Received: 1 December 2017 | Accepted: 24 February 2018

DOI: 10.1111/ina.12459

Check for updates

ORIGINAL ARTICLE

WILEY

Gene expression of indoor fungal communities under damp building conditions: Implications for human health

B. Hegarty[1] | K. C. Dannemiller[2,3] | J. Peccia[1]

[1]Department of Chemical and Environmental Engineering, Yale University, New Haven, CT, USA

[2]Department of Civil, Environmental, and Geodetic Engineering, College of Engineering, The Ohio State University, Columbus, OH, USA

[3]Division of Environmental Health Sciences, College of Public Health, The Ohio State University, Columbus, OH, USA

Correspondence
Jordan Peccia, Department of Chemical and Environmental Engineering, Yale University, New Haven, CT, USA.
Email: Jordan.Peccia@Yale.edu

Funding information
National Science Foundation; Alfred P. Sloan Foundation

Abstract

Dampness and visible mold growth in homes are associated with negative human health outcomes, but causal relationships between fungal exposure and health are not well established. The purpose of this study was to determine whether dampness in buildings impacts fungal community gene expression and how, in turn, gene expression may modulate human health impacts. A metatranscriptomic study was performed on house dust fungal communities to investigate the expression of genes and metabolic processes in chamber experiments at water activity levels of 0.5, 0.85, and 1.0. Fungi at water activities as low as 0.5 were metabolically active, focusing their transcriptional resources on primary processes essential for cell maintenance. Metabolic complexity increased with water activity where communities at 1.0 displayed more diverse secondary metabolic processes. Greater gene expression at increasing water activity has important implications for human health: Fungal communities at 1.0 a_w upregulated a greater number of allergen-, mycotoxin-, and pathogenicity-encoding genes versus communities at 0.85 and 0.5 a_w. In damp buildings, fungi may display increases in secondary metabolic processes with the potential for greater per-cell production of allergens, toxins, and pathogenicity. Assessments in wet versus dry buildings that do not account for this elevated health impact may not accurately reflect exposure.

KEYWORDS

allergens, building dampness, gene expression, health effects, metatranscriptomics, mold

1 | INTRODUCTION

Dampness and the presence of visible mold inside buildings have been consistently associated with respiratory and allergic health effects in infants, children, and adults.[1,2] These connections implicate exposure to fungi grown on damp building materials as the mediator of health effects; however, causal relationships between exposure to microbial concentrations and disease are not well established.[3] Traditional approaches for assessing fungal exposure in buildings incorporate counts of whole cells or spores, culturable spores, or fungal indicators, such as glucans and ergosterol. These assessments do not account for fungal ecology or the variable physiological status of fungal cells between damp and dry buildings.

Fungi present on building materials with elevated water availability (measured as water activity, a_w) may express genes leading to an enhanced, but largely unaccounted for, ability to cause disease in humans. Prior pure culture fungal studies demonstrate that environmental factors including temperature and the elevated presence of NO_2, CO_2, and O_3 during spore development result in physiological changes, differential gene expression, and modulation in pathogenicity and allergenicity.[4-10] Increasing water activity levels in growth media have resulted in elevated emissions of mycotoxins and volatile organic compounds (VOCs) from *Aspergillus* spp. and *Muscodor albus*, respectively.[11-13]

Metatranscriptomics, or profiling community-wide gene expression, provides a snapshot of the genetic activity of the organisms present in a particular sample. As changes in gene expression are

© 2018 John Wiley & Sons A/S. Published by John Wiley & Sons Ltd

typically indicative of changes in metabolism and physiology,[7,11,14] transcriptomic methods may broadly reveal associations between water activity and the expression of toxicity, pathogenicity, and allergenicity genes. Such information allows for quantitative observations on the metabolic status and potential human health implications of fungal communities in the built environment. Thus, researchers can investigate whether, on a per-spore or per-biomass basis, exposure to fungal communities present in damp buildings may result in greater human health impacts in comparison with fungi present in dry buildings. We performed a de novo metatranscriptomic study on dust embedded in carpets incubated at water activities of 0.5, 0.85, and 1.0. Building upon previous studies that focused on identifying fungal taxa that grow on building materials,[15-20] this work uses transcriptomics to uniquely extend the study of fungal growth under damp building conditions toward understanding the biochemical and pathogenic status of fungal communities.

Practical Implications

- Traditional approaches for assessing fungal exposure in buildings incorporate counts of whole cells or spores, culturing spores, or the use of fungal indicators.
- These assessments do not account for the variable physiological status of fungal cells.
- The data presented here demonstrate that at higher water activities, fungi grown in house dust may upregulate more allergen-encoding genes, general pathogenicity pathways, and secondary metabolites, suggesting multiple ways in which the health effects of fungi may depend on the water activity at which they are grown.

2 | MATERIALS AND METHODS

2.1 | Water activity chamber experiments

Used carpet and dust were collected from a home in eastern Massachusetts, USA, in September 2014, and dust, carpet, and fungal communities were initially characterized.[15] The carpet was a 10-year-old, medium-pile, nylon material. Exposure of the carpet and associated dust to elevated moisture conditions resulted in the growth of microbes associated with damp buildings; documented fungal growth rates from fungi on these dust samples for 0.85 a_w and 1.0 a_w were 1.1×10^4 and 1.5×10^5 spore equivalents d^{-1} mg $dust^{-1}$, respectively. No growth was observed for fungi at 0.5 a_w. *Wallemia* was prevalent at all water activities, with the highest relative abundance at 0.85 a_w. As water activity increases from 0.85 to 1.0, *Aspergillus* becomes increasingly enriched.[15]

For this study, 500 mg of house dust was embedded in 10 cm × 10 cm carpet coupons using a modification of ASTM method F608-13 and incubated in glass chambers at 25°C and either 0.5, 0.85, or 1.0 a_w. Water activity was controlled by the presence of a NaCl or $MgCl_2$ salt solution placed in each chamber and measured using an AquaLab series 3 water activity meter (Decagon Devices, Pullman, WA, USA). Triplicate coupons were produced for each water activity. After a 1-week incubation, dust was collected from the 9 coupons using a 19 × 90 mm Whatman cellulose thimble attached to a Eureka Mighty Mite vacuum. Upon dust collection, RNA was immediately extracted to prevent degradation.

2.2 | RNA extraction and nucleic acid sequencing

RNA was extracted from each dust sample using the MoBio PowerMicrobiome RNA Extraction Kit (MoBio, Carlsbad, CA, USA). The manufacturer's protocol was modified to use 10× the recommended concentration of beta-mercaptoethanol to further inhibit RNA degradation. Following extraction, RNA was reverse-transcribed and cDNA library preparation was carried out at Yale University's Center for Genomic Analysis using their polyA selection protocol to select only eukaryotic DNA. Samples were sequenced using Illumina's HiSeq 4000 sequencer, and 75-base pair paired-end reads were generated. Sequences were archived in the European Nucleotide Archive and can be accessed here: http://www.ebi.ac.uk/ena/data/view/PRJEB25059.

2.3 | Initial processing, transcriptome assembly, and contig annotation

Initial sequence quality assessment was carried out using FastQC.[21] Trimmomatic[22] was used within the Trinity pipeline[23] to remove poor quality reads. A de novo metatranscriptome assembly was then carried out using Trinity's default settings. All contigs with a length less than 300 base pairs were removed from the assembly.[24] Contigs were then clustered using CD-HIT-EST[25,26] with a threshold of 80% similarity. These CD-HIT-EST clusters, representing all expressed contigs, comprise the full transcriptome. Contigs were mapped against the Swiss-Prot (SP)[27] and NCBI's RefSeq non-redundant (NR)[28] databases using Blastx.[29] The NR annotations were supplied to Blast2GO[30] to generate gene ontology (GO) terms.[31,32] KEGG ontology (KO) terms were identified for each contig using GhostKOALA,[33] and global metabolic pathway expression was visualized using iPath2 (map01100).[34] For a list of all parameters used in the bioinformatics pipeline, see Table S1.

2.4 | Quantifying gene expression

Before performing gene expression analysis, ribosomal RNA reads were removed by mapping all reads to a database of all eukaryotic rRNA created in Silva[35] using Bowtie2.[36] Those reads that mapped to an rRNA gene were filtered out using Samtools.[37] All remaining reads were mapped to the full transcriptome, as well as to the *Wallemia sebi* genome,[38] using Bowtie2; read coverage was then quantified using Samtools. Only concurrent read pairs with a MAPQ greater than 2 were counted.

Read count normalization to account for GC-content bias within each replicate was performed in EDASeq[39] (using the full-quantile normalization option) on the read counts mapped to the full transcriptome and the *W. sebi* genome. The EDASeq-normalized values were then provided to DESeq2[40] for between-replicate normalization. Read counts were normalized using both the *W. sebi* genome and the full transcriptome to investigate whether gene expression changes were solely due to up- or downregulation of a gene or whether changes in relative abundance also influenced expression levels. *Wallemia* offers this opportunity because, in identical chamber experiments, it was shown to have a similar absolute abundance, but a pronounced increase in relative abundance, from 1.0 to 0.85 a_w.[15] We used 3 normalization procedures to investigate this effect (Figure S1). First, the reads were mapped to the full transcriptome (Figure S1A) and then normalized with EDASeq and DESeq2. Second, we selected all *W. sebi* genes (Figure S1B) from the full transcriptome and then normalized those read counts in EDASeq and DESeq2. Third, we mapped the reads to the *W. sebi* genome (Figure S1C) and then normalized them with EDASeq and DESeq2. These last 2 procedures allowed us to separate the effects of changes in relative abundance in *Wallemia* from per-spore changes in gene expression.

Contigs with an absolute $\log_2$ fold ($\log_2$F) change greater than 1 as calculated by DESeq2 were considered differentially expressed (Benjamini-Hochberg–adjusted *P*-values ≤.05) for subsequent GO term enrichment analysis. As described by Love et al,[40] the DESeq2 package in R was used to transform the read counts for the full transcriptome using the variance stabilizing transformation in a manner unbiased by the sample conditions (dispersions calculated using only a y-intercept). To assess gene expression similarity between replicates, the top 400 most highly expressed contigs were then clustered using the Euclidian distance matrix.

2.5 | GO term enrichment analysis

Enrichment analysis was then performed to determine which GO terms were differentially abundant among those contigs that were up- or downregulated between each condition. Two methods were used to perform enrichment analysis: GAGE[41] and Blast2GO.[30]

We first ran GAGE to assess the overall direction of $\log_2$F change for gene sets based on GO contig mappings. These gene sets were restricted to a relevant set of GO terms (List S1). The log2F values of those contigs with a statistically significant change in gene regulation, as determined by DESeq2, were provided to GAGE. GAGE then identified those GO terms that are significantly correlated with a given condition (Benjamini-Hochberg–adjusted *P*-values ≤ .05).

Building on the method to find "core enzymes" described by Jiang et al,[42] enrichment analysis was performed using Fisher's exact test in Blast2GO between the set of contigs that were expressed in all 9 replicates and the full transcriptome (all expressed contigs) to characterize the gene expression profile of those contigs in common between all replicates. Only GO terms with a FDR ≤ 0.05 were considered differentially enriched.

2.6 | Analysis of microbial composition and allergen and mycotoxin genes

Taxa assignment was performed by GhostKOALA. These were then loaded into MEGAN6[43] and used to perform taxa principal component analysis (PCA) using the Bray-Curtis method to calculate distances between samples. All contigs, irrespective of their taxonomic assignment, were used for our analysis of differential gene expression.

Swiss-Prot and GenProt IDs were obtained from the WHO/IUIS Allergen Nomenclature Database[44] for all known allergens. This information was used to annotate the dataset with allergen identity, and then calculate the expression of each allergen-annotated contig. The number of differentially expressed genes was compared between each of the water activities to assess any changes in allergen gene regulation.

The UniProt database[27] was used to identify potential mycotoxin biosynthetic genes. Searches focused on genes associated with the production of 5 common mycotoxins: aflatoxin, fumonisins, ochratoxin, trichothecenes, and zearalenone.[45] The $\log_2$F changes calculated by DESeq2 were used to investigate changes in gene regulation between each water activity.

2.7 | IgE binding assay for allergenicity

An IgE binding capacity assay was conducted on pure cultures of *Aspergillus fumigatus* grown at variable water activities to determine whether allergenicity varied with water activity. We selected *A. fumigatus* for this analysis because it is an abundant species based on our dataset and has an allergenicity assay that has been described previously.[7] Briefly, pooled serum was used from 7 individuals sensitized to *A. fumigatus* and used in an enzyme-linked immunosorbent assay (ELISA) to measure IgE binding capacity of different fungal extracts. *A. fumigatus* (ATCC 34506) was grown on agar where water activity was regulated at 0.95, 0.964, 0.985, and 1.0 using glycerol and verified using an AquaLab series 3 water activity meter. Spores were collected after 14 days of growth at 25°C using a nylon-flocked swab in 2 mL PBS. These growth conditions were selected to reflect the temperature possibly found in the indoor environment in conjunction with known water activity and germination duration conditions required for this species.[46] Spore concentration was quantified using a hemocytometer, and allergens for the assay were extracted by bead beating for 2 minutes using 0.1 g 0.5-mm and 0.3 g 0.1-mm glass beads and retrieving the supernatant. Final color change of the ELISA was quantified on a spectrophotometer across a dilution series. Allergenicity per 10^7 spores was estimated by linear regression.

3 | RESULTS

3.1 | Overview of the metatranscriptomic dataset

Sequencing produced a total of 185 113 832 paired-end reads. Trinity assembled all high-quality reads into 128 296 contigs with median length of 497 base pairs with a minimum size threshold of 300 bp. Out of the read pairs that passed quality filtering, on average

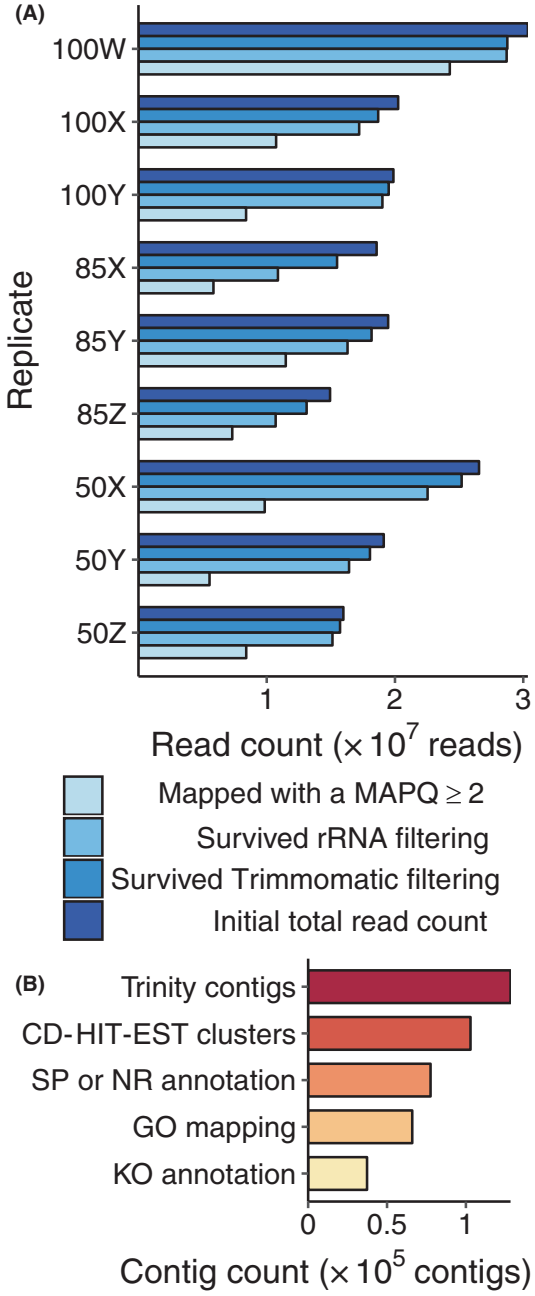


FIGURE 1 Summary of quality statistics. A, Colors indicate the number of properly paired reads that exist after sequencing, filtering, and mapping. B, Colors indicate the number of contigs assembled by Trinity and with an annotation. From top to bottom: contig count for clusters formed by Trinity, CD-HIT-EST clusters, contigs with a Swiss-Prot or NCBI's RefSeq non-redundant database annotation, contigs with a gene ontology mapping, and contigs with a KEGG ontology annotation

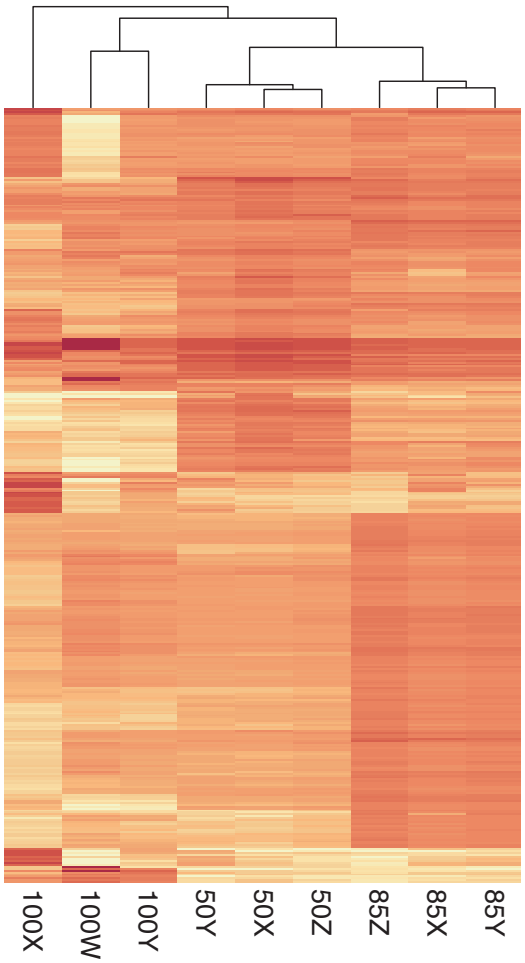


FIGURE 2 Heat map of variance-stabilized read counts for the top 400 most highly expressed contigs. Rows and columns are clustered based on the Euclidian distance, and the dendrogram is only plotted for the conditions

73 ± 12% of reads mapped back to the full transcriptome. However, many of these reads mapped either discordantly or to multiple locations in the transcriptome. After requiring a MAPQ greater than or equal to 2 and only retaining reads that were part of a properly mapping pair, a total of 48 ± 16% of quality-filtered reads were mapped. The percent of reads that survived the quality filtering and mapping is similar to other metatranscriptomic studies, which often have a small percentage of reads that survive quality filtering or map back to the transcriptome. For example, a metatranscriptomic comparison study by Jiang et al[42] using publically available mRNA datasets found that between 19% and 72% of reads mapped back to their transcriptomes. Read and cluster quality statistics are further summarized in Figure 1.

Replicate samples at 1.0, 0.85, and 0.5 a_w cluster by condition based on the heat map of variance-stabilized read counts indicating significant differences in gene expression as a function of water activity (Figure 2). These differences can also be seen in the MA plots (Figure S2). These plots also reveal that there are many more differentially expressed genes between 1.0 and either 0.5 or 0.85 a_w than between 0.5 and 0.85 a_w (6 and 8 times more, respectively).

3.2 | Taxonomic analysis of metabolically active genera

Wallemia and *Aspergillus* were identified as the central fungal genera after taxonomic analysis of annotated sequences. A total of

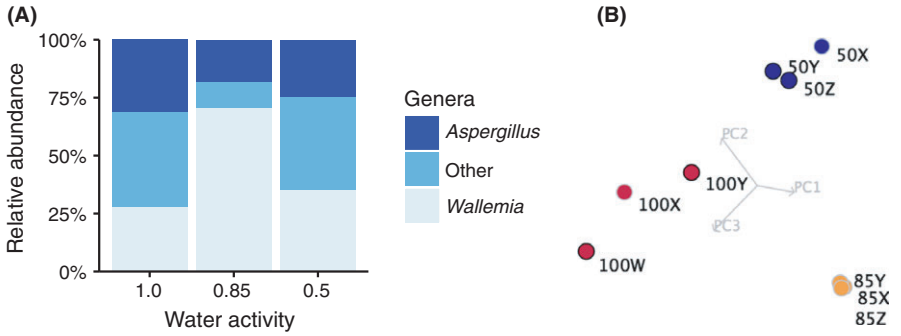


FIGURE 3 A, Taxonomic relative abundance of fungal genera and B, PCA plots. 55.0% of the variation is from PC1, 29.0% from PC2, and 12.7% from PC3

44% of all contigs with a taxonomic annotation came from a fungal source, 30% from a plant source, and 21% from a bacterial source. Based on GhostKOALA annotations across all conditions, *Wallemia* and *Aspergillus* were the most metabolically active fungal genera (Figure 3A). *Wallemia*'s relative abundance was highest at 0.85 compared to either 0.5 or 1.0 a_w (*P*-value ≤ .05). Overall, 82 fungal genera were identified, of which 88% were found across all replicates. PCA of all metabolically active genera shows that the ecology clusters into distinct groups based on water activity condition (Figure 3B).

A previous study by Dannemiller et al[15] found that *Wallemia* had a pronounced change in relative abundance, but a comparable absolute abundance, between 1.0 and 0.85 a_w. This trend was additionally seen in our results, thus offering an opportunity to separate the influence of taxa relative abundance changes from true gene expression changes. When normalized without accounting for the change in relative abundance in *Wallemia* (normalization carried out for all contigs across the full transcriptome), nearly all differentially expressed *Wallemia* contigs between 1.0 and 0.85 a_w were found to be downregulated (Figure S3A). However, many more genes were found to be upregulated at 1.0 compared to 0.85 a_w (Figure S3B,C) when gene expression for *Wallemia* contigs was normalized separately (either by normalizing all *Wallemia* contigs apart from the full transcriptome or by mapping all reads to the *W. sebi* genome and then normalizing). These comparisons demonstrate, at least in the case of *Wallemia*, that changes in gene expression can be driven by water activity and are not simply a reflection of changes in relative abundance.

3.3 | Genes found in all replicates are predominately related to primary metabolism

Gene expression at each water activity indicates that cells are metabolically active. We performed enrichment analysis in Blast2GO to investigate the transcriptomic profile of those contigs that are shared between all 9 replicates. We compared those GO terms that are more prevalent in the set of contigs expressed in all replicates to the full transcriptome (or all expressed contigs), and found that contigs with GO mappings related to the basic metabolic processes necessary for life predominate among those contigs expressed in all replicates (Figure 4). This gene enrichment analysis demonstrates that cells at all 3 water activities are metabolically active. Overenriched GO terms include: "cellular respiration," "binding," and "transport," as well as numerous descendants of "primary metabolic process," such as "translation" and "lipid metabolic process." Critical cellular components, such as the ribosome, nucleus, and mitochondrion, are also overenriched among those contigs shared among all replicates. The more complete coverage of carbohydrate, protein, and lipid metabolism, as well as the TCA cycle, is also clear by comparing the coverage of the KEGG global metabolic pathways map at 1.0 compared to either 0.85 or 0.5 a_w (1.0 vs 0.5 a_w; Figure 5; see Video S1 for an animation of the change in coverage from 0.5 to 0.85 to 1.0 a_w). Thus, the overlap in coverage between each condition is concentrated among contigs related to primary metabolic processes.

3.4 | Microorganisms present in dust are more metabolically active with increasing water activity

Examining the KEGG pathway maps (Figure 5 and Video S1) in more detail illustrates that metabolic coverage is more complete at a water activity of 1.0. While 4369 unique KO terms were identified at 1.0 a_w, only 3901 and 3678 were found at 0.85 and 0.5 a_w, respectively. The degree of overlap in the number of KO terms present in the replicates of a given condition decreases with increasing water activity: Replicates at 0.5 a_w have 92 ± 6% of their KO terms in common with each other, replicates at 0.85 a_w have 88 ± 0.5% in common, and the 1.0 a_w replicates only share 84 ± 5% of their KO terms (Figure S4). For this reason, an increased diversity of KO terms is identified at 1.0 compared with either 0.85 or 0.5 a_w.

Much of this increase in metabolic activity at high water activity can be explained by an increase in secondary metabolite production. The GO term secondary metabolic process is defined by the Gene Ontology Consortium as "the chemical reactions and pathways resulting in many of the chemical changes of compounds that are not necessarily required for growth and maintenance of cells, and are often unique to a taxon."[47] These include the production of certain toxins and the utilization of unusual nutrients. There are 1.4 times more contigs with a "secondary metabolic process" GO term mapping at 1.0 compared to 0.85 a_w and 1.5 times more contigs at 1.0 than 0.5 a_w. Although most contigs with a "secondary metabolic process" GO mapping are not differentially expressed (15% of contigs at 1.0 vs 0.5 a_w and 22% of contigs at 1.0 vs 0.85 a_w are differentially expressed), the majority of those that are differentially expressed are upregulated at 1.0 vs 0.85 or 0.5 a_w (Figure 6). Furthermore, GAGE gene set enrichment shows that "secondary metabolic process" is

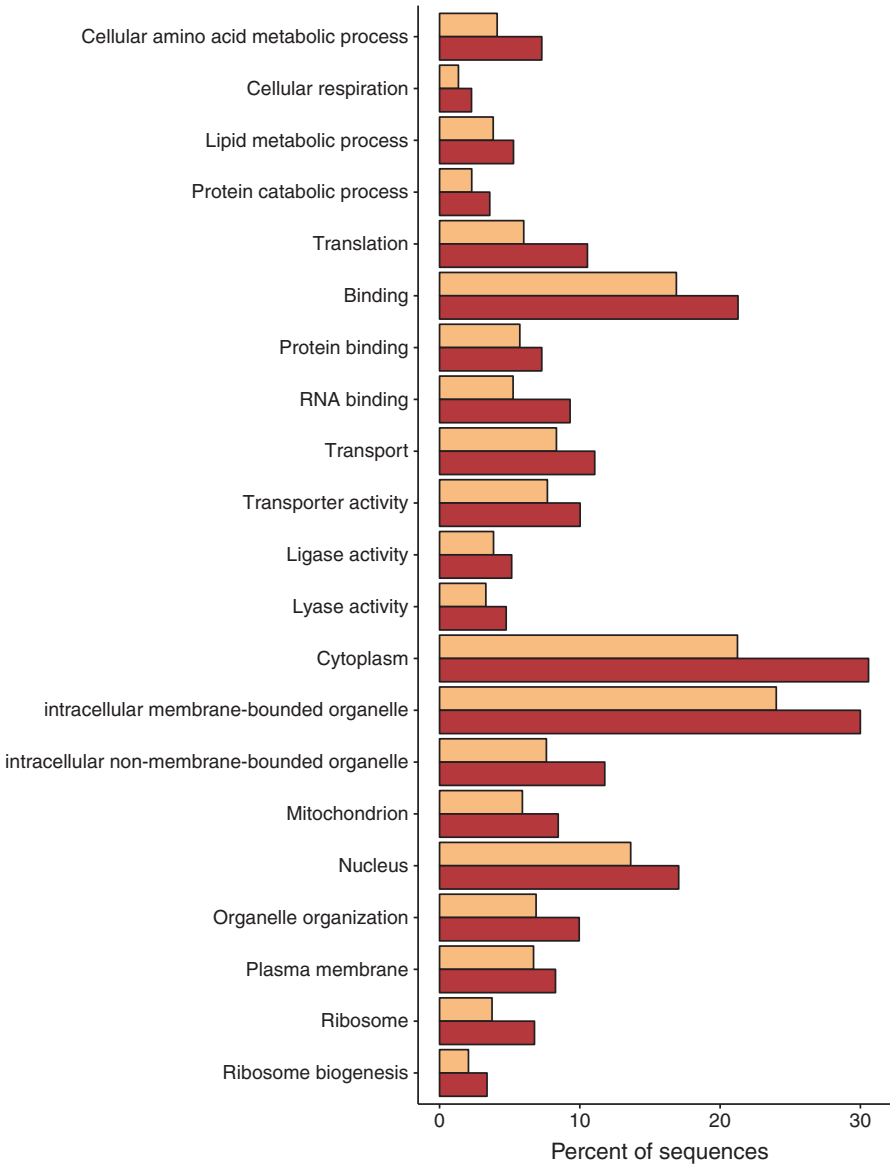


FIGURE 4 Enrichment analysis for core enzymes. All GO terms shown here are overenriched (Fisher's exact test, FDR ≤ 0.05) among those contigs that were found across all 9 replicates compared to the full transcriptome. The percent of sequences displayed are the percentage of contigs with the specified GO term mapping found in the set of contigs expressed in all replicates (red) and in the full transcriptome (orange)

upregulated at 1.0 compared to 0.5 a_w ($p_{adj} \leq 0.05$) (see Table S2 for all GO terms identified as differentially expressed by GAGE). Taken together, these results suggest an overall lower metabolic activity at 0.5 a_w compared to higher water activities.

3.5 | Evidence for increased health effects by microorganisms at a water activity of 1.0

The GO term pathogenesis can be used as a benchmark for the expression of genes that generate the ability of an organism to induce an abnormal or detrimental state in another organism.[47] Differences in gene regulation with water activity emerge when focusing on those contigs mapping to a pathogenesis GO term (Figure 7A). There are 6 times as many upregulated as downregulated contigs at 1.0 compared to 0.5 a_w, and 3 times as many upregulated as downregulated contigs at 1.0 compared to 0.85 a_w. Very few contigs have a statistically significant change in gene regulation between 0.85 and 0.5 a_w. GAGE gene set enrichment for pathogenesis shows that it is upregulated at 1.0 compared to 0.5 a_w ($p_{adj} = 0.06$), although no difference was found between 1.0 and 0.85 a_w or between 0.85 and 0.5 a_w.

Gene expression of allergens and toxins was investigated to examine possible effects on human health. A large diversity of genes encoding for fungal allergens was found to be present in our data. A total of 308 allergens from the WHO/IUIS Allergen Nomenclature Database[44] were identified (see Table S3 for a list of all allergens found), including 60 fungal allergens from 13 genera of fungi (the 30 most highly expressed allergens are shown in Figure 7C) and 22 allergens from dust-mites (Figure S5).

Examining changes in gene regulation, approximately one-third of contigs with an allergen notation were found to be differentially expressed between 1.0 and either 0.85 or 0.5 a_w (Figure 7D). Of these differentially expressed contigs, there are nearly double the number of upregulated allergen contigs at 1.0 compared to either

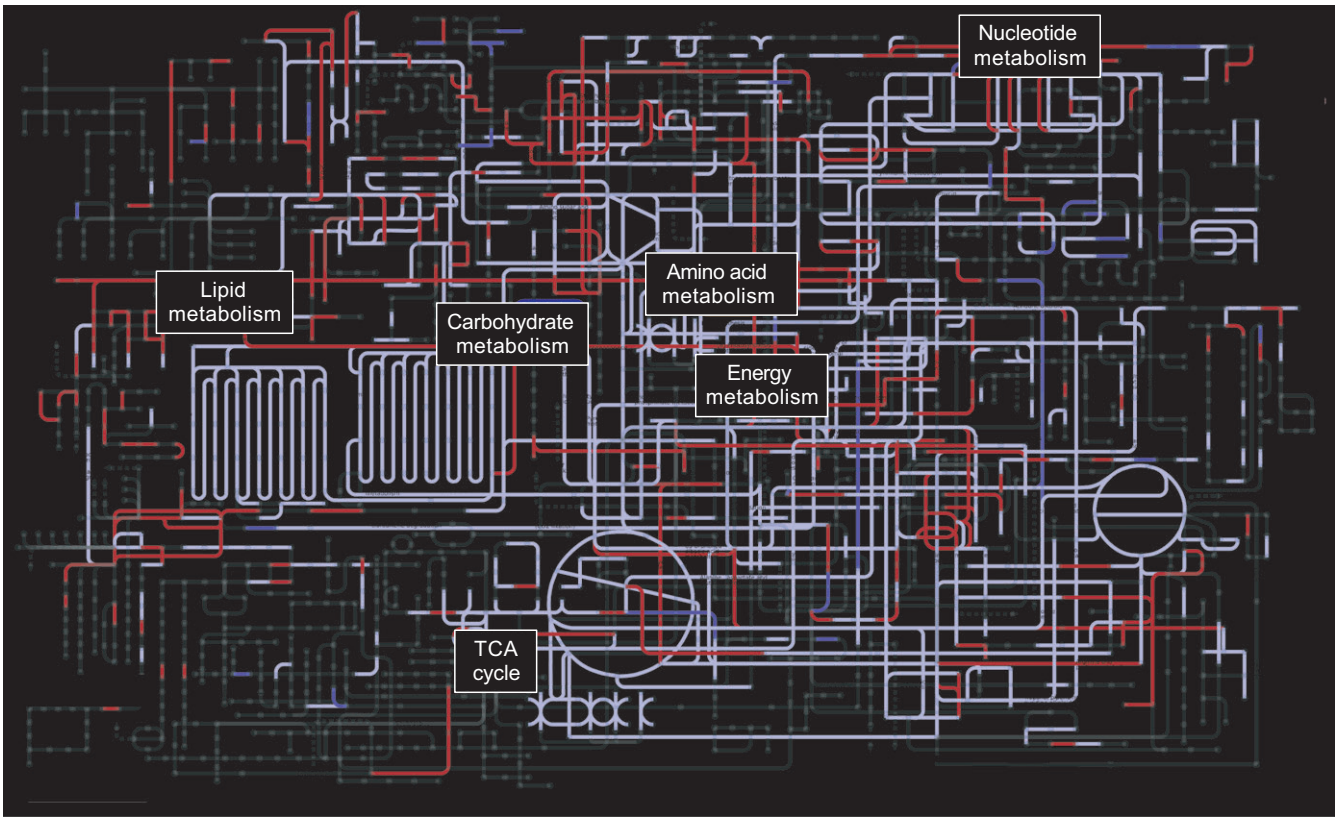


FIGURE 5 Metabolic expression map comparing gene expression between 0.5 and 1.0 a_w. Genes present only at 1.0 a_w are in red, genes present only at 0.5 a_w are in blue, and genes present at both water activities are in light purple

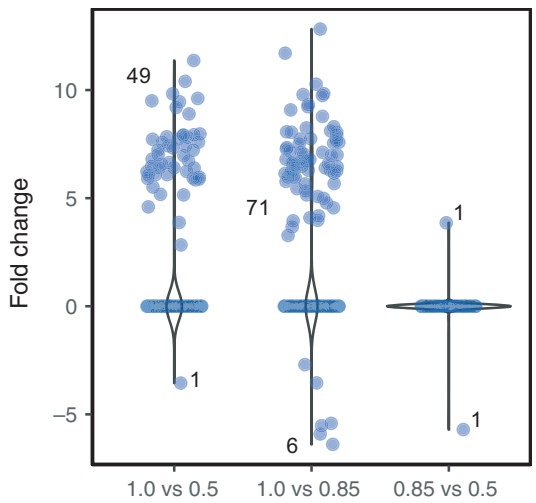


FIGURE 6 Violin plot overlaid with a scatterplot of contigs with "secondary metabolic process" GO mapping. All contigs with $\log_2$F changes not equal to zero are considered differentially expressed ($p_{adj} \leq 0.05$). All contigs with $\log_2$F changes not found to be statistically significant are plotted at zero. Log_2F changes greater than zero indicate that contigs at water activities of 1.0 are more highly expressed than at 0.5 (1.0 vs 0.5) or 0.85 (1.0 vs 0.85). Similarly, $\log_2$F changes greater than zero indicate that contigs at water activities of 0.85 are more highly expressed than at 0.5 (0.85 vs 0.5). The violin plot is a mirrored density plot of the distribution of contig $\log_2$F changes

0.50 or 0.85 a_w. These data suggest that cells at higher water activities are increasing the production of key genes related to allergen production. This trend is further supported by IgE binding assays where allergenicity steadily increases from 0.95 a_w to 1.00 a_w for *Aspergillus fumigatus* (Figure 7B).

Looking specifically at contigs annotated as aflatoxin, gene expression is highly upregulated at 1.0 compared to either 0.5 or 0.85 a_w (Figure 7E). Approximately 16 times more aflatoxin contigs are upregulated at 1.0 vs 0.5 a_w and 7 times more contigs are upregulated at 1.0 compared to 0.85 a_w. Further, there are very few differentially expressed contigs at 0.85 compared to 0.5 a_w. We also looked for transcriptional evidence of production of other mycotoxins, including fumonisins, ochratoxin, trichothecenes, and zearalenone. No evidence of ochratoxin or zearalenone production was found. Limited transcriptional evidence of trichothecenes was found only at water activities of 1.0. Very low production of fumonisins was found at all 3 water activity levels, but no fumonisin-annotated contigs were found to be differentially expressed.

4 | DISCUSSION

Metatranscriptomics is increasingly being applied to investigate the activity and function of fungal communities in environmental

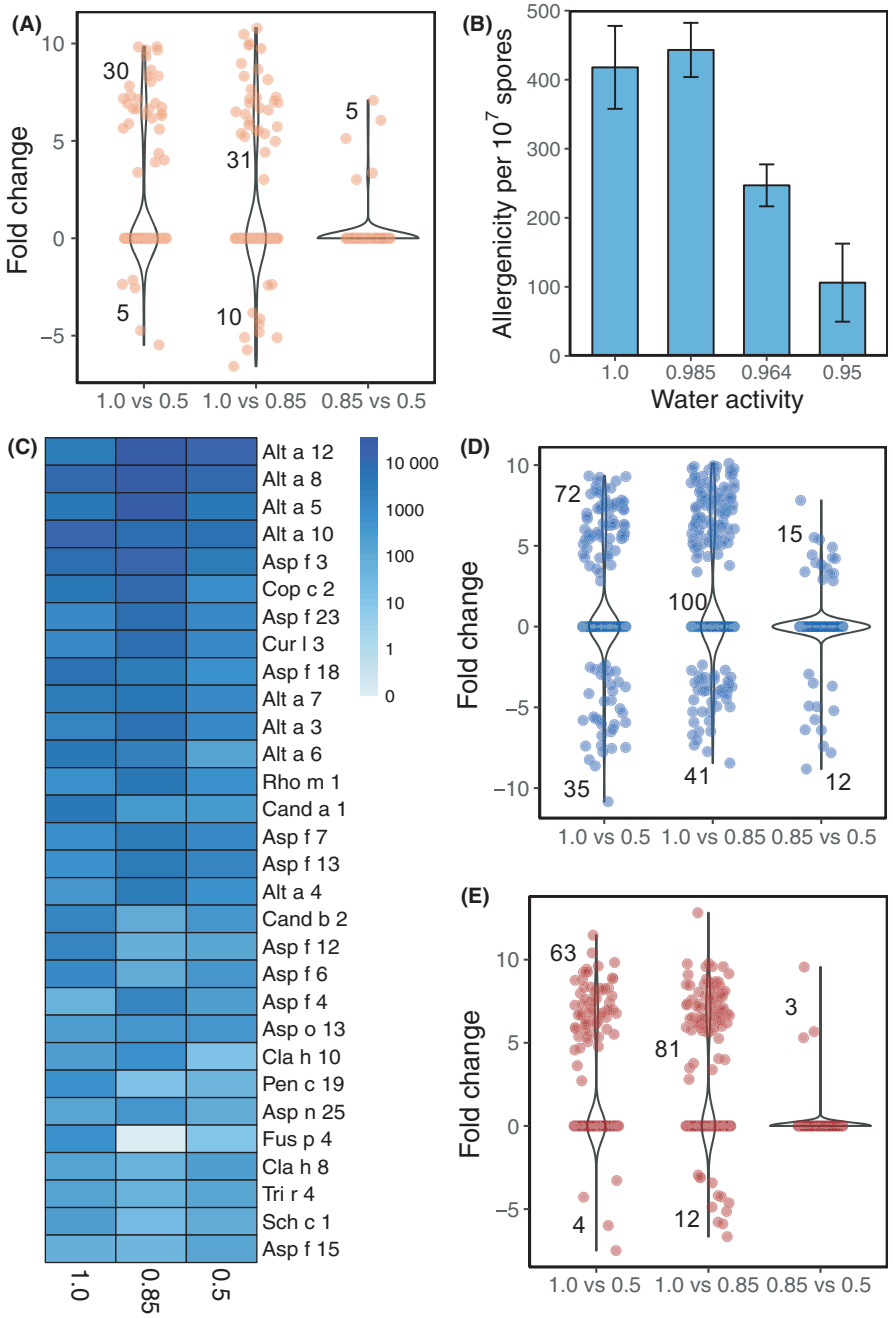


FIGURE 7 Overview of health effects. A, Violin plot of contigs with a "pathogenesis" GO mapping. B, Allergenicity of *Aspergillus fumigatus* spores over various water activities. Error bars represent the standard error. C, Heat map of the mean expression of the 30 most highly expressed fungal allergens found in the dataset for each of the 3 water activities. D, Violin plot of $\log_2$F changes of allergen-annotated contigs. E, Violin plot of $\log_2$F changes of aflatoxin-annotated contigs. For A,D,E, each point is an annotated contig, and all contigs with $\log_2$F changes not equal to zero are considered differentially expressed ($p_{adj} \leq 0.05$). All contigs with $\log_2$F changes not found to be statistically significant are plotted at zero. $\log_2$F changes greater than zero indicate that contigs at water activities of 1.0 are more highly expressed than at 0.5 (1.0 vs 0.5) or 0.85 (1.0 vs 0.85); similarly, $\log_2$F changes greater than zero indicate that contigs at water activities of 0.85 are more highly expressed than at 0.5 (0.85 vs 0.5)

samples.[48-50] This work uniquely studies fungal community gene expression on built environment-relevant substrates at different water activities. Results should catalyze a better understanding of how building material water activity can influence the physiological status and metabolic activity of fungi in the built environment, and has important implications for human health. Two important significant insights were produced in this study: (i) Cells are metabolically active at water activities ranging from 0.5 to 1.0, with significant and large changes in gene expression between all 3 conditions, and (ii) genes associated with potentially negative human health effects are generally upregulated with increasing water activity.

4.1 | Metabolic complexity increases with water activity

Previous studies have shown that fungi are capable of growing on building materials at water activities as low as 0.75.[16-18,20] These xerophilic fungi are able to survive at lower water activities in part by modulating solute permeability and transport in order to maintain homeostasis. The high osmolarity glycerol signaling pathway, in particular, is thought to be critical for controlling the internal solute concentrations of fungi.[51-55] Although xerophilic fungi are capable of growing at water activities below 0.9, their growth is inhibited as a_w decreases.[17,56] Prior measurements of organic carbon, nitrate, and phosphate indicate that

water activity, and not nutrients, is the major limitation to microbial growth in dust.[15] Our results demonstrate that even at water activities as low as 0.5, where growth has not been previously detected, some of the fungi present in dust remain metabolically active. Analyzing their gene expression indicates that these organisms focus their metabolic resources on areas essential for cell maintenance.

We found that cells at water activities of 1.0 are the most transcriptionally active. Much of this increased expression is accounted for in additional genes related to secondary metabolism being expressed. Furthermore, enrichment analysis suggests that those genes that are upregulated at 1.0 a_w are similar in metabolic function to the full transcriptome. In addition to being more metabolically active, we observed increased variability in gene expression between the replicates at 1.0 compared to either 0.85 or 0.5 a_w. This may be due, in part, to the greater variability in ecology that we observed at 1.0 compared to either 0.85 or 0.5 a_w. Previous ITS-based sequencing methods corroborate the increase in variability between replicates at 1.0 a_w that we observed through transcriptomic methods.[15]

Two common, xerophilic fungal genera, *Aspergillus* and *Wallemia*, were found to be the dominant taxa in our dust samples. Both strains have similar germination times when grown on agar plates at 0.86 and 0.82 a_w.[57] However, we found that *Wallemia* dominates the transcriptionally active taxa found at 0.85 a_w, suggesting greater metabolic activity around 0.85 a_w compared to *Aspergillus*. The optimal water activity for *W. sebi* growth has been reported at 0.94 a_w.[58] At higher water activities, *Wallemia's* competitive advantage appears to be lost, as its relative abundance decreases substantially between 0.85 and 1.0 a_w. This corroborates previous ITS sequence-based studies of the fungal community present in dust where the relative abundance of *Wallemia* is elevated at 0.85 compared to either 1.0 or 0.5 a_w.[15]

4.2 | Implications for health

Allergen, toxin, and pathogenicity gene expression results demonstrate a potential for increased health impacts from fungi at elevated water activities. Log_2F changes between the various dampness levels reveal that fungal cells at 1.0 a_w upregulate more allergen-encoding genes than cells at either 0.85 or 0.5 a_w. The overall increase in upregulated allergen genes follows the trend we observed with the increased *A. fumigatus* allergenicity at elevated water activity.

Pathogenesis, or virulence, is defined as "the set of specific processes that generate the ability of an organism to induce an abnormal, generally detrimental state in another organism."[47] These genes are not limited to those that cause a negative response in humans, but include any gene that an organism produces that induces an abnormal or detrimental state in another organism. Three-quarters of contigs with a pathogenesis mapping are from *Aspergillus*. Complementing the allergen data, those contigs with a pathogenicity GO mapping are more likely to be overexpressed at 1.0 a_w compared to lower water activities.

An additional class of molecules that cause negative health effects in humans are mycotoxins. We investigated the gene expression of the following mycotoxins: aflatoxin, fumonisins, ochratoxin, trichothecenes, and zearalenone. Aflatoxin, which has been previously detected in indoor air (along with trichothecenes and rare occurrence of ochratoxin A),[16,59] was found to be differentially expressed with water activity. Fumonisins and zearalenone are most commonly found in agricultural settings;[60] therefore, it is to be expected that very low (fumonisins) or no (zearalenone) gene expression was detected in our samples. More aflatoxin-producing contigs were upregulated between 1.0 and either 0.85 or 0.5 a_w. This result mirrors prior pure culture work that demonstrated that water activity can influence mycotoxin production. Previous work using *Aspergillus flavus* pure cultures has demonstrated that *A. flavus* undergoes a significant transcriptional response to growth at 0.99 vs 0.93 a_w, including upregulation of aflatoxin biosynthesis genes with increasing water activity, and a greater measured production of aflatoxin.[11] Additionally, elevated ochratoxin production in *A. carbonarius* and *A. niger* has been found in water activities ranging from 0.95 to 0.98 compared to 0.92.[12] Although prior research does not provide information on water activity and allergen content, other factors, including temperature and atmospheric gases, have been shown to influence the allergen potency of *A. fumigates* spores.[6,7]

The trends in gene expression observed here reflect current literature that has associated dampness in building materials with human health impacts. Prior meta-analyses of health studies have linked qualitative assessments of building dampness or visible mold with respiratory and allergenic health effects including asthma development, asthma exacerbation, current and past asthma cases, dyspnea, wheeze, cough, respiratory infections, bronchitis, allergic rhinitis, eczema, and upper respiratory tract symptoms.[3] Interestingly, while previous epidemiological investigations have shown consistent relationships between building dampness and visible mold and respiratory health effects, conventional approaches for microbiological analysis of fungal concentrations in dust and air have not ascertained the underlying causal relationship between fungal exposures and health.[3] This work demonstrates that physiological status and taxonomic composition are critical factors in assessing fungal toxicity.

4.3 | Limitations of our approach

One inherent challenge of fungal transcriptomics is the lack of fully annotated fungal genomes.[14] For that reason, many contigs will be identified as the closest homolog present in the database. Furthermore, many of our contigs are not annotated or simply identified as "hypothetical proteins." Performing high-level GO term analysis allows us to eliminate some of the ambiguity resulting from incomplete genome databases, as most homologs will have a similar function. The lack of fully annotated fungal genomes becomes particularly apparent when performing taxonomic analysis. For example, at 0.5 a_w, a high diversity of taxa was previously identified by ITS methods,[15] but is not captured in transcriptomic annotations. For water activities above 0.5, we know that transcriptomics better represents ecology as those genera that were found to dominate based on prior ITS approaches (*Aspergillus*, *Penicillium*, and *Wallemia*) are better represented in genome sequence databases.

An additional challenge to any transcriptomic analysis is normalization. In metatranscriptomic studies where the relative abundance of various taxa is changing, normalization is further complicated. For that reason, a few sources recommend that normalization take into account changes in taxa composition between conditions to allow for changes in gene regulation to be separated from increases or decreases in the relative abundance of a given taxon.[14,61] For *Wallemia*, we investigated the effect of normalizing by taxon on the number of up- and downregulated genes between 1.0 and 0.85 a_w. These results illustrate that accounting for changes in relative abundance is important for separating the effects of ecology from gene expression. Unfortunately, fully normalizing fungal transcriptomes for taxonomy is impractical because of the incomplete nature of the fungal genome databases. Thus, our results are indicative of changes in gene expression of the community as a whole, but cannot be used to separate changes in gene expression based on ecology from those based on gene regulation.

A strength of transcriptomic analysis is the broad, comprehensive assessment of the genetic activity in a microbial community. Current limitations in fungal genome databases relegate this tool to generating insights and hypotheses. Further, this study is only capable of providing associations between water activity and gene expression and cannot elucidate the complex regulatory systems and mechanisms that underlie the expression of specific genes. Additional experiments focused on measuring metabolites or health impacts are needed to corroborate gene expression results.

5 | CONCLUSIONS

Fungi in house dust at water activities ranging from 0.5 to 1.0 are metabolically active and strongly regulate gene expression as a function of water activity. The data presented here demonstrate that at higher water activities, fungi grown in house dust may upregulate more allergen-encoding genes, general pathogenicity pathways, and secondary metabolites, suggesting multiple ways in which the health effects of fungi may depend on the water activity at which they are grown. In damp buildings, measurements of fungal concentrations that do not account for ecology or potential health effects due to growth at elevated water activities may not accurately represent human exposure.

ACKNOWLEDGEMENTS

We acknowledge the Alfred P. Sloan Foundation Postdoctoral Fellowship Program and the National Science Foundation Graduate Research Fellowship Program for support of K.C.D. and B.H., respectively.

CONFLICT OF INTERESTS

None.

ORCID

J. Peccia http://orcid.org/0000-0002-6482-2084

REFERENCES

1. Institute of Medicine (IOM). *Damp Indoor Spaces and Health*. Washington, DC: National Academies Press; 2004.
2. World Health Organization (WHO). *WHO Guidelines for Indoor Air Quality: dampness and Mould*. Copenhagen, Denmark: World Health Organization; 2009.
3. 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;119:748-756.
4. Misaghi IJ, Grogan RG, Duniway JM, Kimble KA. Influence of environment and culture media on spore morphology of *Alternaria alternata*. *Phytopathology*. 1978;68:29-34.
5. Phillips D. Changes in conidia of *Monilinia fructicola* in response to incubation temperature. *Phytopathology*. 1982;72:1281-1283.
6. Lang-Yona N, Shuster-Meiseles T, Mazar Y, Yarden O, Rudich Y. Impact of urban air pollution on the allergenicity of *Aspergillus fumigatus* conidia: outdoor exposure study supported by laboratory experiments. *Sci Total Environ*. 2016;541:365-371.
7. Low SY, Dannemiller K, Yao M, Yamamoto N, Peccia J. The allergenicity of *Aspergillus fumigatus* conidia is influenced by growth temperature. *Fungal Biol*. 2011;115:625-632.
8. Abbas HK, Egley GH, Paul RN. Effect of conidia production temperature on germination and infectivity of *Alternaria helianthi*. *Phytopathology*. 1995;85:667-682.
9. Bertolini P, Tian SP. Effect of temperature of production of *Botrytis allii* conidia on their pathogenicity to harvested white onion bulbs. *Plant Pathol*. 1997;46:432-438.
10. Wolf J, O'Neill NR, Rogers CA, Muilenberg ML, Ziska LH. Elevated atmospheric carbon dioxide concentrations amplify *Alternaria alternate* sporulation and total antigen production. *Environ Health Perspect*. 2010;118:1223-1228.
11. Zhang F, Guo Z, Zhong H, et al. RNA-Seq-based transcriptome analysis of aflatoxigenic *Aspergillus flavus* in response to water activity. *Toxins*. 2014;6:3187-3207.
12. Leong SL, Hocking AD, Scott ES. Effect of temperature and water activity on growth and ochratoxin A production by Australian *Aspergillus carbonarius* and *A. niger* isolates on a simulated grape juice medium. *Int J Food Microbiol*. 2006;110:209-216.
13. Corcuff R, Mercier J, Tweddell R, Arul J. Effect of water activity on the production of volatile organic compounds by *Muscodor albus* and their effect on three pathogens in stored potato. *Fungal Biol*. 2011;115:220-227.
14. Kuske CR, Hesse CN, Challacombe JF, et al. Prospects and challenges for fungal metatranscriptomics of complex communities. *Fungal Ecol*. 2015;14:133-137.
15. Dannemiller KC, Weschler CJ, Peccia J. Fungal and bacterial growth in floor dust at elevated relative humidity levels. *Indoor Air*. 2017;27:354-363.
16. Andersen B, Frisvad JC, Søndergaard I, Rasmussen IS, Larsen LS. Associations between fungal species and water-damaged building materials. *Appl Environ Microbiol*. 2011;77:4180-4188.
17. Korpi A, Pasanen AL, Pasanen P, Kalliokoski P. Microbial growth and metabolism in house dust. *Int Biodeterior Biodegr*. 1997;40:19-27.
18. Wheeler KA, Hocking AD, Pitt JI. Effects of temperature and water activity on germination and growth of *Wallemia sebi*. *Trans Br Mycol Soc*. 1988;90:365-368.
19. Pitkäranta M, Meklin T, Hyvärinen A, et al. Molecular profiling of fungal communities in moisture damaged buildings before and

after remediation - a comparison of culture-dependent and culture-independent methods. *BMC Microbiol*. 2011;11:235.
20. Nielsen KF, Holm G, Uttrup LP, Nielsen PA. Mould growth on building materials under low water activities. Influence of humidity and temperature on fungal growth and secondary metabolism. *Int Biodeterior Biodegr* 2004;54:325-336.
21. Andrews S. FastQC: a quality control tool for high throughput sequence data. 2010. http://www.bioinformatics.babraham.ac.uk/projects/fastqc.
22. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*. 2014;30:2114-2120.
23. Grabherr MG, Haas BJ, Yassour M, et al. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol*. 2011;2011:644-652.
24. Li HZ, Gao X, Li XY, Chen QJ, Dong J, Zhao WC. Evaluation of assembly strategies using RNA-Seq data associated with grain development of wheat (*Triticum aestivum* L.). *PLoS ONE*. 2013;8:e83530.
25. Li W, Godzik A. Cd-hit : a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics*. 2006;22:1658-1659.
26. Fu L, Niu B, Zhu Z, Wu S, Li W. CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics*. 2012;28:3150-3152.
27. Wasmuth EV, Lima CD. UniProt: the universal protein knowledgebase. *Nucleic Acids Res*. 2016;45:1-12.
28. Pruitt KD, Tatusova T, Maglott DR. NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res*. 2007;35:61-65.
29. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol*. 1990;215:403-410.
30. Conesa A, Götz S, García-Gómez JM, et al. Blast2GO : a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics*. 2005;21:3674-3676.
31. Consortium TGO. Gene Ontology : tool for the unification of biology. *Nat Genet*. 2000;25:25-29.
32. Consortium TGO. Gene ontology consortium: going forward. *Nucleic Acids Res*. 2015;43:D1049-D1056.
33. Kanehisa M, Sato Y, Morishima K. BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences. *J Mol Biol*. 2016;428:726-731.
34. Yamada T, Letunic I, Okuda S, Kanehisa M, Bork P. IPath2.0: interactive pathway explorer. *Nucleic Acids Res*. 2011;39:412-415.
35. Pruesse E, Quast C, Knittel K, Fuchs BM, Glo FO, Ludwig W. SILVA : a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic Acids Res*. 2007;35:7188-7196.
36. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods*. 2012;9:357-359.
37. Li H, Handsaker B, Wysoker A, et al. The sequence alignment/map format and SAMtools. *Bioinformatics*. 2009;25:2078-2079.
38. Padamsee M, Kumar ATK, Riley R, et al. The genome of the xerotolerant mold *Wallemia sebi* reveals adaptations to osmotic stress and suggests cryptic sexual reproduction. *Fungal Genet Biol*. 2012;49:217-226.
39. Risso D, Schwartz K, Sherlock G, Dudoit S. GC-content normalization for RNA-Seq data. *BMC Bioinformatics*. 2011;12:480.
40. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*. 2014;15:550.
41. Luo W, Friedman MS, Shedden K, Hankenson KD, Woolf PJ. GAGE: generally applicable gene set enrichment for pathway analysis. *BMC Bioinformatics*. 2009;10:161.
42. Jiang Y, Xiong X, Danska J, Parkinson J. Metatranscriptomic analysis of diverse microbial communities reveals core metabolic pathways and microbiome-specific functionality. *Microbiome*. 2016;4:2.
43. Huson DH, Beier S, Flade I, Górska A, El-Hadidi M. MEGAN community edition - interactive exploration and analysis of large-scale microbiome sequencing data. *PLoS Comput Biol* 2016;12:1-12.
44. Radauer C, Nandy A, Ferreira F, Goodman RE, Larsen JN, Lidholm J. Update of the WHO/IUIS allergen nomenclature database based on analysis of allergen sequences. *Allergy*. 2014;69:413-419.
45. Morath SU, Hung R, Bennett JW. Fungal volatile organic compounds: a review with emphasis on their biotechnological potential. *Fungal Biol Rev*. 2012;26:73-83.
46. Ayerst G. The effects of moisture and temperature on growth and spore germination in some fungi. *J Stored Prod Res*. 1969;5:127-141.
47. Carbon S, Ireland A, Mungall CJ, et al. AmiGO: online access to ontology and annotation data. *Bioinformatics*. 2009;25:288-289.
48. Urich T, Lanzén A, Qi J, Huson DH, Schleper C, Schuster SC. Simultaneous assessment of soil microbial community structure and function through analysis of the meta-transcriptome. *PLoS ONE*. 2008;3:e2527.
49. Coolen MJL, Orsi WD. The transcriptional response of microbial communities in thawing Alaskan permafrost soils. *Front Microbiol*. 2015;6:1-14.
50. Hesse CN, Mueller RC, Vuyisich M, et al. Forest floor community metatranscriptomes identify fungal and bacterial responses to N deposition in two maple forests. *Front Microbiol*. 2015;6:1-15.
51. Beuchat LR. Influence of water activity on growth, metabolic activities and survival of yeasts and molds. *J Food Prot*. 1983;46:35-141.
52. Vinnere Pettersson O, Leong SL. Fungal Xerophiles (Osmophiles). In: Pettis G, ed. *eLS*. Chichester: John Wiley & Sons, Ltd; 2011;1-9.
53. Gunde-Cimerman N, Ramos J, Plemenitaš A. Halotolerant and halophilic fungi. *Mycol Res*. 2009;113:1231-1241.
54. Hohmann S. Control of high osmolarity signalling in the yeast *Saccharomyces cerevisiae*. *FEBS Lett*. 2009;583:4025-4029.
55. Hohmann S. Osmotic stress signaling and osmoadaptation in yeasts. *Microbiol Mol Biol Rev*. 2002;66:300-372.
56. Kralj Kunčič M, Zajc J, Drobne D, Pipan Tkalec Ž, Gunde-Cimerman N. Morphological responses to high sugar concentrations differ from adaptation to high salt concentrations in the xerophilic fungi *Wallemia* spp. *Fungal Biol*. 2013;117:466-478.
57. Gock MA, Hocking AD, Pitt JI, Poulos PG. Influence of temperature, water activity and pH on growth of some xerophilic fungi. *Int J Food Microbiol*. 2003;81:11-19.
58. Patriarca A, Vaamonde G, Fernández Pinto V, Comerio R. Influence of water activity and temperature on the growth of *Wallemia sebi*: application of a predictive model. *Int J Food Microbiol*. 2001;68:61-67.
59. Brasel TL, Douglas DR, Wilson SC, Straus DC. Detection of airborne *Stachybotrys chartarum* macrocyclic trichothecene mycotoxins in the indoor environment. *Appl Environ Microbiol*. 2005;71:114-122.
60. Bennett JW, Klich M. Mycotoxins. *Clin Microbiol Rev*. 2003;16:497-516.
61. Klingenberg H, Meinicke P. How to normalize metatranscriptomic count data for differential expression analysis. *PeerJ*. 2017;5:e3859.

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

How to cite this article: Hegarty B, Dannemiller KC, Peccia J. Gene expression of indoor fungal communities under damp building conditions: Implications for human health. *Indoor Air*. 2018;28:548–558. https://doi.org/10.1111/ina.12459