

Supplementary Results

Community compositions were not found to be associated with any indoor or material environmental covariates

In addition to the analyses described in *Results*, we employed many other tests to identify associations between the office microbiome composition and indoor and material environment covariates. We tested for correlation between microbiome characteristics and building science data based on single timepoints of building science data (e.g. the equilibrium relative humidity measurement at the time of sampling) but also extrapolated this to equilibrium relative humidity measurements from the hour immediately prior to when a sample was taken, as well as data from the week and the month prior. We followed a similar approach with temperature and illumination, though even with multiple different types of comparisons we did not observe any significant associations. To further investigate possible relationships, we looked specifically between offices which exhibited large differences in building measurements. One of the largest differences in building data was observed in San Diego during the summer period. During this period, one San Diego office saw a “time of wetness” (the amount of time a given surface had a equilibrium relative humidity measurement over a certain percent) over 60% of ~2500 hours while a second San Diego office saw a time of wetness less than 10 hours. If a single building measurement were associated with microbial community differences, we expected that this would be a likely candidate. Despite the large differences in equilibrium relative humidity, similar differences in community composition and richness did not exist (two-sample monte carlo t-test with 999 permutations, test-statistic = -1.419, p-value = 0.158). Additionally, of the taxa that were identified to be the most different across cities, the one most significantly correlated with

equilibrium relative humidity, the order *Sphingomonadales*, did not show significant differences between these two San Diego offices, suggesting that its correlation with equilibrium relative humidity may be due to chance. Further analysis including interactions between building parameters, building parameters and sample metadata, log functions, and higher order interactions evaluated with a Lasso regression also did not reveal any discernable patterns. There are several possible explanations for these lack of associations, but one likely important factor is that all of the offices in the study were quite consistent in important factors: they were generally occupied based on the traditional work week, they were all mechanically heated and cooled to similar temperature and relative humidity conditions, they had similar occupant densities, and none experienced extreme conditions, such as a moisture problem, over the course of the investigation.

Fungal results

In addition to the 16S rRNA, we sequenced the ITS-1 region of the fungal genome (a non-functional region between the 18S and 5.8S rRNA genes). Many of the patterns that we report for bacterial community composition and richness were also observed for fungal communities, but there was large variability in the quality and quantity of fungal sequence data. We suspect that this may be due to an inability to collect sufficient fungal biomass from surfaces using the dry swabbing technique we employed in this study, or issues with DNA extraction from fungi as has been recently observed.

Our fungal community taxonomic profiles are similar to those previously reported in the BE. For example, the most dominant taxa identified by [\(Fouquier et al. 2015\)](#), Ascomycota (phylum) and Basidiomycota (phylum), were also the highest abundance identified phyla in our study, representing 32.8% and 17.1% of our fungal community composition. Our most dominant classes, Eurotiomycetes, Dothideomycetes, and Agaricomycetes, respectively, were present in

high abundance by [\(Fouquier et al. 2015\)](#). They report that 59% of their OTUs could not be taxonomically assigned at the phylum level, while 45% (a similar proportion) of our fungal OTUs could be assigned taxonomy at the phylum level.

The mean and median Bray-Curtis distance between all fungal samples following open reference OTU were 0.991 and 0.999, respectively, suggesting little to no overlap in the composition of the communities on these surfaces. These distances are much higher than what has been previously reported (24, 25). Closed reference OTU picking results in lower mean (0.949) and median (0.998) distances, however these are still significantly higher distances than what has been observed in other studies. We present our observations from these data here, but are cautious to avoid over-interpretation of what may be spurious associations.

As in bacterial communities, fungal communities had lower richness in all offices in the summer sampling period than in all subsequent periods (two-sample monte carlo t-test with 999 permutations, p-values less than 0.05 for all comparisons, except for Flagstaff summer and fall, and Toronto summer and fall). This is likely due to the summer sampling period starting immediately after the sampling materials were sterilized. Floor samples had significantly higher richness in all cities, in all periods (two-sample monte carlo t-test with 999 permutations, p-values less than 0.05 for all comparisons, except for San Diego fall and, and Toronto spring; [Supplementary Figure 1](#)). When comparing the Flagstaff fall and winter samples (which were sequenced in the same sequencing run), the floor samples were significantly different in composition than the wall/ceiling samples (Permanova with Bray-Curtis pseudo-F=1.099, p-value = 0.015; Binary-Jaccard pseudo-F = 1.11431, p-value < 0.001) however sample composition did not differ with surface material (Permanova with Bray-Curtis pseudo-F = 1.01254, p-value = 0.308; Binary-Jaccard pseudo-F = 1.01402, p-value = 0.199).

Sampling frequency resulted in noticeable but not statistically significant differences in fungal community composition ([Supplementary Figure 3](#)) for paired samples from frequently and infrequently sampled rows (Permanova with Bray-Curtis pseudo-F=1.06308, p-value = 0.092; Binary-Jaccard pseudo-F=1.0128, p-value = 0.279). As with bacterial communities, this effect was small relative to the biologically relevant effects observed ([Supplementary Figure 3](#)).

We were able to distinguish cities using SVMs ([Supplementary Figure 4](#)), however the accuracy of the predictions was lower than those for the bacterial communities (F-1 score 0.65). The SVM classifier performed as well in predicting office (F-1 score 0.29) with the fungal data as with bacterial data. In the context of previous research (26),(26)t(26)i(26) suggests that higher quality ITS sequence data could provide a powerful vehicle for predicting geographic origin of samples.

While we did observe a run effect in fungal communities, this effect was not as strong as for the bacterial data ([Figure 5](#), [Supplementary Figure 5](#)). As with the other effects presented here, due to the large average distance between samples, it would be not be appropriate to extrapolate this finding to future studies.

Supplementary Methods

Fungal sequencing

ITS-1 (the non-coding region between the 18S and 5.8S rRNA genes) sequencing was used to profile fungal communities using the ITS1-F (CTTGGTCATTTAGAGGAAGTAA) and ITS2 (GCTGCGTTCTTCATCGATGC) primers. The fungal raw sequence data contained 39,173,141 reads for the three runs. After demultiplexing and quality filtering, 30,516,432 fungal reads remained.

All bioinformatics analysis was performed on the 5' reads as described in *Methods*, unless otherwise noted here. The sequences were assigned to operational taxonomic units (OTUs) using QIIME's uclust-based (35), open-reference OTU picking protocol (36) with the UNITE (38) ver6_dynamic_s_30.12.2014 reference set at 97% similarity. ghost-tree was tested for phylogenetic diversity calculations (41), however ghost-tree currently supports closed-reference OTUs only. Due to the low number of our fungal reads that hit the reference database, we chose to focus our ITS analyses on open-reference OTUs and non-phylogenetic diversity metrics. The fungal OTU table contained 2151 samples with a median of 1987 sequences per sample.

Beta diversity calculations were performed using QIIME's implementations Bray-Curtis (43) and Jaccard distance with exactly 1000 randomly selected sequences per sample (44). Samples with less than 1000 sequences were not included in the calculations. Community richness was computed using Observed OTUs.

Supplementary figure captions

Supplementary Figure 1: Experimental design. Analog of Figure 1, illustrating fungal community results.

Supplementary Figure 2: Fungal community dissimilarity. Analog of Figure 2, illustrating fungal community results.

Supplementary Figure 3: Disturbance due to repeat sampling, though detectable, is small compared to other variables. Analog of Figure 3, illustrating fungal community results.

Supplementary Figure 4: Confusion matrices illustrating the performance of city and office SVM classifiers. Analog of Figure 4, illustrating fungal community results.

Supplementary Figure 5: Investigation of sequencing run effect on observed bacterial community composition. Analog of Figure 5, illustrating fungal community results.

Supplementary table captions

Supplementary Table 1: Descriptions of offices included in this study.

Supplementary dataset captions

Supplementary Dataset 1: Differentially abundant OTUs across office locations, as determined by ANCOM.

Supplementary Dataset 2: Differentially abundant OTUs across frequently and infrequently sampled sites, as determined by ANCOM.

Supplementary Dataset 3: Differentially abundant OTUs across cities, as determined by ANCOM.