Environmental Science & Technology

pubs.acs.org/est

Article

DNA Sequence-Based Approach for Classifying the Mold Status of Buildings

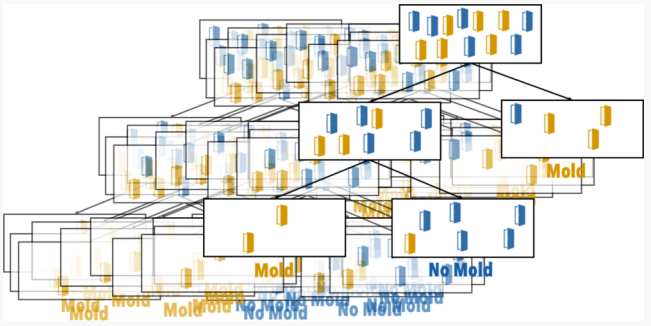
Bridget Hegarty, Annabelle Pan, Ulla Haverinen-Shaughnessy, Richard Shaughnessy, and Jordan Peccia*

Cite This: *Environ. Sci. Technol.* 2020, 54, 15968−15975

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Dampness or water damage in buildings and human exposure to the resultant mold growth is an ever-present public health concern. This study provides quantitative evidence that the airborne fungal ecology of homes with known mold growth ("moldy") differs from the normal airborne fungal ecology of homes with no history of dampness, water damage, or visible mold ("no mold"). Settled dust from indoor air and outdoor air and direct samples from building materials with mold growth were examined in homes from 11 cities across dry, temperate, and continental climate regions within the United States. Community analysis based on the sequence of the internal transcribed spacer region of fungal ribosomal RNA encoding genes demonstrated consistent and quantifiable differences between the fungal ecology of settled dust in homes with inspector-verified water damage and visible mold versus the settled dust of homes with no history of dampness, water damage, or visible mold. These differences include lower community richness (p_{adj} = 0.01) in the settled dust of moldy homes versus no mold homes, as well as distinct community taxonomic structures between moldy and no mold homes (ANOSIM, R = 0.15, p = 0.001). We identified 11 *Ascomycota* taxa that were more highly enriched in moldy homes and 14 taxa from *Ascomycota*, *Basidiomycota,* and *Zygomycota* that were more highly enriched in no mold homes. The indoor air differences between moldy versus no mold homes were significant for all three climate regions considered. These distinct but complex differences between settled dust samples from moldy and no homes were used to train a machine learning-based model to classify the mold status of a home. The model was able to accurately classify 100% of moldy homes and 90% of no mold homes. The integration of DNA-based fungal ecology with advanced computational approaches can be used to accurately classify the presence of mold growth in homes, assist with inspection and remediation decisions, and potentially lead to reduced exposure to hazardous microbes indoors.



INTRODUCTION

Building dampness caused by a water inundation event or high humidity is a common, global occurrence. Climate-driven shifts in regional and seasonal precipitation patterns and sea level rise are expected to compound this problem in many areas of the world.[1,2] The presence of dampness, mold odors, or visible mold growth on building materials has been consistently associated with adverse health outcomes, including respiratory symptoms and the exacerbation and development of asthma.[3,4]

Our current understanding of the specific taxa or ecological characteristics associated with health impacts attributable to water damage and mold indicators (e.g. visible mold, mold odor) may be improved by identifying the fungal ecologies differentially present in the air of "dry" versus "damp/moldy" buildings. These differences can then be leveraged to reveal if water damage and mold growth result in detectable airborne exposures that diverge from a building's normal fungal ecology[5] or indicate if a home's indoor air has been impacted by mold growth from hidden spaces. Such evidence is essential for homeowners, as well as environmental building inspectors and public health researchers when making decisions regarding the health-based need for remediation.

Prior culture- and DNA sequence-based studies on mold and dampness have revealed a useful baseline of mold taxa that commonly grows on a variety of building materials under variable water activities. Important examples of fungi known to grow on damp wood, sheetrock, and ceramic materials include members of the genera *Acremonium*, *Penicillium*, *Stachybotrys*, *Ulocladium*, *Arthrinium*, *Aureobasidium*, *Aspergillus,* and *Mucor*.[6,7] However, detection of these taxa in the air of buildings, where exposure occurs, has not reliably indicated the presence of dampness and mold.[8,9] All of the above taxa naturally occur in the outdoor environment and can thus be

Received: June 29, 2020
Revised: November 19, 2020
Accepted: November 20, 2020
Published: December 1, 2020

transported into the indoor environment.[10] Moreover, the list of mold taxa observed on building materials may extend far beyond those originally identified in culture-based experiments. Setting concentration thresholds for specific taxa or groups of taxa is hampered by the dearth of human dose−response data for mold as well as the temporally dynamic concentrations of fungi in indoor air that are caused by patterns of occupancy, human activity, and building operation.[11,12] Thus, databases or descriptions on what constitutes a normal fungal ecology of a building have not been rigorously established. These must include natural fungal ecology at different climates[10,13] and reflect the full taxonomic resolution that is afforded by DNA-sequencing approaches.

The purpose of this study was to determine the differences in fungal ecologies between the settled dust of inspector-confirmed homes with reported water damage and visible mold observations ("moldy" homes) versus homes with no known history of dampness or mold observations ("no mold" homes). A home sampling campaign was conducted from diverse geographic and climatic regions throughout the U.S. using high-throughput DNA sequencing to identify quantitative fungal ecological differences between the indoor air of no mold versus moldy homes. The resultant microbial community databases were used to develop and validate a machine learning approach to classify a home's fungal ecology as moldy or normal (no mold).

METHODS

Detailed methods are also presented in the Supporting Information.

Sampling Campaign. Samples were collected from diverse climatic regions across the U.S. by local, professional building inspectors. Homes in 11 cities (Atlanta, GA; Orlando, FL; Tulsa, OK; Denver, CO; Phoenix, AZ; Minneapolis, MN; Portsmouth, NH; Portland, OR; Chicago, IL; Boulder, CO; Philadelphia, PA) were further binned into continental (Dfa, Dfb), temperate (Cfa, Csb), and dry climates (Bsh, Bsk, and Bwh) based on their known membership in Köppen climate groups.[14]

Between 5 and 10 single family homes were selected by a building inspector in each city and included both homes with inspector-confirmed water damage and visible mold ("mold" samples), and homes with no history of dampness, water damage or visible mold ("no mold" samples). For each home, the following metadata were collected: home age, home location, home size, brief description of room type where sample was collected, distance between the sample and nearest direct mold (mold homes only), type of building material sampled, observation of odor, current moisture condition of moldy building, area of direct mold, and area of water damage. Evaluation of each home was commensurate in scope with a qualified home inspection by an experienced mold investigation firm. Moldy homes all had verified water damage with visible mold, whereas no mold homes had no signs or known history of dampness, water damage, mold growth, or mold odor.

All samples were collected from surfaces using sterile cotton tipped swabs moistened with filter-sterilized 0.15 M NaCl, 0.1% Tween solution. Two types of samples were taken, and settled dust samples from the tops of doorframes and surface swab samples directly from mold growth on materials were included. A total of four samples were collected in each "no mold" home, with three swab samples collected from the tops of door frames within the home and one sample collected from the outdoor side of a door frame. In "mold" homes, up to nine swab samples were collected in each home, including three indoor doorframe samples, one sampled collected from the outdoor side of a doorframe, and up to five "direct mold" samples collected from the surface of the material(s) with mold growth. Samples were stored at −80 °C within 24 h of receipt and sequenced within 6 months of the sample date.

Fungal Community (Amplicon) Sequencing and Processing. The DNAeasy PowerSoil Kit (Qiagen Inc., Germantown, MD) was used to extract DNA from the cotton swabs.[15] The fungal internal transcribed spacer (ITS) region was amplified using ITS-1F and ITS2 primers.[16] The University of Texas Genomic Sequencing and Analysis Facility (UT GSAF, Austin, TX, USA) carried out library preparation, sequencing, and demultiplexing of fungal sequences, generating 250 base-pair paired end reads. Prior to sequencing, sample DNA concentrations were normalized to an equal concentration, the PCR step in sequence preparation was performed in triplicate, and resulting amplicons were pooled. No template negative controls were included with all sample submissions, and no sequences were returned from these negative controls.

Raw reads were trimmed and filtered and amplicon sequence variants (ASVs) were created using the R software package DADA2.[17] Overall, 288 samples collected from 67 homes passed filtering and were utilized in this study. This included 58 outdoor air-settled dust samples, 59 indoor air-settled dust samples from no mold homes, 58 indoor air-settled dust samples from mold homes, and 113 direct mold samples (Table S1). ASV sequences were then BLASTed against the UNITE database and taxonomic identifications assigned using FHiTINGS, version 1.4.[18,19] A modified DESeq protocol was used to identify taxa that are differentially abundant between mold and no mold indoor air samples.[20]

Random Forest Classification. To classify each indoor door frame sample based on whether its fungal community was more similar to one from a home with or without mold damage, Random Forest (RF) models were created in R using default parameters and 20-fold cross-validation (see the supplemental methods for more details). The first set of RF models was created using the relative abundance data for all taxa. For each iteration, 70 % of the data was used for model development and 30 % of the data was retained for validation to assess the sensitivity of the prediction to the training set. A second RF model using only those taxa that were found to be differentially abundant between mold and no mold was then built using the same process. This second model was used for all subsequent analyses because of improved accuracy for no mold classification. Minimal depths were calculated for each of the 20 cross-validation iterations of the RF model. Any taxa with a minimal depth under the mean minimal depth in at least 19 of the 20 RF iterations was considered an important taxa for the RF classifier.[21]

RESULTS

Homes across six Köppen climate zones (Cfa, Dfa, Dfb, Bsh, Bsk, and Bwh) were sampled. Four categories of samples were collected, including (1) indoor air-settled dust from homes with no history of water damage or visible mold ("no mold"), (2) indoor settled dust from homes with inspector documented water damage and visible mold ("mold"), (3) outdoor settled dust from all homes ("outside"), and (4) direct surface samples from building materials (e.g., wood, ceramic,

https://dx.doi.org/10.1021/acs.est.0c03904
Environ. Sci. Technol. 2020, 54, 15968−15975

sheet rock) with visible fungal growth ("direct mold"). See Figure S1 for a map illustrating the sampling campaign locations.

Richness and Dominant Taxa. We observed differences in α-diversity between the sample types, with mold and direct mold-settled dust samples having a lower richness than no mold and outdoor air-settled dust. Statistically significant differences in the median richness levels (Tukey's test, $p_{adj} < 0.001$) were found between the settled dust samples (outside, no mold, and mold samples) and the direct mold samples (Figure 1). Within the settled dust samples, median richness in mold was lower than no mold and outdoor air samples (Tukey's test, $p_{adj} < 0.05$).

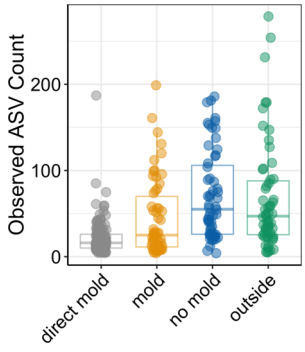


Figure 1. Fungal richness scatterplot based on ASV count. Boxplots (25th quartile, median, 75th quartile) are overlaid with each scatterplot. Samples are colored by their sample type: direct mold—gray, mold—orange, no mold—blue, and outside—green.

Direct mold samples tend to be dominated by a single, highly abundant taxa (ASV) (Figure S2a). The median relative abundance of the top-ranked ASV is 69% for direct mold and 64% for mold, dropping to 45% for outside and 32% for no mold. A total of 54 different genera were the top taxa in at least one sample. Across all sample types, *Aspergillus*, *Cladosporium*, *Penicillium*, *Alternaria*, and *Stachybotrys* are the five most common top ASV genus annotations (Figure S2b). In addition to commonly being the top ASV within a sample, these five taxa were also the top five most common genus annotations among all samples (Figure S3). For the average sample, 32% of the direct mold reads, 35% of mold reads, 14% of no mold reads, and 20% of outside reads are represented by these five genera. The proportion of reads annotated as these five genera was statistically lower in no mold than either direct mold or mold (Tukey's test, $p_{adj} = 0.006$ and $p_{adj} = 0.002$, respectively). *Stachybotrys* was the only genera that was the top annotation in only one sample type (direct mold). *Aspergillus* and *Penicillium* were more commonly (by ~ 2 times) the top genera in mold compared to no mold samples. The difference in relative abundance between mold and no mold is statistically significant for the samples where *Aspergillus* was the top genera (Tukey's test, $p_{adj} = 0.05$), but not when *Penicillium* was the top genera (Tukey's test, $p_{adj} = 0.31$).

A quantitative comparative analysis revealed several ASVs that were differentially enriched between no mold and mold-settled dust (Figures 2 and S4, Tables S2 and S3). A total of 11 ASVs, all *Ascomycota*, were statistically overabundant in the mold homes, while 14 distinct ASV, covering *Ascomycota*, *Basidiomycota*, and *Zygomycota*, were statistically overabundant in no mold homes. The genera *Aspergillus*, *Penicillium*, and

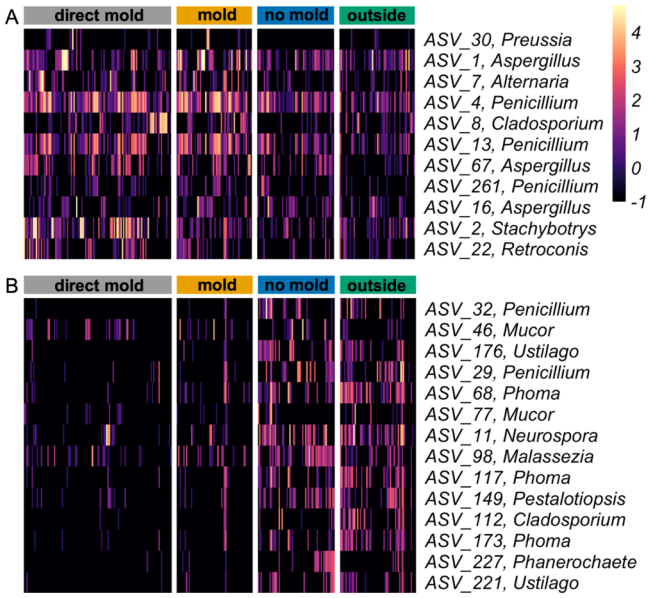


Figure 2. Heatmap of the statistically significant $\log_2$ fold ($\log_2$F) changes between mold and no mold indoor air-settled dust ($p_{adj} \leq 0.01$): (A) ASVs that were more abundant among the mold samples than the no mold samples and (B) ASVs that were more abundant among the no mold samples than the mold samples. Heatmap is colored by the $\log_{10}$ relative abundances with the most highly abundant taxa in yellow and those taxa not present in a sample in black.

Cladosporium have ASVs that are differentially expressed in opposite directions (Figure S4). For *Penicillium*, ASVs 4 and 13 (ambiguous at the species level) and ASV 261 (top BLAST hit *Penicillium aurantiogriseum*) were found to be more common in mold than no mold, with relative abundance (mold/no mold) $\log_2$ fold changes ($\log_2$F) of 3.4, 3.5, and 3.1, respectively. ASVs 29 and 32 (*Penicillium oxalicum* top BLAST hit for both) were found to be more common in no mold than in mold, with (mold/no mold) $\log_2$F changes of −5.9 and −11.0, respectively. For *Cladosporium*, ASV 8 (top BLAST hit *Cladosporium halotolerans*) was more highly abundant in mold than no mold, with a 5.3 (mold/no mold) $\log_2$F change, while ASV 112 (top BLAST hit *Cladosporium sphaerospermum*) was more abundant in the no mold than the mold samples with a (mold/no mold) −6.4 $\log_2$F change. For *Aspergillus*: ASVs 1, 16, and 67 (top BLAST hits *Aspergillus niger*, *Aspergillus piperis*, and *Aspergillus subversicolor*, respectively) were found to be more common in mold than no mold, with relative abundance (mold/no mold) $\log_2$F changes of 8.4, 2.9, and 4.3, respectively.

Fungal Communities Cluster by Sample Type. Ordination plots based on Bray−Curtis dissimilarities demonstrate differences in fungal community composition for different sample types (ANOSIM $R = 0.11$, $p = 0.001$) (Figure S5). No mold, outdoor, and direct mold samples also demonstrated distinct differences based on Köppen Region (dry—Bsh, Bsk, Bwh, temperate—Cfa, and continental—Dfa, Dfb) (ANOSIM $p \leq 0.001$); however, mold samples did not (ANOSIM $p = 0.215$) (Figure S6).

Bray−Curtis dissimilarity ordination plots (Figure S7A) also revealed community differences between mold and no mold samples (ANOSIM $R = 0.15$, $p = 0.001$). Additional comparisons based on settled dust comparisons using only those taxa that were differentially abundant between mold and

https://dx.doi.org/10.1021/acs.est.0c03904
Environ. Sci. Technol. 2020, 54, 15968−15975

no mold (Tables S2 and S3) improved the clustering (Figure S7B) (ANOSIM R = 0.17, p = 0.001). When separating by Köppen climate region, distinction between mold and no mold communities retained their statistical significance (ANOSIM p < 0.01) for each region (Figure S7 C−E).

Intrahome Comparisons. Intrahome comparisons control for climate and the myriad home-specific factors including but not limited to occupancy, construction type, home age, cleaning practices, and ventilation. The indoor air-settled dust ecology of moldy homes is expected to be influenced by both direct mold and outdoor air mold, while the indoor air-settled dust of no mold homes should have no influence from direct mold taxa. Figure 3A demonstrates that the number of ASVs in common between a given inside sample and that home's outside sample is higher for no mold compared to mold homes (Wilcoxon rank sum test, p < 0.001). Within mold homes, there are more ASVs in common between mold and direct mold-settled dust than between outside air and direct mold (Wilcoxon rank sum test, p = 0.03) (Figure 3B). A linear regression model of the number of ASVs in common between direct mold and mold-settled dust samples versus proximity of a mold sample to direct mold reveals that the effect of direct mold on settled dust decreases only slightly with distance from mold damage (R^2 = 0.05; $F(1,53)$ = 3.8; p = 0.05) (Figure 3C). Finally, multiple direct mold, mold, and no mold samples were taken for each home, allowing for intra- versus interhome comparisons of variability. Based on Bray−Curtis dissimilarities between samples, variability between similar sample types from the same home (intrahome) is statistically lower (t-test, p < 0.001) than the variability between similar sample types in different homes (interhome) (Figure S8).

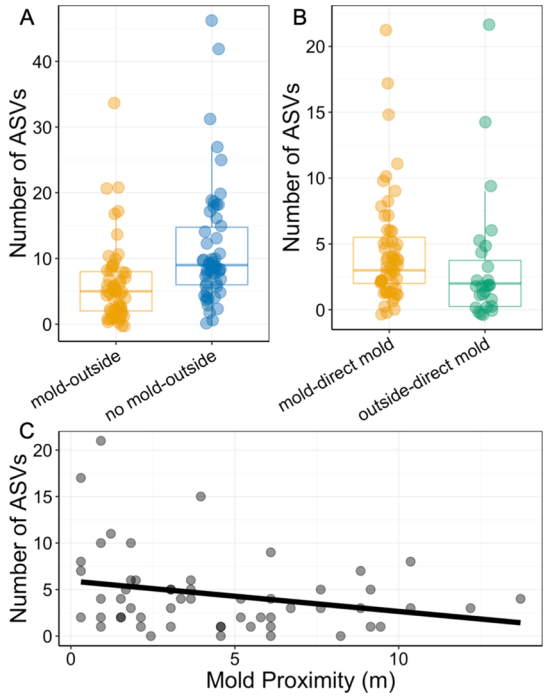


Figure 3. Comparing the influence of direct mold and outside air on indoor fungal communities. (A) Boxplot overlaid with a scatterplot of the number of ASVs in common for each indoor dust sample with the corresponding outdoor dust sample from the same home. There was a statistically significant difference between mold-outside and no mold-outside (Wilcoxon p < 0.001). (B) Boxplot overlaid with a scatterplot of the number of ASVs in common between mold and direct mold, as well as outside and direct mold for each house with mold damage. There was a statistically significant difference between mold-direct mold and outside-direct mold (Wilcoxon p = 0.03). (C) Scatter plot of the number of mold home air taxa also found in a direct mold sample from that home versus proximity to the nearest mold sample. The plot was overlaid with a line representing a linear model of the number of ASVs in common versus mold proximity. The equation of the line was found to be $y = -0.3x + 5.9$ (adjusted R^2 = 0.05, $F(1,53)$ = 3.8, p = 0.05).

RF Model Development and Optimization. The large ecology data sets and the potentially subtle differences that have been identified in fungal ecology between homes with visible mold and no mold homes suggest that the task of classifying a home as a moldy ecology or a normal (no mold) ecology is well-suited to machine learning algorithms, such as RF.[22,23] An initial cross-validated RF model built using all taxa had an average classification accuracy of 82 ± 10% across all samples, and the majority (68%) of the samples are accurately assigned each time they are in the test set during the 20-fold cross-validation (Figure 4A). The greatest source of error originates from misclassifying no mold samples as mold; the average accuracy for the mold samples is 86 ± 8 and 70 ± 12% for the no mold samples. Using only the differentially abundant taxa in the machine learning training sets improves the classification accuracy of the RF model (Figure 4B) for no mold homes from 70 ± 12 to 83 ± 9%.

Of the total 22 homes that contained at least three settled dust samples, 21 are correctly classified as moldy or no mold using the RF model based on the differentially abundant taxa and the requirement that at least two out of the three settled dust samples for that home are correctly classified. All (12 of 12) of the moldy homes were correctly classified when using this benchmark (Figure 5A), while most (9 of 10) of the no mold homes were correctly classified (Figure 5B). Focusing on the no mold home (Minneapolis-3) that was misclassified in more than 50% of the 20-fold validation RF models for two of the three samples in that home, the misclassified samples have an abundance of ASVs (e.g., ASV 4 and ASV 13) ascribed to mold homes.

Minimal depth of variables was used to determine which taxa were most important for classifying between mold and no mold samples (Figure S9). Eight ASVs were deemed important in sufficient model iterations: *Pestalotiopsis* (ASV 149), *Neurospora* (ASV 11), *Penicillium* (ASV 29), and *Malassezia* (ASV 98) which were more abundant in no mold than mold and *Aspergillus* (ASV 1), *Penicillium* (ASVs 13 and 4), and *Cladosporium* (ASV 8) which were more abundant in mold than no mold.

The RF model output sufficient results (Figure 4A) to explore ecological patterns that may underlie these incorrect predictions. Classification accuracy is influenced by patterns for top ASV count, number of ASVs in a sample, and degree of overlap with outside settled dust. For the samples accurately classified in more than 75% of the cross-validation models, the relative abundance of the top ASV (top ASV count) was higher in mold than no mold (t-test, p = 0.001). Conversely, a higher, but not statistically significant (t-test, p = 0.12) top ASV count was observed in no mold than mold for the samples that the RF model predicts accurately less than 25% of the time. Richness was higher in the no mold versus mold for the samples that the RF model predicts correctly more than 75% of the time (t-test, p < 0.001), while no difference richness was

https://dx.doi.org/10.1021/acs.est.0c03904
Environ. Sci. Technol. 2020, 54, 15968−15975

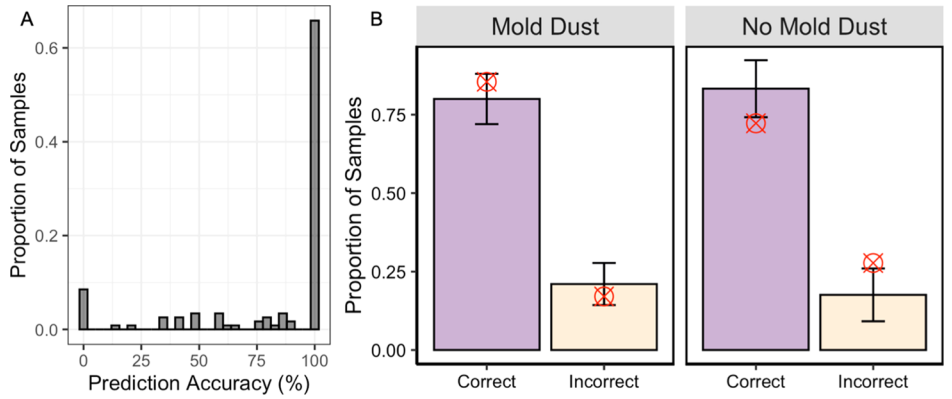


Figure 4. (A) Histogram of the number of correct predictions (converted to prediction accuracy %) for each sample based on 20-fold cross-validation of the RF model. (B) Performance of the 20-fold cross-validated RF model on the subset of ASVs that had a statistically significant change in abundance from mold to no mold. The red circles overlaid with an X show the mean percent accuracy for the RF model using all ASVs within samples.

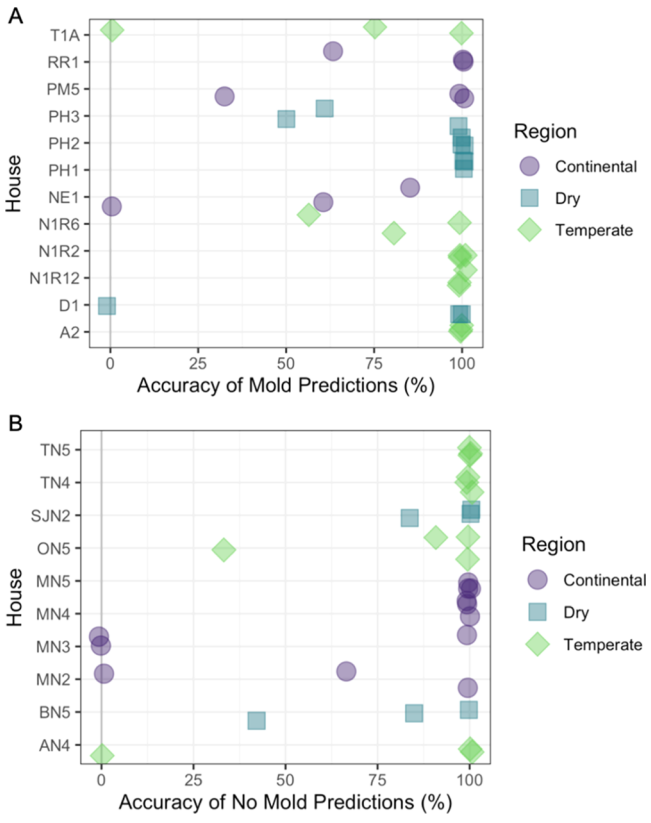


Figure 5. Prediction accuracy for homes with three samples. (A) Mold homes. (B) No mold homes. Samples are colored based on Köppen Region.

detected for the samples that the RF model classifies accurately less than 25% of the time (*t*-test, $p = 0.68$). Regarding overlap with outdoor air, a higher number of ASVs are in common between the inside dust sample and corresponding outside sample in the no mold-settled dust than in the mold-settled dust samples, when considering samples that the RF model predicts correctly more than 75% of the time (*t*-test, $p < 0.001$). Mold and outside are equally similar for those samples that the RF accurately classifies less than 25% of the time (*t*-test, $p = 0.66$). Each of these trends in RF prediction reflects observations from the comparative DNA sequence analysis.

DISCUSSION

This study uniquely integrates DNA sequence-based ecological approaches with modern computational biology and a multiclimate sampling design to determine the ecological differences between fungi settled from air of inspector-verified moldy and not moldy homes. Quantitative ecological differences between mold and no mold indoor settled dust samples were leveraged to train a RF machine learning model that accurately classified a home's airborne fungal exposure as moldy versus normal. This study design and resultant finding on the ecological differences between inspector-identified moldy and not moldy homes is novel and represents two significant advances for preventing human health impacts from damp or water-damaged buildings. First, our comparative approach identified the complex airborne microbial community differences between homes with visible mold and homes with no water damage or visible mold and also defined characteristics of the normal fungal ecology in home with no water damage or visible mold. Second, the tools and insights derived can be practically applied by mold inspection and the remediation industry for improved guidance to determine if the fungal exposures in buildings are associated with hidden or visible mold growth on damp or water-damaged building materials.

Unique Fungal Community Characteristics Common to Moldy Homes. Results from this study demonstrate how mold growth on damp or water-damaged building materials is reflected in indoor air, where human exposure occurs and how these exposures differ from a normal fungal ecology. Indoor air samples in mold homes had a lower richness than no mold samples and were more likely to contain highly abundant taxa. These patterns are consistent with recent studies on the indoor air of a single homes[24] and direct mold from damp building materials,[8,9,25] that suggest that the presence of mold growth on damp materials depresses the richness of airborne fungi that occupants are exposed. Exposure to low fungal and bacterial richness in early life has been empirically associated with asthma development.[26−28] The presence of visible mold, dampness, and mold odors has also been associated with asthma development.[3,29] The observation herein of reduced fungal richness in moldy versus not moldy homes links the low diversity−asthma associations to mold−asthma associations.

The dominant genera in mold samples belonged mostly to fungi that have commonly been identified in prior culture-

https://dx.doi.org/10.1021/acs.est.0c03904
Environ. Sci. Technol. 2020, 54, 15968−15975

based studies and include *Aspergillus*, *Penicillium*, *Stachybotrys*, *Cladosporium*, and *Alternaria*.[30] ASV-level analysis allowed for deeper insights: the three most common genera (*Aspergillus*, *Penicillium*, and *Cladosporium*) were highly abundant in all sample types, but specific ASVs were differentially abundant between mold and no mold samples. Many of the taxa highly enriched in mold homes (all *Ascomycota*) have demonstrated public health significance; they are known allergens,[31] produce mycotoxins,[32,33] and (in the case of *Penicillium*[34]) have been implicated as important for asthma development. The taxa highly abundant in no mold homes included members of *Ascomycota*, *Basidiomycota*, and *Zygomycota*. No mold ASV's are commonly found on human skin (*Malassezia*) or are known to be common in the outdoor environment (e.g., *Phoma*,*Cladosporium*, and *Pestalotiopsis*), thus comprising important members of a building's normal ecology.

Beta diversity analysis demonstrated fungal community membership differences between mold and no mold homes, strong differences between direct mold and outdoor air, and the importance of considering climate zone. Climate (temperate, dry, and continental) appeared to impact fungal ecology, not only in outdoor air, but for indoor air in homes without mold damage, and even direct mold. There is precedent for observing different fungal communities in buildings and outdoor air based on geography.[10,11,13] These differences are believed to exist because of climate-specific outdoor plants and their associations with fungal endophytes, mycorrhizae, pathogens, or saprophytes. The community differences between mold and no mold homes were consistent, even when accounting for the three climate regions considered. However, differences between mold and no mold communities were slightly stronger when the data were stratified into continental, temperate, and dry climates.

Tools for Classifying Homes as Moldy or Not Moldy. The use of a small, inflexible set of indicator taxa to classify the moldiness of a building can lead to misclassifications. For example, building inspectors commonly utilize the abundance of *Aspergillus* or *Penicillium/Aspergillus* to assess whether a harmful mold exposure exists. Xerophilic members of these genera grow on building materials with elevated water activity, have known allergic impacts on humans, and can be identified via culture or direct microscopy. However, these tests typically supply genus level identifications, and neither the *Aspergillus* (*t*-test, $p = 0.98$) nor *Penicillium* genera (*t*-test, $p = 0.42$) abundances were different between the mold and no mold homes surveyed in this study. While *Aspergillus* and *Penicillium* genera abundances are indistinguishable between mold and no mold homes, both contain ASVs that are differentially abundant between mold and no mold homes. The RF model constructed here illustrates the benefit of considering the entire variety of fungal taxa—either directly in the RF model or to tune the RF model—and the necessity of moving beyond genera level identification. Recent DNA-barcoding studies have revealed that the fungal diversity of the built environment is more diverse than previously accounted for by prior culture approaches,[13,35,36] and RF is a common, highly robust machine learning strategy particularly well-suited to classification problems in ecology.[22] The use of multiple predictions per sample, multiple samples per home, and a selected group of taxa allowed for independently and correctly identifying 100% of mold homes and 90% of no mold homes. RF can be applied to a large number of taxa and other sample characteristics simultaneously and allows for the continued evolution of a classification tool as additional training data and fungal ecology insights are made available.

Study Limitations. An important limitation of this study is the inherent environmental and ecological differences from home to home. We observed that the variability of intrahome samples were significantly smaller that the interhome variability. Multiple environmental variables, including the presence and use of AC, pets, occupancy levels, flooring materials, and outdoor land use[37−39] may influence the mycology of a home, and these likely confound and may have weakened the effect we observed between mold and no mold. At the current level of understanding of how these different factors influence fungal communities, accounting for all of these potential dependencies in a classification model is aspirational.

The RF model presented here, while developed based on a diverse subset of homes, relies on underlying patterns in taxa that may have different abundance patterns for regions not yet surveyed. Whether the classification model presented has broader utility beyond the cities sampled depends on its performance on samples from additional locations, particularly ones that are from Köppen regions not represented in this sampling campaign. The studies of comparative analyses and investigation of incorrect classifications also suggest that incorporating ecological characteristics, such as richness or indoor air taxa overlap with outdoor air or direct mold, may lead to a more robust and broadly applicable platform by leveraging the additional ecological differences observed between mold and no mold indoor air that do not rely on specific taxa identification and comparisons. Finally, we acknowledge some inherent uncertainty and ambiguity in the inspector-based classification of homes as moldy or not moldy. While inspection can accurately classify a home as moldy based on visible mold, the history or extent of that mold may not be well documented. The classification of a no mold homes relies upon occupant interviews and inspection at the time of sampling. The complete history of water damage, especially in an older home, may not be known.

Summary. Historically, it has been difficult to classify the mold status or health risks of a home based on measurements of indoor air fungal concentrations or characteristics. This study used DNA-sequencing approaches to reveal the complex ecological differences in fungal communities in professionally inspected moldy and not moldy homes. These differences motivated the development a machine learning-based tool to classify the mold status of a home with approximately 95% accuracy. This tool, coupled with an improved understanding of the airborne fungal ecology characteristics that are associated with the presence of building dampness or water damage comprises a novel contribution toward identifying buildings with fungal communities that may pose a health hazard.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.est.0c03904.

Detailed methods, number of samples by region, differentially abundant taxa tables, U.S. climate classifications, top ASV abundances, fungal genera heatmap, fold-changes in abundances, NDMS plots of Bray−

https://dx.doi.org/10.1021/acs.est.0c03904
Environ. Sci. Technol. 2020, 54, 15968−15975

Curtis dissimilarities, intra- versus interhome comparisons, and minimal depth plot (PDF)

AUTHOR INFORMATION

Corresponding Author

Jordan Peccia − *Department of Chemical and Environmental Engineering, Yale University, New Haven, Connecticut 06520-8286, United States;* orcid.org/0000-0002-6482-2084; Phone: +1 (203) 432-4385; Email: Jordan.Peccia@Yale.edu

Authors

Bridget Hegarty − *Department of Chemical and Environmental Engineering, Yale University, New Haven, Connecticut 06520-8286, United States*

Annabelle Pan − *Department of Chemical and Environmental Engineering, Yale University, New Haven, Connecticut 06520-8286, United States*

Ulla Haverinen-Shaughnessy − *Indoor Air Program, The University of Tulsa, Tulsa, Oklahoma 74101-9700, United States*

Richard Shaughnessy − *Indoor Air Program, The University of Tulsa, Tulsa, Oklahoma 74101-9700, United States*

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.est.0c03904

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This study was funded by the U.S. Department of Housing and Urban Development. We thank the team of professional building inspectors for their contributions to sample collection and home evaluation.

REFERENCES

(1) Groisman, P. Y.; Knight, R. W.; Easterling, D. R.; Karl, T. R.; Hegerl, G. C.; Razuvaev, V. N. Trends in intense precipitation in the Climate Record. *J. Clim.* **2005**, *18*, 1326−1350.

(2) Hirabayashi, Y.; Mahendran, R.; Koirala, S.; Konoshima, L.; Yamazaki, D.; Watanabe, S.; Kim, H.; Kanae, S. Global flood risk under climate change. *Nat. Clim. Change* **2013**, *3*, 816−821.

(3) Mendell, M. J.; Mirer, A. G.; 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*, 748−756.

(4) Institute of Medicine. *Damp Indoor Spaces and Health*; National Academies Press: Washington, DC, 2004.

(5) American National Standards Institute/Institute of Inspection Cleaning and Restoration Certification. *ANSI/IICRC S520: 2015 Reference Guide for Professional Mold Remediation*; Institute of Inspection, Cleaning and Restoration Certification, (S520). 2015.

(6) Andersen, B.; Frisvad, J. C.; Søndergaard, I.; Rasmussen, I. S.; Larsen, L. S. Associations between fungal species and water-damaged building materials. *Appl. Environ. Microbiol.* **2011**, *77*, 4180−4188.

(7) Gravesen, S.; Nielsen, P. A.; Iversen, R.; Nielsen, K. F. Microfungal contamination of damp buildings–examples of risk constructions and risk materials. *Environ. Health Perspect.* **1999**, *107*, 505−508.

(8) An, C.; Yamamoto, N. Fungal compositions and diversities on indoor surfaces with visible mold growths in residential buildings in the Seoul Capital Area of South Korea. *Indoor Air* **2016**, *26*, 714−723.

(9) Lax, S.; Cardona, C.; Zhao, D.; Winton, V. J.; Goodney, G.; Gao, P.; Gottel, N.; Hartmann, E. M.; Henry, C.; Thomas, P. M.; Kelley, S. T.; Stephens, B.; Gilbert, J. A. Microbial and metabolic succession on common building materials under high humidity conditions. *Nat. Commun.* **2019**, *10*, 1767.

(10) Barberán, A.; Dunn, R. R.; Reich, B. J.; Pacifici, K.; Laber, E. B.; Menninger, H. L.; Morton, J. M.; Henley, J. B.; Leff, J. W.; Miller, S. L.; Fierer, N. The ecology of microscopic life in household dust. *Proc. R. Soc. London, Ser. B* **2015**, *282*, 20151139.

(11) Yamamoto, N.; Hospodsky, D.; Dannemiller, K. C.; Nazaroff, W. W.; Peccia, J. Indoor Emissions as a Primary Source of Airborne Allergenic Fungal Particles in Classrooms. *Environ. Sci. Technol.* **2015**, *49*, 5098−5106.

(12) Licina, D.; Tian, Y.; Nazaroff, W. W. Emission rates and the personal cloud effect associated with particle release from the perihuman environment. *Indoor Air* **2017**, *27*, 791−802.

(13) Amend, A. S.; Seifert, K. A.; Samson, R.; Bruns, T. D. 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*, 13748−13753.

(14) Kottek, M.; Grieser, J.; Beck, C.; Rudolf, B.; Rubel, F. World Map of the Köppen-Geiger climate classification updated. *Meteorol. Z.* **2006**, *15*, 259−263.

(15) Yamamoto, N.; Bibby, K.; Qian, J.; Hospodsky, D.; Rismani-Yazdi, H.; Nazaroff, W. W.; Peccia, J. Particle-size distributions and seasonal diversity of allergenic and pathogenic fungi in outdoor air. *ISME J.* **2012**, *6*, 1801−1811.

(16) White, T. J.; Bruns, T.; Lee, S.; Taylor, J. Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. *PCR Protoc.* **1990**, *8* (1), 315−322.

(17) Callahan, B. J.; McMurdie, P. J.; Rosen, M. J.; Han, A. W.; Johnson, A. J. A.; Holmes, S. P. DADA2: High-resolution sample inference from Illumina amplicon data. *Nat. Methods* **2016**, *13*, 581−583.

(18) Kõljalg, U.; Nilsson, R. H.; Abarenkov, K.; Tedersoo, L.; Taylor, A. F. S.; Bahram, M.; Bates, S. T.; Bruns, T. D.; Bengtsson-Palme, J.; Callaghan, T. M.; Douglas, B.; Drenkhan, T.; Eberhardt, U.; Dueñas, M.; Grebenc, T.; Griffith, G. W.; Hartmann, M.; Kirk, P. M.; Kohout, P.; Larsson, E.; Lindahl, B. D.; Lücking, R.; Martín, M. P.; Matheny, P. B.; Nguyen, N. H.; Niskanen, T.; Oja, J.; Peay, K. G.; Peintner, U.; Peterson, M.; Põldmaa, K.; Saag, L.; Saar, I.; Schüßler, A.; Scott, J. A.; Senés, C.; Smith, M. E.; Suija, A.; Taylor, D. L.; Telleria, M. T.; Weiss, M.; Larsson, K.-H. Towards a unified paradigm for sequence-based identification of fungi. *Mol. Ecol.* **2013**, *22*, 5271−5277.

(19) Dannemiller, K. C.; Reeves, D.; Bibby, K.; Yamamoto, N.; Peccia, J. Fungal High-throughput Taxonomic Identification tool for use with Next-Generation Sequencing (FHiTINGS). *J. Basic Microbiol.* **2014**, *54*, 315−321.

(20) Love, M. I.; Huber, W.; Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **2014**, *15*, 550.

(21) Ishwaran, H.; Kogalur, U. B.; Gorodeski, E. Z.; Minn, A. J.; Lauer, M. S. High-dimensional variable selection for survival data. *J. Am. Stat. Assoc.* **2010**, *105*, 205−217.

(22) Cutler, D. R.; Edwards, T. C., Jr.; Beard, K. H.; Cutler, A.; Hess, K. T.; Gibson, J.; Lawler, J. J. Random Forests for Classification in Ecology. *Ecology* **2007**, *88*, 2783−2792.

(23) Richardson, M.; Gottel, N.; Gilbert, J. A.; Lax, S. Microbial similarity between students in a common dormitory environment reveals the forensic potential of individual microbial signatures. *mBio* **2019**, *10*, e01054.

(24) Hegarty, B.; Haverinen-Shaughnessy, U.; Shaughnessy, R. J.; Peccia, J. Spatial gradients of fungal abundance and ecology throughout a damp building. *Environ. Sci. Technol. Lett.* **2019**, *6*, 329−333.

(25) Dannemiller, K. C.; Weschler, C. J.; Peccia, J. Fungal and bacterial growth in floor dust at elevated relative humidity levels. *Indoor Air* **2017**, *27*, 354−363.

(26) Dannemiller, K. C.; Mendell, M. J.; Macher, J. M.; Kumagai, K.; Bradman, A.; Holland, N.; Harley, K.; Eskenazi, B.; Peccia, J. Next-generation DNA sequencing reveals that low fungal diversity in house

dust is associated with childhood asthma development. *Indoor Air* **2014**, *24*, 236−247.
(27) Ege, M. J.; Mayer, M.; Normand, A.-C.; Genuneit, J.; Cookson, W. O. C. M.; Braun-Fahrländer, C.; Heederik, D.; Piarroux, R.; von Mutius, E. Exposure to environmental microorganisms and childhood asthma. *N. Engl. J. Med.* **2011**, *364*, 701−709.
(28) Tischer, C.; Weikl, F.; Probst, A. J.; Standl, M.; Heinrich, J.; Pritsch, K. Urban dust microbiome: Impact on later atopy and wheezing. *Environ. Health Perspect.* **2016**, *124*, 1919−1923.
(29) Mendell, M. J.; Adams, R. I. The challenge for microbial measurements in buildings. *Indoor Air* **2019**, *29*, 523−526.
(30) Andersson, M. A.; Nikulin, M.; Köljalg, U.; Andersson, M. C.; Rainey, F.; Reijula, K.; Hintikka, E. L.; Salkinoja-Salonen, M. Bacteria, molds, and toxins in water-damaged building materials. *Appl. Environ. Microbiol.* **1997**, *63*, 387.
(31) Kurup, V. P. Fungal allergens. *Curr. Allergy Asthma Rep* **2003**, *3*, 416.
(32) Brasel, T. L.; Martin, J. M.; Carriker, C. G.; Wilson, S. C.; Straus, D. C. Detection of Airborne Stachybotrys chartarum Macrocyclic Trichothecene Mycotoxins in the Indoor Environment. *Appl. Environ. Microbiol.* **2005**, *71*, 7376.
(33) Jarvis, B. B.; Miller, J. D. Mycotoxins as harmful indoor air contaminants. *Appl. Microbiol. Biotechnol.* **2005**, *66*, 367−372.
(34) Kanchongkittiphon, W.; Mendell, M. J.; Gaffin, J. M.; Wang, G.; Phipatanakul, W. Indoor environmental exposures and exacerbation of asthma: an update to the 2000 review by the Institute of Medicine. *Environ. Health Perspect.* **2015**, *123*, 6−20.
(35) Adams, R. I.; Miletto, M.; Taylor, J. W.; Bruns, T. D. Dispersal in microbes: fungi in indoor air are dominated by outdoor air and show dispersal limitation at short distances. *ISME J.* **2013**, *7*, 1262.
(36) An, C.; Yamamoto, N. Fungal compositions and diversities on indoor surfaces with visible mold growths in residential buildings in the Seoul Capital Area of South Korea. *Indoor Air* **2016**, *26*, 714−723.
(37) Adams, R. I.; Bateman, A. C.; Bik, H. M.; Meadow, J. F. Microbiota of the indoor environment: a meta-analysis. *Microbiome* **2015**, *3*, 1−18.
(38) Dannemiller, K. C.; Gent, J. F.; Leaderer, B. P.; Peccia, J. Indoor microbial communities: Influence on asthma severity in atopic and nonatopic children. *J. Allergy Clin. Immunol.* **2016**, *138*, 76−83.
(39) Dunn, R. R.; Fierer, N.; Henley, J. B.; Leff, J. W.; Menninger, H. L. Home life: Factors structuring the bacterial diversity found within and between homes. *PloS One* **2013**, *8*, No. e64133.

https://dx.doi.org/10.1021/acs.est.0c03904
Environ. Sci. Technol. 2020, 54, 15968−15975