Moving beyond species: fungal function in house dust provides novel targets for potential indicators of mold growth in homes

Neeraja Balasubrahmaniam[1,2,3], Jon C. King[1,2,3], Bridget Hegarty[4] and Karen C. Dannemiller[2,3,5]*

Abstract

Background Increased risk of asthma and other respiratory diseases is associated with exposures to microbial communities growing in damp and moldy indoor environments. The exact causal mechanisms remain unknown, and occupant health effects have not been consistently associated with any species-based mold measurement methods. We need new quantitative methods to identify homes with potentially harmful fungal growth that are not dependent upon species. The goal of this study was to identify genes consistently associated with fungal growth and associated function under damp conditions for use as potential indicators of mold in homes regardless of fungal species present. A de novo metatranscriptomic analysis was performed using house dust from across the US, incubated at 50%, 85%, or 100% equilibrium relative humidity (ERH) for 1 week.

Results Gene expression was a function of moisture (adonis2 $p < 0.001$), with fungal metabolic activity increasing with an increase in moisture condition (Kruskal–Wallis $p = 0.003$). Genes associated with fungal growth such as sporulation ($n = 264$), hyphal growth ($n = 62$), and secondary metabolism ($n = 124$) were significantly upregulated at elevated ERH conditions when compared to the low 50% ERH (FDR-adjusted $p \leq 0.001$, log2FC ≥ 2), indicating that fungal function is influenced by damp conditions. A total of 67 genes were identified as consistently associated with the elevated 85% or 100% ERH conditions and included fungal developmental regulators and secondary metabolite genes such as brlA (log2FC = 7.39, upregulated at 100% compared to 85%) and stcC (log2FC = 8.78, upregulated at 85% compared to 50%).

Conclusions Our results demonstrate that moisture conditions more strongly influence gene expression of indoor fungal communities compared to species presence. Identifying genes indicative of microbial growth under damp conditions will help develop robust monitoring techniques for indoor microbial exposures and improve understanding of how dampness and mold are linked to disease.

Keywords Moisture, Dampness, Fungi, Microbiome, Gene expression, Housing

*Correspondence:
Karen C. Dannemiller
Dannemiller.70@osu.edu
Full list of author information is available at the end of the article

Introduction

Exposure of asthmatics to mold in housing costs $22.4 billion per year in the United States alone [1, 2]. Water-damaged and moldy homes are consistently associated with asthma, respiratory and allergic health outcomes, in both children and adults [3–12]. These health effects disproportionately affect low-income and minority communities, including those with substandard housing conditions [6, 13]. These residents are often renters and/or may not have the resources for appropriate remediation of mold-damaged homes [14].

Health effects from damp homes are most strongly associated with subjective measures of mold such as visual inspection and detection of moldy odor compared to any available quantitative mold measure [3]. Repeated evidence suggests that it is microbial growth occurring in response to indoor dampness that mediates the link between exposures and health effects [3, 15–17]. Traditional methods to measure indoor mold using counts of microbial spores [18] and fungal indicators like glucans and ergosterol have not shown consistent associations with health effects [19]. Next-generation DNA sequencing-based tools using sequence analysis of microbes [20] have not yet been able to identify specific species as a consistent microbial signature of dampness [21–25]. Solely analyzing species composition changes in response to moisture is not sufficient to quantify microbial growth due to the influence of sampling site [21, 26, 27].

There is an urgent research need to develop new tools that improve building diagnosis and clearance certification for mold industry practitioners [23]. Microbial communities that grow in response to damp conditions express genes and have specific metabolic pathways and functional changes regardless of species that are present that may be strongly associated with negative health outcomes. Analysis of gene expression and metabolic changes in microbial communities have repeatedly acted as early and sensitive predictors of environmental shifts in other systems [28, 29]. Changing environmental factors like temperatures and moisture result in fungal growth with increased production of volatile organic compound emissions (VOCs) and mycotoxins [30–36]. Damp conditions lead to increased fungal allergen potency and metabolic activity that can result in degradation of chemicals such as phthalate esters in the dust [37, 38]. Growing fungal communities in house dust at elevated moisture conditions results in increased expression of genes encoding secondary metabolites, allergens, and pathogenicity factors [38, 39]. Fungal growth results in increased allergen release [40] and is also associated with proteins like hydrophobins and proteases that have direct impacts on evading host immune system response during exposure and correlate to asthma severity [41, 42].

Analyzing gene expression in the fungal communities in dust may yield promising options to help identify the best associations between potential microbial indicators of damp indoor environments and health effects.

The goal of this study is to identify genes consistently associated with fungal growth in indoor dust under damp conditions. These genes may be used in the future to inform the design of better indicators of moisture damage in homes that may be associated with human health effects. We performed a de novo metatranscriptome assembly on dust collected from different homes across the US and incubated them at 3 different relative humidity levels in laboratory chambers. We investigated processes associated with fungal growth enriched at elevated moisture conditions and discovered upregulated fungal genes from these growth pathways. The final set of genes are potential targets to use in homes to indicate harmful fungal growth regardless of the species present. Such genes and their products, after further validation, can be used as diagnostic indicators of moisture damage in homes. The results of this work, through the use of novel tools, identify microbial targets of moisture signature in homes and can provide a novel perspective to further the understanding of the health implications of dampness exposures.

Methods

Participant recruitment and dust collection

Floor dust samples were collected from nine different non-moisture damaged homes across the US from May 2021 to November 2021. Overall methods are shown in Fig. 1. Three homes were from Ohio, and the remaining six were homes from six different states in the US (Table S1, Fig. 2). Due to COVID-19 restrictions, we used an online approach for participant recruitment and instructions for dust collection by participant. Using participant-collected dust as a surrogate for collection by a project staff has shown to be equally effective for studies reporting allergen concentration in dust [43, 44]. Participants were initially recruited via social media, and additional recruitment and screening were completed over email.

A Qualtrics survey (Qualtrics, Provo, UT) containing the consent form, as well as questions on relevant home and indoor environmental measures, was used for screening participants. Participants were asked if there was any evidence of present water damage, moisture, leaks (such as damp carpet or leaky plumbing), or visible mold inside their homes. If participants answered in the affirmative, then these homes were not recruited for our study. The survey also contained information about the floor area and flooring type that was vacuumed, the frequency of vacuuming, types of floor cleaning, the number
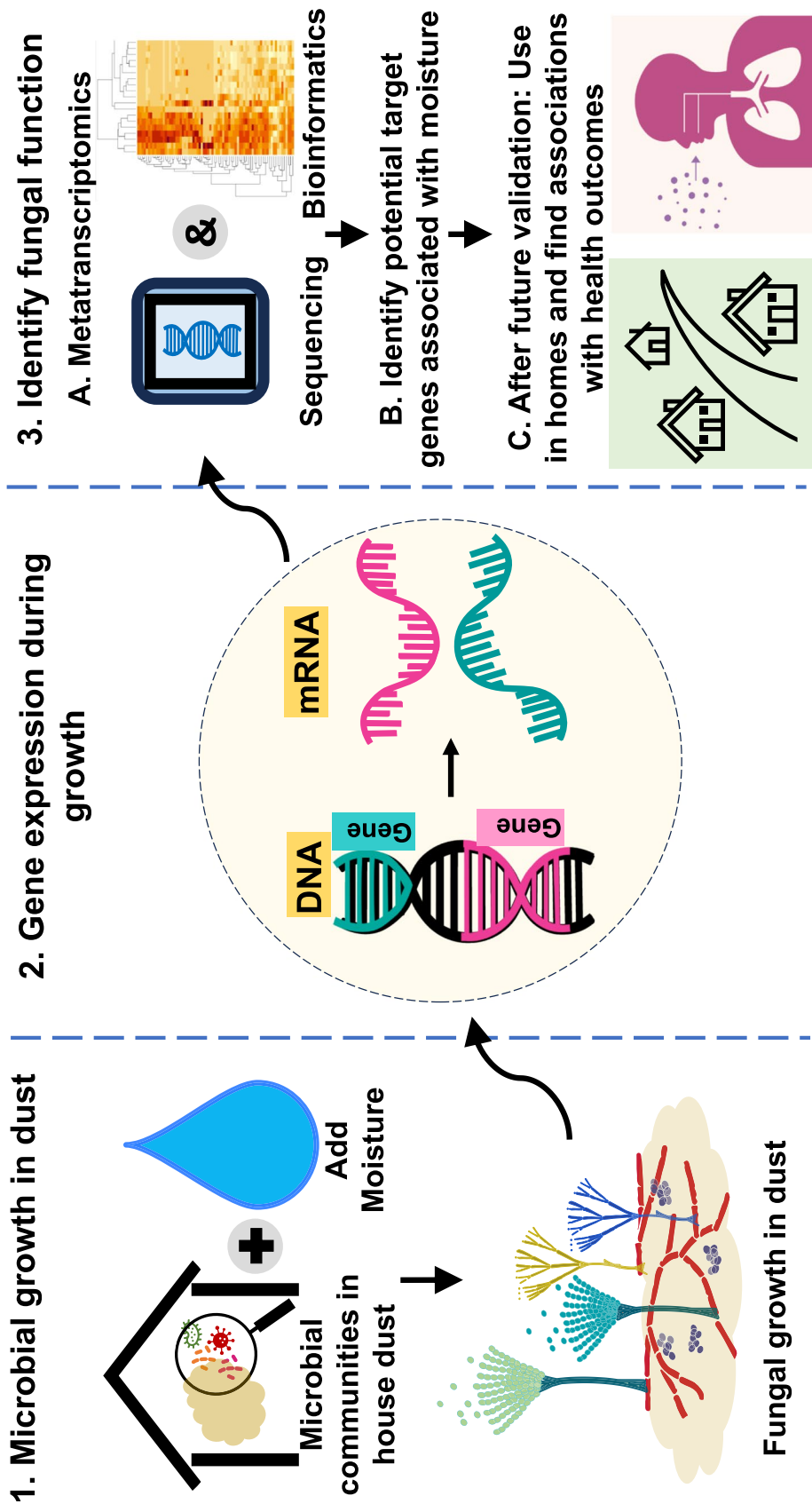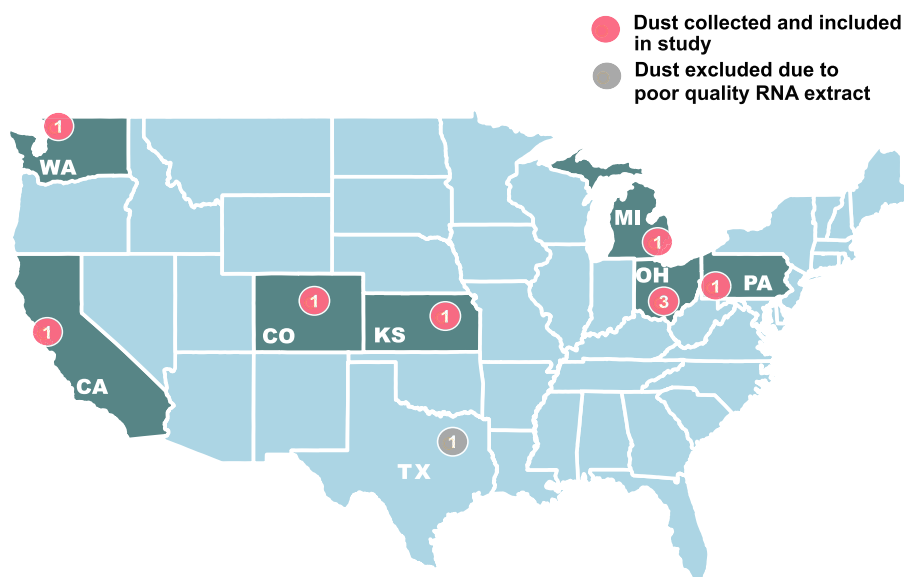
Fig. 1 Overview of methods

Fig. 2 Locations of participating homes that donated dust to the study. Dust from 9 homes, indicated by the pink colored dots, was collected and included for all experiments, sequencing, and downstream analysis. Dust from one home, indicated by the grey dot (located in Texas), was collected but excluded before sequencing due to low quality of extracted RNA. Number of collection sites (homes) within each location are shown by the numbers within the pink and grey colored dots. There were 3 collection sites in Ohio

of occupants (adults and children), number of pets (dogs, cats, birds, and other furry pets) as well as any prior history of moisture damage and mold in participants' homes within the last 5 years.

One home located in Texas was initially recruited but due to consistently low quality of the extracted RNA, the dust was not included in sequencing and was excluded from this study. Two of the homes had potential moisture damage even though the participants answered in the negative to "Is there evidence of water damage, moisture, or leaks (such as damp carpet or leaky plumbing)?" (Table S1). One home (dust sample ID: KS, Table S1) reported to have a temporary leak that occurred after heavy rains and was gone within 24 h, and the other reported to have a leak more than 10 years ago (dust sample ID: WA, Table S1). These two samples were not excluded because they did not meet the extent of moisture damage necessary for exclusion criteria due to the extent of the damage and length of time since the damage, respectively.

Dust collection instructions were sent to the participants over email. Participants were asked to collect floor dust (>25 g), emphasizing collection from the main living areas inside their homes (living room and bedroom) using their home vacuum. If the home vacuum did not contain a vacuum bag, participants were asked to remove dust from the canister and place it in a zip top bag. Participants were then asked to ship their collected dust to our lab or have it dropped

off to a designated location for us to pick up. Once we received the dust, all dust was screened to eliminate for the presence of SARS CoV-2, using a previously described protocol [45], and no dust samples were excluded. Recruitment and dust collection procedures were approved by The Ohio State University Behavioral Institutional Review Board (IRB) under study number 019B0457 for the duration of the study.

The collected dust was then hand-sieved to 300 μm to remove larger-sized dust particles. Sieved dust samples were stored in enclosed jars within a dark chamber covered in parafilm at 25 °C prior to chamber experiments. Dust for RNA analysis were sieved within 10 days of receipt and chamber experiments were performed within 3 weeks after sieving. Dust used for DNA extractions were stored for up to several months after sieving prior to incubations. DNA and RNA extractions were performed exactly 1 week after incubation began. Dust was transferred directly into the first extraction step without freezing. Dust was never frozen to maintain intact microbial communities.

Chamber experiments
For the chamber experiments, 100 mg aliquots of sieved dust were incubated in glass chambers at 25 °C for a period of 1 week, at relative humidities of 50%, 85%, and 100% ERH [17]. A total of 27 dust samples were incubated (9 sites×3 ERH conditions). Additional samples or replicates were not included due to the increased

computational time and cost needed to include more samples in our bioinformatics pipeline and the costs associated with RNA sequencing. Previous studies have shown when using replicate samples of dust, gene expression of samples cluster by moisture condition [38]. Relative humidity levels in the glass chambers were maintained using salt solutions or distilled water, as detailed in previous work [21]. Fifty percent and 85% ERH were maintained by using salt solutions with water activities of 0.5 a_w and 0.85 a_w, respectively, and distilled water was used to maintain an ERH of 100%. The water activities of these salt solutions were tested for accuracy using an AquaLab™ Dew Point Water Activity Meter (Decagon 125 Devices) with a margin of error of ± 0.005.

RNA extractions and nucleic acid sequencing

Immediately following the 1-week incubation, RNA was extracted from incubated dust using a previously used modified protocol of the Qiagen RNeasy PowerMicrobiome extraction kit (Qiagen, Hilden, Germany) [38]. To prevent RNA degradation from RNases, the manufacturer's protocol was modified to use 10× the concentration of β-mercaptoethanol in the first step and 70% ethanol in place of PM4 in the RNA binding step. Extracted RNA was immediately frozen at − 80 °C prior to use and transported on dry ice.

To ensure high RNA quality and integrity, all RNA extracts were analyzed using the High Sensitivity RNA ScreenTape analysis on the Agilent 4200 TapeStation Bioanalyzer (Agilent, Santa Clara, CA, USA) at The Genomics Shared Resource Center (The Ohio State University Comprehensive Cancer Center Shared Resources, Columbus, OH, USA).

RNA extracts were then sent to the Yale Center for Genomic Analysis (Yale University, New Haven, CT, USA) where they were reverse transcribed and then sequenced on a NovaSeq 2×100 lane with 25 million reads per sample. RNASeq library preparation was performed using the NEB Next Single Cell/Low Input RNA Library Prep Kit (New England Biolabs, USA) and the NEB Ultra II FS (New England Biolabs, USA) workflow for Illumina. The polyA selection protocol was used to select for eukaryotic mRNA. Sequence data was submitted to GenBank under accession number PRJNA1072816.

Initial processing, metatranscriptome assembly, and transcript quantification

Processing of sequenced reads followed protocols previously described [37, 38]. FastQC (v.0.12.0) was used for quality assessment of sequences [46]. rCorrector (v.1.0.6) was utilized to correct erroneous k-mers created due to sequencing errors [47]. After correction, reads deemed unfixable by rCorrector were filtered out using the TranscriptomeAssemblyTools package [48].

De novo metatranscriptome assembly was conducted using Trinity (v.2.12.0) [49] with default settings and was run on the Ohio Supercomputer (Ohio Supercomputer Center, Ohio). Trimmomatic within the Trinity pipeline was used to remove poor-quality reads and contigs with a length less than 300 base pairs (bp) [50, 51]. Contigs from the Trinity assembly were clustered using CD-HIT-EST (v.4.8.1) based on 80% sequence similarity [52, 53]. These clusters from CD-HIT-EST represent all expressed contigs and constitute the full transcriptome.

Abundance estimation and alignment were run within the Trinity pipeline with default parameters. RSEM (v.1.3.3) was used to estimate transcript abundance in each sample and to determine transcript-level expression counts of the RNA-Seq fragments for each transcript using alignment-based quantification [54]. Bowtie2 was used to align the quality trimmed paired-end reads after Trimmomatic to the full transcriptome created using CD-HIT-EST [55]. Read coverage was then quantified using Samtools to capture read alignment statistics for concordant read pairs (yielding concordant alignments 1 or more times to the CD-HIT-EST transcriptome) with a MAPQ greater than 2.

Transcript-level abundance estimates were used to construct a matrix of counts and a matrix of normalized expression values. Normalized expression values include Counts Per Million (CPM), Transcripts per Million (TPM) [56], and Trimmed Mean of M-values (TMM) [57] and account for transcript length, number of reads mapped to a transcript, total number of reads over all transcripts, and library size (sequencing depth). Gene-level count and gene-level normalized expression matrices were calculated using txImport [58] implemented directly in the Trinity pipeline.

Differential expression analysis

DESeq2 was used within the Trinity pipeline to perform Differential Gene Expression (DGE) analysis of expressed genes [59]. DGE performed using gene-level counts was used for downstream target gene identification. Performing differential expression analysis on gene levels, in addition to contig or transcript levels, improves interpretation of annotated contigs and potentially increases statistical power [60]. Pairwise comparisons between the three ERH conditions (50%, 85%, and 100%) were performed, giving rise to six pairwise ERH comparisons. Genes that were most differentially expressed based on the most significant False Discovery rate (FDR) [61] (FDR-adjusted $p \leq 0.001$) and log2FC ($\log_2$ fold change) values (log2FC ≥ 2) were extracted and used for subsequent Gene Ontology (GO) enrichment analysis.

Functional annotation and Gene Ontology enrichment

Transcripts were annotated using Trinotate (v.3.2.2), designed for comprehensive functional annotation of de novo transcriptomes [62]. Trinotate integrates all functional annotation data into an SQLite database, which is used to create a whole annotation report for the transcriptome. For functional annotation, Trinotate used BLAST + sequence homology search of transcripts and predicted coding regions against the SwissProt database [63, 64] and protein domain identification using a HMMER (v.3.3.2) search against the PFAM database [65, 66]. Predicted coding regions were identified using TransDecoder (v.5.5.0) that utilizes a minimum length open reading frame (ORF) found in a transcript sequence [67]. The TrEMBL/SwissProt database was used for Gene Ontology (GO) and KEGG assignments of transcripts using Trinotate [62, 68, 69]. KEGG assignments for genes were analyzed using the KEGG Mapper tool to identify the number of metabolic pathways [70] and visualized using the iPath3 tool [71] as metabolic pathway maps.

GOseq, developed specifically to account for gene length bias in RNA-seq data, was used within the Trinity pipeline to perform functional GO enrichment testing [72]. Results from the GO enrichment were analyzed for enriched GO categories based on the significance of enrichment using FDR values and the number of DE genes within these GO categories at each pairwise ERH comparison.

Identifying potential target genes associated with fungal growth at high moisture

GO enrichment was performed on the most highly significant and differentially expressed genes with a cutoff of FDR-adjusted $p \leq 0.001$ and log2FC ≥ 2. GO enrichment results were then analyzed for GO terms associated with fungal growth that were significantly enriched at higher moisture conditions (FDR < 0.05). Higher moisture conditions comprised of GO terms enriched at 100% compared to 85% ERH, enriched at 100% compared to 50% ERH, and enriched at 85% compared to 50% ERH. Finally, genes upregulated within these GO categories associated with fungal growth at higher ERH and having a known fungal annotation (BLASTX) were used to identify genes as potential targets that are indicative of mold growth.

Target genes were chosen based on the criteria that (i) genes are upregulated at high ERH conditions: upregulated at 100% compared to 85% ERH, at 100% compared to 50% ERH or at 85% compared to 50% ERH; (ii) genes have a log2FC ≥ 5; (iii) genes are upregulated (expressed) in at least two-thirds of sampling sites ($n \geq 6$, out of a total $n = 9$ locations); and (iv) genes are not expressed at the 50% ERH condition in any sample. Counts in the 0–10 range are usually considered "noise" [73], and therefore,

the target genes were required to have a count < 10 at 50% ERH. Exceptions were made for some genes that did not meet criteria (iii), if the gene was essential for fungal growth based on prior knowledge. Genes upregulated in at least three sampling sites ($n = 15$) were included based on this exception. Criteria (iv) was included because using such genes as a marker for moisture would be simpler in that they can be quantified without being dependent on increases in abundances/counts and would not need to be compared to a baseline level. If a gene was upregulated in more than one ERH comparison (for example, upregulated at 100% compared to 50% and at 100% compared to 85%), then the largest log2FC value was used, and the ERH comparison corresponding to the log2FC value was reported (Table S10).

To analyze how genes performed when compared to fungal taxa (species and genus) as targets of high moisture conditions, we applied similar criteria to fungal taxa. Fungal taxa were analyzed based on the criteria that they are (i) more abundant at high ERH conditions: upregulated at 100% compared to 85% ERH, at 100% compared to 50% ERH or at 85% compared to 50% ERH; (ii) more abundant in at least two-thirds of sampling sites ($n \geq 6$, out of a total $n = 9$ locations); and (iii) not expressed at the 50% ERH condition in any sample.

Species identification in samples

Fifty milligrams of sieved dust, identical to those used for RNA extractions, was used for DNA extractions. Dust samples were incubated for 1 week at 50%, 85%, and 100% ERH (similar to RNA extractions), prior to DNA extraction ($n = 27$, 9 sites × 3 ERH conditions). DNA extractions were performed using the Maxwell RSC PureFood GMO and Authentication Kit (Promega, USA). After addition of CTAB buffer, proteinase K, and RNAse, a 5-min bead beating step (BioSpec Products, Inc., Bartlesville, OK, USA) was added to release spore contents where bead tubes contained 0.3 g of 100 μm glass beads, 0.1 g of 500 μm glass beads, and 1 g of garnett particles (all BioSpec Products Incl., Bartlesville, OK, USA) [74]. Further protocol modification included (1) 30 min of room-temperature incubation between beat beating and centrifugation and (2) reduction of eluate volume to 75 μL. DNA extracts were stored at −20 °C. Fungal concentration in DNA extracts was measured using a qPCR assay targeting the 18 S rRNA gene with the universal fungal primer pair FF2/FR1 [75]. QPCR reagents, standards, and cycling parameters were as previously described [20]. Next-generation DNA sequencing was performed at Research and Testing Laboratory (Lubbock, TX, USA) using an Illumina MiSeq with 2 × 300 bp chemistry. The adapters ITS1F (5′ – CTTGGTCATTTAGAGGAAGTAA – 3′) and ITS2aR (5′ – GCTGCGTTCTTCATCGATGC – 3′)

were selected to amplify ITS1 region [76]. Raw sequence data is archived in GenBank (PRJNA1072816).

A DADA2-based bioinformatics pipeline customized for ITS sequences [77] was run using R [78] on Ohio Supercomputer (Ohio Supercomputer Center, Ohio). Adapters were first removed using *Cutadapt* [79], *BioStrings* [80], and *ShortRead* [81]. Denoising was performed using DADA2 [82] where the *maxEE* and *truncQ* parameters of the *filterAndTrim* function were both set to eight following Rolling et al. [83]. The UNITE version 9.0 database [84] was used for taxonomic identification. Absolute abundance of organisms in the samples was determined as described previously [39, 85].

Statistical analysis

The statistical analysis software R (v.4.2.2) [78] was used to perform statistical testing. To compare gene expression profiles based on moisture condition, relationships between samples were analyzed using Principal Component Analysis (PCA). Gene expression values in Counts Per Million (CPM) that account for library size normalization were used for PCA. Log_2 transformed and mean-centered standardization, typically applied in gene expression studies, were performed prior to analysis to reduce bias towards highly expressed transcripts [59, 86]. PCoA was performed for relative abundance of fungal taxa (Amplicon Sequence Variants (ASVs), species and genus) using Bray–Curtis distances. We also performed PCoA using Aitchison distances to compare gene expression and fungal taxa abundances. Aitchison distance is Euclidean distance after a centered log ratio (clr) transformation of data and uses relative abundances that are suitable for compositional microbiome data [87]. This ensures that the metric for sample distances is similar for both gene expression and taxa, for a more even comparison as PCA is identical to Principal Coordinates Analysis (PCoA) when using Euclidean distances. The adonis2 function in R using the *vegan* package (v.2.6.4) [88] was used to determine the statistical significance of ERH groupings ($p < 0.05$) from the Euclidean, Bray–Curtis, and Aitchison distance matrix. The test employed 10,000 permutations and used FDR [61] to adjust for multiple comparisons. Significance was defined at FDR-adjusted $p < 0.05$ [89]. A 95% confidence ellipse using the stat_ellipse function within the *ggplot2* package (v.3.4.3) [90] was created to compare moisture conditions to each other.

The Spearman rank correlation coefficient was calculated using the *corrplot* package [91] for differentially expressed genes based on moisture condition. False Discovery Rate was used to adjust for multiple comparisons, and only the correlation coefficients that were significant

(FDR-adjusted $p < 0.05$) were considered. The Spearman rank correlation coefficient determines the strength and direction in the relationship between the data where a value of 1 indicates the strongest positive correlation.

Gene expression heatmaps were plotted using the ComplexHeatmap [92] package in R, based on TMM-normalized (Trimmed Mean of M-values normalized) expression values for direct comparison of gene expression across samples [57]. Log_2 transformed and mean-centered standardization were performed prior to analysis to reduce bias towards highly expressed transcripts [59, 86].

To identify species with differences in absolute abundances between the ERH levels, the Kruskal–Wallis test was first performed to determine significant difference ($p < 0.05$), followed by pairwise Wilcoxon rank sum test using FDR [61] to control for multiple comparisons. To determine significant differences between the number of fungal genes present by ERH condition, the Kruskal–Wallis test followed by pairwise Wilcoxon rank sum test was performed, with FDR to adjust for multiple comparisons. These tests were also used for determining differences in absolute fungal concentrations based on ERH condition. Kruskal–Wallis tests were used as a non-parametric alternative to ANOVA when the data was determined to not be normally distributed (Shapiro–Wilk $p < 0.05$) [93].

Visualization

Figures in the manuscript were generated using R scripts (v.4.2.2) [78], scripts within Trinity [49], ggplot2 (v.3.4.3) [90], ComplexHeatmap (v.2.15.1) [92] for heatmaps (Fig. 6 and Supplementary Fig. S13), corrplot (v.0.92) [91] for Supplementary Fig. S6, iPath3 [71] for global metabolic pathway map (Fig. 4, Supplementary Fig. S9 and S10), Canva (https://www.canva.com) for generating a GIF image, Adobe Illustrator (v.28.3) (http://www.adobe.com/au/products/illustrator.html) for Fig. 1, and Inkscape (v.1.3) (https://inkscape.org/) for Fig. 1 and finalizing other figures. The map in Fig. 2 was created using the R packages ggplot2 (v.3.4.3) [90], maps (v.3.4.1) [94], and ggmaps (v.3.0.2) [95] and further finalized using Inkscape.

Results

Overview of metatranscriptomic dataset

RNA sequencing produced a total of 700,682,204 paired-end reads. Trinity assembled all high-quality reads into 1,983,474 contigs and 1,023,948 genes. The median contig length was 556 base pairs (bp) with a minimum size threshold of 300 bp. After quality filtering, on average, 70.07% of reads mapped back to the full transcriptome. When only retaining reads that were part of a properly mapping pair, a total of 54.83% of quality-filtered reads

were mapped (additional read and cluster quality statistics in Supplementary Fig. S1). The percent of reads that survived the quality filtering and mapping is similar to other metatranscriptomic studies, including studies performed using house dust [38, 96]. DNA sequencing yielded 2,122,711 total reads (an average of 78,618.93 reads per sample), and no samples were excluded for insufficient depth.

Moisture is more consistently associated with microbial function than species

Relative humidity condition is significantly associated with both gene expression (Fig. 3A) and species (Fig. 3B) in the samples (adonis2 $R^2 = 0.28$, $p \leq 0.0001$ and adonis2 $R^2 = 0.21$, $p \leq 0.0001$, respectively, Table S2). The difference is more pronounced in gene expression with nonoverlapping ellipses compared to taxa (ASVs in Fig. 3B, Supplementary Fig. S3A, species in Supplementary Figs. S2A and S3B, or genus in Supplementary Figs. S2B and S3C) with overlapping ellipses, indicating that this may be a stronger predictor of moisture in a sample than taxa (Table S2). For fungal taxa (Fig. 3D, Supplementary Figs. S2(C, D) and S3(D, E, F)), some samples cluster together more strongly by site than by ERH condition (CA 50%, CA 85%, and WA 50%, WA 85%), which is not observed for gene expression (Fig. 3C).

Samples cluster by moisture condition based on the gene expression heatmap (Supplementary Fig. S4) indicating gene expression is a function of relative humidity condition. These differences can also be seen in the MA plots (Supplementary Fig. S5) which indicate differentially expressed genes at each pairwise RH comparison, 100% vs 85%, 100% vs 50%, and 85% vs 50%. Differentially expressed genes were significantly correlated within moisture conditions (Spearman correlation, rho > 0.5, $p < 0.05$), and samples grouped by RH condition based on hierarchical clustering (Supplementary Fig. S6).

Many fungal genes are consistently expressed only at elevated moisture conditions (Supplementary Fig. S7, Table 1). Thousands of genes were upregulated at elevated ERH conditions (100% or 85% compared to 50% ERH) in a majority of samples and were not expressed at the low 50% condition (Table 1). A total of 732 genes were upregulated in all sites at either 100% or 85% ERH conditions when compared to 50% and were not expressed in any 50% samples (Table 1). Overall, this indicates that many genes are expressed 100% or 85% ERH or both. In contrast, no fungal taxa (species or genus) were found to be consistently associated with elevated ERH in all samples (Table 1, Supplementary Tables S3 and S4). Species that were more abundant at elevated ERH conditions (100% or 85%) were also present at the low 50% condition, similar to previous studies [21], and these are

difficult to use as indicators of moisture and need to be associated with increases in abundances. Out of the two species that were more abundant at the 85% ERH condition (*Aspergillus ruber* and *Aspergillus intermedius)*, both were found at 50% ERH (Table S3). For instance, *Aspergillus ruber* that is more differentially abundant at 85% ERH (in all sites) was also found in 7 sites at 50% ERH. Only one species (*Chaetomium angustispirale)* that was more abundant at 100% ERH in 8 sites (compared to 50%) was not found in any of the 50% ERH samples (Table S3). Overall, these results suggest that utilizing genes associated with elevated ERH conditions may potentially be able to overcome the inconsistencies associated with using fungal taxa as indicators of moisture.

Fungal metabolic activity increases with increase in moisture

Fungal gene expression (based on the number of fungal annotated genes) increased with increase in relative humidity condition (Kruskal–Wallis $p = 0.003$), with the 100% ERH condition, on an average, having 2.1 times the number of fungal annotated genes present at 50% ERH (Wilcoxon $p = 0.002$, Supplementary Fig. S8, Table S5).

There were a greater number of upregulated fungal genes at the 100% or 85% ERH conditions compared to the lower 50% ERH ($p \leq 0.001$, log2FC ≥ 2, Supplementary Fig. S7, Table S6). There were 1.8 times the number of significantly upregulated vs downregulated genes at 100% ERH when compared to 50% and 3.2 times the number of significantly upregulated genes at 100% when compared to the 85% ERH condition. We also found a greater number of fungal metabolic pathways upregulated at 100% and 85% ERH conditions than at 50% (Fig. 4, Supplementary Figs. S9 and S10). The 100% ERH condition ($n = 383$) had 3.3 times the number of fungal metabolic pathways as 50% ERH ($n = 117$), based on the 100% versus 50% ERH comparison (Table S7).

Similar to previous studies [17], fungal concentration increased with increase in ERH condition (Kruskal–Wallis $p = 0.007$, Supplementary Fig. S11). The fungal taxa present in the dust at the initial 50% ERH condition varied by site, with the majority in most sites being in the phylum Ascomycota (Supplementary Fig. S12).

Genes associated with fungal growth are upregulated at high relative humidity conditions

GO terms associated with fungal growth are enriched at the 100% and 85% ERH conditions compared to 50% ERH (FDR < 0.05) (Fig. 5). No growth-associated GO terms ($n = 0$, FDR < 0.05) were enriched at the low ERH condition (50% ERH as the upregulated condition), indicating that overall, gene expression related to fungal growth is
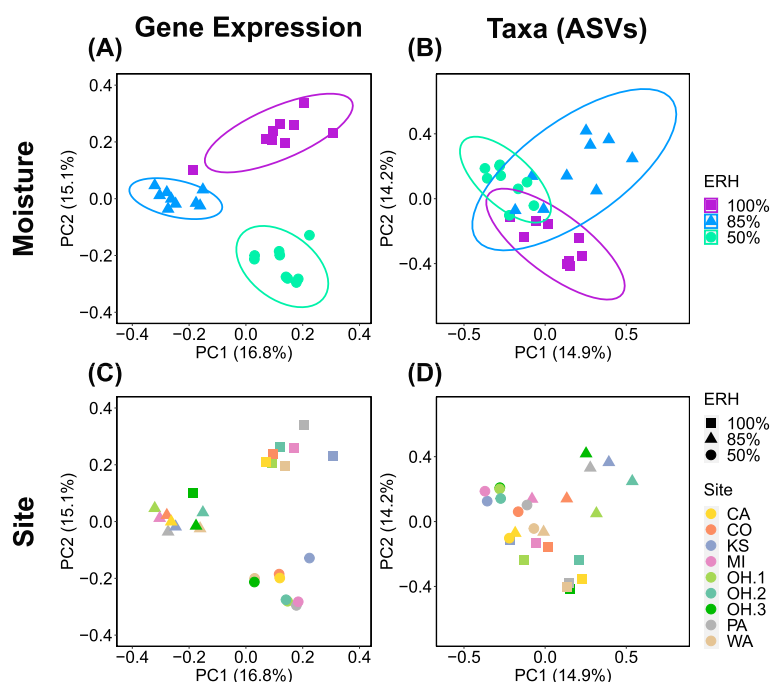
Fig. 3 **A** PCA of gene expression in house dust grouped by ERH, **B** PCoA of the relative abundance of fungal taxa (ASVs) in house dust grouped by ERH using Bray–Curtis distances, **C** PCA of gene expression colored by site, and **D** PCoA of the relative abundance of fungal ASVs colored by site (using Bray–Curtis distances). The color of samples is specific to ERH or site, and shapes are specific to ERH. Overlap between samples indicates greater similarity based on between-sample distance. A 95% confidence ellipse was added for each ERH condition in **A** and **B**

Table 1 Number of upregulated fungal genes and fungal species that are found to be more abundant at 100% compared to 50% and 85% compared to 50% (FDR-adjusted $p < 0.05$)

Gene expression		
	Upregulated at 100% vs 50%	Upregulated at 85% vs 50%
Number of genes upregulated (FDR-adjusted $p < 0.05$)	4141	12,845
Number of genes upregulated in at least 6/9 sites and not expressed at 50%	3188	5437
Number of genes upregulated in at least 8/9 sites and not expressed at 50%	2030	2528
Number of genes upregulated in all sites and not expressed at 50%	324	431
Taxa (species)		
	More abundant at 100% vs 50%	More abundant at 85% vs 50%
Number of species differentially abundant (FDR-adjusted $p < 0.05$)	3	2
Number of species more abundant in at least 6/9 sites and not found at 50%	1	0
Number of species more abundant in at least 8/9 sites and not found at 50%	1	0
Number of species more abundant in all sites and not found at 50%	0	0

associated with higher moisture conditions (Tables S8 and S9).

Morphological processes that occur during fungal growth are significantly enriched at both the 100% and 85% ERH conditions when compared to the low 50% ERH condition. Filamentous fungi begin to grow by elongating the tip of their hyphae [97], which is followed by the formation of reproductive growth structures and the production of spores (sporulation) [98]. Genes associated with the GO term "sporulation" were upregulated at the 100% and 85% ERH conditions when compared to 50% ERH. GO terms associated with hyphal elongation such as "cell septum" and "hyphal tip" were significantly enriched at 85% ERH when compared to 50% (FDR < 10^{-10}, Fig. 5, Table S8). The GO term "anatomical structure formation involved in morphogenesis" had the
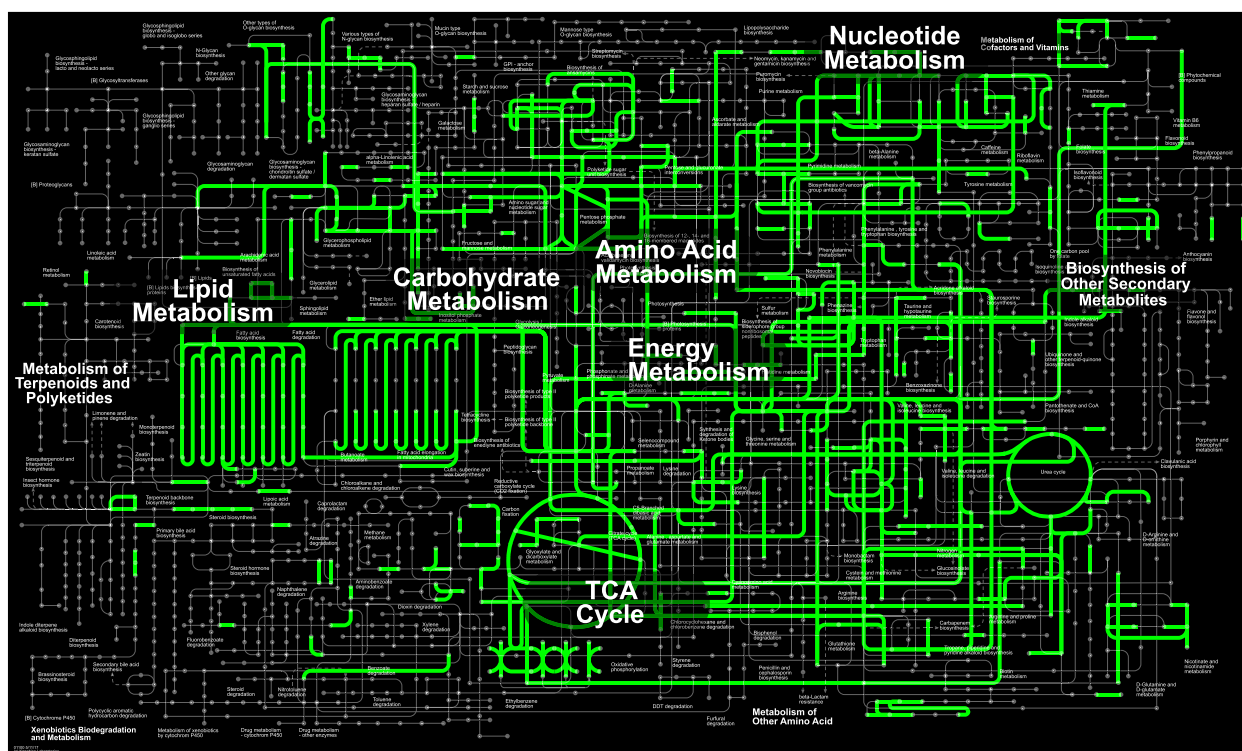
Fig. 4 Metabolic pathways map of fungal genes upregulated at 100% ERH when compared to 50% ERH
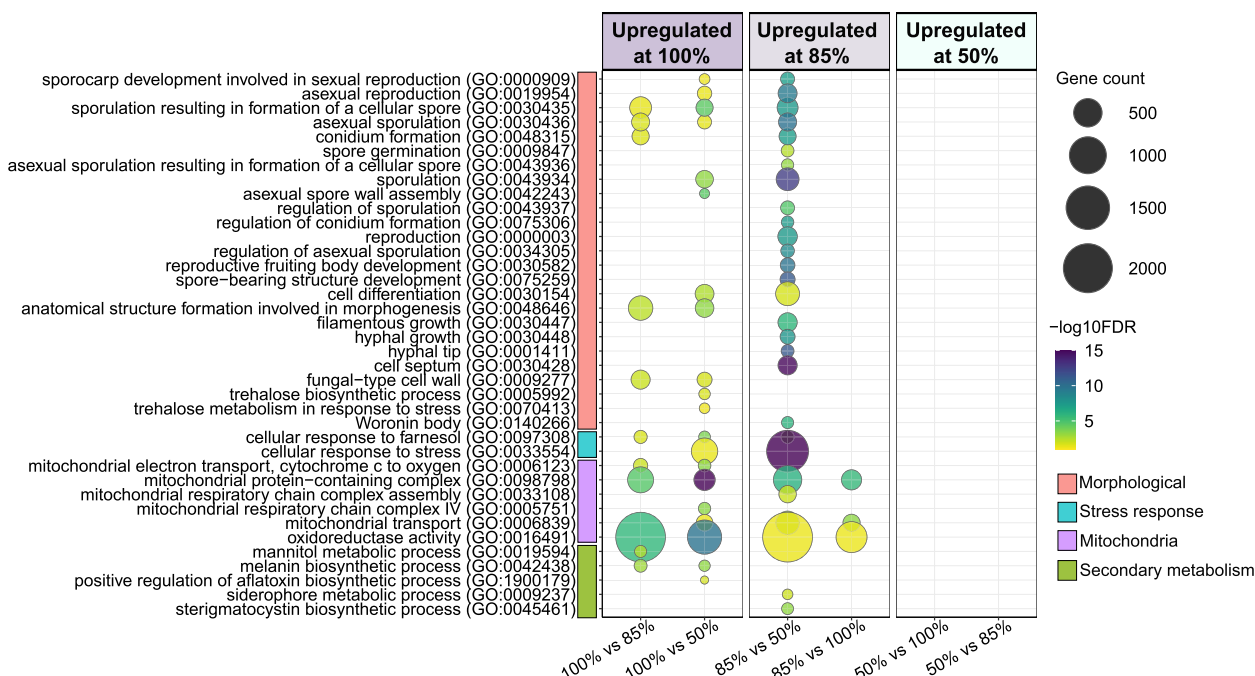

Fig. 5 Bubble plot of representative GO terms associated with fungal growth in all ERH comparisons. Bubble sizes represent the number of upregulated genes within a GO category for a specific ERH comparison. Bubble color values are based on the significance (-$\log_{10}$(FDR)) of the GO term with darker colors representing higher significance of GO enrichment. We grouped GO terms having similar functions into broader categories. Color bars next to GO terms indicate the four broader categories: Morphological, Stress response, Mitochondria and Secondary metabolism. Enriched GO terms along with the number of upregulated genes present can be found in Supplementary Tables S8 and S9

highest number of upregulated genes ($n = 323$) and was significantly enriched at 100% when compared to 85% ERH.

GO terms associated with fungal secondary metabolism are significantly enriched at 100% and 85% ERH conditions when compared to 50% ERH (FDR < 0.05). Secondary metabolic processes are chemical reactions and pathways that are not required for the growth and maintenance of the organism [99]. In filamentous fungi (mold), secondary metabolism includes the production of natural products such as pigments and harmful toxins such as mycotoxins and is often accompanied by fungal morphological growth and virulence [100]. Genes belonging to the term "melanin biosynthetic process" that are associated with the production of the fungal pigment melanin were significantly upregulated at 100% in both the 100% vs 85% and 100% vs 50% ERH comparisons ((FDR = 0.001 and FDR = 0.0005, respectively). Genes associated with fungal mycotoxin production, belonging to GO terms such as "sterigmatocystin biosynthetic process" and "positive regulation of aflatoxin biosynthetic process," were significantly upregulated at 100% and 85% ERH conditions when compared to 50% ERH (FDR < 0.05).

Genes associated with stress response were highly upregulated at 100% and 85% when compared to the low 50% ERH condition. For many filamentous fungi, the act of growing hyphal structures likely places significant stress on the secretory system [101]. A total of 1364 genes belonging to the term "cellular response to stress" (FDR = 2.54×10^{-21}) were upregulated at 85% when compared to 50% ERH condition. These included the bipA and cdc48 genes that function during secretory stress responses and are also required for normal hyphal growth and morphology [102, 103].

Genes associated with mitochondrial respiration and oxidoreductase activity were also found to be upregulated at 100% and 85% when compared to the low 50% ERH condition. Morphological transitions that occur during growth and virulence in fungi have been associated with mitochondrial respiratory activity in fungi [104]. A total of 214 genes belonging to the "mitochondrial protein-containing complex" GO term were found to be significantly upregulated at the 100% condition when compared to 50% ERH.

Hydrophobins, developmental regulators, and secondary metabolite genes are consistently associated with moisture

Overall, fungal growth-associated genes ($n = 67$) fell into 3 groups based on ERH condition, i.e., (1) upregulated at 100% ERH ($n = 29$), (2) upregulated at both 100% and 85% ERH ($n = 11$), and (3) upregulated at 85% ERH

($n = 27$) (Fig. 6, Supplementary Fig. S13, Supplementary Table S10). Across all groups, a majority of the genes ($n = 47$) were associated with fungal morphological processes (Fig. 7). Genes were also associated with stress response ($n = 19$), secondary metabolism ($n = 19$), and mitochondria-related processes ($n = 3$).

The most differentially expressed genes at 100% ERH were the alkaline protease gene, alp1 with log2FC of 25.69 (100% vs 85% ERH comparison, FDR-adjusted $p = 1.28 \times 10^{-27}$), followed by the pigment-related genes arp1 and wA (log2FC = 12.34 and 12, and FDR-adjusted $p = 7.15 \times 10^{-13}$ and 3.54×10^{-17} respectively, both for 100% vs 85% ERH comparison). These genes were predominantly part of morphological growth GO terms such as "conidium formation" (GO:0048315) and "sporulation resulting in formation of a cellular spore" (GO:0030435) (Tables S9 and S10). The genes arp1 and the wA were also associated with pigment biosynthesis, with arp1 associated with the term "melanin biosynthetic process" (GO: GO:0042438). Other morphological growth-associated genes that were highly expressed included the hydrophobin gene, rodA with log2FC = 8.68 at the 100% vs 50% ERH comparison (FDR-adjusted $p = 7.18 \times 10^{-08}$). Developmental regulator genes such as the brlA gene were also highly expressed with log2FC = 7.39 at the 100% vs 85% comparison (FDR-adjusted $p = 3.79 \times 10^{-06}$). The most upregulated genes at both 100% and 85% ERH also included mitochondria-related genes such as mdm10, which showed upregulation at both 85% and 100% with log2FC = 71.87 (FDR-adjusted $p = 8.94 \times 10^{-08}$, at the 100% versus 50% ERH comparison) and was associated with the "mitochondrial protein-containing complex" (GO:0098798) GO term.

The most consistently expressed gene upregulated at both 100% and 85% ERH was the Canx homolog gene (upregulated at 18 out of 18 sites at 100% and 85% ERH, Supplementary Table S10). This is associated with the "cellular response to stress" (GO:0033554) GO term (log2FC = 50.91, FDR-adjusted $p = 1.72 \times 10^{-05}$, at the 100% versus 50% ERH comparison). The hex1 gene associated with Woronin body during hyphal growth (GO:0140266) and the crzA gene associated with sporulation (GO:0043934) and conidium formation (GO:0048315) were also both consistently upregulated at 16 out of the total 18 sites at 100% and 85% ERH (log2FC = 9.09 and 30.39, respectively, FDR-adjusted $p = 8.67 \times 10^{-05}$ and 7.49×10^{-07}, respectively, at the 100% vs 50% ERH comparison).

At the 85% ERH, 20 out of the 27 total upregulated genes were associated with morphological growth processes. Of these, laeA and dop1 function as morphological growth regulators had the highest differential expression with log2FC = 11.65 (FDR-adjusted $p = 1.94 \times 10^{-16}$) and
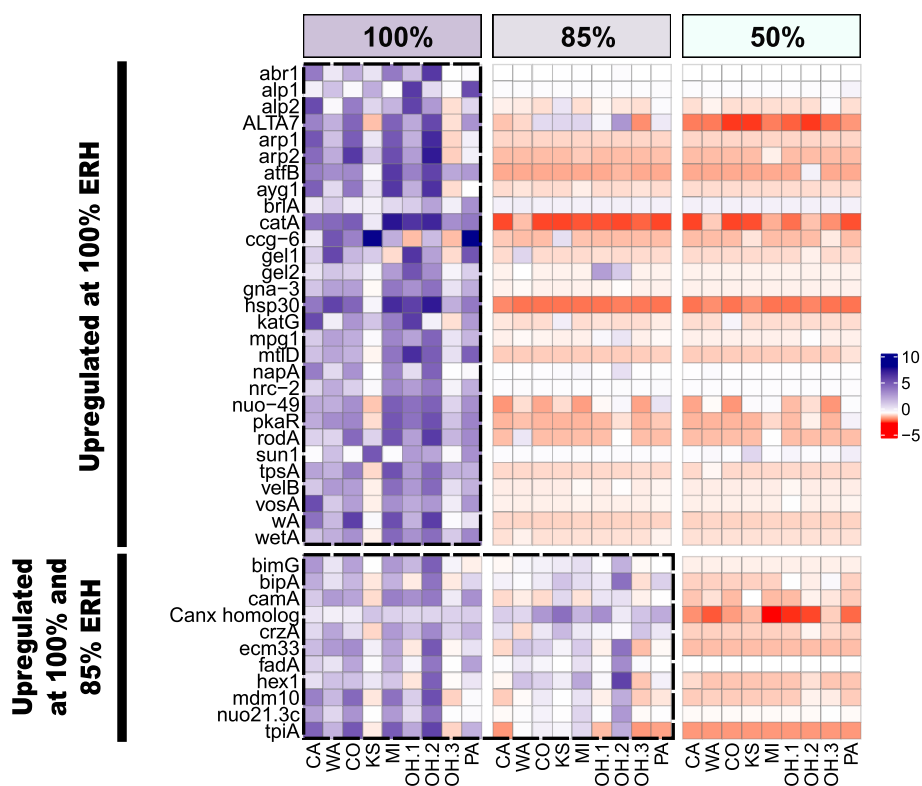
Fig. 6 Heatmap of TMM (Trimmed Mean of M-values)-normalized CPM (Counts Per Million) expression values of fungal target genes upregulated at 100% ERH condition (top group) and upregulated at both 100% and 85% ERH (bottom group). Blue colors represent higher gene expression values. The bottom of the heatmap shows state locations ordered from west to east. Genes are ordered alphabetically within each group
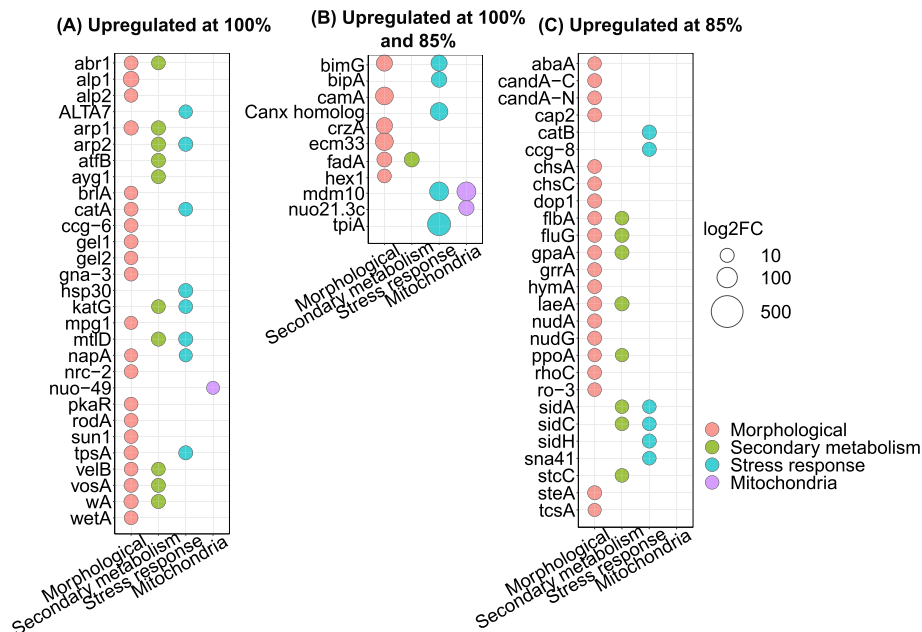

Fig. 7 Bubble plot of log2FC values for target genes in each of the three upregulated fungal gene groups along with their broad functional categories. Bubble colors represent functional categories, and the bubble size represents the magnitude of the log2FC value. Genes that correspond to multiple bubbles indicate that they are associated with those respective multiple functional categories. The log2FC values of genes upregulated at both 100% and 85% ERH are based on the 100% vs 50% comparison. All log2FC values of target genes and the corresponding ERH comparison used are provided in Supplementary Table S10

11.59 (FDR-adjusted $p = 1.33 \times 10^{-13}$) respectively both at 85% vs 50% ERH comparison. The laeA gene additionally functions as a secondary metabolic gene and is associated with the GO term "sterigmatocystin biosynthetic process" (GO:0045461). Similar to the 100% upregulated condition, developmental regulator genes such as flbA and fluG were highly expressed with log2FC = 8.69 (FDR-adjusted $p = 1.74 \times 10^{-11}$) and 10.76 (FDR-adjusted $p = 4.71 \times 10^{-12}$), respectively (85% vs 50% ERH comparison) having both morphological as well as secondary metabolic functions such as "sporulation" (GO:0043934) and "sterigmatocystin biosynthetic process" (GO:0045461).

Discussion

Species-based approaches have yet been unsuccessful in identifying a consistent microbial indicator of moisture damage in buildings that is more associated with health outcomes than subjective measures of visual or odor assessment [3, 12, 22, 25]. Our results demonstrate that gene expression of indoor fungal communities is more strongly driven by moisture condition than taxonomic differences in microbial communities. Genes expressed during growth showed consistent upregulation at elevated moisture conditions and may be used as improved indicators of water damage. The results of this study provide important direction that will be crucial in the search for quantitative indicators of moisture and mold damage in homes that outperform subjective measures in associations with health outcomes.

Function, rather than species, is consistently influenced by moisture

Buildings contain hundreds of different fungal species that vary by geographic location, building use, occupancy, and other factors [105–108]. Because these species vary greatly, the species composition also changes in different ways upon exposure to moisture [21]. However, there are gene clusters shared across the fungal kingdom [109]. We hypothesized that gene expression may be more consistently and clearly associated with ERH condition than species composition, and our results support that. We found 735 fungal annotated genes that were upregulated at elevated ERH conditions (either 100% or 85% ERH or both) across all 9 samples from 6 distinct geographical sites across the US. For instance, growth-associated genes encoding for the hydrophobin rodA (that supports aerial growth and attachment to solid supports [110]) and the sporulation-specific catalase catA were upregulated at 100% ERH in samples from every single sampling site (Fig. 5) and not expressed at 50% ERH condition in any site (Table S10).

Genes associated with a specific metabolic or functional response can span across a wide range of taxa, enabling the measurement of coordinated and multi-species responses to environmental changes. Similar processes occur in other environmental systems such as marine environments, soil, and groundwater microbiomes [111–114]. For instance, marine picoplankton populations exhibit cross-species, synchronous, and tightly regulated patterns of gene expression for many genes, particularly those genes associated with growth and nutrient acquisition [115]. Many microbial functions are conserved across taxa and may contribute to the higher sensitivity of gene expression to environmental changes over taxonomic composition [116, 117].

Gene expression associated with health effects: implications for housing quality

The work in this study also provides advanced insights into the microbial activity that occurs in damp indoor environments that are associated with health effects [3]. We found genes associated with allergens and mycotoxins upregulated at elevated ERH conditions when compared to 50%, similar to prior studies [38]. These included fungal allergen genes such as Alt a 7 upregulated at 100% compared to 50% ERH and secondary metabolite genes associated with mycotoxin production (GO:0045461) such as stcC that was upregulated at 85% when compared to 50% ERH. Genes associated with fungal growth also had associations with negative health effects in prior studies. The fungal alkaline protease gene alp1 (also known as the allergen Asp f 13 gene) was upregulated at 100% compared to 85% ERH and has strong correlations with asthma severity and respiratory dysfunction [42] and potential functions in promoting fungal growth and infection development in the host [118]. The hydrophobin gene associated with fungal spore surfaces, rodA (rodlet protein or rodlet layer), can evade human host immune responses [119, 120]. Genes related to mitochondrial functions, such as mdm10, were upregulated at 85% ERH compared to 50% and have potential associations with fungal virulence by regulating stress responses and mediating morphogenetic transitions [104]. Fungal exposure is linked to asthma exacerbations in both children and adults [105, 121], and these results suggest that the metabolic state, rather than specific taxa, may be more strongly associated than taxa with negative health effects linked to damp buildings. These associations will need to be evaluated in future studies.

Function can help identify novel targets to indicate mold growth indoors

Targeting metabolic functions specific to high moisture conditions may be a more robust approach than species-based indicators to identifying microbial indicators of moisture damage based on these results. Targeting genes

that are upregulated at both the 100% and 85% ERH conditions (compared to 50%) or using multiple genes where some are indicative of the 100% condition and others of 85%, may be better at detecting microbial changes at the onset of dampness. A quantitative microbial indicator of moisture would, at minimum, need to be consistently upregulated in most (if not all) sampling sites at high ERH conditions, but not expressed or expressed only at low levels at the low 50% condition. Such a fungal target could be used in homes similarly to fecal indicators in water systems. For instance, crAssphage is a human gut-associated bacteriophage that can be used as a viral indicator of human fecal pollution which is potentially quantitatively representative of viral pathogen fate and concentration changes in sewage-contaminated waters [122, 123]. The target gene groups reported in our study may be able to measure moisture and mold damage in homes and help correlate these measurements to occupant exposure and health outcomes in a quantitative manner. Future studies will be needed to confirm if these measures outperform current measurement methods. Ultimately, these targets could be integrated into standards and regulations.

Limitations

Here, the computational intensity of the RNA analysis on fungal communities limited us to 9 sites from 6 geographic regions of the US, so targets may need to be further validated in other areas of the country and the rest of the world. This study is also subject to standard limitations associated with metatranscriptomic analyses of fungal communities, where identification of genes is subject to database and sequencing limitations, with fungal metatranscriptomics studies suffering from the lack of completely annotated genomes [124, 125]. Many contigs do not have an identified functional annotation (hypothetical), and insufficient sequencing depth and algorithmic difficulties during the metatranscriptome assembly also cause additional redundancies in functional annotations. Advances in fungal genomics and sequencing technologies can overcome such challenges. This was a laboratory-based incubation of these samples, and results may vary when moisture is elevated in homes. Future studies will be needed to evaluate associations between these targets and negative health outcomes.

Conclusion

Overall, our work improved understanding of the functional processes occuring within indoor fungal communities and demonstrated that high moisture is associated with growth processes, upregulation of secondary metabolic pathways, and increased mitochondrial activity.

Upregulation of these genes was more strongly associated with high moisture than taxonomic measures. Together with other work, our findings strongly suggest that we need to move beyond the assumption that a microbial indicator of moisture in homes must be identified through species-based approaches or that an indicator is solely taxonomic in nature. Ideally, selected target genes or their products from our gene groups after further validation can be used in quantitative measurement systems that can perform sensitive detection of moisture damage in homes. Such a system would address both the substantial financial and health impact of mold growth in our society and be especially important for vulnerable groups such as children with asthma.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s40168-024-01915-9.

Additional file 1: Tables S1 to S10 and Figs. S1 to S13.

Authors' contributions
KCD conceived the idea, obtained funding, obtained IRB approval, and oversaw the project. NB collected and processed the samples and performed the bioinformatics on the RNA data. JCK processed and analyzed the DNA data. BH helped develop the RNA data processing pipeline and consulted on the bioinformatics analysis. NB and KCD wrote the initial version of the manuscript. All authors reviewed, edited, and approved the final version of the manuscript.

Data availability
Metatranscriptomic and ITS1 sequence data were submitted to GenBank under accession number PRJNA1072816. Original R scripts used for analysis and visualizations, including the figures generated are available in GitHub (https://github.com/n-bsub/Mold_metatranscriptomics).

Declarations

Ethics approval and consent to participate
Recruitment and dust collection procedures were approved by The Ohio State University Behavioral Institutional Review Board (IRB) under study number 019B0457 for the duration of the study.

Consent for publication
Not applicable.

Competing interests
All the authors have submitted a patent application on the material included in this manuscript.

Author details
[1]Environmental Sciences Graduate Program, The Ohio State University, Columbus, OH 43210, USA. [2]Department of Civil, Environmental & Geodetic Engineering, College of Engineering, The Ohio State University, 470 Hitchcock Hall, 2070 Neil Ave, Columbus, OH 43210, USA. [3]Division of Environmental Health Sciences, College of Public Health, The Ohio State University, Columbus, OH 43210, USA. [4]Department of Civil & Environmental Engineering, College of Engineering, Case Western Reserve University, Cleveland, OH 44106, USA. [5]Sustainability Institute, The Ohio State University, Columbus, OH 43210, USA.

References
1. Mudarri DH. Valuing the economic costs of allergic rhinitis, acute bronchitis, and asthma from exposure to indoor dampness and mold in the US. J Environ Public Health. 2016;2016:2386596.
2. Mudarri D, Fisk WJ. Public health and economic impact of dampness and mold. Indoor Air. 2007;17:226–35.
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;119:748–56.
4. Jaakkola MS, Quansah R, Hugg TT, Heikkinen SAM, Jaakkola JJK. Association of indoor dampness and molds with rhinitis risk: a systematic review and meta-analysis. J Allergy Clin Immunol. 2013;132:1099-1110.e18.
5. Fisk WJ, Lei-Gomez Q, Mendell MJ. Meta-analyses of the associations of respiratory health effects with dampness and mold in homes. Indoor Air. 2007;17:284–96.
6. IOM (Institute of Medicine). Damp indoor spaces and health. Washington D.C: The National Academic Press; 2004.
7. Fisk WJ, Eliseeva EA, Mendell MJ. Association of residential dampness and mold with respiratory tract infections and bronchitis: a meta-analysis. Environ Health. 2010;9:72.
8. Kanchongkittiphon W, Mendell MJ, Gaffin JM, Wang G, Phipatanakul W. Indoor environmental exposures and asthma exac- erbation: an update to the 2000 review by the Institute of. Med Environ Health Perspect. 2015;123:6–20.
9. Reponen T, Lockey J, Bernstein DI. Infant origins of child- hood asthma associated with specific molds. J Allergy Clin Immunol. 2012;130:639.
10. Iossifova YY, Reponen T, Ryan PH. Mold exposure during in- fancy as a predictor of potential asthma development. Ann Allergy Asthma Immunol. 2009;102:131–7.
11. Shorter C, Crane J, Pierse N. Indoor visible mold and mold odour are associated with new-onset childhood wheeze in a dose dependent manner. Indoor Air. 2018;28:6–15.
12. Mendell MJ, Adams RI. The challenge for microbial measurements in buildings. Indoor Air. 2019;29:523–6.
13. Bryant-Stephens T. Asthma disparities in urban environments. J Allergy Clin Immunol. 2009;123:1199–206 quiz 1207–8.
14. WHO. Guidelines for indoor air quality: dampness and mould. Copenhagen: World Health Organization; 2009.
15. Mendell MJ, Kumagai K. Observation-based metrics for residential dampness and mold with dose–response relationships to health: A review. Indoor Air. 2017;27:506–17.
16. Pasanen A-L, Juutinen T, Jantunen MJ, Kalliokoski P. Occurrence and moisture requirements of microbial growth in building materials. Int Biodeterior Biodegradation. 1992;30:273–83.
17. Dannemiller KC, Weschler CJ, Peccia J. Fungal and bacterial growth in floor dust at elevated relative humidity levels. Indoor Air. 2017;27:354–63.
18. Méheust D, Le Cann P, Reboux G, Millon L, Gangneux J-P. Indoor fungal contamination: health risks and measurement methods in hospitals, homes and workplaces. Crit Rev Microbiol. 2014;40:248–60.
19. Choi H, Byrne S, Larsen LS, Sigsgaard T, Thorne PS, Larsson L, et al. Residential culturable fungi, (1–3, 1–6)-β-d-glucan, and ergosterol concentrations in dust are not associated with asthma, rhinitis, or eczema diagnoses in children. Indoor Air. 2014;24:158–70.
20. Dannemiller KC, Mendell MJ, Macher JM, Kumagai K, Bradman A, Holland N, et al. Next-generation DNA sequencing reveals that low fungal diversity in house dust is associated with childhood asthma development. Indoor Air. 2014;24:236–47.
21. Haines SR, Siegel JA, Dannemiller KC. Modeling microbial growth in carpet dust exposed to diurnal variations in relative humidity using the "Time-of-Wetness" framework. Indoor Air. 2020;30:978–92.
22. Adams RI, Sylvain I, Spilak MP, Taylor JW, Waring MS, Mendell MJ. Fungal signature of moisture damage in buildings: identification by targeted and untargeted approaches with mycobiome data. Appl Environ Microbiol. 2020;86:e01047.https://doi.org/10.1128/AEM.01047-20
23. Peccia J, Haverinen-Shaughnessy U, Täubel M, Gentner DR, Shaughnessy R. Practitioner-driven research for improving the outcomes of mold inspection and remediation. Sci Total Environ. 2021;762:144190.
24. Jayaprakash B, Adams RI, Kirjavainen P, Karvonen A, Vepsäläinen A, Valkonen M, et al. Indoor microbiota in severely moisture damaged homes and the impact of interventions. Microbiome. 2017;5:138.
25. Adams RI, Leppänen H, Karvonen AM, Jacobs J, Borràs-Santos A, Valkonen M, et al. Microbial exposures in moisture-damaged schools and associations with respiratory symptoms in students: a multi-country environmental exposure study. Indoor Air. 2021;31:1952–66.
26. Amend AS, Seifert KA, Samson R, Bruns TD. Indoor fungal composition is geographically patterned and more diverse in temperate zones than in the tropics. Proc Natl Acad Sci U S A. 2010;107:13748–53.
27. Haines SR, Hall EC, Marciniak K, Misztal PK, Goldstein AH, Adams RI, et al. Microbial growth and volatile organic compound (VOC) emissions from carpet and drywall under elevated relative humidity conditions. Microbiome. 2021;9:209.
28. Bettina NS, Webster DG. Microbial indicators as a diagnostic tool for assessing water quality and climate stress in coral reef ecosystems. Mar Biol. 2017;164:1–18.
29. Nowicki EM, Shroff R, Singleton JA, Renaud DE, Wallace D, Drury J, et al. Microbiota and metatranscriptome changes accompanying the onset of gingivitis. MBio. 2018;9:10. https://doi.org/10.1128/mBio.00575-18.
30. Misaghi IJ. Influence of environment and culture media on spore morphology of Alternaria alternata. Phytopathology. 1978;68:29.
31. 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;541:365–71.
32. Low SY, Dannemiller K, Yao M, Yamamoto N, Peccia J. The allergenicity of Aspergillus fumigatus conidia is influenced by growth temperature. Fungal Biol. 2011;115:625–32.
33. Bertolini P, Tian SP. Effect of temperature of production of Botrytis allii conidia on their pathogenicity to harvested white onion bulbs. Plant Pathol. 1997;46:432–8.
34. Wolf J, Neill O, Rogers NR, Muilenberg CA, Ziska ML. Elevated atmospheric carbon dioxide concentrations amplify Alternaria alternate sporulation and total antigen production. Environ Health Perspect. 2010;118:1223–8.
35. Abbas HK, Egley GH, Paul RN. Effect of conidia production temperature on germination and infectivity of Alternaria helianthi. Phytopathology. 1995;85:667–82.
36. Phillips D. Changes in conidia of Monilinia fructicola in response to incubation temperature. Phytopathology. 1982;72:1281–3.
37. Bope A, Haines SR, Hegarty B, Weschler CJ, Peccia J, Dannemiller KC. Degradation of phthalate esters in floor dust at elevated relative humidity. Environ Sci Process Impacts. 2019;21:1268–79.
38. Hegarty B, Dannemiller KC, Peccia J. Gene expression of indoor fungal communities under damp building conditions: implications for human health. Indoor Air. 2018;28:548–58.
39. Nastasi N, Haines SR, Xu L, da Silva H, Divjan A, Barnes MA, et al. Morphology and quantification of fungal growth in residential dust and carpets. Build Environ. 2020;174:106774.
40. Green BJ, Tovey ER, Sercombe JK, Blachere FM, Beezhold DH, Schmechel D. Airborne fungal fragments and allergenicity. Med Mycol. 2006;44(Suppl 1):S245–55.
41. Croft CA, Culibrk L, Moore MM, Tebbutt SJ. Interactions of Aspergillus fumigatus conidia with airway epithelial cells: a critical review. Front Microbiol. 2016;7:472.

42. Basu T, Seyedmousavi S, Sugui JA, Balenga N, Zhao M, Kwon Chung KJ, et al. Aspergillus fumigatus alkaline protease 1 (Alp1/Asp f13) in the airways correlates with asthma severity. J Allergy Clin Immunol. 2018;141:423-425.e7.

43. Sever M, Arbes SJ Jr, Vaughn B, Mehta J, Lynch JT, Mitchell H, et al. Feasibility of using subject-collected dust samples in epidemiological and clinical studies of indoor allergens. J Allergy Clin Immunol. 2005;115:S97.

44. Barnes C, Portnoy JM, Ciaccio CE, Pacheco F. A comparison of subject room dust with home vacuum dust for evaluation of dust-borne aeroallergens. Ann Allergy Asthma Immunol. 2013;110:375–9.

45. Renninger N, Nastasi N, Bope A, Cochran SJ, Haines SR, Balasubrahmaniam N, et al. Indoor dust as a matrix for surveillance of COVID-19. mSystems. 2021;6:e01350.https://doi.org/10.1128/mSystems.01350-20

46. Babraham Bioinformatics - FastQC A quality control tool for high throughput sequence data. Accessed v0.11.9 on 08/14/2022. 2010. Available from: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/. Cited 2022 Aug 14

47. Song L, Florea L. Rcorrector: efficient and accurate error correction for Illumina RNA-seq reads. Gigascience. 2015;4:48.

48. https://github.com/harvardinformatics/transcriptomeassemblytools. Accessed FilterUncorrectabledPEfastq.py on 08/15/2022. 2016]. Available from:https://github.com/harvardinformatics/TranscriptomeAssemblyTools/blob/master/FilterUncorrectablePEfastq.py

49. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, et al. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat Biotechnol. 2011;29:644–52.

50. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014;30:2114–20.

51. 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.

52. Li W, Godzik A. Cd- hit : a fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics. 2006;22:1658–9.

53. Fu L, Niu B, Zhu Z, Wu S, Li W. CD- HIT: accelerated for clustering the next- generation sequencing data. Bioinformatics. 2012;28:3150–2.

54. Bo L, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinform. 2011;12:323.

55. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012;9:357–9.

56. Wagner GP, Kin K, Lynch VJ. Measurement of mRNA abundance using RNA-seq data: RPKM measure is inconsistent among samples. Theory Biosci. 2012;131:281–5.

57. Robinson MD, Oshlack A. A scaling normalization method for differential expression analysis of RNA-seq data. Genome Biol. 2010;11:R25.

58. Soneson C, Love MI, Robinson MD. Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. F1000Res. 2015;4:1521.

59. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014;15:550.

60. Davidson NM, Oshlack A. Corset: enabling differential gene expression analysis for de novo assembled transcriptomes. Genome Biol. 2014;15:410.

61. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc. 1995;57:289–300.

62. Bryant DM, Johnson K, DiTommaso T, Tickle T, Couger MB, Payzin-Dogru D, et al. A tissue-mapped axolotl de novo transcriptome enables identification of limb regeneration factors. Cell Rep. 2017;18:762–76.

63. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, et al. BLAST+: architecture and applications. BMC Bioinformatics. 2009;10:421.

64. Boutet E, Lieberherr D, Tognolli M, Schneider M, Bansal P, Bridge AJ, et al. UniProtKB/Swiss-Prot, the manually annotated section of the UniProt KnowledgeBase: how to use the entry view. Methods Mol Biol. 2016;1374:23–54.

65. Sonnhammer EL, Eddy SR, Durbin R. Pfam: a comprehensive database of protein domain families based on seed alignments. Proteins. 1997;28:405–20.

66. Mistry J, Chuguransky S, Williams L, Qureshi M, Salazar GA, Sonnhammer ELL, et al. Pfam: The protein families database in 2021. Nucleic Acids Res. 2021;49:D412–9.

67. Haas BJ. TransDecoder (Find Coding Regions Within Transcripts). Accessed TransDecoder-v5.5.0 on 10/09/2022. 2017. Available from: https://github.com/TransDecoder/TransDecoder. Cited 2022 Oct 9

68. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium Nat Genet. 2000;25:25–9.

69. Kanehisa M, Goto S, Furumichi M, Tanabe M, Hirakawa M. KEGG for representation and analysis of molecular networks involving diseases and drugs. Nucleic Acids Res. 2010;38:D355–60.

70. Kanehisa M, Sato Y. KEGG Mapper for inferring cellular functions from protein sequences. Protein Sci. 2020;29:28–35.

71. Darzi Y, Letunic I, Bork P, Yamada T. iPath3.0: interactive pathways explorer v3. Nucleic Acids Res. 2018;46:W510-3.

72. Young MD, Wakefield MJ, Smyth GK, Oshlack A. Gene ontology analysis for RNA-seq: accounting for selection bias. Genome Biol. 2010;11:R14.

73. Law CW, Alhamdoosh M, Su S, Dong X, Tian L, Smyth GK, et al. RNA-seq analysis is easy as 1–2–3 with limma Glimma and edgeR. F1000Res. 2016;5:1408.

74. Hospodsky D, Yamamoto N, Peccia J. Accuracy, precision, and method detection limits of quantitative PCR for airborne bacteria and fungi. Appl Environ Microbiol. 2010;76:7004–12.

75. Zhou G, Whong WZ, Ong T, Chen B. Development of a fungus-specific PCR assay for detecting low-level fungi in an indoor environment. Mol Cell Probes. 2000;14:339–48.

76. Bokulich NA, Mills DA. Improved selection of internal transcribed spacer-specific primers enables quantitative, ultra-high-throughput profiling of fungal communities. Appl Environ Microbiol. 2013;79:2519–26.

77. Callahan BJ. DADA2 ITS Pipeline Workflow (1.8). Available from: https://benjjneb.github.io/dada2/ITS_workflow.html. Cited 2024 Jan 9

78. R Core Team (2022). R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria; 2022. Available from: URL https://www.R-project.org/

79. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet J. 2011;17:10.

80. Hervé Pagès, Patrick Aboyoun, Robert Gentleman, Saikat DebRoy. Biostrings: efficient manipulation of biological strings. 2024; Available from: https://bioconductor.org/packages/Biostrings.

81. Morgan M, Anders S, Lawrence M, Aboyoun P, Pagès H, Gentleman R. ShortRead: a bioconductor package for input, quality assessment and exploration of high-throughput sequence data. Bioinformatics. 2009;25:2607–8.

82. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP. DADA2: High-resolution sample inference from Illumina amplicon data. Nat Methods. 2016;13:581–3.

83. Rolling T, Zhai B, Frame J, Hohl TM, Taur Y. Customization of a DADA2-based pipeline for fungal internal transcribed spacer 1 (ITS1) amplicon data sets. JCI Insight. 2022;7:e151663.https://doi.org/10.1172/jci.insight.151663

84. Abarenkov K, Nilsson RH, Larsson K-H, Taylor AFS, May TW, Frøslev TG, et al. The UNITE database for molecular identification and taxonomic communication of fungi and other eukaryotes: sequences, taxa and classifications reconsidered. Nucleic Acids Res. 2024;52:D791–7. https://pubmed.ncbi.nlm.nih.gov/37953409/.

85. Dannemiller KC, Lang-Yona N, Yamamoto N, Rudich Y, Peccia J. Combining real-time PCR and next-generation DNA sequencing to provide quantitative comparisons of fungal aerosol populations. Atmos Environ. 1994;2014(84):113–21.

86. Ringnér M. What is principal component analysis? Nat Biotechnol. 2008;26:303–4.

87. Gloor GB, Macklaim JM, Pawlowsky-Glahn V, Egozcue JJ. Microbiome datasets are compositional: and this is not optional. Front Microbiol. 2017;8:2224. https://doi.org/10.3389/fmicb.2017.02224.

88. Dixon P. VEGAN, a package of R functions for community ecology. J Veg Sci. 2003;14:927–30.

89. Storey JD, Taylor JE, Siegmund D. Strong control, conservative point estimation and simultaneous conservative consistency of false discovery rates: a unified approach. J R Stat Soc Series B Stat Methodol. 2004;66:187–205.

90. Wickham H. ggplot2. Cham: Springer International Publishing; 2016. https://link.springer.com/book/10.1007/978-3-319-24277-4.

91. Wei T, Simko V, Levy M, Xie Y, Jin Y, Zemla J. Package 'corrplot.' Statistician. 2017;56:316–24.

92. Gu Z, Eils R, Schlesner M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. Bioinformatics. 2016;32:2847–9.

93. Shapiro SS, Wilk MB. An analysis of variance test for normality (complete samples). Biometrika. 1965;52:591–611.

94. Original S code by Richard A. Becker and Allan R. Wilks. R version by Ray Brownrigg. Enhancements by Thomas P Minka and Alex Deckmyn. maps: draw geographical maps. 2022. Available from: {https://CRAN.R-project.org/package=maps}

95. Kahle D, Wickham H. Ggmap: Spatial visualization with ggplot2. R J. 2013;5:144.

96. 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.

97. Steinberg G, Peñalva MA, Riquelme M, Wösten HA, Harris SD. Cell biology of hyphal growth. Microbiol Spectr. 2017;5:10. https://doi.org/10.1128/microbiolspec.funk-0034-2016.

98. Riquelme M, Aguirre J, Bartnicki-García S, Braus GH, Feldbrügge M, Fleig U, et al. Fungal morphogenesis, from the polarized growth of hyphae to complex reproduction and infection structures. Microbiol Mol Biol Rev. 2018;82:e00068.https://doi.org/10.1128/MMBR.00068-17

99. Carbon S, Ireland A, Mungall CJ, Shu S, Marshall B, Lewis S, et al. AmiGO: online access to ontology and annotation data. Bioinformatics. 2009;25:288–9.

100. Keller NP, Turner G, Bennett JW. Fungal secondary metabolism - from biochemistry to genomics. Nat Rev Microbiol. 2005;3:937–47.

101. Cramer RA. Secretion stress and fungal pathogenesis: a new, exploitable chink in fungal armor? Virulence. 2011;2:1–3.

102. Richie DL, Feng X, Hartl L, Aimanianda V, Krishnan K, Powers-Fletcher MV, et al. The virulence of the opportunistic fungal pathogen Aspergillus fumigatus requires cooperation between the endoplasmic reticulum-associated degradation pathway (ERAD) and the unfolded protein response (UPR). Virulence. 2011;2:12–21.

103. Morita Y, Kikumatsu F, Higuchi Y, Katakura Y, Takegawa K. Characterization and functional analysis of ERAD-related AAA+ ATPase Cdc48 in Aspergillus oryzae. Fungal Biol. 2020;124:801–13.

104. Calderone R, Li D, Traven A. System-level impact of mitochondria on fungal virulence: to metabolism and beyond. FEMS Yeast Res. 2015;15:fov027.

105. Dannemiller KC, Gent JF, Leaderer BP, Peccia J. Influence of housing characteristics on bacterial and fungal communities in homes of asthmatic children. Indoor Air. 2016;26:179–92.

106. Meadow JF, Altrichter AE, Kembel SW, Kline J, Mhuireach G, Moriyama M, et al. Indoor airborne bacterial communities are influenced by ventilation, occupancy, and outdoor air source. Indoor Air. 2014;24:41–8.

107. Lax S, Smith DP, Hampton-Marcell J, Owens SM, Handley KM, Scott NM, et al. Longitudinal analysis of microbial interaction between humans and the indoor environment. Science. 2014;345:1048–52.

108. Frankel M, Bekö G, Timm M, Gustavsen S, Hansen EW, Madsen AM. Seasonal variations of indoor microbial exposures and their relation to temperature, relative humidity, and air exchange rate. Appl Environ Microbiol. 2012;78:8289–97.

109. Marcet-Houben M, Gabaldón T. Evolutionary and functional patterns of shared gene neighbourhood in fungi. Nat Microbiol. 2019;4:2383–92.

110. Linder MB, Szilvay GR, Nakari-Setälä T, Penttilä ME. Hydrophobins: the protein-amphiphiles of filamentous fungi. FEMS Microbiol Rev. 2005;29:877–96.

111. He Z, Zhang P, Wu L, Rocha AM, Tu Q, Shi Z, et al. Microbial functional gene diversity predicts groundwater contamination and ecosystem functioning. MBio. 2018;9:e02435.https://doi.org/10.1128/mBio.02435-17

112. Ottesen EA, Young CR, Gifford SM, Eppley JM, Marin R 3rd, Schuster SC, et al. Ocean microbes. Multispecies diel transcriptional oscillations in open ocean heterotrophic bacterial assemblages. Science. 2014;345:207–12.

113. Burke C, Steinberg P, Rusch D, Kjelleberg S, Thomas T. Bacterial community assembly based on functional genes rather than species. Proc Natl Acad Sci U S A. 2011;108:14288–93.

114. Muyzer G, Stams AJM. The ecology and biotechnology of sulphate-reducing bacteria. Nat Rev Microbiol. 2008;6:441–54.

115. Ottesen EA, Young CR, Eppley JM, Ryan JP, Chavez FP, Scholin CA, et al. Pattern and synchrony of gene expression among sympatric marine microbial populations. Proc Natl Acad Sci U S A. 2013;110:E488–97.

116. Ma X, Zhang Q, Zheng M, Gao Y, Yuan T, Hale L, et al. Microbial functional traits are sensitive indicators of mild disturbance by lamb grazing. ISME J. 2019;13:1370–3.

117. Escalas A, Hale L, Voordeckers JW, Yang Y, Firestone MK, Alvarez-Cohen L, et al. Microbial functional diversity: from concepts to applications. Ecol Evol. 2019;9:12000–16.

118. Satala D, Bras G, Kozik A, Rapala-Kozik M, Karkowska-Kuleta J. More than just protein degradation: the regulatory roles and moonlighting functions of extracellular proteases produced by fungi pathogenic for humans. J Fungi (Basel). 2023;9:121.

119. Aimanianda V, Bayry J, Bozza S, Kniemeyer O, Perruccio K, Elluru SR, et al. Surface hydrophobin prevents immune recognition of airborne fungal spores. Nature. 2009;460:1117–21.

120. Zhang Z, Reponen T, Hershey GKK. Fungal exposure and asthma: IgE and non-IgE-mediated mechanisms. Curr Allergy Asthma Rep. 2016;16:86.

121. Agarwal R. Severe asthma with fungal sensitization. Curr Allergy Asthma Rep. 2011;11:403–13.

122. Stachler E, Bibby K. Metagenomic evaluation of the highly abundant human gut bacteriophage CrAssphage for source tracking of human fecal pollution. Environ Sci Technol Lett. 2014;1:405–9.>

123. Sabar MA, Honda R, Haramoto E. CrAssphage as an indicator of human-fecal contamination in water environment and virus reduction in wastewater treatment. Water Res. 2022;221:118827.

124. Kuske CR, Hesse CN, Challacombe JF, Cullen D, Herr JR, Mueller RC, et al. Prospects and challenges for fungal metatranscriptomics of complex communities. Fungal Ecol. 2015;4(14):133–7.

125. Kress A, Poch O, Lecompte O, Thompson JD. Real or fake? Measuring the impact of protein annotation errors on estimates of domain gain and loss events. Front Bioinform. 2023;3:1178926.

Publisher's Note