doi:10.1111/j.1574-6941.2012.01437.x

Table S4. Supplementary information about the top-200 most abundant OTUs.

Excel table "mec15916-sup-0002-tables4.xlsx" is available for download on the publication website.



Paper III

The indoor mycobiome of daycare centers is affected by occupancy and climate

Eva Lena F. Estensmo[1]*, Synnøve Smebye Botnen[2], Sundy Maurice[1], Pedro M. Martin-Sanchez[1], Luis Morgado[3], Ingeborg Bjorvand Engh[4], Klaus Høiland[1], Inger Skrede[1], Håvard Kauserud[1]*

1. Section for Genetics and Evolutionary Biology (Evogene), Department of Biosciences, University of Oslo, P.O. Box 1066 Blindern, 0316 Oslo, Norway.
2. Oslo Metropolitan University, P.O. Box 4, St. Olavs plass, 0130 Oslo, Norway.
3. Naturalis Biodiversity Center, Darwinweg 2, 2333 CR Leiden, Nederland.
4. Mycoteam AS, P.O. Box 5 Blindern, 0313 Oslo, Norway.

Abstract

Many children spend considerable time in daycare centers and may here be influenced by indoor microorganisms, including fungi. In this study, we investigate the indoor mycobiome of 125 daycare centers covering strong gradients of environmental conditions throughout Norway. Dust samples were collected on doorframes outside and inside buildings using a citizen science sampling approach. The fungal communities in the dust samples were analyzed using DNA metabarcoding of the ITS2 region. We observed a marked difference between the outdoor and indoor mycobiomes. The indoor mycobiome included considerably more yeasts and molds compared to the outdoor samples, with *Saccharomyces*, *Mucor*, *Malassezia* and *Penicillium* among the most dominant indoor fungal genera. The indoor fungal richness and composition could be ascribed to numerous outdoor and indoor variables. There was a clear geographic signal in the indoor mycobiome composition that mirrored the outdoor climate, ranging from humid areas in western Norway to drier and colder areas in eastern Norway. Moreover, the number of children in the daycare centers and various building features also influenced the indoor mycobiome composition. We conclude that the indoor mycobiome of Norwegian daycares are dominated by yeasts and molds, and that a multitude of factors structure the mycobiome. This study exemplifies how citizen science sampling enables DNA-based analyses of a high number of samples covering wide geographic areas in a relatively short timeframe.

Importance

With an alarming increase in prevalence of chronic diseases like childhood asthma and allergies, there is an increased focus on the exposure of young children to indoor biological and chemical air pollutants. Our study of 125 daycares throughout Norway demonstrates that the indoor mycobiome not just reflect co-occurring outdoor fungi, but includes a high abundance of yeast and mold fungi with a clear affinity to indoor environment. A multitude of factors influence the indoor mycobiome in daycares, including building type, inhabitants, as well as the outdoor environment. Many of the detected yeasts and molds are associated with the human body, and some have been coupled to allergies and respiratory problems. A better characterization and understanding of the indoor mycobiome, and the factors influencing the community composition, is important for children's health. Our study calls for further studies investigating the potential impact of daycare mycobiome on children health.

Introduction

Over a few thousand years, humans have moved from a largely outdoor life-style to now spending a large part of their lifetime in indoor environments. Although the diversity of other co-occurring organisms is considerably lower in indoor environments, humans are not alone. If moisture and organic materials are available indoor, extremotolerant fungi will grow, sporulate and disperse spores. Some of the most prevalent fungi that are able to grow and sporulate in houses include various ascomycete molds, such as *Cladosporium*, *Penicillium* and *Aspergillus* (1, 2). Fungal growth in indoor environments can lead to poor indoor air quality, and some of these fungi are associated with allergic reactions (3-5) and respiratory disease symptoms (6, 7), which may have long-term impacts on human health. Furthermore, unfortunate combinations of indoor fungi and bacteria in moisture damaged buildings may also cause negative health effects, also in low concentrations (8).

In many countries, children spend considerable time in daycare centers, where they are exposed to co-occurring microorganisms, including fungi. Since small children often vector organic material such as soil and litter from nature, daycare centers may accumulate extra organic substrates promoting fungal growth, as compared to other indoor environments. In line with this, it has previously been shown that the concentration of fungi in daycares is higher compared to private homes (9). In several studies, the outdoor environment has been found to be the main source of indoor fungi (10-13), due to the influx of

fungal spores through windows, entrances and the ventilation system. Hence, the outdoor vegetation and climate that structure the outdoor fungi will therefore indirectly also structure the indoor mycobiome (11). In correspondence with this, in a recent DNA-based metabarcoding study performed in 271 houses across Norway, we showed that outdoor climate was one of the main drivers of the indoor dust mycobiome (12). A similar observation was done by Barberán et al. (2015) in North America (11).

Although several studies have concluded that the indoor mycobiome largely reflects outdoor fungi (13, 14), the inhabitants and their diverse activities, the presence of pets and plants, as well as various building features, may also contribute and structure the indoor mycobiome (15, 16). Many yeast fungi, such as *Malassezia* and *Candida,* are associated with the human body and may therefore be prevalent indoors (17-20). Which fungi that are associated with the human body may to some extent be age-dependent. For instance, the basidiomycete yeast *Malassezia* seems particularly prevalent on adults (21), while children tend to have a more diverse skin-associated mycobiome, including genera like *Aspergillus*, *Epicoccum*, *Cladosporium*, *Cryptococcus* and *Phoma*, in addition to *Malassezia* (19).

The indoor mycobiome can be analyzed in different ways, including isolation and cultivation of fungi, microscopy of spores and hyphal remains and by different molecular analyses. DNA metabarcoding, based on high throughput sequencing of PCR amplified markers, is established as an effective approach to survey fungal communities (22). In buildings, DNA-metabarcoding of collected dust samples, integrating spores and hyphal remains that have accumulated over time, has proven to be an effective mean for exploring the indoor mycobiome (10, 12, 13, 23, 24). However, it might be difficult to get access and obtain samples from many buildings. By providing detailed instructions, dust samples can alternatively be collected by the inhabitants themselves, from where DNA can be extracted and analyzed further (12, 25). This type of community-based research, where networks of non-professionals help to collect data as part of a research project, is regarded as citizen science (26-28). Sampling through citizen science is a powerful approach, where sample equipment can be sent out by post, returning hundreds or even thousands of samples covering large geographic areas.

Given the long-term impact that some indoor fungi potentially can have on human health, there is a need to better characterize the indoor mycobiome to

which we are exposed, from an early age. In this study, we aim to analyze the indoor mycobiome associated with daycare centers. We first ask (1) which outdoor and indoor factors drive the daycare mycobiome? Second, we ask (2) which fungal groups dominate in the daycare centers, as compared to outdoor samples? To address these research questions, we obtained 572 dust samples from doorframes inside (bathroom and main room) and outside (main entrance) 125 daycare centers throughout Norway (Fig. 1).

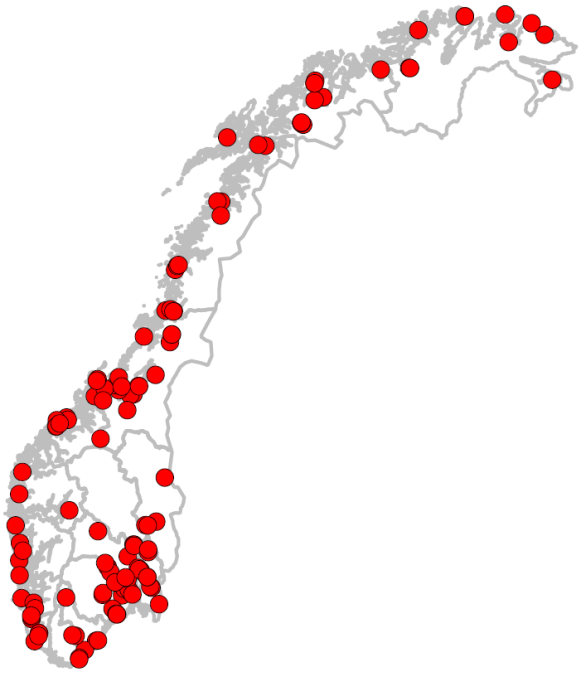


Figure 1. Map of Norway showing the geographical location of the 125 daycares.

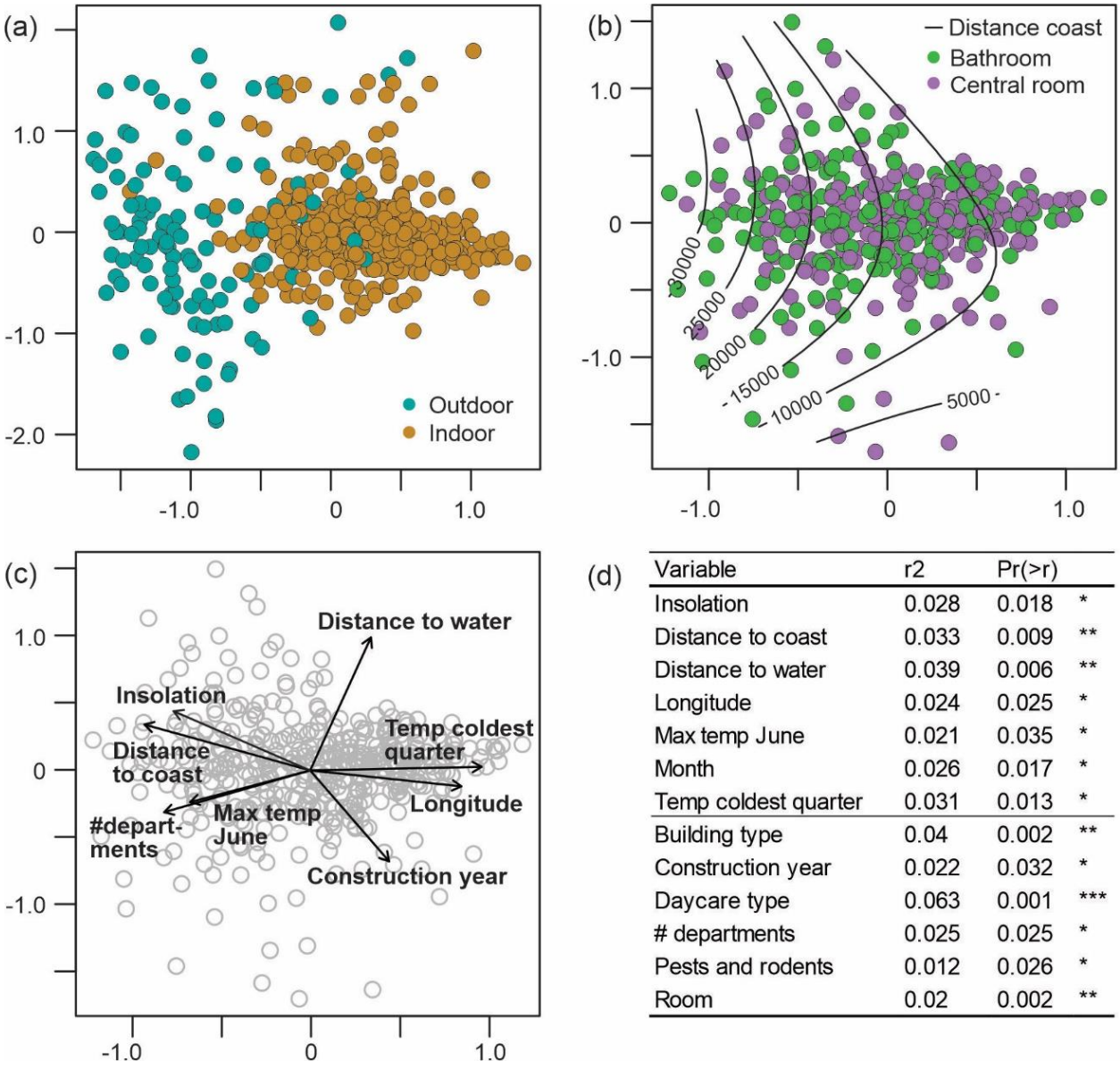
The dust samples were obtained using a citizen science approach, where daycare personnel performed sampling according to our instructions. Norway spans extensive gradients in climate and other environmental drivers, enabling us to evaluate the influence of outdoor climate on the indoor mycobiome, in addition to building features and inhabitant characteristics. The obtained dust samples were analyzed by DNA metabarcoding of the rDNA ITS2 region.

Results

Factors influencing the indoor mycobiome

Our final dataset from the 125 daycare centers included 748,836 sequences, with 1342 sequences in each of the 558 samples of indoor and outdoor environments. A total of 5946 fungal OTUs appeared in the dataset. In a multivariate (NMDS) analysis, we observed a relatively clear separation between the outdoor and indoor dust mycobiomes (Fig. 2a). However, the two types of indoor samples, main room versus bathroom, overlapped fully in fungal community composition (Fig. 2b).

Through a questionnaire to the citizen scientists (daycare personnel), we obtained information about different building and occupancy variables (Table 1). In addition, information about the local climate and vegetation were extracted based on the geographic coordinates of the daycares (29). Considered individually, numerous of these variables correlated significantly with the compositional



Variable	r2	Pr(>r)	
Insolation	0.028	0.018	*
Distance to coast	0.033	0.009	**
Distance to water	0.039	0.006	**
Longitude	0.024	0.025	*
Max temp June	0.021	0.035	*
Month	0.026	0.017	*
Temp coldest quarter	0.031	0.013	*
Building type	0.04	0.002	**
Construction year	0.022	0.032	*
Daycare type	0.063	0.001	***
# departments	0.025	0.025	*
Pests and rodents	0.012	0.026	*
Room	0.02	0.002	**

Figure 2. Fungal community composition in daycare centers. (a-c) Ordination plots displaying compositional variation in the dust mycobiome, where each point indicates one dust sample. (a) NMDS plot displaying both outdoor (cyan) and indoor (brown) samples. (b) NMDS plot of only indoor samples, displaying samples from bathrooms (green) and central room (purple). The isolines represent the distance to coast. (c) The indoor samples with vectors representing numeric variables showing significant associations to the compositional variation in the indoor mycobiome ($p<0.05$). Categorical variables are not shown. (d) Goodness-of-fit statistics (r2) for variables that significantly ($p<0.05$) account for variation in the composition of the indoor mycobiome. Variables related to regional climate are listed above the horizontal line, while variables related to the specific daycares are listed below.

variation in the indoor mycobiome (Fig. 2c), including variables related to the daycare centers such as daycare type, construction year, number of departments, pests and building type. Climatic variables such as temperature and total insolation were also significantly correlated to the indoor mycobiome composition, as well as spatial variables that likely mirror additional regional environmental variability (Fig. 2c, d). Many of the inferred variables were associated with the major climate gradient stretching from humid, oceanic areas in western Norway, to inland, continental areas in eastern Norway (Fig. 2c, d).

Table 1. Climatic and building metadata. The upper part of the table include the six first climatic variables extracted from a database (29) using georeferences of the daycare centers. The variables provided by volunteers in each daycare are listed in the lower part of the table.

Variables	Categories
Area	Categorical: Urban/Rural
Avg temperature of the coldest quarter	Numeric
Max June temperature	Numeric
Max May temperature	Numeric
Proximity to all water bodies	Numeric
Proximity to coast	Numeric
Longitude	Numeric
Sampling month	Categorical: March-May
Age of children in the sampled department	Numeric
Building material	Categorical: Wood/Brick/Concrete
Building type	Categorical: Detached house/Semi-detached house/Block/Collection of buildings
Building year	Numeric
Construction year	Numeric
Moisture problems	Categorical: Yes/No
Number of children	Numeric
Number of departments	Numeric
Pest / rodents	Categorical: No/Mouse/Rat/Grey silverfish/Other
Ventilation type	Categorical: Natural/Mechanical/Balanced
Water damage	Categorical: Yes/No

Evaluating the relative contribution of variables together in a CCA analysis (Table 2), revealed that longitude, presence of pest/rodents, construction year of the daycare center and number of children were the main drivers of the fungal community composition, with very low interaction effects (<0.01%). These factors accounted altogether for only 7% of the variation in mycobiome composition (Table 2). Longitude in the CCA analyses mirrors the regional climate gradient mentioned above.

The indoor fungal richness, calculated on a sample-basis, was significantly higher in the bathroom compared to the central room, and there was a significantly positive relationship between indoor fungal richness and maximum temperature during May at the sampling location, as well as proximity to coast (see the Mixed Effect Model presented in Table 3).

Table 2. Variables with explanatory power in the CCA. Note that these variables may reflect and represent other variables.

Variables	Variation explained
Longitude	0.0159
Pests / rodents	0.0187
Construction year	0.0181
Number of children	0.0156
Interaction effects	0.0001
Unexplained variation	0.9316

Table 3. Richness analyses using a mixed effect model with number of OTUs per sample as response and with daycare as a random effect. Room type Bathroom is in the baseline of the model, the estimate for Room represents the difference from bathroom to central room.

Variable	Estimate	Std error	t-value	p-value
Room (bathroom = baseline)	-3.0773	1.310563	-2.348083	0.0195
Proximity to coast	0.000095	0.000043	2.193266	0.0291
Max May temperature	1.671905	0.55933	2.98912	0.003

Taxonomic composition of daycare mycobiome

The indoor mycobiome were dominated by *Saccharomycetales* and *Mucorales*, in contrast to the outdoor mycobiome that were mainly dominated by *Pucciniales*, *Capnodiales*, *Agaricales* and *Chaetothyriales* (Fig. 3a). The true yeasts of *Saccharomycetales* where considerably more abundant in the indoor environments. *Malasseziales* and *Tremellales*, both groups likely representing basidiomycete yeasts, where also somewhat more abundant in the indoor samples (Fig. 3a). We annotated the 1253 most abundant OTUs (OTUs with >20 sequences) into different growth and life forms, which revealed that yeasts, dimorphic yeasts and molds were considerably more abundant in indoor environments, while litter and wood saprotrophs, plant pathogens and lichens dominated relatively more in the outdoor samples (Fig. 3b).

Among the top 30 genera detected in this study, measured in sequence abundance in a balanced indoor/outdoor dataset (where the two indoor samples were averaged), many had a clear affinity towards either indoor or outdoor environments (Fig. 4).

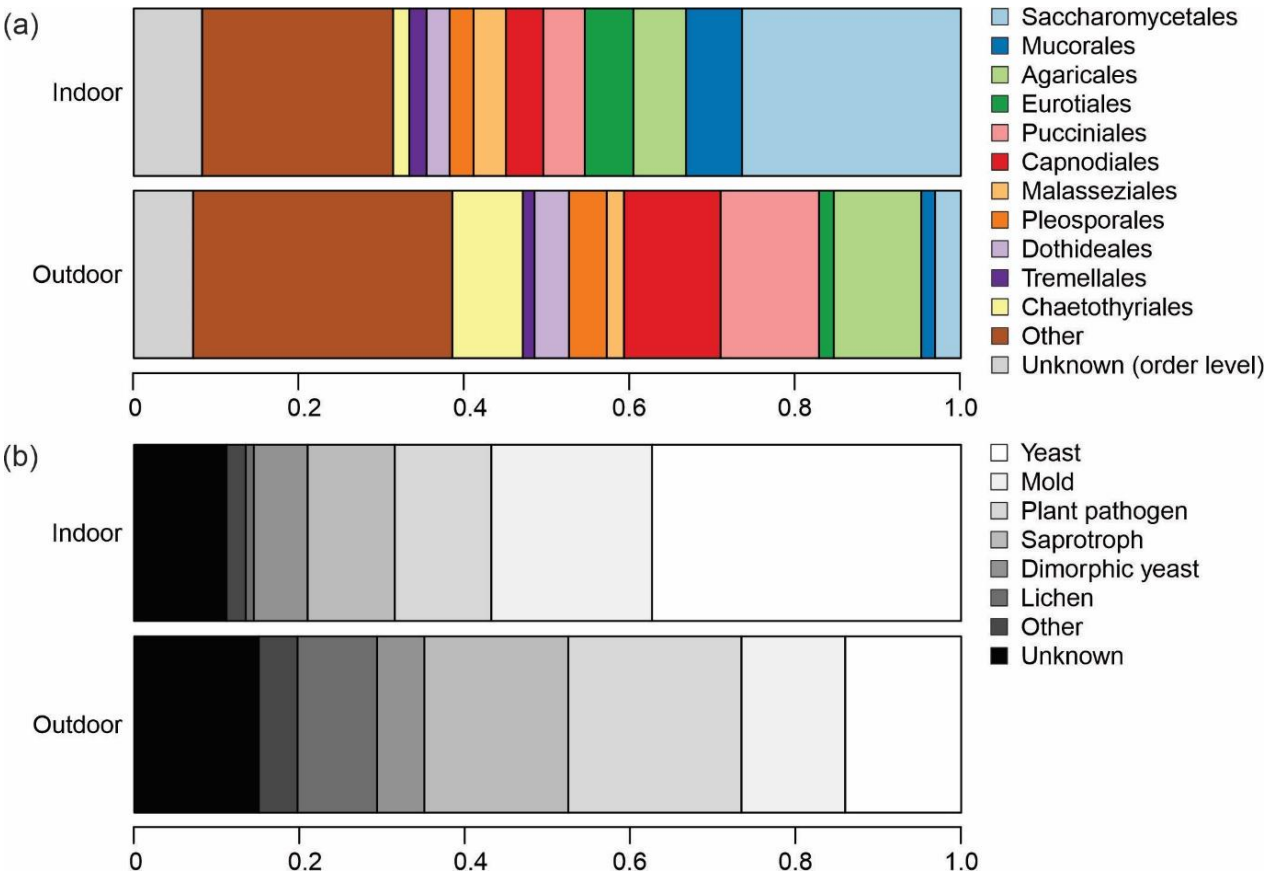


Figure 3. Taxonomic distribution in outdoor and indoor dust samples from the daycare centers reflecting sequence numbers. (a) Relative abundance of the main fungal orders in outdoor and indoor samples. (b) Relative abundances of fungi with different growth forms / nutritional modes in the indoor and outdoor samples. The category saprotrophs represent litter and wood decay fungi.

Ten genera, namely, *Aspergillus, Candida, Debaryomyces, Filobasidium, Malassezia, Morierella, Mucor, Penicillium, Rhodotorula, Saccharomyces* and *Wallemia,* have a distinct affinity towards indoor environments. *Saccharomyces* was by far the most abundant genera in the indoor environment, with about 12.5 time's higher abundance in indoor compared to outdoor samples. In contrast, plant pathogens like *Melampsora, Puccinastrum* and *Melampsoridium* were relatively more common in the outdoor samples, but also present indoor, likely due to airborne spore dispersal from outdoor sources. Interestingly, some genera with affinity to the outdoor environment, like *Verrucocladosporium, Scoliciosporum* and *Sordaria* were almost exclusively present in the outdoor samples, while others like

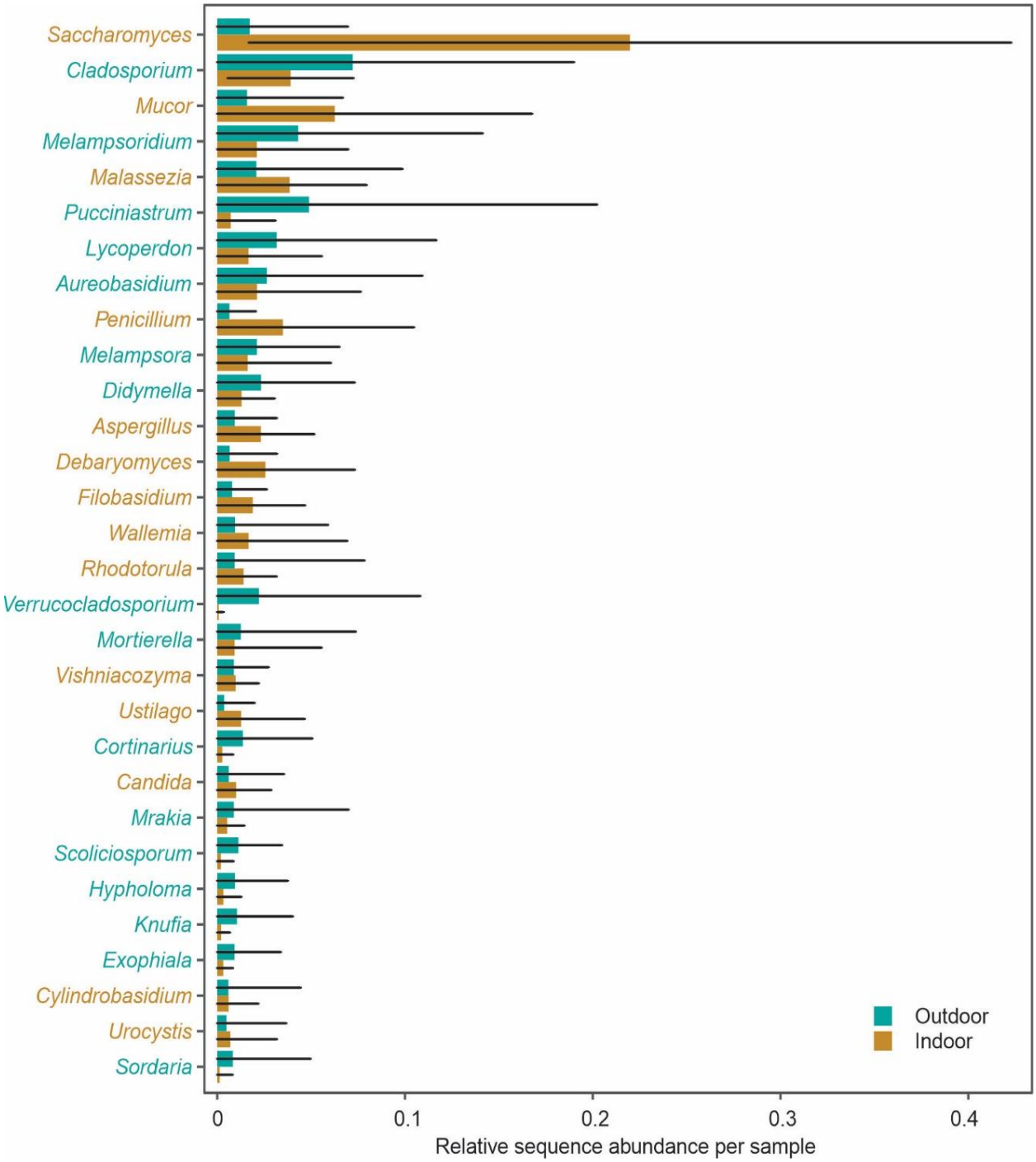


Figure 4. The 30 most abundant genera in the dataset, displaying their average sequence abundance across indoor and outdoor samples in the 125 daycare centers. For the indoor samples, an average value from the bathroom and central room was used for the calculations, providing a balanced indoor-outdoor dataset. Genera with higher indoor abundance are displayed in brown color, while genera with higher outdoor abundance are shown in cyan. The black lines indicate standard error.

Cladosporium, *Melampsoridium* and *Lycoperdon*, were also abundant in the indoor environment.

Discussion

Factors influencing the indoor mycobiome

We observed a clear separation between the outdoor and indoor mycobiome across the 125 Norwegian daycare centers, and that numerous variables both associated with the outdoor climate and the indoor environment together influenced the indoor mycobiome. We observed a similar pattern in a study of private houses across the same climatic gradients in Norway (12). Likewise, Barberán et al. (2015) reported a similar trend from North America, where they analyzed indoor and outdoor dust mycobiomes throughout the USA (25). However, other preceding studies have concluded that indoor air and dust merely consist of outdoor fungi that have spread into buildings through the ventilation system, windows or doors (13, 14, 30). Shin et al. (2015) concluded that human activity had little influence on the indoor fungal community composition in daycare centers in Seoul, South Korea (30). Similarly, in a study investigating indoor fungi in a housing facility in California, Adams et al. (2013) concluded that the outdoor air and not the residents structured the indoor mycobiome (13). Interestingly, in our recent study of seasonality of the indoor mycobiome, we observed that the indoor mycobiome is more influenced by the outdoor fungal diversity during summer and fall (9). Thus, as the citizen scientists in the present study did the sampling during early spring, we may have detected a stronger influence of indoor variables in the current study compared to e.g. Shin et al. (2015), where samples were collected from August to October in a comparable climate in South Korea.

According to our analyses, the number of children in daycares accounted for some of the overall variation in the indoor mycobiome composition, together with construction year and the occurrence of pests/rodents (CCA analysis). In addition, the variables building type, number of departments, room (main room versus bathroom), and type of daycare correlated significantly with the mycobiome composition in single factor analyses. Taken together, this indicates that how daycares are organized and in which buildings daycares are localized, influence their mycobiome composition. A more balanced, factorial study design (focusing on fewer variables) is necessary to gain a better insight into the influence of the different variables. In addition to these local scale variables, regional climate related factors such as maximum temperature in June, mean temperature of coldest quarter and total insolation also

correlated significantly with the indoor mycobiome composition. These findings mirrors the observations by Barberán et al. (2015) and Martin-Sanchez et al. (2021), where regional climate also were found to be important for the indoor mycobiome. Longitude, an approximation for regional climate variability, also had explanatory power. Throughout most of Norway, longitude mirrors a climate gradient from oceanic and humid areas in the west, to areas with dryer, colder and high temperature seasonality conditions in the east. The climate factors most likely have indirect effects on the indoor fungi, as they probably influence and structure the outdoor fungi that spread into buildings.

Despite several of the assessed variables were significantly related to the composition of the indoor mycobiome, only a small fraction of the variation in indoor mycobiome composition was accounted for (7%). However, the low level of explanatory power is not a unique feature distinct to this study, but rather a common trend across most studied fungal communities (12). Fungal communities are largely assembled through colonization by spore dispersal, which to a large extent is a random process. Because of this, it is generally difficult to account for the fungal community composition.

Taxonomic composition of daycare mycobiome

The most marked taxonomic difference in the indoor and outdoor dust mycobiome was the predominance of yeasts and molds inside the daycare centers. *Saccharomyces* was by far the most abundant genus in our study, and had a clear affinity to indoor environments. *Saccharomyces* may partly be derived from food, but has also been found as one of the most abundant genera in the human gut (31) and on children's skin (32). Other true yeasts, like *Debaryomyces* and *Candida*, had also a clear affinity to indoor environments in the studied daycares. *Candida* is one of the most widespread fungi associated with external (skin) and internal (mouth, digestive tract) parts of the human body (33). It is well documented that *Candida* is particularly associated with children, commonly resulting in oral thrush (mouth and throat) in the first years of life (34). The lipophilic basidiomycete yeast *Malassezia*, a widespread genus on human skin (19), and *Rhodotorula*, another basidiomycete yeast associated with the human body (33), were also prevalent in the daycares. *Malassezia*, as well as *Candida*, are known to be associated with inflammatory skin disorders such as seborrheic dermatitis and atopic dermatitis in childhood as well as in adulthood (35, 36). However, *Malassezia* most often has a commensal role, as they

are widespread on healthy skin. For instance, 11 of the 14 known *Malassezia* species were associated with different parts of the skin of 14 healthy adults (21), indicating that human skin is colonized with a wide range of *Malassezia*. On children´s skin, a dominance of the species *Malassezia globosa* has been observed (19). We hypothesize that the yeasts dominating the indoor daycare mycobiome are mainly derived from different parts of the human body. The high density of children and close physical contact may lead to easy and fast transmission of yeasts in daycares, possibly explaining the up-concentration of these species indoor.

In addition to an up-concentration of yeasts in the indoor mycobiome, several extremotolerant molds, such as *Mucor, Penicillium, Aspergillus* and *Wallemia* also showed a clear preference for the indoor environment. These genera are widespread and abundant members of the indoor mycobiome and detected in most indoor studies (12, 25, 37). In addition to rapid growth on organic materials indoor, some of these taxa are often detected on and in the human body as well (19). *Cladosporium*, another abundant mold in indoor environments, was prevalent both indoor and outdoor and might largely be dispersed from outdoor sources. Though no direct cause-effect relationship has been established, some of these mold taxa were abundant in houses with children with allergies and respiratory diseases (16, 38). Generally, higher fungal richness in the indoor environment during early life are associated with respiratory diseases (39). Thus, the fungal diversity in daycares centers can potentially have negative effect on children health if present in high abundance (40).

Concluding remarks

For the current study, dust samples were obtained during a relative short time window during the spring 2018. From other studies, we know there is an extensive temporal variability (10), which is not accounted for here. Moreover, sampling at approximately the same time throughout Norway, a country that spans a wide range of latitudes and longitudes, means that the outdoor climate, vegetation and fungal communities are in different (phenological) growth phases, also influencing which fungi we recovered. Indeed, the variable (sampling) month was significantly correlated to the fungal community composition, but it only accounted for a small amount of the variation. Most likely, indoor fungi dominated by yeasts and molds, can be sampled in higher proportions during winter and spring in the Norwegian climate, while outdoor fungi will influence more strongly during the growth and sporulation period of most

mushrooms (summer and fall). Hence, a sampling time during the winter period may be even more representative of the specific indoor fungal community in future studies. However, we conclude that the indoor mycobiome of Norwegian daycares are dominated by yeasts and molds, and that a multitude of factors structure the mycobiome.

In this study, we carried out a citizen science sampling approach for obtaining our study material. Only a few outlier samples occurred, and the indoor and outdoor dust samples were largely separated, indicating a low influence of sampling bias. Moreover, very few samples were discarded due to low DNA yields. Altogether, this study suggests that citizen science sampling can be a powerful approach to obtain samples from a widespread geographic area during a short time span. We advocate for further citizen science studies for evaluating biological and chemical air pollutants, which will also help to raise public awareness on air quality problems in buildings.

Materials and methods

Sampling

A list of Norwegian daycares was retrieved from the Norwegian ministry of health (Helsedirektoratet). The list was sorted alphabetically after counties and municipalities, and the first five municipalities in each county were selected for the study. The first 3-4 daycares in the list in each of these municipalities were chosen as candidate sites for dust sampling. Sampling kits containing five floq swabs (Copan Italia spa, Brescia, Italy) and a questionnaire were sent to the selected daycares asking them to perform dust sampling on doorframes on specific locations: (1) outdoor, (2) central room and (3) bathroom. If the daycare had two different departments, we asked them to repeat the sampling in (4) the central room and (5) the bathroom of the second department as well. Overall, 572 samples were retrieved from a total of 125 studied daycare centers (Fig. 1), and the swabs were stored at -80 ºC until DNA extraction.

DNA extraction and metabarcoding

Samples were prepared and DNA was extracted using the E.Z.N.A Soil DNA kit (Omega Biotek, Norcross, GA, USA). The tips of the swabs were placed in disruptor tubes by using a sterilized scissor. The empty swab tubes were filled with 800 µL SLX-Mlus Buffer to collect remaining dust before being transferred to the disruptor tubes. The samples were homogenized for 2 x 1 min at 30 Hz using TissueLyser (Qiagen, Hilden, Germany) and stored at -20 °C until further processing.

DNA extraction and metabarcoding library preparation were performed

according to Estensmo et al., in review (9). Briefly, samples were thawed at 70 °C, followed by an incubation of 10 minutes at the same temperature and homogenized for 2 x 1 min at 30 Hz using a TissueLyser (Qiagen, Hilden, Germany). The samples were then cooled on ice before adding 600 µL chloroform, vortexed and centrifuged at 13 000 rpm for 5 min at RT. The aqueous phase was transferred to a new 1.5 mL tube and an equal volume of XP1 Buffer was added before vortexing. The extract was transferred to the HiBind DNA Mini Column and further processed following the manufacturer's guidelines. The DNA was eluted in 50 µL Elution Buffer.

We targeted the ITS2 region with the forward primer ITS4: 5′-xCTCCGCTTATTGATATG (41) and the modified reverse primer gITS7: 5′-xGTGARTCATCGARTCTTTG (42), barcodes x ranging from 6-9 base pairs. The amplification mix contained 2 µl DNA template, 14.6 µl Milli-Q water, 2.5 µl 10x Gold buffer, 0.2 µl dNTP's (25 nM), 1.5 µl reverse and forward primers (10 µM), 2.5 µl MgCl2 (50 mM), 1.0 µl BSA (20 mg/ml) and 0.2 µl AmpliTaq Gold polymerase (5 U/µl). DNA was amplified by initial denaturation at 95 °C for 5 min, followed by 32 cycles of denaturation at 95 °C for 30 s, annealing at 55 °C for 30 s, and elongation at 72 °C for 1 min. An elongation step was included at 72 °C for 10 min. Amplicons were normalized using the SequalPrep Normalization Plate Kit (Invitrogen, Thermo Fisher Scientific, Waltham, MA, USA) and eluted in 20 µL Elution Buffer. The resulting PCR products were processed into seven libraries of 96 samples using a combination of 96 tagged primers. Technical replicates and controls were included in each library. The technical replicates included DNA from 12 dust samples, one mock community (artificial fungal community composed of DNA in 1 ng/µL equimolar concentration from *Mycena belliarum*, *Pycnoporellus fulgens*, *Serpula similis* and *Pseudoinonotus dryadeus*), negative DNA controls (using a clean swab as starting material) and negative PCR controls. The 96 PCR products within each library were pooled, concentrated and purified using Agencourt AMPure XP magnetic beads (Beckman Coulter, CA, USA). The quality of the purified pools was measured using Qubit (Invitrogen, Thermo Fisher Scientific, Waltham, MA, USA). The seven libraries were barcoded with Illumina adapters, spiked with PhiX and sequenced in three Illumina MiSeq (Illumina, San Diego, CA, USA) lanes with 2 x 250 bp paired-end reads at Fasteris SA (Plan-les-Ouates, Switzerland).

Bioinformatics

The bioinformatics analyses were performed according to Estensmo et al., in review (9). Basically, raw sequences were demultiplexed independently using CUTADPT (43) allowing no miss-matches between barcode tags and sequence primer, and sequences shorter than 100 bp where discarded. DADA2 (44) was used to filter low quality reads and error correction. We then merged the error corrected sequences using a minimum overlap of five bp. Chimeras were removed using the bimera algorithm, using default parameters implemented in DADA2. The resulting ASV table were further clustered into 10955 operational taxonomic units (OTUs) using VSEARCH (45) at 97% similarity. LULU (46) was used with default settings to correct for potential OTU over-splitting. Taxonomy was assigned using BLAST (47) to the final OTU table using the UNITE database (48). Sequences with no match to any known fungal sequence and samples with less than 10 OTUs were discarded from downstream analyses. The final dataset (without technical controls) contained 7 399 OTUs and 22 655 516 reads from 572 samples. The number of reads per sample varied from 19 to 182 266 with a mean value of 39 608. The number of OTUs per sample varied from 10 to 863, with a mean value of 257.

Environmental variables

Metadata about building features and occupancy of each daycare were provided by the volunteers in a questionnaire that were delivered together with the samples (Table 1). The location of daycares with complete addresses were provided, and corresponding geographic coordinates (latitude and longitude) were retrieved. Relevant environmental variables were kindly provided by the authors from a recent study modelling the vegetation types in Norway (29). These variables were extracted based on the geographic coordinates of the daycares. From this extensive set of environmental variables (>30), a subset of non-collinear variables (cor > 0.6) was selected for the further analyses (Table 1).

Annotation of fungal (OTUs) growth characteristics

We annotated the 1253 most abundant OTUs, those with > 20 sequences and taxonomic annotation at a species, genus or family level, into growth forms / nutritional mode based on literature surveys. Species/genera/families having unknown, dubious or multiple growth forms/nutritional modes, were not included.

Statistics

The statistical analyses were all preformed in R (49). First, the similarity of the technical replicates was evaluated by

nonmetric multidimensional scaling (NMDS). Since DNA-metabarcoding analyses of samples with low DNA yields may introduce biases during the wet-lab analyses and sequencing, we wanted to control the consistency of our results. For this we used the metaMDS function from the vegan package version 2.4-2 (50) and the results were visualized by ggplot2 (51) (Fig. S1). As visualized in Fig. S1, the distances between biological replicates are generally markedly higher than between the technical replicates. Then, all the samples in the complete dataset were rarefied to 1 342 sequences using the function rrarefy (vegan). Fourteen samples were discarded from downstream statistical analyses due to too low sequencing depth.

To visualize and investigate patterns in OTU composition in relation to environmental variables we performed a global non-metric multidimensional scaling (GNMDS) using the VEGAN package (52), using the settings as recommended by (53). To ensure reliability of the results a detrended correspondence analyses (DCA) was performed in parallel. Extreme outliers that were the same in both ordinations, were manually inspected and subsequently removed from the dataset before the analyses were repeated. Both ordination analyses revealed the same overall pattern (data not shown) and we hereafter focus on the GNMDS analyses. The GNMDS was scaled into half change (HC) units and subjected to varimax rotation using principal component analyses (PCA). To confirm convergence the two best solutions of the GNMDS were compared using Procrustes comparisons with 999 permutations (corr: 0.99, $p = 0.001$). The ordinations were first conducted on the entire dataset containing both indoor and outdoor samples, where a clear pattern was observed. Thereafter, a dataset containing only indoor samples from bathrooms and central rooms was extracted, and the ordinations were conducted on this dataset using the same settings and had the same correlation in the Procrustes comparisons. The following analyses were only conducted on the indoor dataset. The envfit function in VEGAN (i.e. the fit (R^2) of each variables assessed with a Monte-Carlo analyses of 999 permutation) was used to fit the environmental variables: Building type, average construction year, June temperature, longitude, mean temp of the coldest quarter, month, number of departments and children, presence of rodents (pests), proximity to all types of water, proximity to coast, room type, and type of daycare, to the GNMDS. The numerical variables were visualized using the vectors from the output from the envfit function. We further did a variation partitioning with CCA (canonical correspondence analysis) with 999

permutations, to quantify the components of variation by the variables mentioned above, with forward selection, as implemented in vegan.

To investigate OTU richness trends, a linear mixed effect model was applied using the nlme package (54) in the statistical environment R, including daycare ID as a random contribution. Colinear variables were excluded as described above (cor > 0.6), however, to further avoid multicollinearity in the mixed effect model the corvif function described in Zuur et al (2009) was employed, using a threshold of 2.5 (55). Backwards stepwise model selection was preformed based on Akaike information criterion (AIC). The distribution of the 30 most abundant genera across indoor and outdoor samples were visualized. To obtain a balanced indoor/outdoor dataset we used the average values from the indoor samples.

Data accessibility

Our initial dataset, as well as the final rarefied dataset, are available at Dryad together with metadata, information about taxonomic annotations and growth form/nutritional mode annotations.

Acknowledgements

We would like to acknowledge the daycare centers for the sampling and for providing metadata of the building features. Mycoteam AS contributed to the sampling and provided sampling equipment. The University of Oslo and the Norwegian Asthma and Allergy Association (NAAF) financially supported the research. PMMS was funded by the European Union's Horizon 2020 research and innovation program (Marie Skłodowska-Curie Individual Fellowship; grant agreement MycoIndoor No 741332).

References

1. Horner WE. 2003. Assessment of the indoor environment: evaluation of mold growth indoors. Immunology and Allergy Clinics 23:519-531.

2. Nevalainen A, Täubel M, Hyvärinen A. 2015. Indoor fungi: companions and contaminants. Indoor Air 25:125-156.

3. Bornehag CG, Blomquist G, Gyntelberg F, Jarvholm B, Malmberg P, Nordvall L, Nielsen A, Pershagen G, Sundell J. 2001. Dampness in buildings and health. Nordic interdisciplinary review of the scientific evidence on associations between exposure to "dampness" in buildings and health effects (NORDDAMP). Indoor Air 11:72-86.

4. Mendell MJ, Mirer AG, Cheung K, Tong M, Douwes J. 2011. Respiratory and allergic health effects of dampness, mold, and dampness-related agents: A review of the epidemiologic evidence. Environmental Health Perspectives 119:748-756.

5. Portnoy JM, Kwak K, Dowling P, VanOsdol T, Barnes C. 2005. Health effects of indoor fungi. Annals of Allergy, Asthma & Immunology 94:313-320.

6. Madureira J, Paciência I, Rufo J, Ramos E, Barros H, Teixeira JP, de Oliveira Fernandes E. 2015. Indoor air quality in schools and its relationship with children's respiratory symptoms. Atmospheric Environment 118:145-156.

7. Vandenborght L-E, Enaud R, Urien C, Coron N, Girodet P-O, Ferreira S, Berger P, Delhaes L. 2020. Type 2–high asthma is associated with a specific indoor mycobiome and microbiome. Journal of Allergy and Clinical Immunology doi:https://doi.org/10.1016/j.jaci.2020.08.035.

8. Korkalainen M, Täubel M, Naarala J, Kirjavainen P, Koistinen A, Hyvärinen A, Komulainen H, Viluksela M. 2017. Synergistic proinflammatory interactions of microbial toxins and structural components characteristic to moisture-damaged buildings. Indoor Air 27:13-23.

9. Madureira J, Paciência I, Rufo JC, Pereira C, Teixeira JP, de Oliveira Fernandes E. 2015. Assessment and determinants of airborne bacterial and fungal concentrations in different indoor environments: Homes, child day-care centres, primary schools and elderly care centres. Atmospheric Environment 109:139-146.

10. Estensmo EL, Morgado L, Maurice S, Martin-Sanchez PM, Engh IB, Mattsson J, Kauserud H, Skrede I. 2021. Spatiotemporal variation of the indoor mycobiome in daycare centers. Research Square preprint doi:10.21203/rs.3.rs-147547/v1.

11. Barberán A, Ladau J, Leff JW, Pollard KS, Menninger HL, Dunn RR, Fierer N. 2015. Continental-scale distributions of dust-associated bacteria and fungi. Proceedings of the National Academy of Sciences 112:5756-61.

12. Martin-Sanchez PM, Estensmo ELF, Morgado LN, Maurice S, Engh IB, Skrede I, Kauserud H. 2021. Analyzing indoor mycobiomes through a large-scale citizen science study in Norway. Molecular Ecology https://doi.org/10.1111/mec.15916.

13. Adams RI, Miletto M, Taylor JW, Bruns TD. 2013. Dispersal in microbes: fungi in indoor air are dominated by outdoor air and show dispersal limitation at short distances. The ISME Journal 7:1262-1273.

14. Adams RI, Miletto M, Taylor JW, Bruns TD. 2013. The diversity and distribution of fungi on residential surfaces. PLOS ONE 8:e78866.

15. Prussin AJ, Marr LC. 2015. Sources of airborne microorganisms in the built environment. Microbiome 3:78.

16. Dannemiller KC, Gent JF, Leaderer BP, Peccia J. 2016. Influence of housing characteristics on bacterial and fungal communities in homes of asthmatic children. Indoor Air 26:179-192.

17. Ghannoum MA, Jurevic RJ, Mukherjee PK, Cui F, Sikaroodi M, Naqvi A, Gillevet PM. 2010. Characterization of the oral fungal microbiome (mycobiome) in healthy individuals. PLOS Pathogens 6:e1000713.

18. Dupuy AK, David MS, Li L, Heider TN, Peterson JD, Montano EA, Dongari-Bagtzoglou A, Diaz PI, Strausbaugh LD. 2014. Redefining the human oral mycobiome with improved practices in amplicon-based taxonomy: discovery of malassezia as a prominent commensal. PLOS ONE 9:e90899.

19. Jo J-H, Deming C, Kennedy EA, Conlan S, Polley EC, Ng W-I, Segre JA, Kong HH. 2016. Diverse human skin fungal communities in children converge in adulthood. Journal of Investigative Dermatology 136:2356-2363.

20. White TC, Findley K, Dawson TL, Scheynius A, Boekhout T, Cuomo CA, Xu J, Saunders CW. 2014. Fungi on the skin: dermatophytes and *Malassezia*. Cold Spring Harbor Perspectives in Medicine 4:a019802.

21. Findley K, Oh J, Yang J, Conlan S, Deming C, Meyer JA, Schoenfeld D, Nomicos E, Park M, Kong HH, Segre JA. 2013. Topographic diversity of fungal and bacterial communities in human skin. Nature 498:367-70.

22. Nilsson RH, Anslan S, Bahram M, Wurzbacher C, Baldrian P, Tedersoo L. 2019. Mycobiome diversity: high-throughput sequencing and identification of fungi. Nature Reviews Microbiology 17:95-109.

23. Weikl F, Tischer C, Probst AJ, Heinrich J, Markevych I, Jochner S, Pritsch K. 2016. Fungal and bacterial communities in indoor dust follow different environmental determinants. PLOS ONE 11:e0154131-e0154131.

24. Yamamoto N, Hospodsky D, Dannemiller KC, Nazaroff WW, Peccia J. 2015. Indoor emissions as a primary source of airborne allergenic fungal particles in classrooms. Environmental Science & Technology 49:5098-106.

25. Barberán A, Dunn RR, Reich BJ, Pacifici K, Laber EB, Menninger HL, Morton JM, Henley JB, Leff JW, Miller SL, Fierer N. 2015. The ecology of microscopic life in household dust. Proceedings of the Royal Society B: Biological Sciences 282:20151139.

26. Gura T. 2013. Citizen science: amateur experts. Nature 496:259-261.

27. Cohn JP. 2008. Citizen science: can volunteers do real research? BioScience 58:192-197.

28. Dickinson JL, Zuckerberg B, Bonter DN. 2010. Citizen science as an ecological research tool: challenges and benefits. Annual Review of Ecology, Evolution, and Systematics 41:149-172.

29. Horvath P, Halvorsen R, Stordal F, Tallaksen LM, Tang H, Bryn A. 2019. Distribution modelling of vegetation types based on area frame survey data. Applied Vegetation Science 22:547-560.

30. Shin S-K, Kim J, Ha S-m, Oh H-S, Chun J, Sohn J, Yi H. 2015. Metagenomic insights into the bioaerosols in the indoor and outdoor environments of childcare facilities. PLOS ONE 10:e0126960.

31. Raimondi S, Amaretti A, Gozzoli C, Simone M, Righini L, Candeliere F, Brun P, Ardizzoni A, Colombari B, Paulone S, Castagliuolo I, Cavalieri D, Blasi E, Rossi M, Peppoloni S. 2019. Longitudinal survey of fungi in the human gut: ITS profiling, phenotyping, and colonization. Frontiers in Microbiology 10:1575.

32. Zhu T, Duan Y-Y, Kong F-Q, Galzote C, Quan Z-X. 2020. Dynamics of skin mycobiome in infants. Frontiers in Microbiology 11.

33. Cui L, Morris A, Ghedin E. 2013. The human mycobiome in health and disease. Genome Medicine 5:63.

34. Jean J, Goldberg S, Khare R, Bailey LC, Forrest CB, Hajishengallis E, Koo H. 2018. Retrospective Analysis of *Candida*-related conditions in infancy and early childhood caries. Pediatric Dentistry 40:131-135.

35. Nakabayashi A, Sei Y, Guillot J. 2000. Identification of *Malassezia* species isolated from patients with seborrhoeic dermatitis, atopic dermatitis, pityriasis versicolor and normal subjects. Medical Mycology 38:337-41.

36. Zinkeviciene A, Vaiciulioniene N, Baranauskiene I, Kvedariene V, Emuzyte R, Citavicius D. 2011. Cutaneous yeast microflora in patients with atopic dermatitis. Central European Journal of Medicine 6:713.

37. Shelton BG, Kirkland KH, Flanders WD, Morris GK. 2002. Profiles of airborne fungi in buildings and outdoor environments in the United States. Applied and Environmental Microbiology 68:1743-1753.

38. O'Connor G T, Walter M, Mitchell H, Kattan M, Morgan WJ, Gruchalla RS, Pongracic JA, Smartt E, Stout JW, Evans R, Crain EF, Burge HA. 2004. Airborne fungi in the homes of children with asthma in low-income urban communities: The inner-city asthma study. Journal of Allergy & Clinical Immunology 114:599-606.

39. Adams RI, Tian Y, Taylor JW, Bruns TD, Hyvärinen A, Täubel M. 2015. Passive dust collectors for assessing airborne microbial material. Microbiome 3:46.

40. Korkalainen M, Täubel M, Naarala J, Kirjavainen P, Koistinen A, Hyvärinen A, Komulainen H, Viluksela M. 2017. Synergistic proinflammatory interactions of microbial toxins and structural components characteristic to moisture-damaged buildings. Indoor Air 27:13-23.

41. White TJ, Bruns T, Lee S, Taylor J. 1990. Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. PCR Protocols: a Guide to Methods and Applications 18:315-322.

42. Ihrmark K, Bödeker I, Cruz-Martinez K, Friberg H, Kubartova A, Schenck J, Strid Y, Stenlid J, Brandström-Durling M, Clemmensen KE. 2012. New primers to amplify the

fungal ITS2 region–evaluation by 454-sequencing of artificial and natural communities. FEMS Microbiology Ecology 82:666-677.

43. Martin M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet journal 17:10-12.

44. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP. 2016. DADA2: high-resolution sample inference from Illumina amplicon data. Nature Methods 13:581.

45. Rognes T, Flouri T, Nichols B, Quince C, Mahé F. 2016. VSEARCH: a versatile open source tool for metagenomics. PeerJ Preprints 4:e2409v1.

46. Frøslev TG, Kjøller R, Bruun HH, Ejrnæs R, Brunbjerg AK, Pietroni C, Hansen AJ. 2017. Algorithm for post-clustering curation of DNA amplicon data yields reliable biodiversity estimates. Nature Communications 8:1188.

47. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. Journal of Molecular Biology 215:403-410.

48. Koljalg U, Larsson KH, Abarenkov K, Nilsson RH, Alexander IJ, Eberhardt U, Erland S, Hoiland K, Kjoller R, Larsson E, Pennanen T, Sen R, Taylor AF, Tedersoo L, Vralstad T, Ursing BM. 2005. UNITE: a database providing web-based methods for the molecular identification of ectomycorrhizal fungi. New Phytologist 166:1063-8.

49. Team RC. 2019. R: A language and environment for statistical computing. R Foundation for Statistical Computing. Austria: Vienna, https://www.R-project.org/.

50. Oksanen J, Blanchet FG, Friendly M, Kindt R, Legendre P, McGlinn D, Minchin PR, O'hara R, Simpson GL, Solymos P. 2016. vegan: community ecology package. R package version 2.4-2. Vienna: R Foundation for Statistical Computing.

51. Wickham H. 2016. ggplot2: elegant graphics for data analysis. Springer.

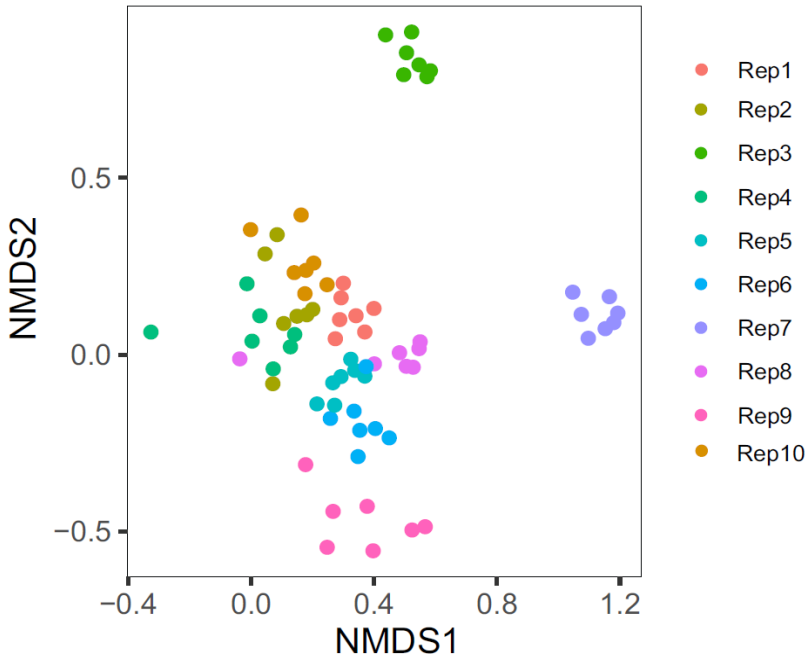
52. Oksanen J, Kindt R, Legendre P, O'Hara B. 2019. The vegan package.

53. Liu H, Økland T, Halvorsen R, Gao J, Liu Q, Eilertsen O, Bratli H. 2008. Gradients analyses of forests ground vegetation and its relationships to environmental variables in five subtropical forest areas, S and SW China. Sommerfeltia 32:3-196.

54. Pinheiro J, Bates D, DebRoy S, Sarkar D, Team RC. 2019. nlme: Linear and nonlinear mixed effects models. R package version 3:109.

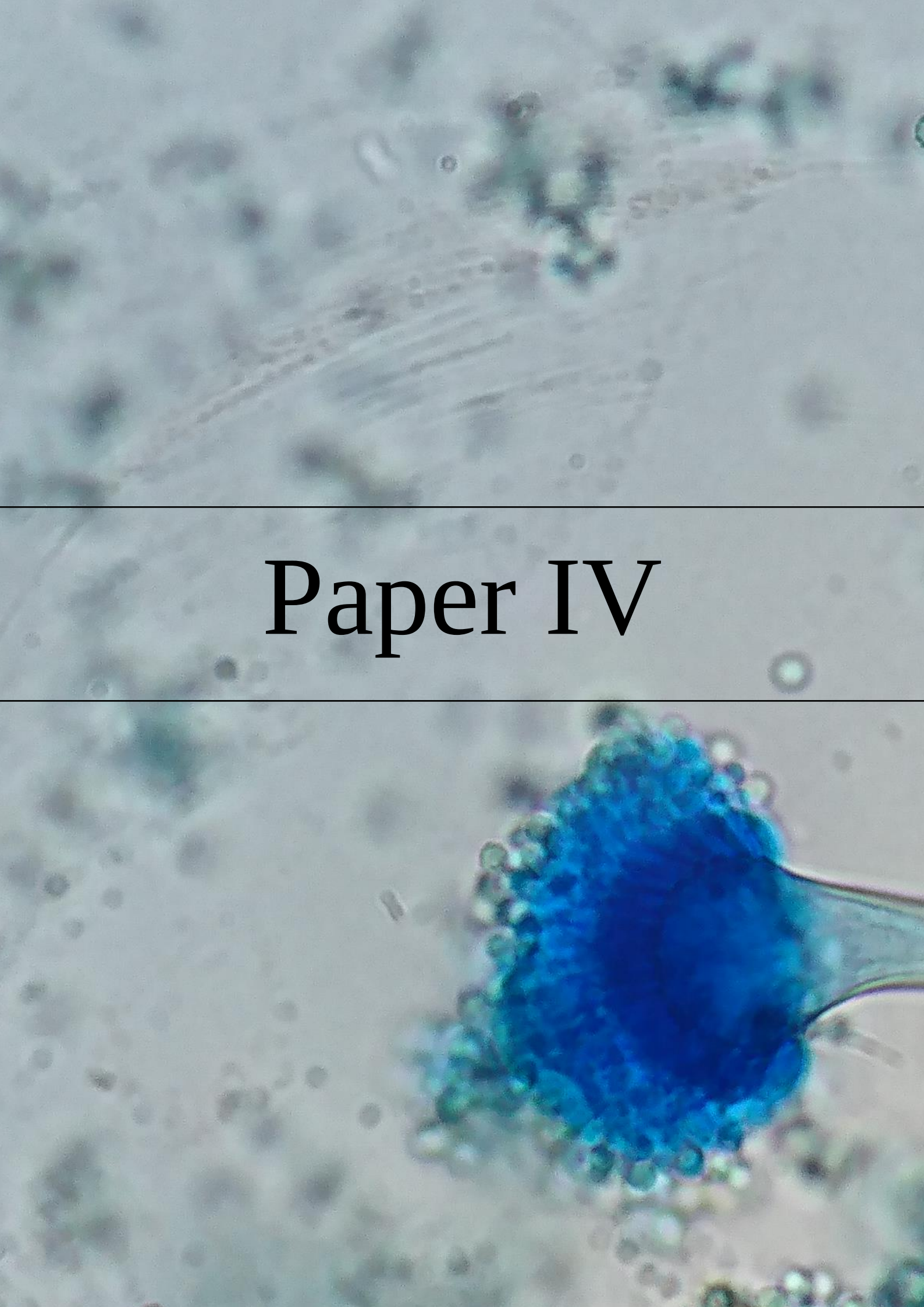
55. Zuur A, Ieno EN, Walker N, Saveliev AA, Smith GM. 2009. Mixed effects models and extensions in ecology with R. Springer Science & Business Media.

Supplementary



Supplementary figure 1. Nonmetric multidimensional scaling (NMDS) ordination plot of technical controls. Each point represents one sample, and the color separates the different replicates. The plot illustrates that the technical replicates cluster together (with a few exceptions) and that the distance between biological replicates are generally higher than between technical replicates.

Paper IV



Spatiotemporal variation of the indoor mycobiome in daycare centers

Eva Lena F. Estensmo[1], Luis Morgado[1,2], Sundy Maurice[1], Pedro M. Martin-Sanchez[1], Ingeborg B. Engh[3], Johan Mattsson[3], Håvard Kauserud[1], Inger Skrede[1]

1. Section for Genetics and Evolutionary Biology (Evogene), Department of Biosciences, University of Oslo, P.O. Box 1066 Blindern, 0316 Oslo, Norway.
2. Naturalis Biodiversity Center, Darwinweg 2, 2333 CR Leiden, Nederland.
3. Mycoteam AS, P.O. Box 5 Blindern, 0313 Oslo, Norway.

Abstract

Background: Children spend considerable time in daycare centers in parts of the world, and are exposed to the indoor micro- and mycobiomes of these facilities. The level of exposure to microorganisms varies within and between buildings, depending on occupancy, climate and season. In order to evaluate indoor air quality, and the effect of usage and seasonality, we investigate the spatiotemporal variation in the indoor mycobiomes of two daycare centers. We collected dust samples from different rooms throughout a year and analyzed their mycobiomes using DNA metabarcoding.

Results: The fungal community composition in rooms with limited occupancy (auxiliary rooms) was similar to the outdoor samples, and clearly different from the rooms with higher occupancy (main rooms). The main rooms had higher abundance of *Ascomycota*, while the auxiliary rooms contained comparably more *Basidiomycota*. We observed a strong seasonal pattern in the mycobiome composition, mainly structured by the outdoor climate. Most markedly, basidiomycetes of the orders *Agaricales* and *Polyporales*, mainly reflecting typical outdoor fungi, were more abundant during summer and fall. In contrast, ascomycetes of the orders *Saccharomycetales* and *Capnodiales* were dominant during winter and spring.

Conclusions: Our findings provide clear evidences that the indoor mycobiome in daycare centers are structured by occupancy as well as outdoor seasonality. We conclude that the temporal variability should be accounted for in indoor mycobiome studies and in the evaluation of indoor air quality of buildings.

Background

Humans spend significant amount of time indoors, in private homes, but also in workplaces, schools, daycare centers and hospitals. We share these indoor environments with a variety of microorganisms, including microscopic fungi that may affect our health in different ways. In moist conditions, fungi can propagate and act as sources of indoor pollutants leading to poor indoor air quality. This has been associated with adverse health effects, such as allergies, asthma and other respiratory symptoms [1, 2]. The indoor microorganisms originate from both indoor and outdoor sources and are potentially structured by numerous factors, including building features, building usage, the number and type of occupants, and not least, our behavior [3, 4]. The bacterial indoor microbiome is known to be highly affected by the occupants and their activities, and often directly related to the human body [5, 6]. However, indoor fungi, which can be referred to as the indoor mycobiome, are known to be highly influenced by the outdoor air and climate [5, 7, 8]. Previous studies at large geographical scale in the US and Norway, have demonstrated that the composition of the indoor mycobiomes significantly correlates with variables of the outdoor environment (i.e. climate, soil and vegetation) [9, 10]. The most important indoor sources of fungi include occupants, pets, food, waste, plants, plumbing systems, mold damages, heating, ventilation and air conditioning [11]. Different rooms in buildings may have different mycobiome composition due to different occupancy and exposure to outdoor air [12, 13]. For example, central rooms with higher activity, like the kitchen and living room, promote dust resuspension in the air that facilitate dispersal of fungi from occupants, their activities and outdoor sources. Similarly, floor dust of high activity rooms contains higher levels of skin-associated yeasts of the genera *Rhodotorula*, *Candida*, *Cryptococcus*, *Malassezia*, and *Trichosporon* [14].

The indoor mycobiomes may not only differ in space, but also in time. Previous culture-based studies have been reviewed by Nevalainen et al. [15], where they found a general pattern of seasonal variation with lower concentrations of airborne fungi in winter than in summer. This review included studies from different climatic regions in countries like Australia [16], Denmark [17], and Taiwan [18]. DNA-based studies have also reported a clear seasonal variation of fungal richness, diversity and community composition in indoor environments, in both dust and air

samples [7, 19]. By analyzing dust samples from a university housing facility in California, Adams et al. [7] reported higher fungal richness in winter than in summer. Likewise, Weikl et al. [19] showed a drop of the fungal diversity in summer, based on floor dust samples from 286 houses in Munich. This latter observation was explained by the high prevalence of a few dominant taxa during summer [19]. Hence, observed temporal trends in indoor mycobiomes are not uniform.

In boreal and temperate climatic regions, the fungal spore diversity and composition in outdoor air are expected to vary significantly more throughout the year because of clear seasons. For example, Karlsson et al. [20] reported lowest richness of fungi and bacteria for air samples collected during winter in two climatic zones from Sweden. It can be expected that this variation influences the indoor mycobiome, due to an influx of spores into buildings. Many fungi, especially basidiomycetes, produces fruit bodies during the fall leading to a relatively higher spore abundance during this period [21]. Plant pathogens, dominated by ascomycetes, may have a wider temporal distribution since many spread asexual spores during the entire plant growth season [22]. Indoor fungi originating from indoor sources, here growing on available organic materials, can be expected to have a year-round growth and sporulation connected to human activity.

A particularly interesting environment to study the spatiotemporal variation of the mycobiome is daycare centers, where children, at least in parts of the world, spend a considerable amount of time. For example, in Norway, 92.2% of children between 1-5 years old are in daycares. This particular built environment is characterized by a high occupancy with high levels of activity, and higher fungal concentrations have been detected here compared to private homes [23]. Exploring the indoor mycobiome and revealing the factors driving this spatiotemporal variation are important not only to understand the ecological context of indoor fungi, but also to recognize the effect that some fungal species may have on children's health. To what degree the mycobiome associated with daycares affect the children's health is still unknown.

The overarching aim of this study is to reveal the indoor mycobiomes spatiotemporal dynamics in daycare centers in order to improve evaluations of air quality in indoor air. We expect rooms with different occupancy to differ in mycobiome composition (Hypothesis 1; H1), with

frequently accessed rooms being dominated by indoor fungi derived from the occupants and their activities. Given that part of the indoor mycobiome originates from outdoor sources, we hypothesize that indoor mycobiomes fluctuate with seasons (H2). In seasons with optimal fungal growth conditions outdoors, as in summer and fall, we expect that a higher proportion of the indoor mycobiome is derived from outdoor sources, with Basidiomycota dominating during the fall season (H3). In contrast, we expect that a higher proportion of the mycobiome has an indoor origin with increased amount of time spend inside during winter and spring (H4). To test these hypotheses, we collected indoor dust and outdoor air samples from two daycare centers bi-weekly during a year and performed DNA metabarcoding of the rDNA ITS2 region. Two daycare centers located in Oslo, Norway, were selected for the study. We collected dust swab samples every second week from different rooms and stores in the daycare centers (Fig. S1), as well as outdoor air samples every week. Fungi present in the samples were surveyed through DNA metabarcoding analyses of the rDNA ITS2 region.

Results

Mycobiome composition

A variation partitioning analysis of the indoor dust mycobiome (Fig. 1a) revealed that 37% of the compositional variation could be ascribed to assessed factors, including outdoor climate, time (i.e. the biweekly sampling point), space and occupant characteristics. Most of the explained variation was accounted for by the combined effects of occupants and room type (19%). These two factors are correlated, as the activity of both staff and children are considerably lower in the basement and loft (hereafter called the auxiliary rooms) than in the remaining indoor rooms (hereafter referred to as the main rooms). Nine percent of the variation was accounted for by time alone, likely reflecting other unmeasured environmental factors changing with time, while 8% was accounted for by the combined effect of time and climate, which again are tightly coupled.

These findings were also displayed in multivariate analyses (Fig. 1b-d), with a distinct difference in mycobiome composition between samples from the auxiliary rooms and the main rooms, the latter used more frequently by the staff and children (Fig. 1b). The dust mycobiomes from the auxiliary rooms were similar to the mycobiome obtained from the outside environment (Fig. S2).

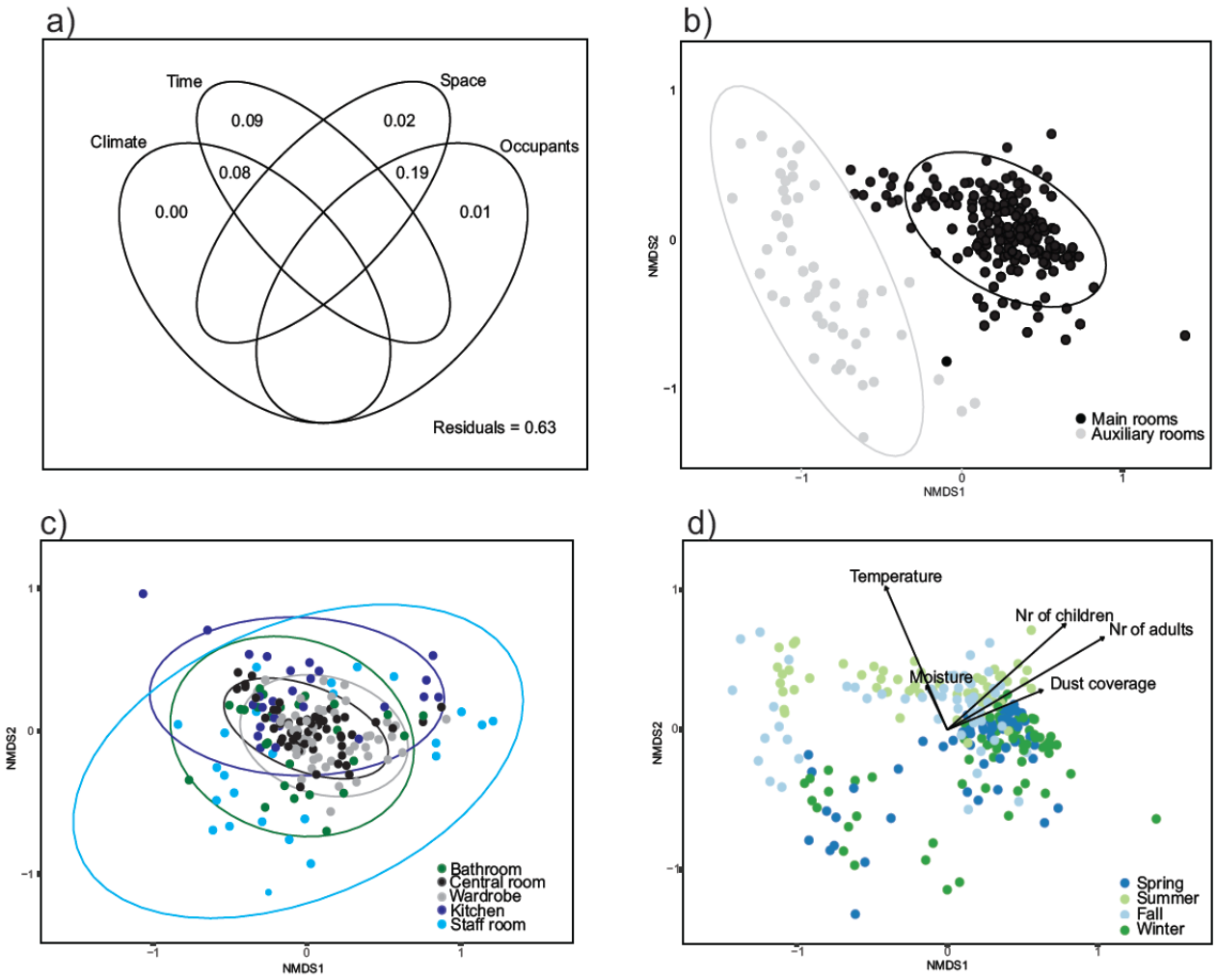


Fig. 1. Fungal community composition in indoor dust samples from two daycare centers in Oslo, Norway throughout a full year. (a) Variation partitioning analysis (VPA) for the indoor dataset (including auxiliary and main rooms), summarizing the effects of four groups of variables: Climate = Temperature (PCA1) and Moisture (PCA2), Time = Month and Season, Space = Daycare and Room, Occupants = Number of adults, Age of children and Number of children. (b) NMDS ordination plot of indoor dust samples displaying their compositional variation in the mycobiome. The color differentiates the main rooms from the auxiliary rooms. (c) NMDS ordination plot of main rooms, with colors differentiating between the rooms. (d) NMDS ordination plot of the indoor samples differentiated by season, including numerical variables with significant association ($p < 0.05$).

The main rooms from both daycare centers overlapped in mycobiome composition, although the compositional variation (beta-diversity) across samples was higher in some rooms, like the kitchen and staff room (Fig. 1c). In addition to the compositional differences related to room type and occupancy, we observed a clear

temporal variation in the indoor mycobiome composition (Fig. 1d). Although there was some overlap, the winter and spring samples were more similar in fungal community composition, as were the samples from summer and fall. The temporal trend in mycobiome composition correlated with the yearly variation in temperature and moisture, as could be seen from the vectors fitted in Fig. 1d. Room, time and occupants were the most important factors structuring the fungal community composition, whereas the impact of which daycare the samples were from, was low (Table 1).

Table 1. Adonis test results showing the influence of the variables on the composition of the dust mycobiome from the complete indoor dataset (auxiliary and main rooms), and from the main rooms of two daycare centers sampled in Oslo, Norway throughout a full year.

Variable	**Auxiliary and main rooms**		**Main rooms**	
	R squared	p value	R squared	p value
Room	0.218	0.001	0.092	0.001
Month	0.196	0.001	0.233	0.001
Children age	0.133	0.001	0.056	0.001
Season	0.115	0.001	0.124	0.001
Nr of adults	0.114	0.001	0.032	0.001
Nr of children	0.097	0.001	0.027	0.001
Temperature	0.071	0.001	0.082	0.001
Daycare	0.07	0.001	0.033	0.001
Dust coverage	0.039	0.001	0.008	0.029
Moisture	0.02	0.001	0.022	0.001

Taxonomic variation

Overall, *Ascomycota* was more prominent in the main room while *Basidiomycota* was far more abundant in samples from the auxiliary rooms (Fig. 2a). *Mucoromycota* was more prevalent in the main rooms, as well as other fungal phyla (mainly *Chytridiomycota*). Ascomycete yeasts affiliated to *Saccharomycetales* were more abundant in the main rooms, while basidiomycetes from the orders *Agaricales* and *Polyporales* were dominating the samples from auxiliary rooms. We observed a clear temporal trend in the composition of fungal taxonomic groups (displayed at order level in Fig. 2a). Most markedly, the proportion of basidiomycete sequences from the orders *Agaricales*, *Polyporales* and *Hymenochaetales*, mainly reflecting outdoor fungi, were higher during the growth season (May-November) than in winter, a pattern far more prominent in the

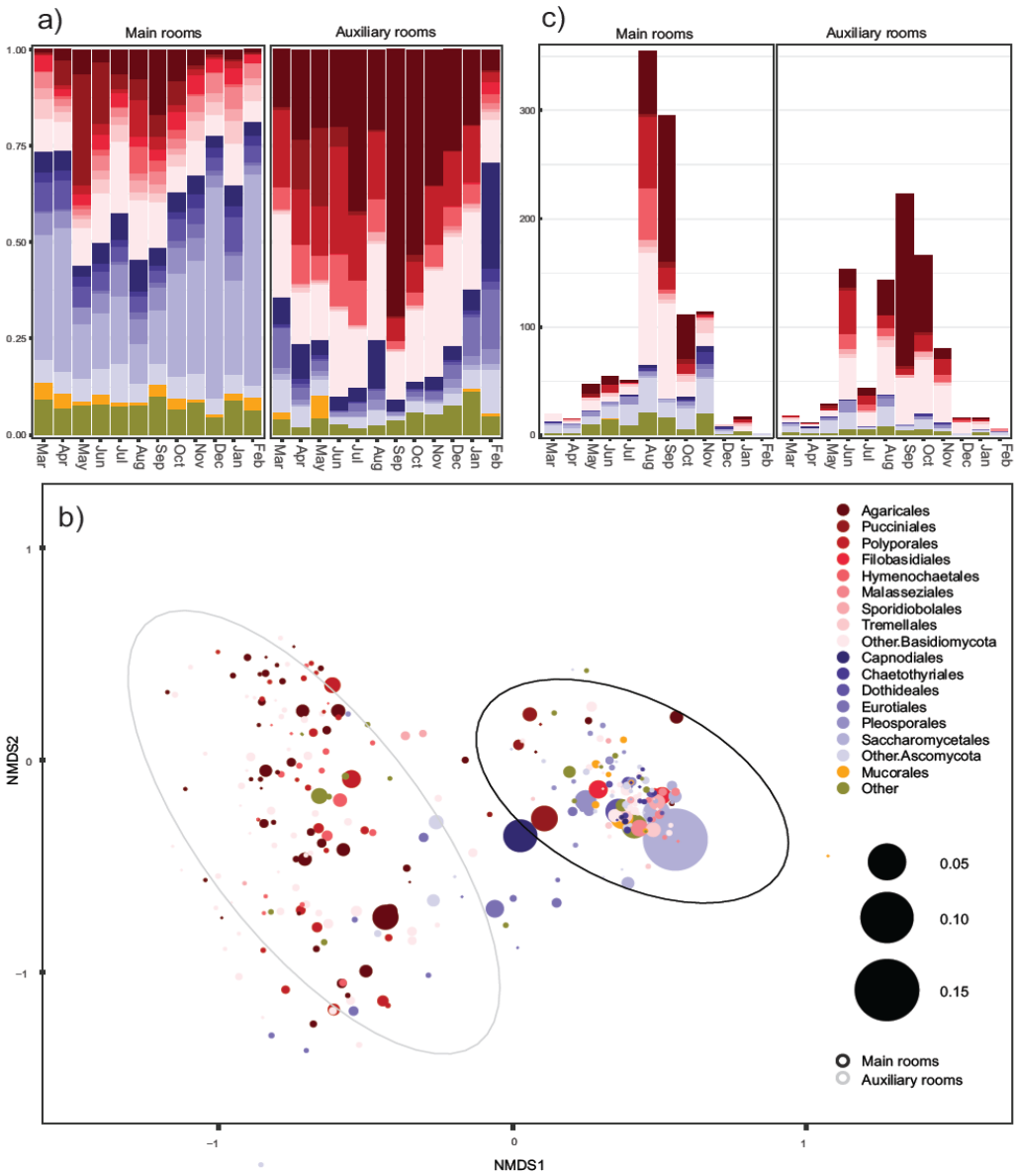


Fig. 2. Temporal taxonomic variation in indoor dust samples from two daycare centers in Oslo, Norway sampled throughout a full year. (a) Relative species abundance of the main fungal orders. (b) NMDS ordination plot of the 300 most abundant fungal OTUs. Point size indicates relative abundance and colors indicates their taxonomical order. Colors in red = Basidiomycota, blue = Ascomycota, yellow = Mucorales and green = species belonging to another order. The ellipses represent the main rooms and the auxiliary rooms, as shown in Fig. 1b. (c) Number of indicator species detected for each month, as well as their taxonomic affiliation at order level (only OTUs present in at least 3 samples per month were included). Seasons: winter from December to February, spring from March to May, summer from June to August and fall from September to November.

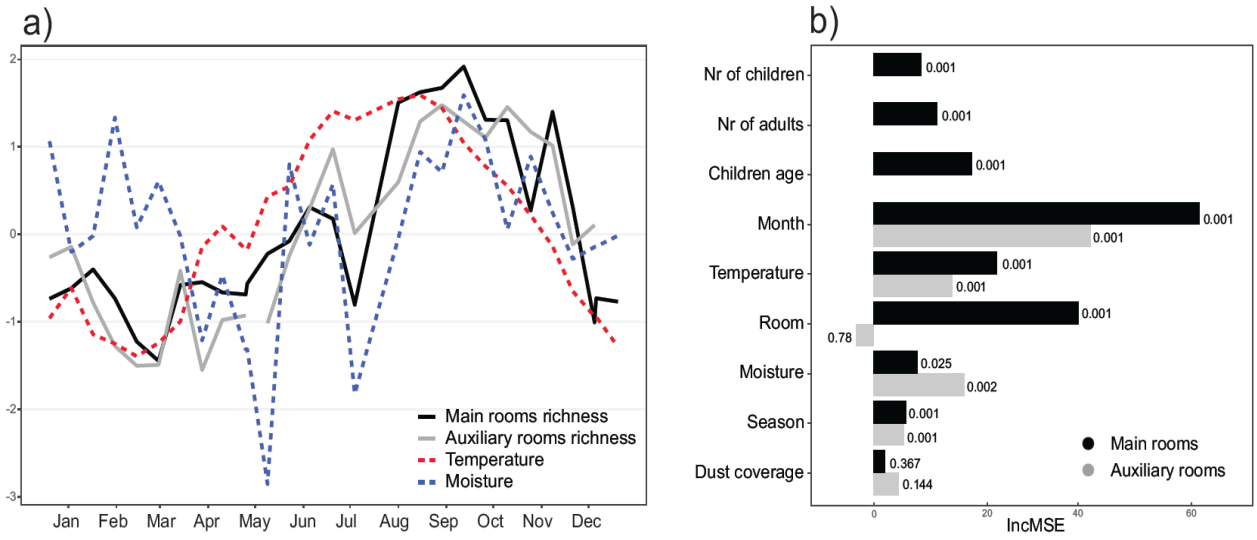


Fig. 3. (a) Richness time series for the main room and auxiliary rooms of two daycare centers in Oslo, Norway sampled throughout a full year. The gap in auxiliary rooms richness in May is due to samples being excluded from the analysis because of low number of reads. (b) Random forest model showing the importance (percentage of increase in mean squared errors - IncMSE) of each variable for richness of the two indoor datasets, auxiliary and main rooms. Numbers on the bars indicate statistical significance, which was obtained through bootstrapping.

auxiliary rooms than in the main rooms. The *Ascomycota* was proportionally more abundant in colder periods. However, the *Saccharomycetales*, likely derived from indoor sources, was prevalent year-round in the main room. The order *Eurotiales*, including fungal genera with allergenic potentials, such as *Penicillium* and *Aspergillus*, was relatively more prevalent in the main rooms in the colder season. Similar trends were observed in the OTU ordination plot (Fig. 2b), where the dominant *Ascomycota* OTUs are associated with the main rooms, while the *Basidiomycota* OTUs with the auxiliary rooms. Further, the main rooms are dominated by OTUs of *Saccharomycetales*, *Mucorales*, *Malasseziales* and *Filobasidiales*.

Indicator species analyses, assessing which fungal OTUs followed a significant temporal trend on a monthly basis, revealed that numerous OTUs in the already mentioned orders of *Agaricales*, *Polyporales* and *Hymenochaetales* increased considerably during their expected fruiting season, independently of space (i.e., room) (Fig. 2c).

Richness and evenness trends

The main and auxiliary rooms had comparable fungal richness that largely followed a similar temporal trend, with higher richness in the summer and fall (June-November) (Fig. 3a). The richness followed the variation in annual temperature. In winter, the richness deviated more from the moisture gradients. The evenness followed a similar trend as the richness (Fig. S3).A random forest model, which was used to assess the contribution of each factor in the observed richness patterns (Fig. 3b), revealed that month and season (both enclosing various environmental factors), as well as, temperature and moisture, accounted for much of the variation in both the main and the auxiliary rooms datasets. In addition, the factor rooms was highly important in the main rooms dataset, where presence of children and adults also contributes to the richness.

Discussion

In this study, we observed that the indoor mycobiomes of two daycare centers were strongly structured by room type and occupancy (hypothesis H1) and, further, that the mycobiomes changed systematically throughout the seasons (hypothesis H2). No marked difference in mycobiome composition was observed between the two studied daycare centers, indicating a common pattern of indoor mycobiomes in daycare centers from the same local geographic region.

Spatial distribution

We observed a clear separation in mycobiome composition of the main rooms and auxiliary rooms, which likely can be explained by the number of people accessing and using the rooms. These results strongly support our hypothesis (H1) and further suggest that occupancy is an important factor shaping the indoor mycobiome, in addition to the outdoor air. The outdoor samples (air sampling) were collected as point samples in one day, while the auxiliary room samples represent a collection of dust accumulated within two weeks. These different sampling methods may influence the recovered mycobiomes. Nevertheless, the mycobiome composition of the auxiliary rooms and the outdoor mycobiome were highly similar, which supports that occupancy strongly affects the indoor mycobiome. Previous studies of indoor environments suggest that the indoor mycobiomes are highly affected by outdoor air [7, 9, 10, 24, 25]. Most of these studies have not accounted for indoors environments with different levels of activity. However, in a recent study in private homes in Norway, we demonstrated

that the number of inhabitants affected the indoor mycobiome composition [10].

The highest fungal richness was found in the main rooms. This may be explained by that the indoor air of the main rooms includes outdoor taxa, in addition to the more specific indoor fungi derived from the occupants and their activities. Higher fungal richness in indoor environments than in outdoor air has also been found in private houses and schools [5, 10, 14]. It should be noted, though, that richness analyses based on DNA-metabarcoding is vulnerable to various biases. For example, if some dominant species are present, they may mask the remaining richness during the PCR process, since their DNA templates may outcompete the rarer species during PCR amplification. However, the evenness follows largely the same trend for both types of rooms and are therefore probably not causing significant biases for the richness analyses.

OTUs of the phylum *Basidiomycota* were overrepresented in the auxiliary rooms, whereas there were relatively more OTUs of *Ascomycota* in the main rooms. Likewise, previous studies have demonstrated a predominance of *Ascomycota* in indoor samples, while *Basidiomycota* prevails in outdoor samples [26, 27]. As the auxiliary rooms were more similar to the outdoor air, we expected basidiomycetes to be more prevalent in these rooms, especially the mushroom-forming *Agaricales* and *Polyporales*. In the main rooms, the high abundance of ascomycetes can be explained by their high tolerance towards environmental stressors, such as high temperature and low water availability, typical conditions in indoor environments. The orders *Saccharomycetales* and *Capnodiales* were the most abundant ascomycetes. *Saccharomycetales* are yeasts including the well-known genera *Saccharomyces*, associated with foods, and the potential human pathogen *Candida*. *Capnodiales*, with the widespread genus *Cladosporium*, includes both plant and human pathogens [28]. In addition, the basidiomycete orders *Malasseziales* and *Filobasidiales*, together with the order *Mucorales* were abundant in the main rooms. These orders includes yeasts and molds, and were also more abundant in indoor mycobiomes than outdoor air in our previous study of private homes [10].

Seasonality

We observed a clear seasonal pattern in the indoor mycobiomes, supporting our hypothesis H2. Collection month was best able to explain the variation in fungal

richness in all rooms. This seasonal pattern is further supported by the evenness and richness analyses of time series, which follows the shift of temperature and moisture throughout the year. Our observed patterns mirror those found in seasonal studies on outdoor mycobiomes. For example, in northern Sweden, the outdoor fungal communities shifted throughout the season [20]. Since the outdoor fungal community has a strong impact on indoor mycobiomes, it is expected that seasonal changes in the outdoor environment also affect which fungi occur indoor. During the spring, summer and fall, with temperatures above zero, fungal activity and sporulation are clearly linked to the level of precipitation (i.e. rainfall). However, during winter, the precipitation manifests largely as snow, which has less effect on the fungal communities at sub-zero temperatures. It has been suggested that during the winter, when the ground is frozen and covered by snow, the impact of the outdoor fungal community on the indoor mycobiome is limited [29]. This can explain the drop of richness during the winter observed in our study.

Although our results demonstrate that dust sampling can be used to reveal the seasonal variation in the indoor mycobiome, there are methodological constraints that should be taken into consideration when analyzing samples with relatively small amount of DNA. All samples were treated equally in the laboratory, nevertheless, some steps could represent sources of heterogeneity in the dataset e.g. variability in DNA extraction efficiency among organisms, primer bias to different fungal taxonomic group, PCR bias and sequencing errors, which might affect the fungal community. Some of the rooms with limited occupancy, such as the auxiliary rooms and the staff room, had a considerably lower amount of dust and potentially lower amount of DNA. Likewise, the outdoor air samples collected during winter contained lower amount of DNA, most likely due to considerably lower number of fungal spores in the air compared to other seasons.

We observed higher abundances of basidiomycetes during summer and fall in all rooms, with a predominance of *Agaricomycetes*, confirming our hypothesis H3. *Agaricomycetes* cover the mushroom-forming species that typically disperse spores during the summer and fall in high latitude ecosystems. In addition, more indicator species, showing a distinct temporal pattern, were found in these two seasons, in particular from *Agaricales* and *Polyporales*. Thus, high outdoor spore

production of basidiomycetes during the summer and fall affects the indoor mycobiome. A high outdoor aerial abundance of basidiomycetes during summer and fall was also observed in northern Sweden [20]. However, our findings are rather opposite to what has been found in seasonal studies in Munich (Germany) and California [7, 19]. Weikl and colleagues explained their observed decline in diversity during the summer in houses in Munich with a few highly abundant OTUs, and not necessarily of lower diversity [19]. Further, in California the summers are warm and dry, and the mushroom-forming species of the *Agaricomycetes* often fruit during late fall and winter. Thus, all studies may show the same pattern of higher richness of indoor mycobiome during the outdoor sporulation period of basidiomycetes.

The ascomycetes, especially prevalent in indoor conditions, were proportionally more abundant indoor during winter and spring compared to summer and fall. This confirms hypothesis H4. At sub-zero temperatures during winter and spring, fungal growth and sporulation outdoor is reduced in the study area (Oslo, Norway), which will limit the input of basidiomycetes to the indoor environment. Instead, ascomycetes of indoor origin will be more prevalent during this time-period. Similar findings were reported in a seasonal study of indoor mycobiomes of four office complexes, where ascomycete molds and basidiomycete yeasts were more common in the spring and winter [25]. In contrast, in another study monitoring airborne fungi in four daycares centers over 12 months through culturing, viable counts of major indoor fungi were significantly lower in the winter [30]. Overall, they found the ascomycetes *Cladosporium, Penicillium, Alternaria,* and *Aspergillus* to be the most dominating genera. These genera, considered to be some of the most allergenic fungi normally present indoors and outdoors, have also been reported as abundant in other studies [31, 32]. De Ana et *al.* investigated the seasonal distribution of these species, and found that the highest presence of *Aspergillus*, *Cladosporium* and *Penicillium* in the indoor environment was registered in fall, whereas *Alternaria* was more frequent in the summer [32]. In our study, the order Eurotiales, including *Penicillium* and *Aspergillus*, was relatively more prevalent in the main rooms in the colder season. In addition, the genera *Saccharomyces*, *Cladosporium* and *Didymella,* often encountered in indoor environments in other studies [31], were also especially prevalent in the winter.

Numerous indoor ascomycetes are known to cause allergies and disease in humans, and it is a concern if these species have a higher prevalence during the winter when the children spend more time inside. In addition, in a previous study of school environments [14], they showed that occupancy contributed more to the allergenic fungal populations in indoor air than outdoor fungi. Understanding this spatiotemporal variation of the indoor mycobiome is important as the time spent inside during the different seasons varies, and will reflect how the children are affected by these fungal species.

Conclusions

In conclusion, our study demonstrates clear differences in the dust mycobiome composition in daycare centers between rooms with different occupancy. The more human activity, the more the indoor mycobiome differs from the outdoor mycobiome composition. To our knowledge, this is the first study that monitors the same rooms and buildings continuously over a full year using a DNA metabarcoding approach. Thus, our results demonstrate how the mycobiome composition follows a strong seasonal trend, mirroring outdoor weather conditions. Knowledge about the seasonal trends will have important implications for monitoring and evaluation of indoor air quality.

Methods

Sampling

Dust samples from the daycare centers were collected with floq swabs (Copan Italia spa, Brescia, Italy) and adhesive tapes (Mycotape 2, Mycoteam AS, Oslo, Norway) from 30 x 40 cm^2 glass plates located 1-2 m above floor level. The swab collected dust from an area of 30 x 30 cm^2, whereas the tapes sampled dust from 3.8 x 7.5 cm^2 from the remaining area to calculate the percentage of dust coverage. These samples were collected once for every sampling date. The plates were placed in different rooms and stores in the daycares (Fig. S1). Five rooms were sampled in daycare A, and four rooms in daycare B. The plates were sterilized with 85% ethanol after each harvesting, every second week throughout a year. In addition, outdoor air samples were collected every week throughout a year by processing approximately 1800 L air through a 25 mm cassette with a 0.8 µm pore diameter mixed cellulose ester filter (Zefon international, Ocala, FL, USA) by using an air pump. The 294 swab and filter samples were stored at -80 ºC until DNA extraction, whereas the

adhesive tapes were directly scanned for dust coverage using Epson Perfection V850 Pro (Seiko Epson Corporation, Nagano, Japan). The percentage of dust coverage was calculated with the Olympus Stream v 1.9 software.

DNA extraction and fungal metabarcoding

DNA from swabs and filter samples were extracted using the E.Z.N.A Soil DNA kit (Omega Bio-tek, Norcross, GA, USA). The swabs and filters were placed in disruptor tubes using sterilized scissor or forceps, respectively, and 800 µL SLX-Mlus Buffer was added. The samples were homogenized for 2 x 1 min at 30 Hz using TissueLyser (Qiagen, Hilden, Germany) and stored at -20 °C until further processing. The samples were thawed at 70 °C, following an incubation of 10 min and homogenized twice for 1 min at 30 Hz using a TissueLyser. The samples were cooled on ice before 600 µL chloroform was added. Then, the samples were vortexed and centrifuged at 13 000 rpm for 5 min at RT. The aqueous phase was transferred to a new 1.5 mL tube and an equal volume of XP1 Buffer was added before vortexing. The samples were then added to the HiBind DNA Mini Column and further processed by following the manufacturer´s guidelines. The extracted genomic DNA was eluted in 50 µL of elution buffer.

The ITS2 region was targeted by using the forward primer ITS4 (5′-xCTCCGCTTATTGATATG; White et al., 1990) and the reverse primer gITS7 (5′-xGTGARTCATCGARTCTTTG; Ihrmark et al., 2012). The sample barcodes x ranged from 6-9 base pairs. The PCR reaction contained 2 µl DNA template and 23 µl master mix; 14.6 µl Milli-Q water, 2.5 µl 10x Gold buffer, 0.2 µl dNTP's (25 nM), 1.5 µl reverse and forward primers (10 µM), 2.5 µl $MgCl_2$ (50 mM), 1.0 µl BSA (20 mg/ml) and 0.2 µl AmpliTaq Gold polymerase (5 U/µl, Applied Biosystems, Thermo Fisher Scientific). For samples with low DNA concentration (weak gel bands), 5 µl DNA template and 20 µl master mix were used. The DNA was amplified by initial denaturation at 95 °C for 5 min, followed by 32 cycles of denaturation at 95 °C for 30 s, annealing at 55 °C for 30 s, and elongation at 72 °C for 1 min. A final elongation step was included at 72 °C for 10 min. PCR products were normalized by using the SequalPrep Normalization Plate Kit (Invitrogen, Thermo Fisher Scientific, Waltham, MA, USA) and eluted in 20 µL elution buffer. The resulting 345 PCR products, including technical replicates, negative PCR controls and mock

community (1 ng/µL equimolar DNA concentration from an artificial mix of *Mycena belliarum*, *Pycnoporellus fulgens*, *Serpula similis* and *Pseudoinonotus dryadeus*), were processed in a total of four metabarcoding libraries. The technical replicates included DNA from 12 dust samples and were included in each library. The 96 uniquely barcoded PCR products within each library were pooled, and the pools were concentrated and purified using Agencourt AMPure XP magnetic beads.

Bioinformatics

The raw forward and reverse sequences, were demultiplexed independently on a sample basis using CUTADAPT v 2.7 [33], allowing no mismatches between barcode tags and sequence primer, and sequences shorter than 100 bp where discarded. DADA2 [34] was used to filter low quality sequences, with a maximum expected error of 2.5 and to correct read errors based on a machine learning model built from the sequence data. We then merged the error-corrected forward and reverse sequences using a minimum overlap of 5 bp. Chimeras were filtered out using the bimera algorithm, with default parameters implemented in DADA2 v.12. The resulting 28 346 amplicon sequence variants were further clustered into operational taxonomic units (OTUs) using VSEARCH [35] at 97% similarity. LULU [36] was used with default settings to correct for potential OTU over-splitting. Taxonomy was assigned using BLAST [37] to the final OTU table using the UNITE database [38]. All the negative PCR controls and most of the negative DNA controls were automatically removed during the bioinformatics because the number of sequences was too low. The OTUs of the remaining controls were inspected to assess any contamination issues. The final dataset (excluding controls and replicates) contained 6 800 OTUs accounting for 18 694 392 reads from 292 retained samples. The number of reads per sample varied from 470 (from outdoor air during the winter) to 257 599 with a mean value of 65 365. The number of OTUs per sample varied from 3 to 1 259.

Environmental variables

Climatic variables were retrieved from The Norwegian Climate and Service Center (https://klimaservicesenter.no/, accessed March 11th, 2020), recorded by the Meteorological station at Blindern, Oslo, Norway. The daycare centers are located within a 500 m radius to the meteorological station. The climatic variables included: mean air temperature, mean dew point temperature, max air temperature, min air

temperature, mean cloud area fraction, mean water vapor partial pressure, mean surface air pressure, mean wind speed, max relative humidity, mean relative humidity, min relative humidity, humidity mixing ratio, specific humidity, snow coverage, surface snow thickness, amount of precipitation and duration of sunshine. The variables were downloaded for each week throughout the year, and averages for every two weeks prior to sampling were calculated and used for seasonal analyses. These variables were studied with principal component analyses (Fig. S4). The results indicated that the first and second dimensions explained a total of 75.6% of the variance. The first dimension was clearly correlated with variables associated with temperature while the second dimension was associated with variables related to humidity and moisture. The coordinates of dimension 1 and 2 of the PCA analyses were designated as temperature and moisture, respectively, and used as surrogate for all the above-mentioned climatic variables in downstream analyses. Season was also included as a variable, with related data averaged accordingly. The following months were grouped in four different seasons: winter from December to February, spring from March to May, summer from June to August and fall from September to November. In addition, the number of children, age of children and number of working adults (staff) having access to each daycare center and room between two sampling dates were recorded and included as variables. Continuous variables were scaled using the *scale* function in R.

Statistical analyses

All statistical analyses were performed in R version 3.6.2 [39] through RStudio (version 1.3.959) unless stated otherwise. We first confirmed the similarity of the technical replicates by nonmetric multidimensional scaling (NMDS) using the *metaMDS* function from the vegan package version 2.4-2 [40], and visualized by ggplot2 [41] (Fig. S5). Then, the complete dataset was rarefied to 1 649 sequences sample-wise, using the function *rrarefy* (vegan). This led to three samples being discarded for downstream statistical analyses, because of shallow sequencing depth in these samples. We then transformed the abundance of OTU per sample table (OTU table) into Hellinger abundance, using the *decostand* function (vegan). The community structure was analyzed using NMDS as described above. A stable solution, for NMDS, was searched with a maximum number of 200 random starts and iterations with the

convergence criteria set to stress and/or scale factor of the gradient below 1 x 10e-7, using a Bray-Curtis dissimilarity distance. The community structure was visualized using ggplot2 [41] with the axes transformed into half-change units.

The results showed a clear distinction between outdoor samples and indoor samples, with the exception of auxiliary rooms (the indoor samples belonging to rooms with very low frequency of occupancy), which showed very similar patterns to the outdoor samples. Considering that the outdoor samples were collected in a different way and time-frame and that outdoor air seasonality was not the main focus of our hypotheses, we decided to focus on the indoor space and refrained for downstream analyses regarding the outdoor samples. Since the indoor samples showed a clear segregation between auxiliary rooms and the main rooms, we decided to analyze the indoor data in two separate sets; auxiliary and main rooms together, and only the main rooms. For both datasets we rarefied all the samples to the sample with the lowest number reads in the respective dataset, 2 657 sequences for the auxiliary and main rooms dataset and 3 381 sequences in the main rooms dataset. We used the same procedure described above to analyze and visualize community composition. The function *envfit* (vegan) was used to regress the environmental variables onto the Bray-Curtis dissimilarity matrix. Significance of the regression was assessed using 999 permutations. The variables with significant effect were overlaid as vectors in the ordination (NMDS) graphic with arrows pointing in the increasing direction. In addition, we used the function *adonis2* (vegan) with 999 permutations to perform a permutational multivariate analysis of variance to assess the contribution of each environmental variable in explaining variability in the community structure. Additionally, we performed variation partitioning analysis using *varpart* (vegan) to assess the interaction and total variability explained by the following groups of variables: climate (temperature and moisture), time (month and season), space (daycare and room), and occupants (number of adults and children, and age of children).

The taxonomic compositional summary was achieved by summing all the rarefied reads, at the order level, within a sample and averaged across the time period. Richness, Shannon-Weaver and evenness indices were determined using the functions, *specnumber*, *diversity* (vegan) and $\frac{Shannon-Weaver}{\log(richness)}$, respectively. Richness and Shannon-Weaver were strongly

correlated; we therefore retained richness as a representative of alpha diversity. To estimate the effects of temperature, moisture, season, month, room, dust coverage, number of children and adults and children age on richness, we conducted linear models followed by analyses of variance as implemented by *lm* and *anova* functions in R [39]. Random forest models with permutations, as implemented in the R package rfPermute [42], with all predictor variables randomly sampled at each tree node, 500 trees and 999 permutations were applied to determine the significance and importance of each variable. In all models the squared-root of richness was used to normalize the response variable.

We further identified indicator OTUs on a monthly basis for indicator species analyses using the function *multipatt* in the R package indicspecies [43]. We then retained only OTUs with a $p < 0.05$ and present in at least three samples per month. The results were summarized by number of OTUs per order per month. The full lists of indicator OTUs of the auxiliary rooms and the main rooms are provided in Table S1 and Table S2, respectively.

Declarations

Ethics approval and consent to participate

Personal data of the occupants in the day care centers is not collected, the addresses of the day care centers are omitted from the metadata.

Consent for publication

Not applicable.

Availability of data and materials

All sequencing data, metadata, and scripts used in the bioinformatic analyses and statistics are deposited in Dryad Digital Repository (http://dx.doi.org/) and will be made available upon publication or earlier by request.

Competing interests

The authors declare no competing interests.

Funding

The research was financially supported by the University of Oslo and the Norwegian Asthma and Allergy Association (NAAF). PMMS was funded by the European Union's Horizon 2020 research and innovation programme (Marie Skłodowska-Curie Individual Fellowship; grant agreement MycoIndoor No 741332).

Authors´ contributions

ELFE, IBE, HK and IS planned and initiated the study, ELFE and IS collected

the dust samples, ELFE did all lab and bioinformatic analyses, LM did the statistical analyses. All authors drafted the manuscript, discussed the results and agreed about the final version.

Acknowledgements

We would like to thank the daycare centers for allowing us to perform sampling and for providing metadata of the usage of the daycare centers. Mycoteam AS contributed to the sampling and provided sampling equipment.

References

1. Bornehag CG, Blomquist G, Gyntelberg F, Jarvholm B, Malmberg P, Nordvall L, Nielsen A, Pershagen G, Sundell J: Dampness in buildings and health. Nordic interdisciplinary review of the scientific evidence on associations between exposure to "dampness" in buildings and health effects (NORDDAMP). Indoor air 2001, 11:72-86.

2. Mendell MJ, Mirer AG, Cheung K, Tong M, Douwes J: Respiratory and allergic health effects of dampness, mold, and dampness-related agents: A review of the Epidemiologic Evidence. Environ Health Pers 2011, 119:748-756.

3. Adams RI, Bhangar S, Dannemiller KC, Eisen JA, Fierer N, Gilbert JA, Green JL, Marr LC, Miller SL, Siegel JA et al: Ten questions concerning the microbiomes of buildings. Building Environ 2016, 109:224-234.

4. Horve PF, Lloyd S, Mhuireach GA, Dietz L, Fretz M, MacCrone G, Van Den Wymelenberg K, Ishaq SL: Building upon current knowledge and techniques of indoor microbiology to construct the next era of theory into microorganisms, health, and the built environment. J Expo Sci Environ Epidemiol 2020, 30:219-235.

5. Barberán A, Dunn RR, Reich BJ, Pacifici K, Laber EB, Menninger HL, Morton JM, Henley JB, Leff JW, Miller SL et al: The ecology of microscopic life in household dust. Proc Biol Sci 2015, 282.

6. Lax S, Smith DP, Hampton-Marcell J, Owens SM, Handley KM, Scott NM, Gibbons SM, Larsen P, Shogan BD, Weiss S et al: Longitudinal analysis of microbial interaction between humans and the indoor environment. Science 2014, 345:1048-1052.

7. Adams RI, Miletto M, Taylor JW, Bruns TD: Dispersal in microbes: fungi in indoor air are dominated by outdoor air and show dispersal limitation at short distances. ISME 2013, 7:1262-1273.

8. Amend AS, Seifert KA, Samson R, Bruns TD: Indoor fungal composition is geographically patterned and more diverse in temperate zones than in the tropics. PNAS 2010, 107:13748-13753.

9. Barberan A, Ladau J, Leff JW, Pollard KS, Menninger HL, Dunn RR, Fierer N: Continental-scale distributions of dust-associated bacteria and fungi. PNAS 2015, 112:5756-5761.

10. Martin-Sanchez PM, Estensmo ELF, Morgado LN, Maurice S, Engh IB, Skrede I, Kauserud H: Analyzing indoor mycobiomes through a large-scale citizen science study of houses from Norway. Research Square 2020.

11. Prussin AJ, Marr LC: Sources of airborne microorganisms in the built environment. Microbiome 2015, 3:78.

12. Adams RI, Miletto M, Lindow SE, Taylor JW, Bruns TD: Airborne bacterial communities in residences: Similarities and differences with fungi. PloS one 2014, 9:e91283.

13. Hegarty B, Haverinen-Shaughnessy U, Shaughnessy RJ, Peccia J: Spatial gradients of fungal abundance and ecology throughout a damp building. Environ Sci Technol Lett 2019, 6:329-333.

14. Yamamoto N, Hospodsky D, Dannemiller KC, Nazaroff WW, Peccia J: Indoor emissions as a primary source of airborne allergenic fungal particles in classrooms. Environ Sci Technol 2015, 49:5098-5106.

15. Nevalainen A, Täubel M, Hyvärinen A: Indoor fungi: companions and contaminants. Indoor air 2015, 25:125-156.

16. Garrett MH, Rayment PR, Hooper MA, Abramson MJ, Hooper BM: Indoor airborne fungal spores, house dampness and associations with environmental factors and respiratory health in children. Clin Exp Allergy 1998, 28:459-467.

17. Frankel M, Bekö G, Timm M, Gustavsen S, Hansen EW, Madsen AM: Seasonal variations of indoor microbial exposures and their relation to temperature, relative humidity, and air exchange rate. Appl Environ Microbiol 2012, 78:8289-8297.

18. Su HJ, Wu PC, Chen HL, Lee FC, Lin LL: Exposure assessment of indoor allergens, endotoxin, and airborne fungi for homes in southern Taiwan. Environ Res 2001, 85:135-144.

19. Weikl F, Tischer C, Probst AJ, Heinrich J, Markevych I, Jochner S, Pritsch K: Fungal and bacterial communities in indoor dust follow different environmental determinants. PloS one 2016, 11:e0154131.

20. Karlsson E, Johansson A-M, Ahlinder J, Lundkvist MJ, Singh NJ, Brodin T, Forsman M, Stenberg P: Airborne microbial biodiversity and seasonality in Northern and Southern Sweden. PeerJ 2020, 8:e8424.

21. Peay KG, Bruns TD: Spore dispersal of basidiomycete fungi at the landscape scale is driven by stochastic and deterministic processes and generates variability in plant–fungal interactions. New Phytol 2014, 204:180-191.

22. Chen W, Hambleton S, Seifert KA, Carisse O, Diarra MS, Peters RD, Lowe C, Chapados JT, Lévesque CA: Assessing performance of spore samplers in monitoring aeromycobiota and fungal plant pathogen diversity in Canada. Appl Environ Microbiol 2018, 84:e02601-02617.

23. Madureira J, Paciência I, Rufo JC, Pereira C, Teixeira JP, de Oliveira Fernandes E: Assessment and determinants of airborne bacterial and fungal concentrations in

different indoor environments: Homes, child day-care centres, primary schools and elderly care centres. Atmos Environ 2015, 109:139-146.

24. Frankel M, Bekö G, Timm M, Gustavsen S, Hansen EW, Madsen AM: Seasonal variations of indoor microbial exposures and their relation to temperature, relative humidity, and air exchange rate. Appl Environ Microbiol 2012, 78:8289.

25. Pitkäranta M, Meklin T, Hyvärinen A, Paulin L, Auvinen P, Nevalainen A, Rintala H: Analysis of fungal flora in indoor dust by ribosomal DNA sequence analysis, quantitative PCR, and culture. Appl Environ Microbiol 2008, 74:233-244.

26. Hanson B, Zhou Y, Bautista EJ, Urch B, Speck M, Silverman F, Muilenberg M, Phipatanakul W, Weinstock G, Sodergren E et al: Characterization of the bacterial and fungal microbiome in indoor dust and outdoor air samples: a pilot study. Environ Sci Process Impacts 2016, 18:713-724.

27. Hanson B, Zhou Y, Bautista EJ, Urch B, Speck M, Silverman F, Muilenberg M, Phipatanakul W, Weinstock G, Sodergren E et al: Characterization of the bacterial and fungal microbiome in indoor dust and outdoor air samples: a pilot study. Environ Sci Process Impacts 2016, 18:713-724.

28. Crous PW, Schoch CL, Hyde KD, Wood AR, Gueidan C, de Hoog GS, Groenewald JZ: Phylogenetic lineages in the *Capnodiales*. Stud Mycol 2009, 64:17-47S17.

29. Reponen T, Nevalainen A, Jantunen M, Pellikka M, Kalliokoski P: Normal range criteria for indoor air bacteria and fungal spores in a subarctic climate. Indoor air 1992, 2:26-31.

30. Aydogdu H, Asan A: Airborne fungi in child day care centers in Edirne City, Turkey. Environ Monit Assess 2008, 147:423-444.

31. Fairs A, Wardlaw A, Thompson J, Pashley CH: Guidelines on ambient intramural airborne fungal spores. J Investig Allergol Clin Immunol 2010, 20:490-498.

32. de Ana SG, Torres-Rodríguez JM, Ramírez EA, García SM, Belmonte-Soler J: Seasonal distribution of *Alternaria, Aspergillus, Cladosporium* and *Penicillium* species isolated in homes of fungal allergic patients. J Investig Allergol Clin Immunol 2006, 16:357-363.

33. Martin M: Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet 2011, 17:3.

34. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJ, Holmes SP: DADA2: High-resolution sample inference from Illumina amplicon data. Nat Methods 2016, 13:581-583.

35. Rognes T, Flouri T, Nichols B, Quince C, Mahé F: VSEARCH: a versatile open source tool for metagenomics. PeerJ Preprints 2016, 4:e2409v2401.

36. Frøslev TG, Kjøller R, Bruun HH, Ejrnæs R, Brunbjerg AK, Pietroni C, Hansen AJ: Algorithm for post-clustering curation of DNA amplicon data yields reliable biodiversity estimates. Nat Comm 2017, 8:1188.

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

38. Koljalg U, Larsson KH, Abarenkov K, Nilsson RH, Alexander IJ, Eberhardt U, Erland S, Hoiland K, Kjoller R, Larsson E et al: UNITE: A database providing web-based methods for the molecular identification of ectomycorrhizal fungi. New Phytol 2005, 166:1063-1068.

39. R Core Team: R: A language and environment for statistical computing. R Foundation for Statistical Computing. Austria: Vienna. In. https://www.R-project.org/; 2018.

40. Oksanen J, Blanchet FG, Friendly M, Kindt R, Legendre P, McGlinn D, Minchin PR, O'hara R, Simpson GL, Solymos P: Vegan: Community ecology package. R package version 2.4-2. Vienna: R Foundation for Statistical Computing; 2016.

41. Wickham H: ggplot2: elegant graphics for data analysis: Springer; 2016.

42. Archer E: rfPermute: Estimate permutation p-values for random forest importance metrics. In., R package version 2.1.81 edn. https://CRAN.R-project.org/package=rfPermute; 2020.

43. De Cáceres M, Sol D, Lapiedra O, Legendre P: A framework for estimating niche metrics using the resemblance between qualitative resources. Oikos 2011, 120:1341-1350.

Supplementary information

Supplementary tables

Supplementary table 1. The complete list of indicator species OTUs detected for the different months in the auxiliary rooms of a daycare center in Oslo, Norway.

Excel table “Supplementary table 1”. Available for download on Dryad.

Supplementary table 2. The complete list of indicator species OTUs detected for the different months in the main rooms of two daycare centers in Oslo.

Excel table “Supplementary table 2”. Available for download on Dryad.

Supplementary figures

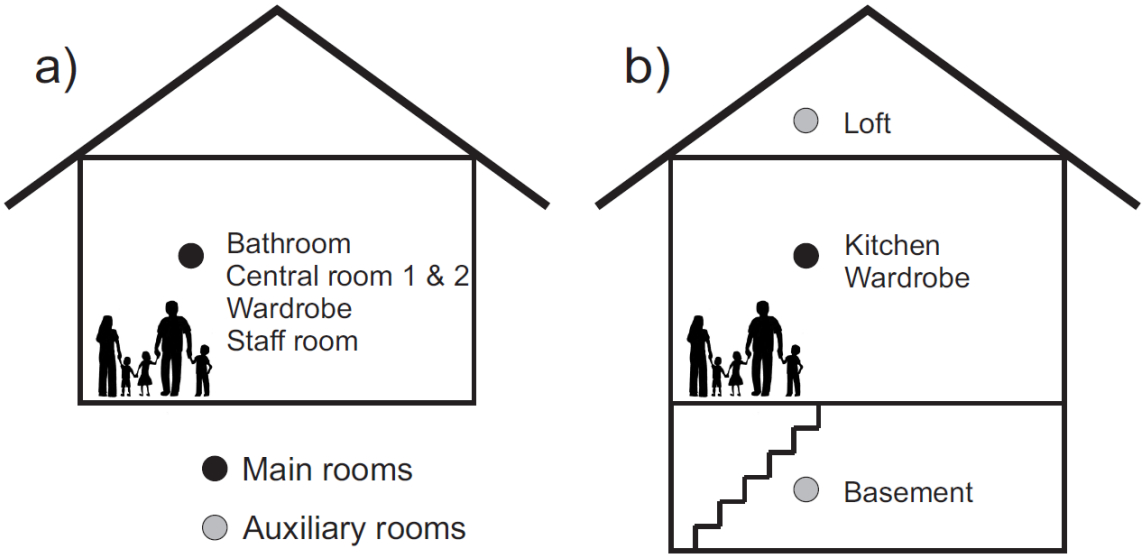


Fig. S1. Overview of different rooms sampled in two daycare centers in Oslo, Norway. In daycare a, only the main rooms with presence of children and adults (parents and staff) were sampled. Daycare b had in addition a loft and a basement (auxiliary rooms), which were sampled besides the main rooms. The auxiliary rooms have limited presence of staff, and no presence of children nor parents.

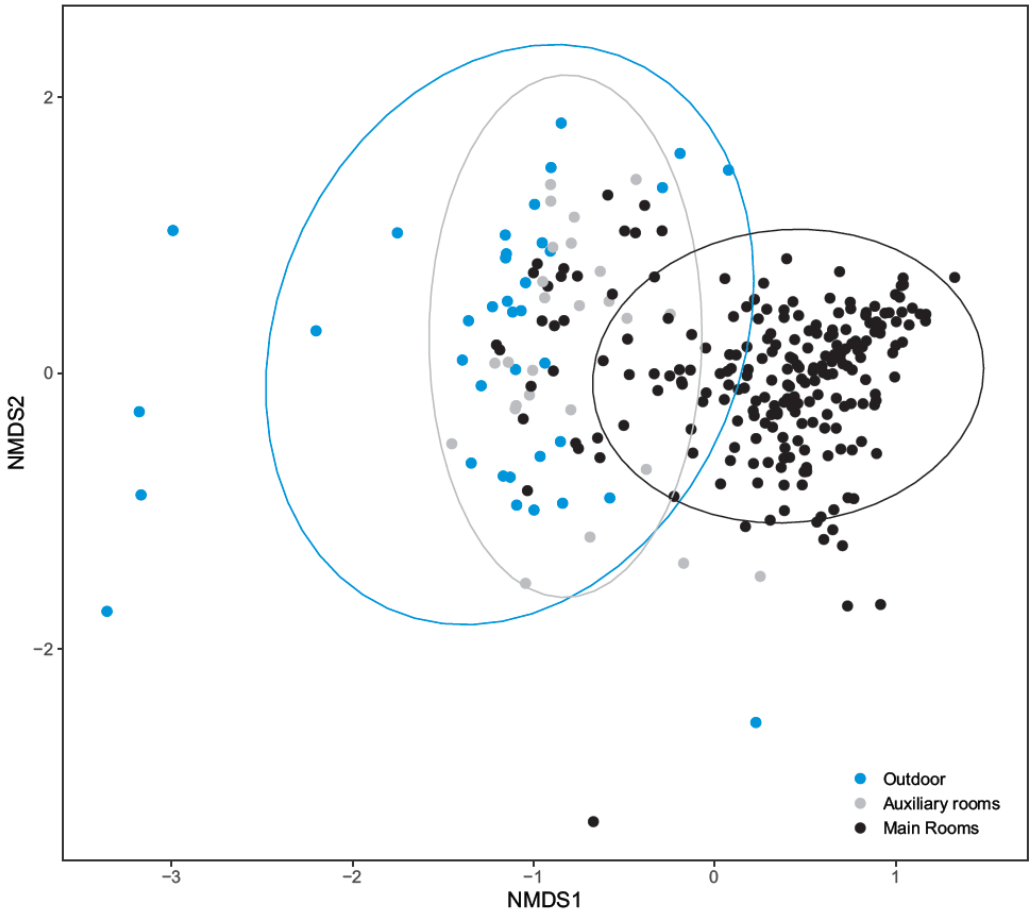


Fig. S2. Nonmetric multidimensional scaling (NMDS) ordination plot of fungal composition from outdoors air samples and indoor dust samples from different room types (main and auxiliary) of two daycare centers in Oslo, Norway sampled throughout a year. Each point represents one sample, and the color separates the samples from the outdoor and the auxiliary and main rooms.

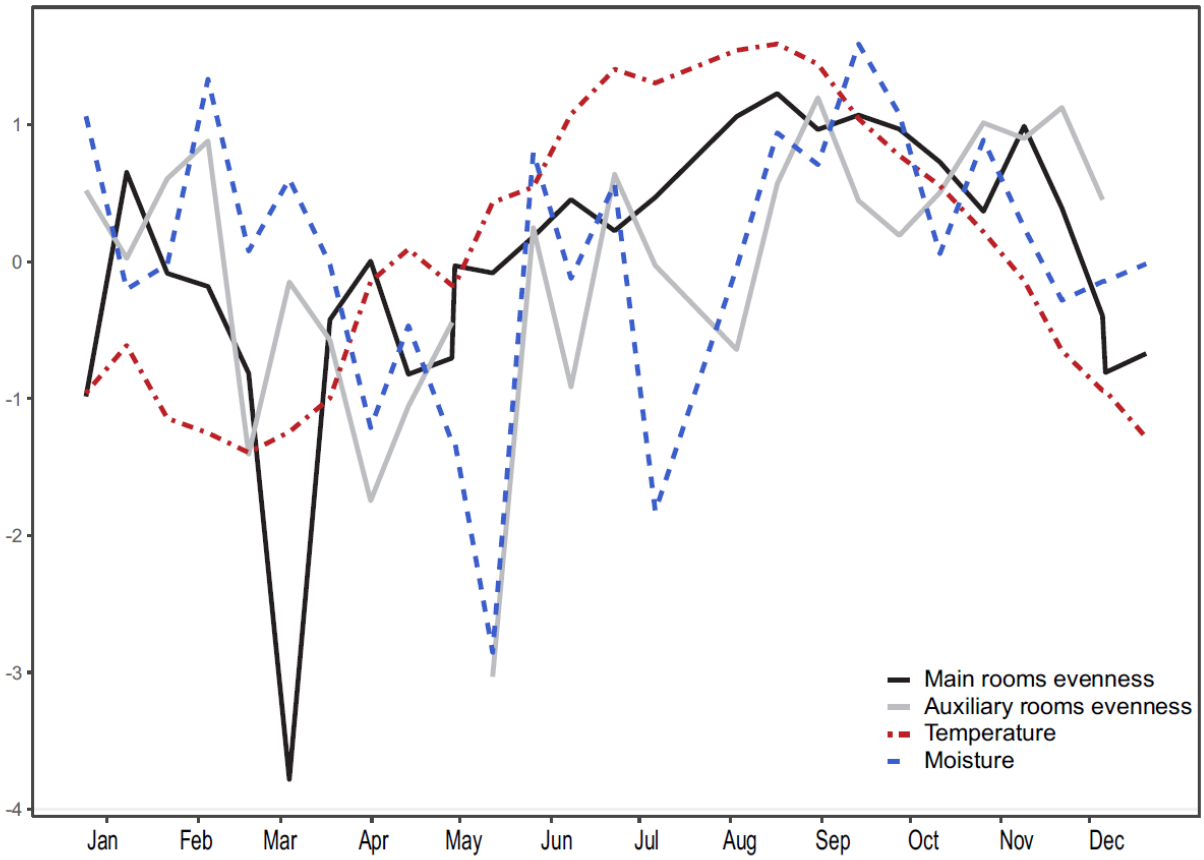


Fig. S3. Monthly evenness time-series of the dust samples in the main rooms (black line) and the auxiliary rooms (grey line) sampled in two daycare centers in Oslo, Norway. The dotted lines represent the monthly fluctuation of temperature (red) and moisture (blue). The gap in auxiliary rooms' evenness in May is due to samples excluded from the analysis because of low number of reads.

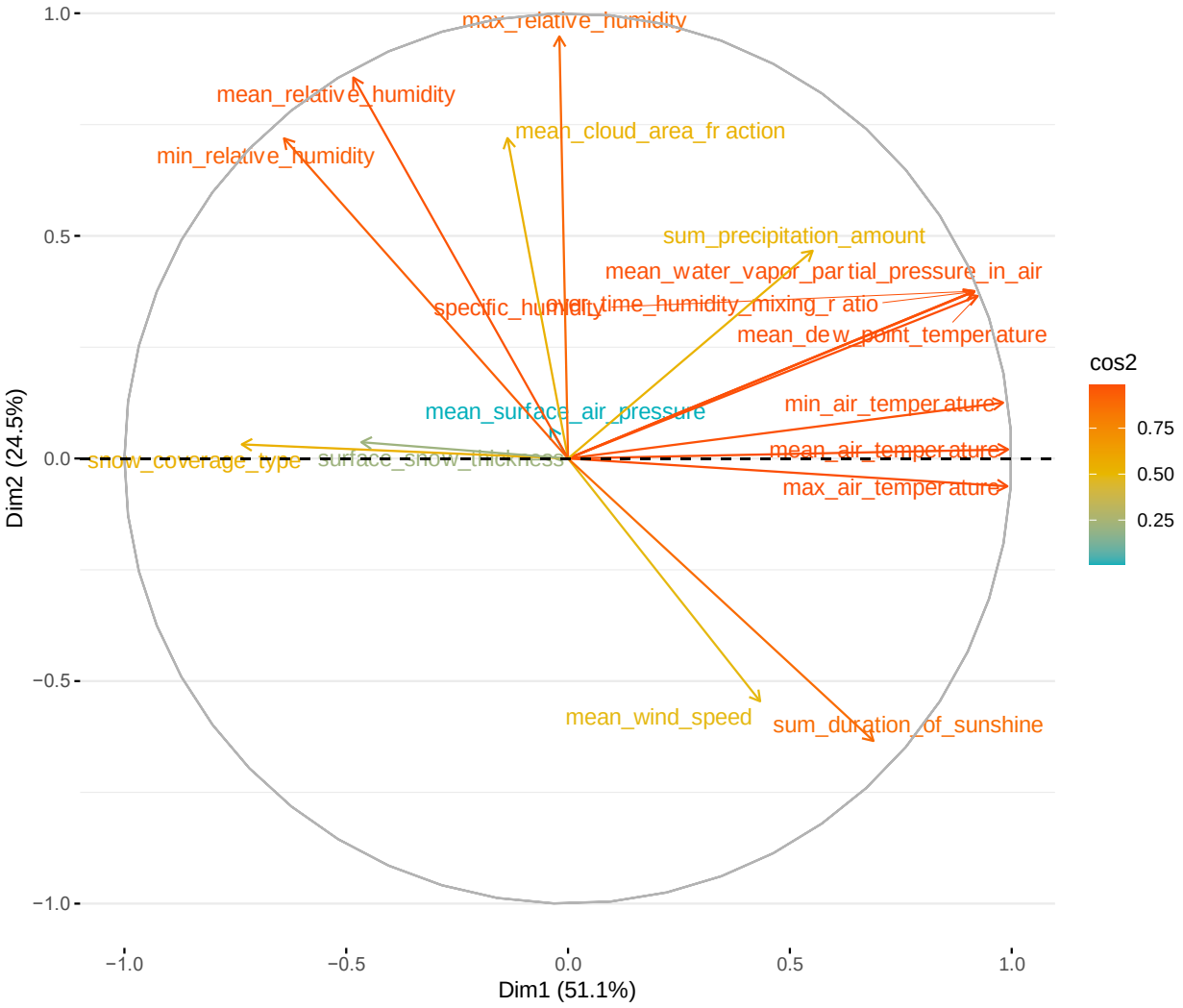


Fig. S4. Principal Component Analysis (PCA; Axes 1 and 2) for data from the meteorological numerical variables explored in this study. Data were recorded by the meteorological station at Blinder, located within 1 km^2 of both daycare centers in Oslo, Norway.

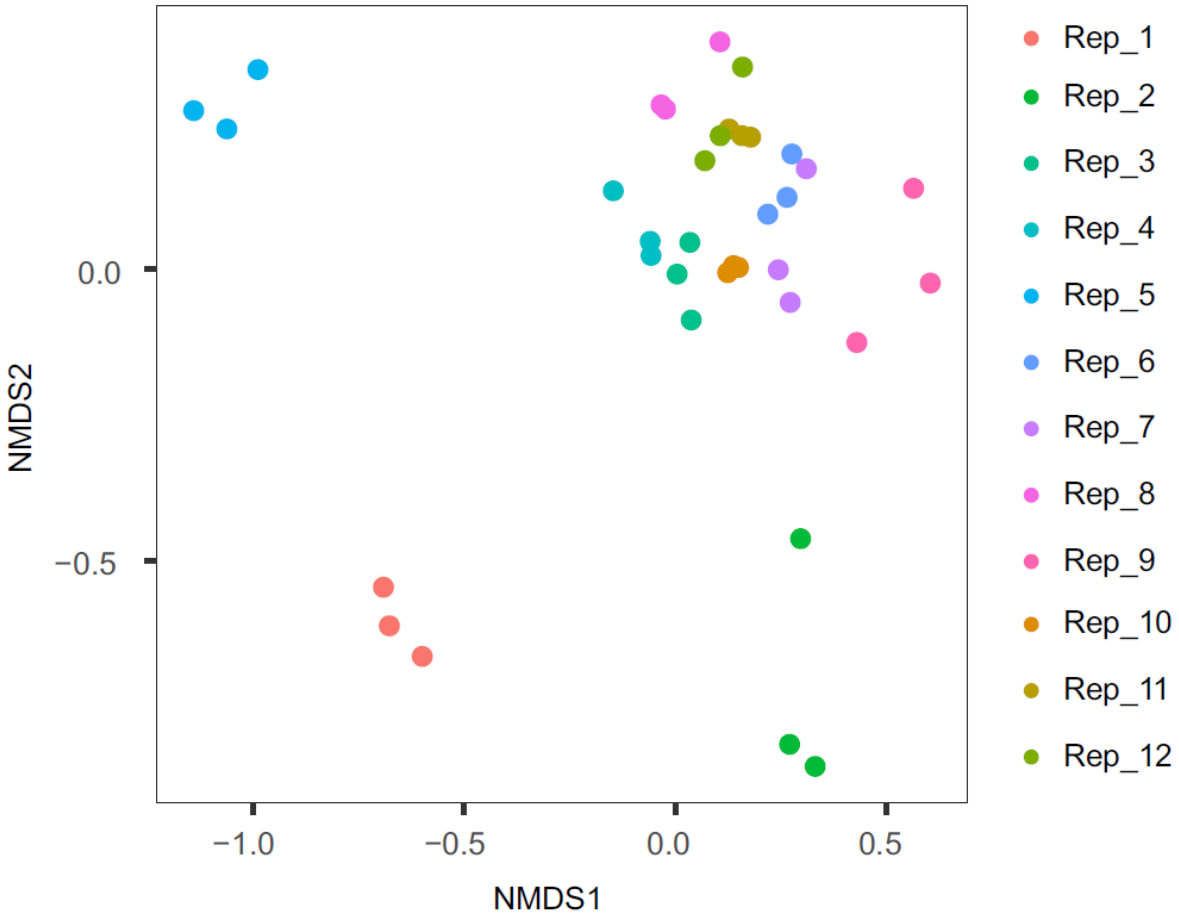


Fig. S5. Nonmetric multidimensional scaling (NMDS) ordination plot of technical PCR replicates included in the library preparation and sequencing. Each point represents one sample, and the color indicates the different replicates.