Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Analyzing indoor mycobiomes through a large-scale citizen science study of houses from Norway

Pedro M. Martin-Sanchez (p.m.martin-sanchez@ibv.uio.no)
University of Oslo https://orcid.org/0000-0003-2674-8650
Eva-Lena F. Estensmo
University of Oslo
Luis N. Morgado
University of Oslo
Sundy Maurice
University of Oslo
Ingeborg B. Engh
Mycoteam AS
Inger Skrede
University of Oslo
Håvard Kauserud
University of Oslo

Research

Keywords: Fungi, built environment, buildings, indoor and outdoor environments, dust, ITS2 metabarcoding, climate, occupants, indicator species

Posted Date: July 7th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-40337/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License

Abstract

Background: In the built environment, fungi can cause important deterioration of building materials and have adverse health effects on occupants. We here investigate what are the most important environmental variables (climate, building features or occupant characteristics) driving the indoor mycobiome structure and diversity in houses. We used Norway as study area since it spans a wide range of environmental gradients such as temperature and precipitation. We organized a large citizen science sampling campaign including 271 houses and 807 dust samples collected from three different house compartments: outside of the building, living room and bathroom. Fungal community composition and diversity were determined by DNA metabarcoding, and correlated to the environmental variables by multivariate statistical analyses.

Results: The fungal community composition was clearly different between indoor and outdoor samples, but there were no significant differences between the two indoor compartments (i.e living room vs. bathroom). 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 most important factor (7.6%). Next, the regional-scale 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 mainly dominated by ascomycetes (≥70%) with *Capnodiales* and *Eurotiales* as the most abundant orders. Compared to the outdoor, 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 molds (mainly *Penicillium* and *Aspergillus*) and skin-associated yeasts (*Malassezia*, *Debaryomyces*, *Candida* and *Rhodotorula*).

Conclusions: We show that indoor mycobiomes represent a mixture of fungi from indoor and outdoor sources, and that a multitude of indoor and outdoor variables structure indoor mycobiomes. We also demonstrate that citizen science is a successful and effective strategy to obtain samples for characterizing the indoor microbiomes at large geographical scales.

Background

Throughout the world, people spend a major part of their lifetime in the built environment, including houses, workplaces, kindergartens and schools. Buildings harbor unique and complex microbial assemblages (fungi, bacteria, archaea and viruses) whose ecological role and impact on human health remain largely unknown [1].

Fungi, one of the most diverse kingdoms of life, with essential ecosystem functions [2], have expanded their niche into the built environment. 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 the ubiquitous airborne mold 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 [3–5]. Dampness- and mold-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 [6, 7].

Microbiological assessments in the built environment mainly focus on air and dust samples indicative of human exposure indoors. The fungal content of these samples can be analyzed using different approaches: microscopy, culturing, chemical and DNA-based analyses [5]. Considering the well-known limitations of culture-based methods, a

shift toward DNA-based methods has taken place in the last decades. High-throughput sequencing (HTS) of amplified markers (DNA metabarcoding) has lately become a key tool for surveying fungal communities in environmental samples [8, 9]. In the last decade, many studies have used DNA metabarcoding to reveal the microbiome of residential buildings in different parts of the world [1], mainly focusing on bacteria [10, 11] but also on fungi [12, 13].

The indoor mycobiome is primarily determined by large-scale environmental gradients, like climate, but local environmental variation within individual buildings, resulting from differences in construction features and building function, can also contribute in shaping the fungal diversity and composition [1, 14, 15]. A first global survey analyzing 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 significant effect [16].

Both culture- and DNA-based studies have demonstrated that outdoor air is the main source of indoor fungi. Adams et al. [12, 17] 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. [18, 19] analyzed dust microbiomes collected inside and outside 1,200 houses across the United States (US) and confirmed that most indoor fungi were derived from outside. They further identified geographic patterns in the indoor mycobiomes that could be explained by climate and soil variables. However, a variety of internal secondary sources must be considered as well, e.g. organic materials (food, waste and potted plants), certain surfaces (drains, windowsills and carpets) and occupants (humans and pets) [3–5, 17, 20]. DNA-based dust studies indicated that various building features and occupant characteristics were also key determinants of the indoor mycobiome [21, 22]. In this regard, Yamamoto et al. [23] 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 largely determined by outdoor environment, while bacteria are to a larger degree determined by the occupants and their activities [1, 10, 14, 15, 18].

Except for the pioneering global study by Amend et al [16], and the continental-scale study across the US by Barberán et al. [18, 19], the majority of existing DNA-based mycobiome studies have focused on specific building units at local scale. A few regional studies have also targeted some big cities, like Munich [24] and Hong Kong [13]. 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 thus constitute the basis for managing indoor air quality and identifying health risks.

Our study area, Norway, possess marked climatic and environmental gradients, making us able to assess to what degree the outdoor environment, versus 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 to 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 human and building microbiomes [18, 25].

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? Related 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.

Results

Citizen scientists were recruited through scientific networks and diverse actions in social and public media. The volunteers provided dust samples and information (metadata) about their houses (Fig. 1a). Following our instructions, dust samples were swabbed from the upper doorframes of three house compartments: outside of the building (main entrance), living room and bathroom. Doorframes act as passive collectors of dust deposited during an unknown amount of time. The dust coverage in the living room was measured in parallel, using one adhesive tape (Mycotape2; Additional file 1: Fig. S1), and included as a variable. To minimize the influence of seasonality effects, all samples were collected in a short time span during spring 2018, mainly in May. In addition to the building features and occupant characteristics provided by the volunteers, variables related to the outdoor environment (climate, topography, geology and land use), specific to each house, were obtained. In total 269 houses were sampled from mainland Norway, covering its major climatic gradients (Fig. 1b). Two houses from Longyearbyen, in the Arctic Archipelago of Svalbard, were also included. To characterize the dust mycobiomes in houses across Norway, we carried out DNA metabarcoding analyses of all collected samples sequenced on Illumina MiSeq platform after amplification of the internal transcribed spacer 2 (ITS2) region.

Data features and overall fungal diversity

After quality filtering, denoising and sequence clustering, the final dataset contained 7,110 operational taxonomic units (OTUs) (22,622,391 reads), distributed among 810 dust samples from 271 houses. The number of reads per sample varied widely, from 1,367 to 245,588 (mean = 27,929), and the number of OTUs per sample (richness) ranged from 7 to 867 (mean = 270) (Additional file 1: Fig. S2). Likewise, the abundance of OTUs varied extensively, from 2 to 2,040,802 reads per OTU (mean = 3,182), while their occurrences ranged from 1 to 807 dust samples (mean = 31). For further statistical analyses, we resampled the data to a relatively low sequencing depth (2,000 reads per sample) in order to keep the majority of samples (three samples were excluded) and houses, representative of a wide geographic area. The rarefied dataset contained 6,632 OTUs distributed across 807 samples.

The average richness and diversity (Shannon) per sample were significantly higher in the indoor compared to outdoor samples (Fig. 2a). The evenness as well as other diversity indices (Simpson and Chao1) followed a similar trend (Additional file 1: Fig. S3). The two indoor compartments, i.e. 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 generalized linear model (GLM), (p = 7.48e-05). Several other variables (mean temperatures of 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 ($p < 0.05$). However, taken together with the strongest predictor (indoor vs. outdoor), they showed 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 (Additional file 1: Fig. S4). In contrast to richness and diversity, the turnover in species composition (beta diversity) was higher among the outdoor samples (Fig. 2b).

Factors driving the community composition

We observed a marked compositional difference between indoor and outdoor mycobiomes, as revealed by NMDS ordination of all dust samples (Fig. 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; Fig. 4). While, there was no difference between the two indoor compartments, living room versus bathroom (Fig. 3a). Noteworthy, only a relatively low proportion of the variation in fungal community composition was explained by the assessed factors (Table 1), altogether accounting for about 15% of the variation (Fig. 4). Climatic variables were, next to house compartment, also of importance for the fungal community composition in the dust samples; as seen in the ordination plot. Various climatic variables were correlated with the second ordination axis (Fig. 3b, Additional file 1: Table S2). Together, climatic variables accounted for 4.18% of the variation among all dust samples, which increased to 5.54% for the outdoor samples when analyzed separately (Fig. 4). The four most important climatic variables were annual temperature variation (Temperature seasonality; BIO4), mean temperature of warmest (BIO10) and driest (BIO9) quarter, as well as annual precipitation (BIO12) (Table 1). There was a clear geographic 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 (Additional file 1: Fig. S5a and c, Table S2), which again relate to the regional climate variability in the study area (Fig. 1). Building features and occupant characteristics could only account for 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 analyzed on indoor samples exclusively (Fig. 4, Table 1). Interestingly, the more occupants staying in the houses, the more similar the indoor samples were to outdoor samples in fungal community composition (Fig. 3b).

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

Variables (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 warmest quarter (BIO10)	Climate	1.82 ***	2.73 ***	2.22 ***
Mean temperature of 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 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 ***
# of children	Occupants	0.56 ***	0.46	1.02 ***
# of people	Occupants	0.55 ***	0.43	1.03 ***

[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 datasets: All samples (n = 807), Outdoor (n = 266) and Indoor (n = 541). Asterisks indicate significance levels: $p < 0.001$ (***), $p < 0.01$ (**) and $p < 0.05$ (*).

Variables (source) [a]	VPA group	All samples R^2 (%) [b]	Outdoor R^2 (%) [b]	Indoor R^2 (%) [b]
Urban/rural area		0.50 ***	0.90 ***	0.64 ***
# 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 *

[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 datasets: All samples (n = 807), Outdoor (n = 266) and Indoor (n = 541). Asterisks indicate significance levels: $p < 0.001$ (***), $p < 0.01$ (**) and $p < 0.05$ (*).

n/a: not applicable.

Dominant fungi in house dust

The taxonomic assignment for the most abundant fungi detected in dust samples is shown in Fig. 5 and Table S1 (Additional file 2: Top-200 most abundant OTUs). 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% (Fig. 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 Fig. 3c, there was a broad-scale structuring of the major taxonomic groups. Both for the indoor and outdoor samples, ascomycetes were in general more associated with areas with higher precipitation, lower mean temperature of warmest quarter and low degree of seasonality in temperature, while the basidiomycetes show an opposite pattern (Fig. 3b and c). Hence, basidiomycetes were to a larger extent associated with more continental climates (warm summers, high annual temperature variation and low precipitation) and the opposite for the ascomycetes.

Looking at the taxonomic composition at the order level, there were some marked differences between indoor and outdoor samples (Fig. 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 occurred 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 (Fig. 5c): *Penicillium*, *Aspergillus*, *Saccharomyces*, *Malassezia* and *Botrytis* were far more abundant indoors, while *Cladosporium*, *Thekopsora*, *Verrucocladosporium*, *Scoliciosporum* and *Hypogymnia* were more abundant outside the buildings. *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 total sequences; Fig. 3d). In addition, the yeast genera *Malassezia* and *Aureobasidium* were particularly abundant in bathrooms.

Indoor versus outdoor mycobiomes

A large proportion of the fungi (36.3% OTUs) were present in all three house compartments, and 50.6% of the OTUs were shared between indoor and outdoor compartments (Fig. 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 (Fig. 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 (Fig. 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 ($p > 0.05$).

The indicator species analysis revealed 241 OTUs (3.6% of the total number of OTUs) significantly ($p < 0.05$) associated with indoor environments, while 550 OTUs (8.3%) were associated with outdoor environments. In line with the taxonomic results (Fig. 5), there were many indoor indicator OTUs in *Eurotiales* (16.9% of indicator OTUs) and *Agaricales* (15.5%), and numerous outdoor indicator OTUs in *Lecanorales* (16.5%), *Chaetothyriales* (16.5%) and *Capnodiales* (13.4%) (Additional file 1: Fig. S6). OTUs with the highest indicator values (IndVal > 50%) for indoor and outdoor environments are detailed in Tables 2 and 3, respectively. 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. When comparing the two indoor compartments, 46 OTUs (0.79%) were significantly ($p < 0.05$) associated with living rooms, and the same number of indicator species for bathrooms. The best indicator fungi for living rooms were OTU498 *Diatrype* sp. and OTU853 *Gibberella* sp. with the IndVal 36.2% and 34.9%, respectively. The best indicator fungi for bathrooms were OTU189 *Malassezia* sp., OTU111 *Ramularia* sp. and OTU288 *Cryptococcus* sp. with the IndVal 52.1%, 50.1% and 32.6%, respectively.

Table 2
Indoor indicator species (IndVal > 50%; $p < 0.05$).

OTU ID	Phylum	Order	Genus	Trait[a]	RA[b] (%)	Occurrence[c] (% houses)	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*		0.51	79.3	73.8
OTU91	*Ascomycota*	*Eurotiales*	*Penicillium*	O, A, S, M	0.22	71.2	70.5
OTU26*	*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

* Top-20 most abundant fungi, also detailed in the ordination plot for OTUs (Fig. 3c and d).

[a] Relevant traits: osmotolerant fungi (O) [3, 26], allergenic fungi (A) [27, 28], skin-associated fungi (S) [29], and material-colonizing fungi (M) [3, 30].

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

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

OTU ID	Phylum	Order	Genus	Trait[a]	RA[b] (%)	Occurrence[c] (% houses)	IndVal (%)
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

* Top-20 most abundant fungi, also detailed in the ordination plot for OTUs (Fig. 3c and d).

[a] Relevant traits: osmotolerant fungi (O) [3, 26], allergenic fungi (A) [27, 28], skin-associated fungi (S) [29], and material-colonizing fungi (M) [3, 30].

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

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

Table 3
Outdoor indicator species (IndVal > 50%; $p < 0.05$).

OTU ID	Phylum	Order	Genus	Trait[a]	RA[b] (%)	Occurrence[c] (% houses)	IndVal (%)
OTU2*	*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
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

* Top-20 most abundant fungi, also detailed in the ordination plot for OTUs (Fig. 3c and d).

[a] Relevant traits: plant pathogenic fungi (P), rock-inhabiting fungi (R) and lichen-forming fungi (L) [26, 31].

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

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

Discussion

Determinants for the indoor dust mycobiome

This is the first study about indoor mycobiomes at a large geographical scale in Europe. We analyzed dust samples from 271 houses throughout Norway to investigate which factors are shaping their mycobiomes. From other studies [12, 18], it is known that outdoor air is a major source for the indoor mycobiomes, otherwise, there is little knowledge on contributing intrinsic and extrinsic factors. We therefore inferred metadata related to outdoor conditions, building features and occupant characteristics to identify the drivers of fungal diversity and composition. The relatively low proportion (15%) of the overall community composition variation explained by the assessed variables, is a common trend in most fungal community studies, as fungal communities to a large degree are assembled through stochastic processes related to spore spread [32]. Although accounting for a small part of the explained variation, the PERMANOVA revealed that all variables were significantly correlated with the community composition. Remarkably, for the indoor samples, all variables showed slightly higher contribution than for the combined indoor and outdoor dataset. As expected, the contribution of the climatic variables, as well as those reflecting other outdoor characteristics (geography, topography and land use), was considerably higher for the outdoor samples.

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) 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 [12, 15, 16, 18, 19]. Amend et al. [16] first suggested that large-scale (extrinsic) factors are driving the fungal composition in buildings, rather than specific building features. Likewise, Barberán et al. [18] 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 5% explained by building features. They also identified continental-scale geographic patterns explained by climate, soil and vegetation variables [19]. In our study, the climatic variables temperature seasonality, mean temperature of warmest and driest quarters, showed better explanatory power than annual precipitation and annual mean temperature. Likewise, a recent meta-study, focused on soil mycobiome data across the world (36 studies covering 3370 soil samples), identified climate as the key driver of the global distribution of common fungi, as well as fungal diversity and community composition [33]. They reported the mean temperature of driest quarter as the most important variable explaining the biography of the most frequent fungi.

Although with limited impact, all building features significantly affected the indoor mycobiome. According to the PERMANOVA, the presence of pests was the most relevant building factor explaining 1.99% of variation for 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 during the last years and is considered as a major nuisance pest in modern buildings of Norway [34]. Madden et al. [35] reported that arthropod and microbial (fungi and bacteria) diversities follow parallel trends in settled-dust samples. Other building factors studied (type, material, ventilation, construction year and moisture-related problems) also explained smaller fractions of the indoor mycobiome variation (R^2 values between 1.24% – 0.44%). Previous studies have identified similar building determinants: For example, Dannemiller et al. [22] reported that the use of air conditioning systems influenced the fungal community composition, and both air conditioning and water leaks increased the species richness. While, Kettleson et al. [21] found age and moldiness of the studied houses to be positively correlated to the fungal richness.

The most relevant occupant-related variable was the presence of allergies (including pollen, food, and skin reactions), explaining 1.61% of the indoor fungal community composition, while the presence of asthma cases accounted for only 0.26% of the variation. We found in indoor samples a striking abundance of taxa with allergenic effects on human. Despite the well-recognized association between dampness/mold problems and negative health effects as allergies and asthma development, the causal agents and mechanisms remain unclear. A previous study associated the asthma severity in children with the fungal community composition in house dust [36]. Further, in our study, the number of occupants and the presence of pets were also significant explanatory variables, with R^2 values 1.46% and 1.03%, respectively. Dannemiller et al. [22] also reported the influence of occupancy (people, children and pets) on the fungal community composition, with an increased richness associated with the presence of pets.

Fungal diversity in Norwegian houses

Fungal richness, evenness and alpha diversity were consistently higher in indoor than outdoor samples. This trend has been reported, with approximately 50% increase of fungal richness/diversity indoors compared to outdoors, in previous studies [18, 23]. As suggested by Barberán et al. [18], this tendency may be partially due to two inter-related phenomena: (i) the dominance of a few taxa in the outdoor communities, and (ii) the higher richness/diversity indoors, where there is a mixture of outdoor and indoor fungi. Both phenomena were likely relevant in our study. Dominant outdoor taxa from the genera *Cladosporium*, *Thekopsora* and *Verrucocladosporium* are in the top-20 OTUs (1st, 2nd and 8th most abundant OTUs) and occurred in more than 80% of the houses. These taxa were also present indoor but with relatively lower proportions.

In contrast, studies of specific building units reported the opposite trend, with higher fungal diversity and richness outdoor [12, 37, 38]. Both fungal and bacterial richness, as well as the fungal biomass were higher in outdoor dust samples from a university housing facility in California [12, 37]. Similar trend was also 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 molds [38]. However, that finding was associated with the influence of a few dominant taxa, which were likely growing and spreading from mold colonies indoors. In this regard, Adams et al. [39] demonstrated that local sources of abundantly sporulating fungi might distort the perception of species richness and community composition assessed by HTS approaches, a few (or unique) abundant species mask the real community. In our study, based on low DNA yields extracted from low-input dust material, this phenomenon might partly explain our results related to the high abundance of some sporulating taxa outdoors (e.g. rust fungi affiliated to *Thekopsora*) and indoors (e.g. mold genera as *Penicillium* and *Aspergillus*), as well as the considerably low evenness found in some outdoor samples. Unfortunately, our dataset does not allow to verify if this distortion occurred and to which extent.

Regardless the house compartment, the fungal richness (mean = 142.8 OTUs per sample) was about ten folds lower than those reported by Barberán et al. [18]. This is most likely due to the methodological differences between both studies, such as sampling materials, DNA preparation and polymerase-chain-reaction (PCR) protocols, barcodes (ITS1 vs. ITS2), HTS set-up and bioinformatics pipelines. Another relevant point is that our sampling campaign was carried out in spring (mostly in May), while the Barberán et al. [18] included samples collected in different seasons for more than one year. In contrast, Adams et al. [12] reported a more similar range of fungal richness to our study (78.5 and 179.7 as mean number of OTUs per indoor and outdoor samples, respectively) for dust settled on sterile Petri dishes during 1-month period. In addition, several studies have reported a global trend for fungal diversity and richness that increase with the latitude [16, 33]. Our study also supports this trend as slightly higher alpha diversity 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 our study 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 major representative genus in our dataset) is particularly rich in extremotolerant species, including saprobes, plant pathogens, endophytes, epiphytes and rock-inhabiting fungi [31, 40], while *Eurotiales* contains many xerophilic fungi (especially *Aspergillus* and *Penicillium* species) that are able to grow on substrates with low water activity ($a_w \leq 0.85$) like household dust [3, 41].

Interestingly, we observed a distinct difference in the overall distribution of *Ascomycota* and *Basidiomycota*; ascomycetes were 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 plant phenology across the study area. During the sampling campaign in May, the plant phenology has likely progressed more in areas with a longer growing season, meaning that a larger proportion of plant-associated ascomycetes (including e.g. pathogens, endophytes and saprotrophs) have become dominant in these areas. Further, 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.

Overlapping between indoor and outdoor mycobiomes

According to previous findings [12, 18], we initially expected that a major part of indoor fungi is originated from outdoor sources (H3). Barberán et al. [18] reported that 65% of OTUs found indoor were also present outdoor. 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 OTUs were present in both outdoor and indoor environments, and 13% of the OTUs in both indoor compartments (living room and bathroom). The lower overlap between compartments in single houses may be due to the limited representativeness of the collected samples (one per house compartment) or/and the influence of indoor fungal sources nearby the sampled surfaces. Considering these results, the hypothesis H3 has been partly refuted, as we cannot conclude that the major fraction of indoor fungi was from outdoor. In agreement with Yamamoto et al. [23], the indoor emissions 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 [5, 13, 17]. Presumably, the indoor mycobiome is assembled by a combination of outdoor and indoor sources and their exact contributions are hard to tease apart.

The indoor core mycobiome

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 molds affiliated to *Eurotiales* (17% of indoor indicator OTUs; mostly to the genera *Penicillium* and *Aspergillus*) and *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 genus *Malassezia* (4%).

Penicillium and *Aspergillus* species are ubiquitous fungi found in dust and air samples, both indoor and outdoor, during all seasons [4, 5, 19, 42, 43]. They are especially abundant indoors, as part of household dust or colonizing building materials and foodstuffs, which become relevant sources for further conidial dispersion [3, 30]. *Wallemia* is

an extreme xerophilic basidiomycete, commonly found in dust due to its ability to grow at low water potential, a_w < 0.75 [3, 44]. The yeast genera *Malassezia*, *Debaryomyces, Candida* and *Rhodotorula* are commensal fungi associated with human skin, showing prevalence in indoor environments [4, 13, 29, 43, 45, 46]. The fourth most-abundant species (OTU3, 5.4% of total reads, present in 96% of houses), with the highest indoor IndVal (91.2%), was identified as *Saccharomyces* sp., a gastronomically relevant genus that includes *S. cerevisiae* (baker's and brewer's yeast) and has previously been reported in indoor environments [18, 43]. In addition, some common edible mushrooms affiliated to *Agaricales* – *Agaricus bisporus* and *Pleurotus ostreatus* – were identified as indoor indicator species in 39% and 5% of study houses, respectively. The majority of these fungi found indoor have been described as potential allergenic taxa [27, 28]. 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 mycobiomes

Mycobiomes detected at the level of the main entrance outside the buildings, showed striking differences compared to the indoor mycobiomes. *Cladosporium* and *Thekopsora* were the most-abundant and widespread genera in the outdoor samples (18% and 16% mean relative abundance per samples, respectively). The genus *Cladosporium* includes many common airborne molds that colonize living and dead plant materials in nature [40] and is commonly found in airborne and dust-associated fungi inside and outside buildings [12, 19, 42]. The second most-abundant OTU correspond to *Thekopsora aerolata*, fungus causing important rust damages in cones of *Picea* spp., especially on *Picea abies*, and leaves of *Prunus* spp. [47]. *P. abies* is widely distributed in Norway (south, east and mid), where the majority of houses in this study are located.

Many of the taxa fruiting in nature and prevalent in the outdoor samples were also frequent indoors (even showing higher relative abundances), including members of *Agaricales* (genera *Strobilurus* – commonly found on pine cones, *Lycoperdon* – puffballs, and *Cylindrobasidium* – corticioid fungi) and *Polyporales* (*Fomitopsis* – bracket wood-decay fungi). Adams et al. [12] also reported a high diversity of typically outdoor fungi – mushrooms, wood-decay polypores, puffballs and lichenized fungi – in both outdoor and indoor environments. Although these characteristic outdoor fungi have higher absolute abundances outdoor, their proportions may be higher in indoor samples, leading to a bias picture of their distributions. Based on former culture-based studies, it is acknowledged that the concentration of fungi in outdoor air are generally higher than in indoor air [42]. Adams et al. [12] also reported much greater (20–100 fold) fungal biomass in outdoor dust samples than indoor samples, resulting in a lower sequencing depth and coverage of the outdoor assemblages. We cannot rule out this impact in our study.

Indicator species analysis revealed that outdoor mycobiomes were distinctly enriched in so call 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 multi-stress tolerance and prevalence in diverse outdoor environments such as rocks and buildings, where they are exposed to stresses like solar radiation, desiccation and rehydration, temperature fluctuations, osmotic stress, pollutants and lack of nutrients [31, 48]. It is not surprising that these fungal groups were especially abundant in the outdoor samples, as mostly collected from the doorframes exposed to the external conditions and prone to colonization by subaerial biofilms.

Conclusions

Our main findings are in line with previous indoor mycobiome studies, identifying climatic variables as the key determinants of indoor mycobiome. Building features and occupant characteristics had 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 indoor. The indoor core mycobiome is characterized by two ecological groups with allergenic potential, xerophilic molds and skin-associated yeasts. In contrast, rock-inhabiting fungi, well-known for their multi-stress tolerance and ability to form biofilms on buildings, were the main outdoor indicator fungi.

Despite methodological limitations related to citizen science sampling (e.g. non-uniform way of collection, small amount of dust collected with subsequently low DNA yields, and low number of samples per house), this approach was a successful strategy in characterizing the indoor mycobiome of a large set of houses throughout Norway during a short period of time. Future large-scale studies on indoor mycobiomes should preferably target other regions, beyond the US and Europe, to reveal whether similar trends are present under different climates.

Methods

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/public media: Facebook website, outreach articles in the Titan newspaper [49] from the University of Oslo (UiO) and the Agarica magazine from the Norwegian Mycological Society, as well as a radio interview at the Norwegian public broadcasting. A total of 359 volunteers signed up in this study and provided relevant information about their houses by filling out an online questionnaire (details in the section Environmental data). Sampling kits (Additional file 1: Fig. S1), including instructions, return envelope, three sterile FLOQSwabs in tubes (Code 552C, Copan Italia spa, Brescia, Italy) and two adhesive tapes (Mycotape2, Mycoteam AS, Oslo, Norway), were sent to volunteers by post.

For DNA metabarcoding analyses, three dry dust samples were swabbed from different compartments in each house: outside (entrance door), living room (main room) and bathroom. The samples were preferentially collected from the upper surface of doorframes, but in cases where this was not possible, similar areas on shelves or windowsills were sampled. As stated in previous studies [18, 19], these selected areas, with little contact from the house occupants, act as passive collectors of dust that was deposited during an unknown amount of time. In addition, one adhesive tape was collected from shelves or windowsills in the living room to calculate the percentage of dust coverage, which was later included as an environmental variable in the study. Samples were sent back to UiO by post, where they were registered and the swabs were stored at -80 °C until DNA extractions. Whereas the adhesive tapes were immediately scanned using a scanner Epson Perfection V850 Pro (Seiko Epson Corporation, Nagano, Japan), and the percentage of dust coverage was calculated on a surface area of 45 × 18 mm by image analysis using the Olympus Stream v 1.9 software (threshold at the maximum value 61100).

This large-scale sampling campaign was mostly conducted in May 2018 (from 27th April to 5th June). Overall, we received 869 swabs from 290 houses. However, 57 samples failed during the DNA laboratory works (extractions, PCR and library preparation). Thus, the HTS was performed on 812 dust samples from 271 houses, including two houses from Svalbard.

Environmental data

Meta data about the study houses and their occupants were provided by the volunteers in the online questionnaire at UiO website. In addition to the location of houses, with the complete addresses and their corresponding geographic coordinates (latitude and longitude), the following variables (with categories for factor variables) were extracted from the questionnaire: building type (detached house/semi-detached 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/mousses/rats/grey silverfish). Data about the location of dust samples in the house were included as two factor variables: house compartment (outside/living room/bathroom) and indoor vs. outdoor (indoor/outdoor).

Based on the geographic coordinates of study houses, data for six relevant WorldClim 2 bioclimatic variables (annual mean temperature BIO1, temperature seasonality BIO4, mean temperature of driest quarter BIO9, mean temperature of warmest quarter BIO10, mean temperature of coldest quarter BIO11, and annual precipitation BIO12), at 30 seconds resolution (~ 1 km^2), were extracted using the R package *dismo* following the authors' instructions [50]. Moreover, data for 116 environmental variables, related to geology, topography, climate and hydrology, were also explored. They were the explanatory variables analyzed in a recent study modelling the vegetation types in Norway [51], and were kindly provided by their authors. The contribution of the numerical variables, 46 from this dataset plus the 6 previously extracted from WorldClim, were evaluated by principal component analysis (PCA; Additional file 1: Fig. S7). Based on PCA results, 10 numerical variables were finally selected for the statistical analyses: the six detailed WorldClim bioclimatic variables, growing season length (The Norwegian Metereological Institute, MET), snow covered area in February (sca-2, MET), snow water equivalent in April (swe-4, MET), and potential incoming solar radiation (Geodata AS). Two additional factor variables: land cover AR50 (developed area/agricultural area/forest/barren land/bog and fen/fresh water; Norwegian Institute of Bioeconomy Research, NIBIO) and bedrock nutrient (poor/average/rich; Norwegian Geological Survey, NGU), were included in the final selection (Fig. 1a).

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, Norcross, GA, USA). Swab tips were transferred to the kit Disruptor tubes that contain glass pearls 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. Irvine, Ca, USA), the samples were frozen at -20ºC for at least 30 min. Afterwards, samples were incubated at 70ºC for 15 min, and again shaken at FastPrep-24 (2 cycles of 30 s at 4.5 $m·s^{-1}$). These successive thermal-shocking and bead-beating steps were carried out to get a proper homogenization of dust samples and lysis of fungal conidia and spores. After adding 600 µl chloroform, samples were vortexed for 30 s and centrifuged at 13,000 rpm for 5 min. DNA from the aqueous top phase was further purified using the HiBind DNA Mini Column following the kit'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, Waltham, MA, USA). Low DNA yields, ranged from 0.05 to 1 $ng·µl^{-1}$, were recovered from the swabs, which were expected considering the small amount of dust collected with dry swabs. Nine blank controls (unused sterile swabs) from different extractions batches were included through the complete DNA metabarcoding protocol.

The internal transcribed spacer 2 (ITS2) region of the nuclear rDNA were amplified using the primers gITS7 5'-GTGARTCATCGARTCTTTG-3' [52] and ITS4 5'-TCCTCCGCTTATTGATATGC-3' [53]. The selected marker and primers

have shown a good species resolution in previous mycobiome studies of diverse environments. Both forward and reverse primers were designed with 96 unique tags (barcodes) of 7–9 bp at the 5'-end, which differed in at least three positions from each other. To avoid tag switching errors [54], samples were combined in pools of 96 samples, each with a unique tag combination (Additional file 1: Table S3). Nine pools (96 samples each) were analyzed in this study, each of them included an extraction blank, a PCR negative and a mock community that was used as positive control. 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. In total, 17 dust samples were duplicates, as technical replicates across different PCR libraries.

PCR reactions in 25 µl contained 1 U of AmpliTaq Gold DNA polymerase (Applied Biosystems, Thermo Fisher Scientific), 0.4 µM of each primer, 0.8 mg/ml bovine serum albumin (BSA; Thermo Scientific, Thermo Fisher Scientific), 1X 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 depending on dust sample). 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. 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, Inc., CA, USA), DNA concentration and length of the final pooled amplicons were checked using Qubit dsDNA HS Assay Kit and Bioanalyzer High Sensitivity DNA chip (Agilent Technologies, Inc., Santa Clara, CA, USA), respectively. Sequencing was carried out at Fasteris SA (Plan-les-Ouates, Switzerland) 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 v3 runs (Illumina Inc., San Diego, CA, USA) were used. Each run included three pooled libraries labelled with specific indexes in their Illumina barcodes. The complete resulting dataset contained 55,568,124 paired reads and are deposited in Dryad Digital Repository [55].

Bioinformatics pipeline

After an initial quality checking of sequencing results, using FastQC [56], samples were demultiplexed independently (R1 and R2) with CUTADAPT v 1.8 [57] allowing zero miss matches in tags and primers; these were simultaneously removed along with sequences shorter than 100 bp. The demultiplexed R1 and R2 reads were kept separate for the next steps using DADA2 v 1.12 [58]: (i) quality filtering and trimming [*filterAndTrim* with options *maxN = 0, maxEE = 2.5, trimLeft = c(10,10), truncLen = c(200,200), rm.phix = TRUE, compress = TRUE, matchIDs = TRUE, multithread = TRUE*], (ii) dereplication [*derepFastq* with option *verbose = TRUE*], (iii) generating error models and denoising [*dada* with options *err = NULL, pool = "pseudo", selfConsist = TRUE*], (iv) merging in contigs [*mergePairs* with options *verbose = TRUE, minOverlap = 5*], (v) creating the table of amplicon sequence variants (ASVs) [*makeSequenceTable*], and removal of chimeras [*removeBimeraDenovo* with options *method = "pooled", verbose = TRUE*]. Additional clustering of sequences in OTUs was done using VSEARCH v 2.11.1 [59] at 98% similarity. This clustering level is similar to the 98.5% level used to define the species hypotheses (SHs) in the UNITE database [60]. 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 [61] with default settings. An initial matchlist (sequence similarity) was created with blastn [62] [options: *-qcov_hsp_perc 80 -perc_identity 84*], for the subsequent LULU run [options: *minimum_match = 84*, *minimum_relative_cooccurence = 0.95*]. Taxonomic assignment of the OTUs was carried out using VSEARCH against the eukaryotic ITS dataset from UNITE v 8.0 [63].

In the resulting OTU table, we initially kept and identified all dust biodiversity captured by this DNA metabarcoding approach, including mostly fungi but also members of the clade Viridiplantae (green plants). Previous studies have reported that gITS7/ITS4 primers can also amplify plant DNA [52]. Two filters were subsequently applied on the table to select the OTUs that contained at least 10 reads, and showed at least 70% of 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 + v 2.8 against both UNITE and NCBI [64] databases. The key steps of this bioinformatics pipeline, as well as the resulting of numbers of reads and OTUs throughout the pipeline, are summarized in the Table S4 (Additional file 1).

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. 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 (Additional file 1: Fig. S8), and the replicate with lower number of reads were discarded. Hence, confirming the reproducibility of the DNA metabarcoding workflow.

Statistical analyses

Statistical analyses were conducted in R v 3.5.2 [65] through RStudio v 1.2.1335. Tidyverse v 1.2.1 [66] and the vegan v 2.5-6 [67] R packages were used for data manipulation and plotting, and ecological analyses, respectively. The most relevant R functions used are detailed below, remarking when they belong to R packages different from vegan. Initially, the OTU table was rarefied (× 10 times resampling with the median value taken per OTU) to 2,000 reads per sample using the function *rrarefy*, and further adapted for the three datasets: all samples (full dataset), indoor samples, and outdoor samples.

Abundance of fungi was estimated using two different ways: (i) relative abundance of OTUs as percentages of rarefied reads on the total count in the dataset, and (ii) relative abundance of a certain taxa (at different levels, e.g.

phylum, order and genus) as percentage of rarefied number of reads per sample. Prevalence of OTUs was also calculated as percentages of samples or houses in which each OTU was detected, based of rarefied tables.

Alpha-diversity of samples was assessed calculating species richness (number of observed OTUs) and evenness (equitability between OTUs), as well as Shannon, Simpson and Chao1 indices, on the rarefied OTU table. Significant differences in the variance of these parameters were evaluated with the analysis of variance (ANOVA) test. Beta-diversity was assessed with NMDS ordination of both dust samples and OTUs using *metaMDS*, Bray-Curtis dissimilarity index and 200 random starts in search of stable solution. NMDS analyses were done on the Hellinger-transformed rarefied OTU tables, after testing three transformations: Hellinger and log using the function *deconstand*, and Cumulative Sum Scaling (CSS) using *cumNorm* of the metagenomeSeq package [68]. Continuous environmental variables and alpha-diversity indices were regressed against NMDS ordination and added as vectors on the ordination plots using the function *gg_envfit* of the package ggordiplot v 0.3.0 [69] to visualize their association with the dust mycobiomes. In addition, beta-diversity was also assessed using the function *betadisper* to test the homogeneity of variance in different groups of samples.

To evaluate the correlation between each environmental variable and the observed variance in fungal community composition, permutational multivariate analysis of variance (PERMANOVA; 999 permutations) with the function *adonis2* was used. The effects of four groups of variables (building, occupants, climate and house compartment), were assessed by variation partitioning analysis (VPA) based on the Bray-Curtis dissimilarities using the function *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 on the total counts in the dataset, and percentages of OTUs per house. Indicator species analysis was performed to reveal the significant associations ($p < 0.05$) between OTUs and some relevant environmental variables related to the house compartments, building features (e.g. construction materials, water damages and moisture problems) and occupants (e.g. presence of allergies and pets). These analyses were performed using the *multipatt* function of the *indicspecies* package [70] in R. Pearson's correlation coefficients and their corresponding *p* values were calculated to explore the associations between OTUs and the continuous variables related to climate and number of occupants.

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 GLM analyses using the *glm* function. A forward selection was performed using AIC in order to assess model improvement in comparison with the null model.

Abbreviations

AIC: Akaike information criterion

ANOVA: analysis of variance

CSS: cumulative sum scaling

GLM: generalized linear model

HTS: High-throughput sequencing

ITS: internal transcribed spacer

NMDS: nonmetric multidimensional scaling

OTU: operational taxonomic units

PCR: polymerase chain reaction

PERMANOVA: permutational multivariate analysis of variance

UiO: University of Oslo

VPA: variation partitioning analysis

Declarations

Ethics approval and consent to participate

Personal data of the citizen scientists, i.e addresses and geographic coordinates of their houses included in the sampling campaign, were omitted from the metadata.

Consent for publication

Not applicable.

Availability of data and material

In addition to the data included in this published article [Additional file 1 – Supplementary Figures and Tables, and Additional file 2 – Top-200 most abundant OTUs], all datasets analyzed during the current study are available in the Dryad Digital Repository [https://doi.org/10.5061/dryad.59zw3r24w] [55], including raw sequences, mapping files, complete metadata file, the final fungal rarefied OTU table, as well as the taxonomic assignment of their OTUs. The scripts used for the analyses will be made available upon e-mail request to the authors.

Competing interests

The authors declare that they have no competing interests.

Funding

This study has received funding from the European Union's Horizon 2020 research and innovation programme through a Marie Skłodowska-Curie Individual Fellowship to PMMS, under the grant agreement *MycoIndoor* No 741332.

Authors' contributions

PMMS, ELFE, IS and HK conceived and designed the study. ELFE organized the citizen science sampling with contribution from IBE. PMMS performed laboratory work. PMMS and LNM analyzed data (bioinformatics and statistics) and prepared figures. SM provided technical advice on laboratory work and contributed to statistical analyses. PMMS wrote the first draft of the manuscript. All authors contributed to the data interpretation, as well as edited and approved the final manuscript.

Acknowledgements

All volunteers who provided samples and data of their houses are thanked for their essential contributions in this citizen science study. Peter Horvath and Anders Bryn, from the Natural History Museum at UiO, are thanked for providing the metadata of environmental variables. Maria Nuñez from Mycoteam AS is acknowledged for her valuable feedback on fungal taxonomy and building ecology.

References

1. Gilbert JA, Stephens B. Microbiology of the built environment. Nat Rev Microbiol. 2018;16:661–70. doi:10.1038/s41579-018-0065-5.
2. Willis KJ, editor. State of the World's Fungi 2018. London: Royal Botanic Gardens, Kew;: Report; 2018.
3. Flannigan B, Miller JD. Microbial growth in indoor environments. In: Flannigan B, Samson RA, Miller JD, editors. Microorganisms in home and indoor work environments: diversity, health impacts, investigation and control. London: CRC Press; 2011. pp. 57–108.
4. Rintala H, Pitkäranta M, Täubel M. Microbial communities associated with house dust. Adv Appl Microbiol. 2012;78:75–120. doi:10.1016/B978-0-12-394805-2.00004-X.
5. Nevalainen A, Täubel M, Hyvärinen A. Indoor fungi: companions and contaminants. Indoor Air. 2015;25:125–56. doi:10.1111/ina.12182.
6. Fisk WJ, Lei-Gomez Q, Mendell MJ. Meta-analyses of the associations of respiratory health effects with dampness and mold in homes. Indoor Air. 2007;17:284–96. doi:10.1111/j.1600-0668.2007.00475.x.
7. 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 Perspect. 2011;119:748–56. doi:10.1289/ehp.1002410.
8. Lindahl BD, Nilsson RH, Tedersoo L, Abarenkov K, Carlsen T, Kjøller R, et al. Fungal community analysis by high-throughput sequencing of amplified markers–a user's guide. New Phytol. 2013;199:288–99. doi:10.1111/nph.12243.
9. Nilsson RH, Anslan S, Bahram M, Wurzbacher C, Baldrian P, Tedersoo L. Mycobiome diversity: high-throughput sequencing and identification of fungi. Nat Rev Microbiol. 2019;17:95–109. doi:10.1038/s41579-018-0116-y.
10. Lax S, Smith DP, Hampton-Marcell J, Owens SM, Handley KM, Scott NM, et al. Longitudinal analysis of microbial interaction between humans and the indoor environment. Science. 2014;345:1048–52. doi:10.1126/science.1254529.
11. Adams RI, Bateman AC, Bik HM, Meadow JF. Microbiota of the indoor environment: a meta-analysis. Microbiome. 2015;3:49. doi:10.1186/s40168-015-0108-3.
12. 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 J. 2013;7:1262–73. doi:10.1038/ismej.2013.28.
13. Tong X, Leung MHY, Wilkins D, Lee PKH. City-scale distribution and dispersal routes of mycobiome in residences. Microbiome. 2017;5:131. doi:10.1186/s40168-017-0346-7.
14. Adams RI, Bhangar S, Dannemiller KC, Eisen JA, Fierer N, Gilbert JA, et al. Ten questions concerning the microbiomes of buildings. Build Environ. 2016;109:224–34.
15. 10.1128/mSystems.00083-16
    Stephens B. (2016). What have we learned about the microbiomes of indoor environments? mSystems. 2016;1:e00083-16. doi: 10.1128/mSystems.00083-16.

16. Amend AS, Seifert KA, Samson R, Bruns TD. Indoor fungal composition is geographically patterned and more diverse in temperate zones than in the tropics. Proc Natl Acad Sci U S A. 2010;107:13748–53. doi:10.1073/pnas.1000454107.
17. Adams RI, Miletto M, Taylor JW, Bruns TD. The diversity and distribution of fungi on residential surfaces. PLoS One. 2013;8:e78866. doi:10.1371/journal.pone.0078866.
18. Barberán A, Dunn RR, Reich BJ, Pacifici K, Laber EB, Menninger HL, et al. The ecology of microscopic life in household dust. Proc R Soc B. 2015;282:20151139.
19. Barberán A, Ladau J, Leff JW, Pollard KS, Menninger HL, Dunn RR, Fierer N. Continental-scale distributions of dust-associated bacteria and fungi. Proc Natl Acad Sci U S A. 2015;112:5756–61. doi:10.1073/pnas.1420815112.
20. Haines SR, Adams RI, Boor BE, Bruton T, Downey J, Ferro AR, et al. Ten Questions Concerning the Implications of Carpet on Indoor Chemistry and Microbiology. Build Environ. 2019;170:1–16. doi:10.1016/j.buildenv.2019.106589.
21. 10.1016/j.envres.2015.02.003
Kettleson EM, Adhikari A, Vesper S, Coombs K, Indugula R, Reponen T. (2015). Key determinants of the fungal and bacterial microbiomes in homes. Environ Res. 2015;138:130-5. doi:10.1016/j.envres.2015.02.003.
22. Dannemiller KC, Gent JF, Leaderer BP, Peccia J. Influence of housing characteristics on bacterial and fungal communities in homes of asthmatic children. Indoor Air. 2016;26:179–92. doi:10.1111/ina.12205.
23. 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;49:5098–106. doi:10.1021/es506165z.
24. 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. doi:10.1371/journal.pone.0154131.
25. McDonald D, Hyde E, Debelius JW, Morton JT, Gonzalez A, Ackermann G, et al. American Gut: an open platform for citizen science microbiome research. mSystems. 2018;3:e00031-18. doi:10.1128/mSystems.00031-18.
26. Gostinčar C, Zajc J, Lenassi M, Plemenitaš A, de Hoog S, Al-Hatmi AMS, Gunde-Cimerman N. Fungi between extremotolerance and opportunistic pathogenicity on humans. Fungal Divers. 2018;93:195–213. doi:10.1007/s13225-018-0414-8.
27. Esch RE, Hartsell CJ, Crenshaw R, Jacobson RS. Common allergenic pollens, fungi, animals, and arthropods. Clin Rev Allergy Immunol. 2001;21:261–92. doi:10.1385/CRIAI:21:2-3:261.
28. Yamamoto N, Bibby K, Qian J, Hospodsky D, Rismani-Yazdi H, Nazaroff WW, Peccia J. Particle-size distributions and seasonal diversity of allergenic and pathogenic fungi in outdoor air. ISME J. 2012;6:1801–11. doi:10.1038/ismej.2012.30.
29. Findley K, Oh J, Yang J, Conlan S, Deming C, Meyer JA, et al. Topographic diversity of fungal and bacterial communities in human skin. Nature. 2013;498:367–70. doi:10.1038/nature12171.
30. Andersen B, Frisvad JC, Søndergaard I, Rasmussen IS, Larsen LS. Associations between fungal species and water-damaged building materials. Appl Environ Microbiol. 2011;77:4180–8.
31. Ametrano CG, Muggia L, Grube M. Extremotolerant black fungi from rocks and lichens. In: Tiquia-Arashiro S, Grube M, editors. Fungi in extreme environments: ecological role and biotechnological significance. Cham: Springer; 2019. pp. 119–43.

32. 10.1073/pnas.1907414116
Tipton L, Zahn G, Datlof E, Kivlin SN, Sheridan, Amend AS, Hynson NA. (2019). Fungal aerobiota are not affected by time nor environment over a 13-y time series at the Mauna Loa Observatory. Proc Natl Acad Sci U S A. 2019;116:25728-33. doi:10.1073/pnas.1907414116.

33. Větrovský T, Kohout P, Kopecký M, Machac A, Man M, Bahnmann BD, et al. A meta-analysis of global fungal distribution reveals climate-driven patterns. Nat Commun. 2019;10:5142. doi:10.1038/s41467-019-13164-8.

34. Aak A, Rukke BA, Ottesen PS, Hage M. Long-tailed silverfish (*Ctenolepisma longicaudata*) – biology and control. In: Norwegian Institute of Public Health Report. Oslo: Norwegian Institute of Public Health; 2019. pp. 1–43.

35. 10.1111/mec.13900
Madden AA, Barberán A, Bertone MA, Menninger HL, Dunn RR, Fierer N. (2016). The diversity of arthropods in homes across the United States as determined by environmental DNA analyses. Mol Ecol. 2016;25:6214-24. doi:10.1111/mec.13900.

36. Dannemiller KC, Gent JF, Leaderer BP, Peccia J. Indoor microbial communities: Influence on asthma severity in atopic and nonatopic children. J Allergy Clin Immunol. 2016;138:76–83.e1. doi:10.1016/j.jaci.2015.11.027.

37. 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. doi:10.1371/journal.pone.0091283.

38. Sylvain IA, Adams RI, Taylor JW. A different suite: The assemblage of distinct fungal communities in water-damaged units of a poorly-maintained public housing building. PLoS One. 2019;14:e0213355. doi:10.1371/journal.pone.0213355.

39. Adams RI, Amend AS, Taylor JW, Bruns TD. 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. Microb Ecol. 2013;66:735–41. doi:10.1007/s00248-013-0266-4.

40. Crous PW, Schoch CL, Hyde KD, Wood AR, Gueidan C, de Hoog GS, Groenewald JZ. Phylogenetic lineages in the *Capnodiales*. Stud Mycol. 2009;64:17–47S7. doi:10.3114/sim.2009.64.02.

41. Vinnere Pettersson O, Leong SL. Fungal xerophiles (osmophiles). eLS (Encyclopaedia of Life Sciences). Chichester: John Wiley & Sons, Ltd; 2011. doi:10.1002/9780470015902.a0000376.pub2.

42. 10.1128/AEM.68.4.1743-1753.2002
Shelton BG, Kirkland KH, Flanders WD, Morris GK. 2002. Profiles of airborne fungi in buildings and outdoor environments in the United States. Appl. Environ. Microbiol. 2002;68:1743-53. doi:10.1128/AEM.68.4.1743-1753.2002.

43. Flannigan B. Microorganisms in indoor air. In: Flannigan B, Samson RA, Miller JD, editors. Microorganisms in home and indoor work environments: diversity, health impacts, investigation and control. London: CRC Press; 2011. pp. 25–43.

44. Nguyen HDT, Jančič S, Meijer M, Tanney JB, Zalar P, Gunde-Cimerman N, Seifert KA. Application of the phylogenetic species concept to *Wallemia sebi* from house dust and indoor air revealed by multi- locus genealogical concordance. PLoS One. 2015;10:e0120894. doi:10.1371/journal.pone.0120894.

45. Dannemiller K, Mendell M, Macher J, Kumagai K, Bradman A, Holland N, et al. Next generation DNA Sequencing reveals that low fungal diversity in house dust is associated with childhood asthma development. Indoor Air. 2014;24:236–47. doi:10.1111/ina.12072.

46. 10.1186/s40168-018-0407-6

Maestre JP, Jennings W, Wylie D, Horner SD, Siegel J, Kinney KA. (2018). Filter forensics: microbiota recovery from residential HVAC filters. Microbiome. 2018;6:22. doi:10.1186/s40168-018-0407-6.

47. 10.1080/02827581.2012.727024
Kaitera J. (2013) *Thekopsora* and *Chrysomyxa* cone rusts damage Norway spruce cones after a good cone crop in Finland. Scand J For Res. 2013;28:217 – 22. doi: 10.1080/02827581.2012.727024.

48. Gorbushina AA. Life on the rocks. Environ Microbiol. 2007;9:1613–31. doi:10.1111/j.1462-2920.2007.01301.x.

49. Røsjø B. Alle har sopp i huset – men hva slags? Titan newspaper, 2018. https://titan.uio.no/node/2797. Accessed 20 March 2018.

50. Fick SE, Hijmans RJ. Worldclim. New 1-km spatial resolution climate surfaces for global land areas. Int J Climatol. 2017;2:37:4302–15. doi:10.1002/joc.5086.

51. Horvath P, Halvorsen R, Stordal F, Tallaksen LM, Tang H, Bryn A. Distribution modelling of vegetation types based on area-frame survey data. Applied Vegetation Sci. 2019;22:547–60. doi:10.1111/avsc.12451.

52. Ihrmark K, Bödeker ITM, Cruz-Martinez K, Friberg H, Kubartova A, Schenck J, et al. New primers to amplify the fungal ITS2 region–evaluation by 454-sequencing of artificial and natural communities. FEMS Microbiol Ecol. 2012;82:666–77. doi:10.1111/j.1574-6941.2012.01437.x.

53. White TJ, Bruns TD, Lee SB, Taylor JW. Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. In: Innis MA, Gelfand DH, Sninsky JJ, White TJ, editors. PCR – Protocols and Applications – A Laboratory Manual. New York: Academic Press; 1990. pp. 315–22.

54. Carlsen T, Aas AB, Lindner D, Vrålstad T, Schumacher T, Kauserud H. Don't make a mista(g)ke: is tag switching an overlooked source of error in amplicon pyrosequencing studies? Fungal Ecol. 2012;5:747–9. doi:10.1016/j.funeco.2012.06.003.

55. 10.5061/dryad.59zw3r24w
Martin-Sanchez PM, Estensmo ELF, Morgado LN, Maurice S, Engh IB, Skrede I, Kauserud H. (2020), Data from: Analyzing indoor mycobiomes through a large-scale citizen science study of houses from Norway. Dryad Digital Repository. 2020. https://doi.org/10.5061/dryad.59zw3r24w.

56. Babraham Bioinformatics Team. FastQC. Babraham Institute, Cambridge, UK. https://www.bioinformatics.babraham.ac.uk/projects/fastqc/. Accessed 30 June 2020.

57. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet Journal. 2011;17:10–2. doi:10.14806/ej.17.1.200.

58. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJ, Holmes SP. DADA2: High-resolution sample inference from Illumina amplicon data. Nat Methods. 2016;13:581–3. doi:10.1038/nmeth.3869.

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

60. Kõljalg U, Nilsson RH, Abarenkov K, Tedersoo L, Taylor AFS, Bahram M, et al. Towards a unified paradigm for sequence-based identification of fungi. Mol Ecol. 2013;22:5271–7. doi:10.1111/mec.12481.

61. Frøslev TG, Kjøller R, Bruun HH, Ejrnæs R, Brunbjerg AK, Pietroni C, Johansen AJ. Algorithm for post-clustering curation of DNA amplicon data yields reliable biodiversity estimates. Nat Commun. 2017;8:1188. doi:10.1038/s41467-017-01312-x.

62. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol. 1990;215:403–10. doi:10.1016/S0022-2836(05)80360-2.

63. UNITE Community. UNITE USEARCH/UTAX release for eukaryotes. Version 18.11.2018. UNITE Community. 2019. https://doi.org/10.15156/BIO/786346. Accessed 30 June 2020.
64. UNITE Community. Full UNITE + INSD dataset for Fungi. Version 18.11.2018. UNITE Community. 2019. https://doi.org/10.15156/BIO/786347. Accessed 30 June 2020.
65. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. 2018. https://www.R-project.org/. Accessed 30 June 2020.
66. Wickham H. tidyverse: Easily Install and Load the 'Tidyverse'. R package version 1.2.1. 2017. https://CRAN.R-project.org/package=tidyverse.
67. Oksanen J, Blanchet FG, Friendly M, Kindt R, Legendre P, McGlinn D, et al. vegan: Community Ecology Package. R package version 2.5-6. 2019. https://CRAN.R-project.org/package=vegan.
68. Paulson JN, Olson ND, Braccia DJ, Wagner J, Talukder H, Pop M, Bravo HC. metagenomeSeq: Statistical analysis for sparse high-throughput sequencing. Bioconductor package. 2013. http://www.cbcb.umd.edu/software/metagenomeSeq. Accessed 30 June 2020.
69. Quensen J. ggordiplots. Make ggplot versions of vegan's ordiplots. R package version 0.3.0. 2018. https://github.com/jfq3/ggordiplots.
70. 10.1890/08-1823.1
De Cáceres M, Legendre P. 2009. Associations between species and groups of sites: indices and statistical inference. Ecology. 2009;90:3566-74. doi:10.1890/08-1823.1.

Figures

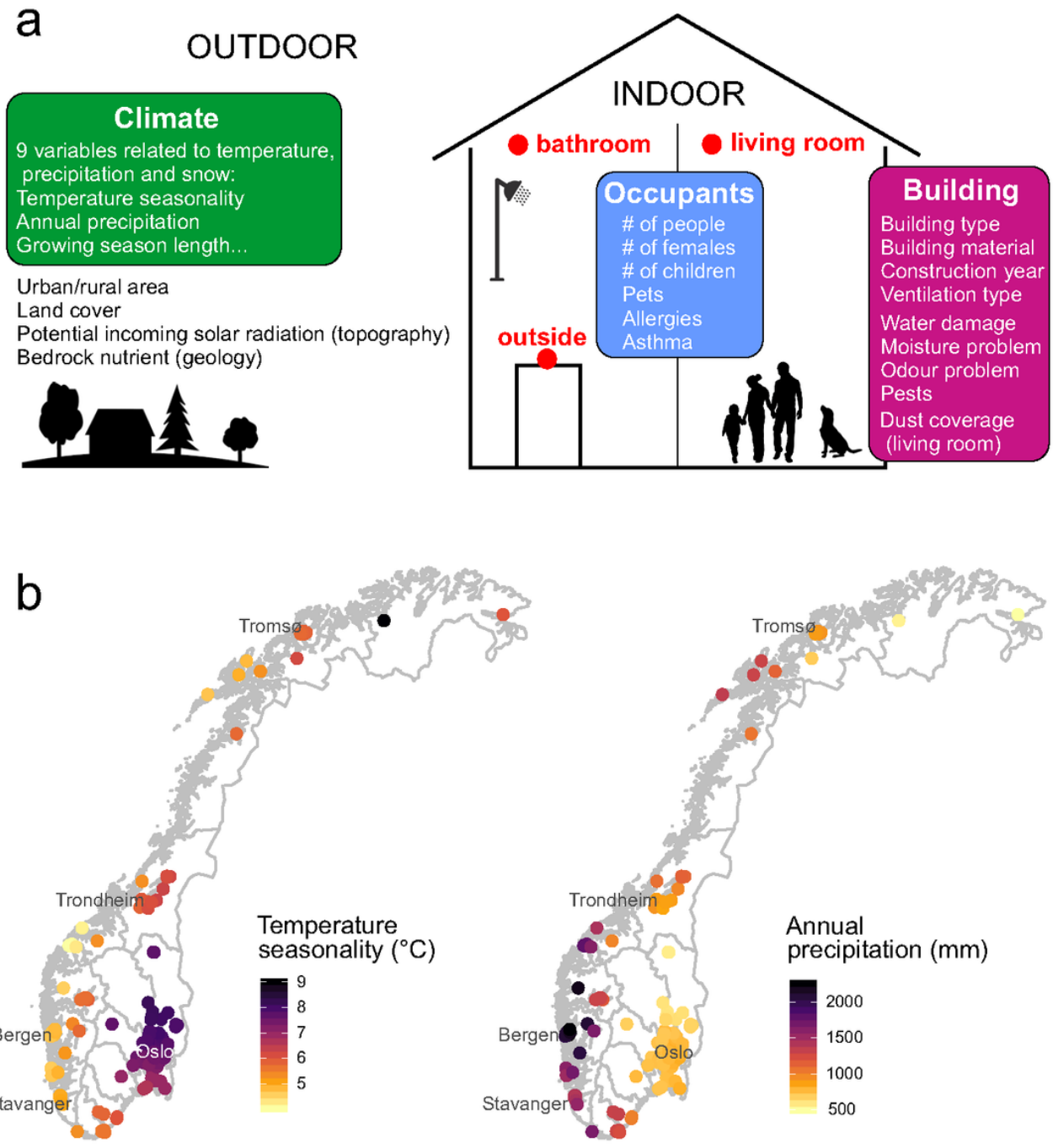


Figure 1

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) colored according to their temperature seasonality (left; standard deviation of mean monthly temperatures = BIO4/100) and the annual precipitation (right; BIO12).

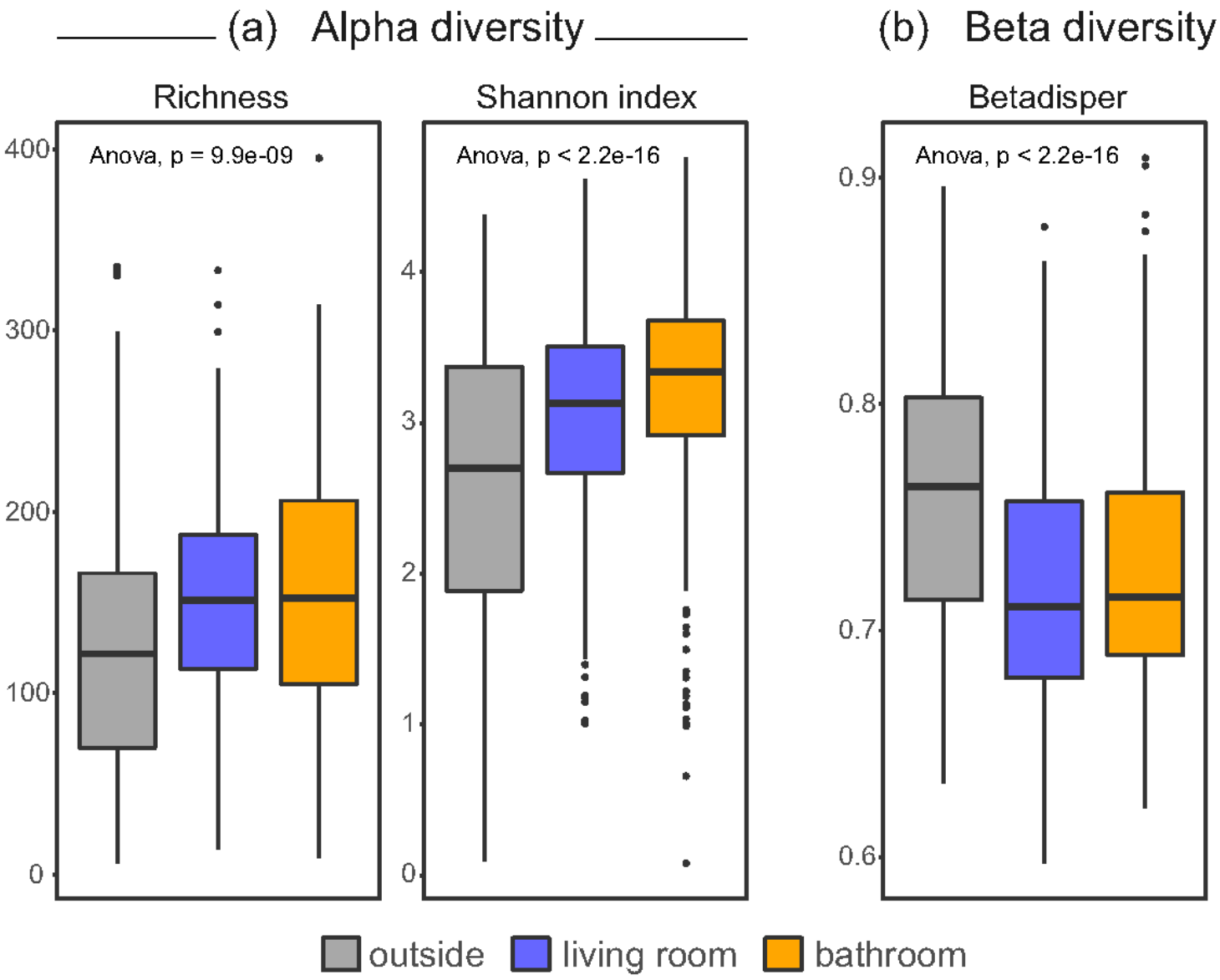
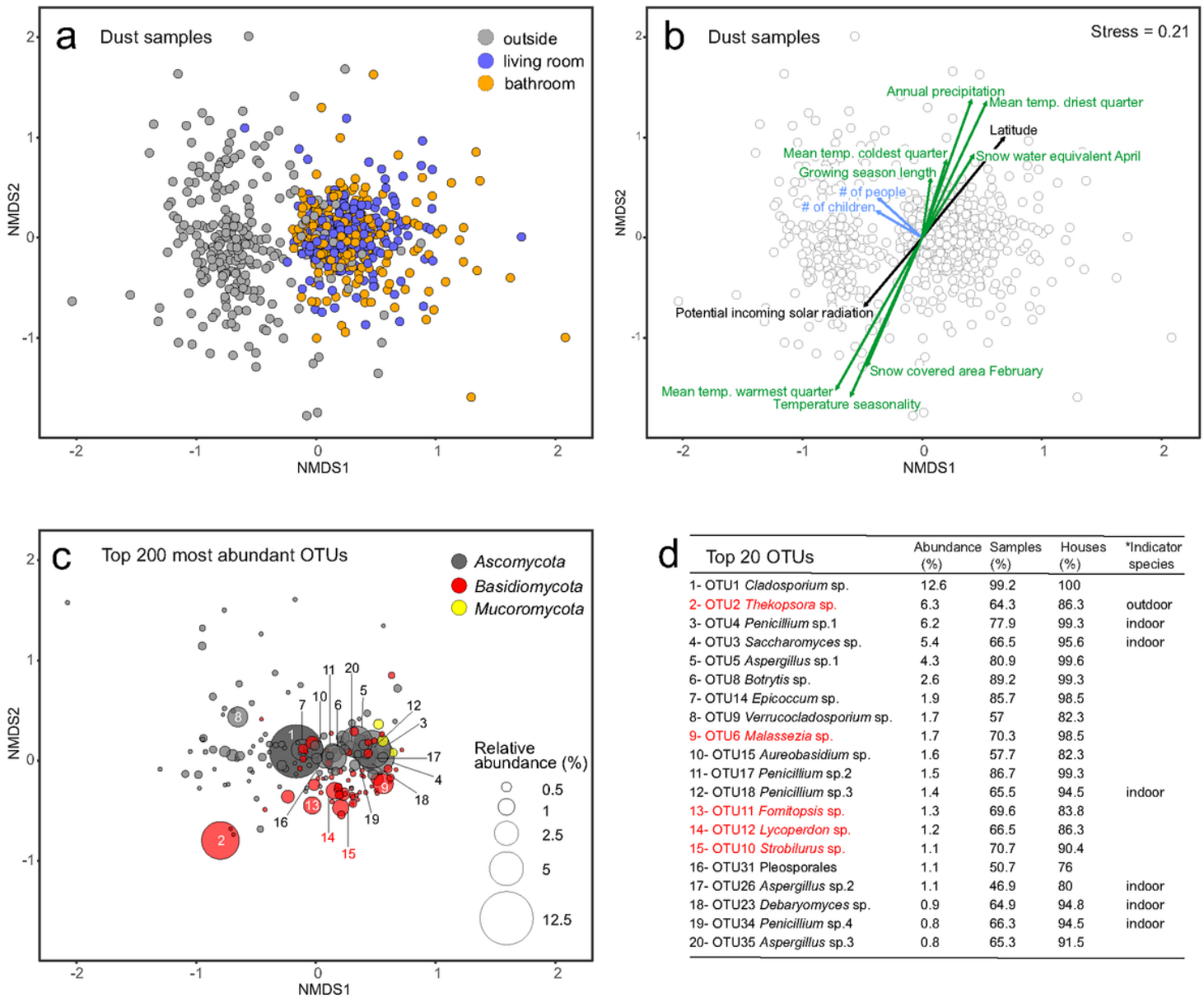


Figure 2

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 (6,632 OTUs).



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 dataset (n = 807). Each point indicates one dust sample. (a) Color indicates the three compartments (outside, living room and bathroom). (b) Linear-regression of continuous variables with significant association ($p < 0.05$) 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 proportion of the total number of rarefied reads and color indicates their phylum assignment. The numbers indicate the taxonomic assignment of the 20 most abundant OTUs, as provided in panel d. (d) OTU id, taxonomical 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 Tables 2 and 3). Ascomycota is indicated in black and Basidiomycota in red.

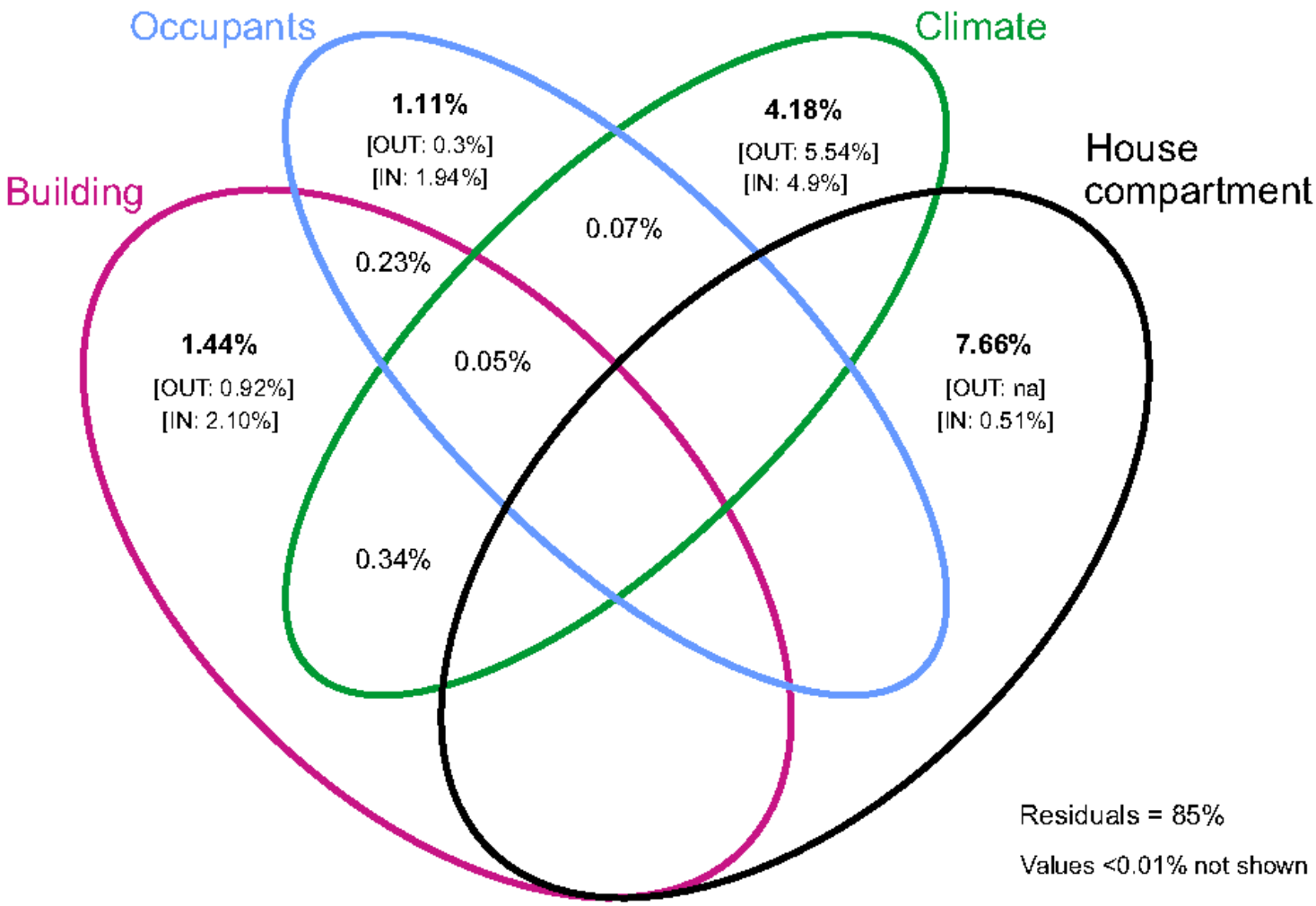


Figure 4

Venn diagram summarizing the variation partitioning analysis (VPA). The four groups of variables are indicated in colors (“Building”, “Occupants”, “Climate” and “House compartment”; see Table 1 for selection). The percentage of variation explained by each group alone are in bold for the complete dataset. VPA values in square brackets were obtained for the partial datasets when analyzed separately: OUT - outdoor dataset, IN - indoor dataset. The unexplained variation (residual) was 85% and variables explaining < 0.01% are not shown in the Venn diagram.

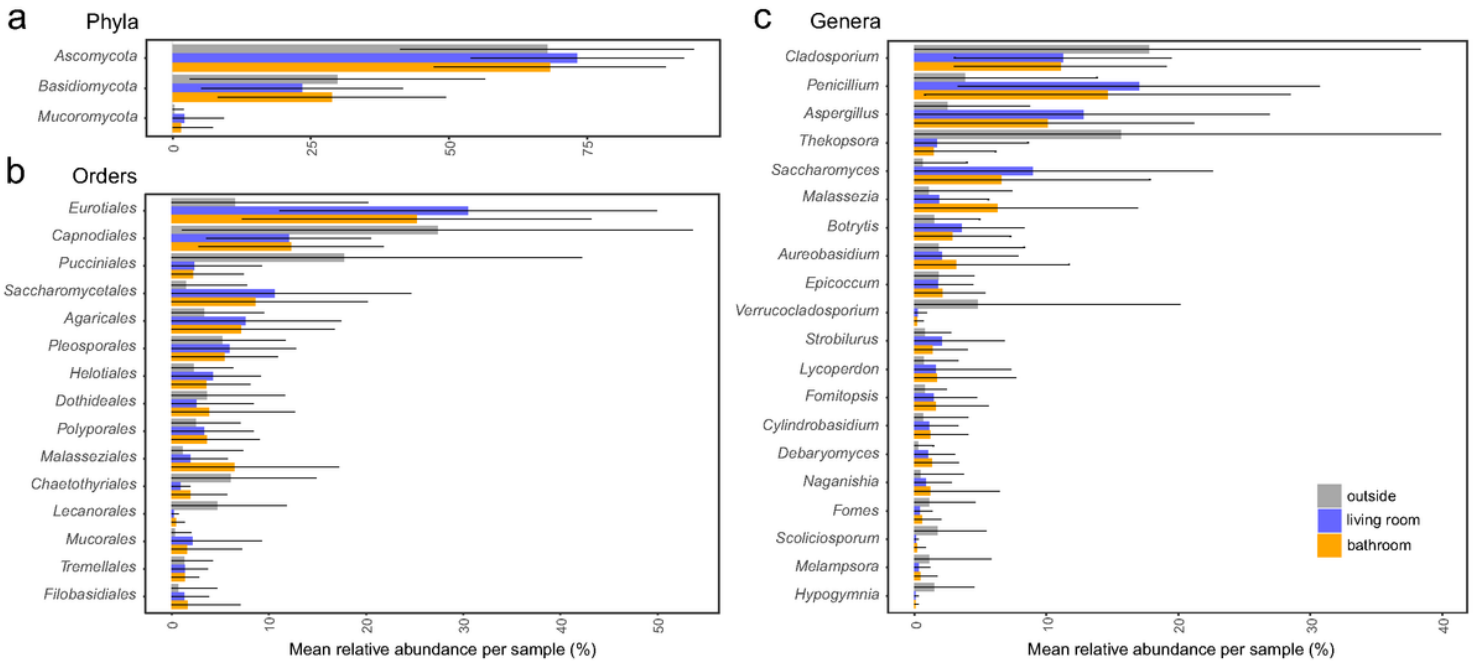


Figure 5

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-3 phyla, (b) top-15 orders, and (c) top-20 genera. Relative abundance of taxa are mean values calculated based on the rarefied datasets. Non-assigned OTUs at the corresponding taxonomic level were excluded.

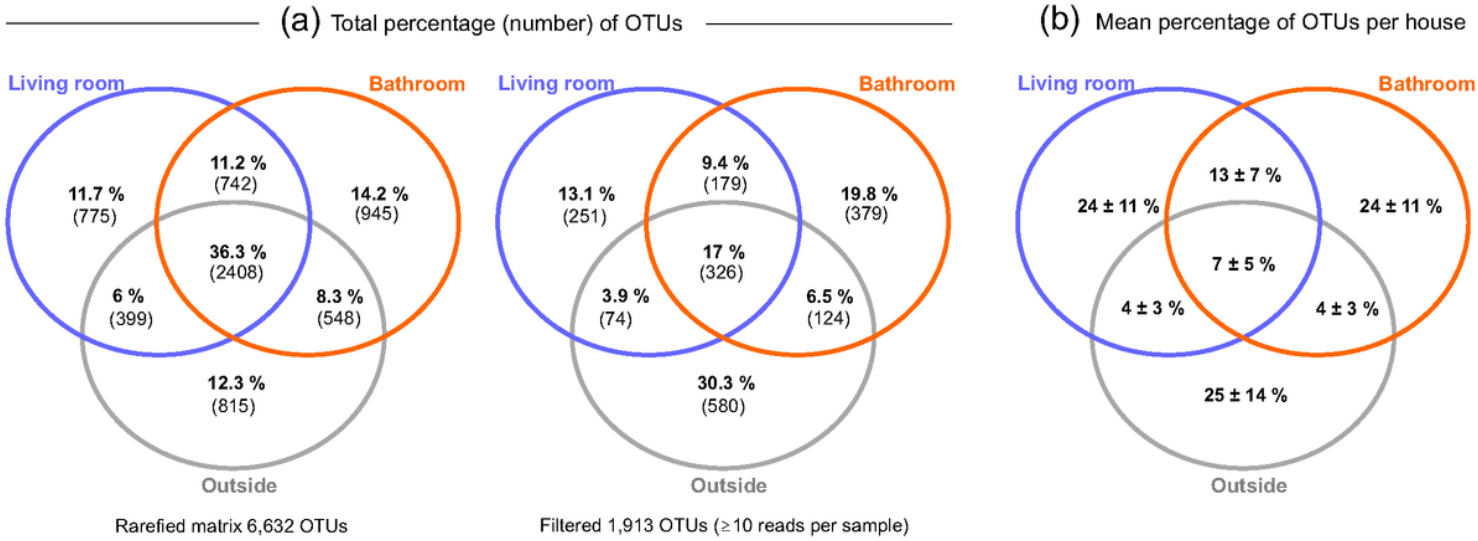


Figure 6

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.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Top200OTUs.xlsx
- Additionalfile1.pdf