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). Damppness- 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; Kettleison 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...
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 (Mycotope2; 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 km²) 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/ricb), 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′-GTGARTCATCAGCTTTTGTGC-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 (Carlson 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 MgCl₂, 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 SequaPrep 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 dada2 version 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 (singleton) 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 rarefy, 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 ggplot2 of the r package ggordiplot version 0.3.0 (Quensen, 2018) to visualize their association with the dust mycobionmes. 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...
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 coldest (BIO19) 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 3b).

**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 (*; extracted from Table 2). Ascomycota are indicated in black and Basidiomycota in red.
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% 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).

#### 3.4 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
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

### 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    | Climate   | 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              | Climate   | 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) | Occupants | 1.20* | 2.19 | 1.76* |
| Allergies             | 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                  | 0.87* | 1.30 | 1.46* |
| Building type         | 0.85* | 1.54* | 1.24* |
| Building material      | Building  | 0.70* | 1.21 | 1.03* |
| Bedrock nutrient (NGU) | Building  | 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 ($n = 541$).
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 allergic 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
| OTU ID | Phylum        | Order             | Genus            | Traits | RA (%) | Occurrence (%) | IndVal (%) |
|--------|---------------|-------------------|------------------|--------|--------|----------------|------------|
| OTU3   | Ascomycota    | Saccharomycetales| Saccharomyces    | O, A, S| 5.42   | 95.6           | 91.2       |
| OTU4   | Ascomycota    | Eurotiales        | Penicillium      | O, A, S,M| 6.23   | 99.3           | 89.5       |
| OTU18  | 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  | Ascomycota    | Saccharomycetales| Debaryomyces     | O, S   | 0.87   | 94.8           | 83.2       |
| OTU34  | 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       |        |        |                |            |
| OTU91  | Ascomycota    | Eurotiales        | Penicillium      | O, A, S,M| 0.22   | 71.2           | 70.5       |
| 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          |        |        |                |            |
| 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       |

(Continues)
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 (Ctenolepis 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 (R² 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 R² 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 (Table 2) and 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 Capnodiiales and Eurotiales as major orders in abundance, corroborating our hypothesis H2. These orders are well known for their stress tolerance; Capnodiiales (with *Cladosporium* 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 *Rhotorula* 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 *Rhotorula* 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.59zw324w (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
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;30:2689–2705. https://doi.org/10.1111/mec.15916