THE UNIVERSITY OF CHICAGO

OUR MICROBIAL INTERACTION WITH BUILT ENVIRONMENTS

A DISSERTATION SUBMITTED TO

THE FACULTY OF THE DIVISION OF THE BIOLOGICAL SCIENCES

AND THE PRITZKER SCHOOL OF MEDICINE

IN CANDIDACY FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

DEPARTMENT OF ECOLOGY AND EVOLUTION

BY

SIMON LAX

CHICAGO, ILLINOIS

DECEMBER 2017

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Over the past 150 years we have become an indoor species. For many of us, our natural ecosystem has been restricted to the built environment, especially in the developed world, where an average of 90% of our lives takes place indoors. Modern buildings are equipped with surfaces and environmental systems designed to reduce the potential for microbial life to flourish. This fundamental shift in our lifestyle is likely impacting the development and function of our immune systems in ways that we are only beginning to understand.

Humans are born mostly sterile (Aagaard *et. al.*, 2014), and the subsequent microbial colonization and succession within and on our bodies is influenced by our interaction with the world into which we are thrust. This interaction can take many shapes, including physical interaction with other humans, ingesting breast milk which contains a complex microbiota (Funkhouser & Bordenstein, 2013), interacting with the outdoor environment, and consuming food and water. All these interactions can act as sources that shape our microbial selves, such that the extraordinary number of permutations of potential microbial interactions with our environments leads to the development of a unique microbiota in every person. Even identical twins do not share any greater microbial similarity than normal siblings (Turnbaugh *et al.*, 2010), which suggests a degree of stochasticity of taxonomic succession that is independent of host genetics. However, this could be misinterpreted as suggesting that the host system does not select for favorable microbial features from the constant bombardment it encounters. Nothing could be further from the truth. The human immune system exhibits a complex and dynamic relationship with the microbiota, that results in microbial taxa with favorable qualities being recruited. The ability of intestinal regulatory T

cells to promote bacterial diversity by controlling the production of IgA is an ideal example of this process (Kawamoto *et al.*, 2014).

Because the diversity of environmental sources we interact with may influence the colonization and succession of our microbiota from birth to a stable adult state (David *et al.,* 2014), it is likely that if the diversity of sources we are exposed to were to decline, so too would the repertoire of our microbial taxonomy. Reducing the humidity in the air through air conditioning, and reducing the availability of complex substrates (e.g., soil) or porous surfaces (e.g., wood) through choices in building materials, are likely to significantly impact not only microbial biomass but also the diversity of the microbiota that can actively grow. This in turn is likely to significantly influence the source of our own microbiome, which could have untold consequences for our physiological, immunological, and neurological development.

When a person grows up within an urban environment they are exposed to a myriad of potential insults to their health, but the impact of being exposed to a reduced diversity of microorganisms has not been properly explored. The urban ecosystem is growing more rapidly than virtually any other environment on Earth, and the microbial ecology of the built and urban environment is a brave new frontier of microbiology. As this new environment emerges microbes will colonize and adapt to the myriad niches provided, and are therefore likely to develop novel ecologies, and potentially evolutionary trajectories, that could interact with humanity in many ways. It is clear that choices in architectural design and building materials can significantly affect the microbial communities we interact with on a daily basis (Kembel *et al.*, 2014, Ramos & Stephens, 2014). Harnessing the true potential of cities to provide services for their developing population density will require a fundamental understanding not only of the social and economic interactions we have with this environment but also of biological relationships.

Recent studies have focused on characterizing the basic ecological structure of the indoor and urban environment (Ramos & Stephens, 2014). From investigations of airplanes, kitchens,

offices, restrooms, and hospital wards, it is apparent that the human microbiome comprises the major microbial source for these indoor systems (Kembel *et al.*, 2014; Lax *et al.*, 2014). This is unsurprising because we are hypothesized to shed millions of microbial cells hourly into our local environment. The built environment usually has limited circulation and exchange with the outdoors, and such dispersal is highly likely to build up in our built spaces. Indeed, recent work on the microbiology of our homes over time suggests that we colonize the microbial niches of an indoor space with our own microbiota within 24 h (Lax *et al.*, 2014). Interestingly, in this same study, the microbial signature within each household was directly attributable to the microbial signature of its occupants. This suggests that, within a given property, the major microbial source you will be exposed to is your own. If an infant is growing up in such a space, its immune system will be trained by, and adapt to, this simplified and self-reinforcing microbial ecosystem. It is possible that this may result in a perpetuation of environmentally driven immunological and neurological conditions within families, such as asthma, allergies, and even depression. Recent work has demonstrated that the reintroduction of key bacteria can help to alleviate the symptoms of asthma and allergies in animal models (Fujimura *et al.*, 2014; Stefka *et al.*, 2014), which suggests hope for the development of clinical probiotics to alleviate these conditions. However, based on the interface we have with the built environment, it is also time we examined ways to manipulate this environment to create a greater diversity of sources that can help to increase our exposure to a wider phylogeny of microorganisms from birth and, in this way, contribute to health. The Hospital Microbiome Project (Lax *et al.*, 2017) was designed to determine how the microbial community of a hospital develops, and how interacts with the human population. By better understanding the successional ecology and the existence of stable ecological states in the hospital microbiome, we will be better placed to develop architectural and engineering solutions to create an environmental microbial ecology that may help to support healthy immune system interaction. This is especially important for hospitals because patients are often under extreme physical stress, as is their microbiota. Treatment with

antibiotics and surgery can lead to the development of a pathobiome with negative health implications for the patient (Zaborin *et al.*, 2014). If the patients were provided a course of designed probiotics and an environment that promoted the colonization of their bodies with an appropriate microbiota, it is possible that patients might experience enhanced recovery.

If we are to characterize the ecology of the unique microbial world we all live in, it is essential that we work with as diverse a range of individuals as possible, and harness the efforts of the growing community of citizen scientists. Observational studies from people and their built environments have played a major role in advancing our understanding of the breadth of microbial diversity across these environments. Citizen scientists have been instrumental in gathering evidence about the human and built environment, providing an exciting opportunity for scientific exploration and educational outreach. Working with the public to implement experimental designs that can address questions we have about built-environment ecology has been highly successful, enabling us to broaden the research framework in which we test hypotheses. Although such investigation modalities are difficult to control, and can result in a higher degree of variance being attributable to sampling error or sample-storage differences than would be expected during laboratory-based studies, the value of distributed data collection, and the potential to integrate research into the community, far outweigh these problems, which can mostly be solved through appropriate statistical design.

In this dissertation, I aim to answer fundamental questions about the microbial interactions we share with built environments. I will begin in chapter two by addressing some of the most basic questions in built environment microbial ecology: where do the microbial communities originate, how unique are the communities associated with different residences, and how stable are these communities over time? In chapters 3 and 4; I will address how our microbial interactions with objects and built surfaces can potentially be leveraged for trace evidence in forensic investigations. Finally, in chapters 5 and 6, I will turn my focus to hospitals, the built environment whose microbial

communities may have the most profound impact on human health and well being. Taken together, this work represents a great step forward in our understanding of the factors structuring microbial communities in built environments, and may provide insight into the design and usage of buildings in the future.

Chapter 2

Longitudinal analysis of microbial interaction between humans and the indoor environment

2.1 Abstract

The bacteria that colonize humans and our built environments have the potential to influence our health. Microbial communities associated with seven families and their homes over 6 weeks were assessed, including three families that moved their home. Microbial communities differed substantially among homes, and the home microbiome was largely sourced from humans. The microbiota in each home were identifiable by family. Network analysis identified humans as the primary bacterial vector, and a Bayesian method significantly matched individuals to their dwellings. After a house move, the microbial community in the new house rapidly converged on the microbial community of the occupants' former house, suggesting rapid colonization by the family's microbiota.

2.2 Introduction

The global trend toward urbanization has increasingly bound humanity, as a species, to the indoor environment (Höppe & Martinac, 1998; Custovic *et al.*, 1994). We spend much of our time in our homes but know little about how microbial transmission influences the home and its occupants. Each human maintains a specific microbial "fingerprint" (Dekio *et al.*, 2005; Gao *et al.*, 2007, Grice *et al.*, 2008; Grice *et al.*, 2009; Grice & Segre, 2011), which should transfer to a new indoor space with skin shedding, respiratory activity, and skin-surface contact (Tringe *et al.*, 2008), the latter of which can transfer millions of microbial cells per event (Dawson *et al.*, 2007). The microbial diversity of

the home likely affects immune defense (Fujimura *et al.*, 2014) and disease transmission (Boone *et al.*, 2007) among its residents, so that tracking how people microbially interact with the indoor environment may provide a "road map" to defining the health in our homes.

In the Home Microbiome Project (www.homemicrobiome.com), we microbially monitored seven ethnically diverse U.S. families and their homes over 6 weeks by sampling their skin- and home-surface bacterial communities. Eighteen participants were trained in the collection of 1625 microbial samples from body and home sites of interest over a 4- to 6-week period from 10 houses, three dogs, and one cat. For three families, samples were taken immediately before and after moving to a new home. Approximately 15 million high-quality 16*S* rRNA V4 amplicons represented 136,957 distinct operational taxonomic units (OTUs) (97% nucleotide identity). We subsampled this database at 2500 sequences per sample, omitting OTUs represented by <10 reads, which yielded 4 million sequences comprising 21,997 OTUs (97% identity) from 1586 samples.

2.3 **Results**

Samples from different sites within the same home differed less than samples from the same site in different homes [analysis of similarities (ANOSIM) $R = 0.210$, $P < 0.0001$ versus $R = 0.408$, $P < 0.0001$]. A density plot of weighted UniFrac distances between all home and human samples (**Figure 1A**) showed that microbial communities of human hands, noses, and bare feet resemble those of home surfaces. However, microbial communities found on home surfaces varied less than those found on humans. In each analyzed home surface, the microbial communities of different houses differed significantly ($P < 0.0001$) (**Figure 1B**), but the extent depended on the surface sampled and was highest for floor environments (ANOSIM $R = 0.757$ and 0.716 for kitchen and bedroom floors, respectively), whereas doorknobs were the most similar ($R = 0.379$ for front and 0.402 for bedroom doorknob). ANOSIM tests of the differences between the microbial community structure (weighted UniFrac) of the surfaces of each of the three pre- and postmove house

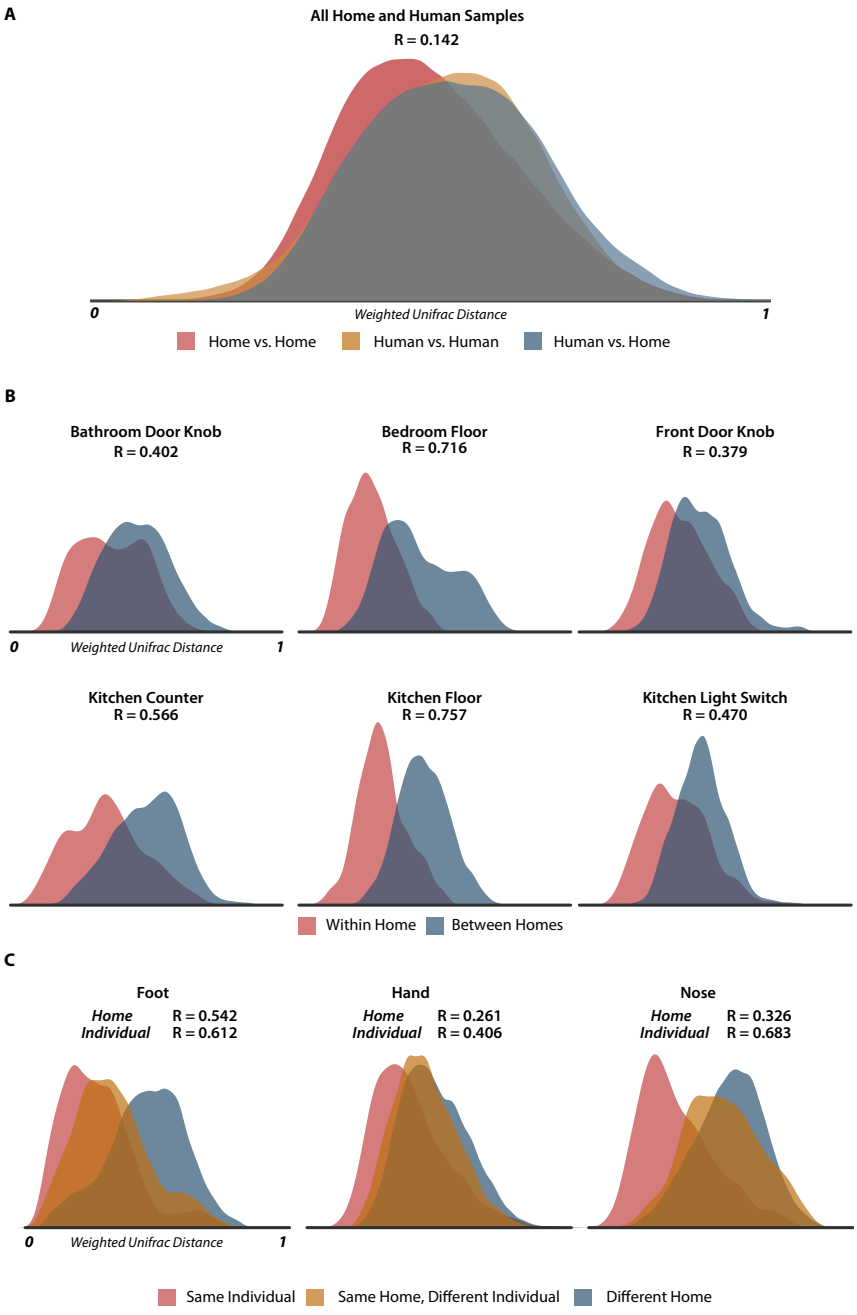

Figure 1: Differentiation in microbial community structure between homes and individuals. Density plots comparing the distributions of weighted UniFrac distances calculated within and between various criteria with accompanying ANOSIM tests of differentiation (all *P* values are less than 0.0001 based on 10,000 permutations of the randomized data set). (**A**) Distribution of distances calculated between two human samples, between two home samples, and between a human sample and a home sample. An ANOSIM test on the effect of source produced a low *R* value of 0.142, suggesting that home and human surfaces share a large degree of their microbial communities. (**B**) Distribution of distances for within-home and between-home comparisons of all samples taken from individual home surfaces. (**C**) Distribution of distances between human samples for the three sampled surfaces. Comparisons are segregated by whether a sample was compared with another from the same person, a sample taken from an occupant of the same house, or a sample taken from a resident of a different home. ANOSIM results are for tests on the effect of the home the sample was taken from (top) and of the individual the sample was taken from (bottom).

combinations (homes 5, 6, and 7) were insignificant, suggesting rapid colonization of the new home by the microbial signature of the family. Strikingly, one of the premove homes was a hotel room.

People sharing a home were more microbially similar than those not sharing a home, with samples taken from the same individual having the greatest similarity (**Figure 1C**). Of the three human environments analyzed in this study, foot samples were differentiated most by home (R = 0.542) and least by hand samples (R = 0.261). Hand samples were also least differentiated by individual (R = 0.406), and nose samples differed most between individuals (R = 0.683). ANOSIM statistics were robust to sequencing depth, with rarefaction to even 100 reads per sample having little effect on the observed strength of differentiation.

A third of all abundant OTUs (564 OTUs with >500 reads) had relative abundances that did not significantly differ between human and inanimate environments [nonparametric t test with false discovery rate (FDR) correction >0.05; ϱ = 0.88] (**Figure 2A**). Abundant OTUs were less likely to differ significantly in abundance between human and home surfaces when homes were analyzed individually and rarer OTUs (100 reads/OTU) were included (average = 60% undifferentiated OTUs). Although relative OTU abundances always correlated tightly between human and home environments, they varied in their correlations with pets.

Only about one third of OTUs were detected in all three sources, yet these 7200 OTUs composed between 93.6 and 97.8% of sequences in each source (**Figure 2B**). OTUs detected exclusively in a single source, although numerous (4137 OTUs), composed <0.6% of sequences in each sample. Relative abundances of dominant bacterial phyla differed among sources (**Figure 2C**). Firmicutes and Actinobacteria were enriched in human samples relative to the home, Proteobacteria dominated home and pet samples, whereas Bacteroidetes were abundant in pets. However, the relative abundances of the nine most abundant bacterial classes had no significant relationship with the number of sources that shared them (ANOSIM $P > 0.05$) (**Figure 2D**). Pairwise comparison of

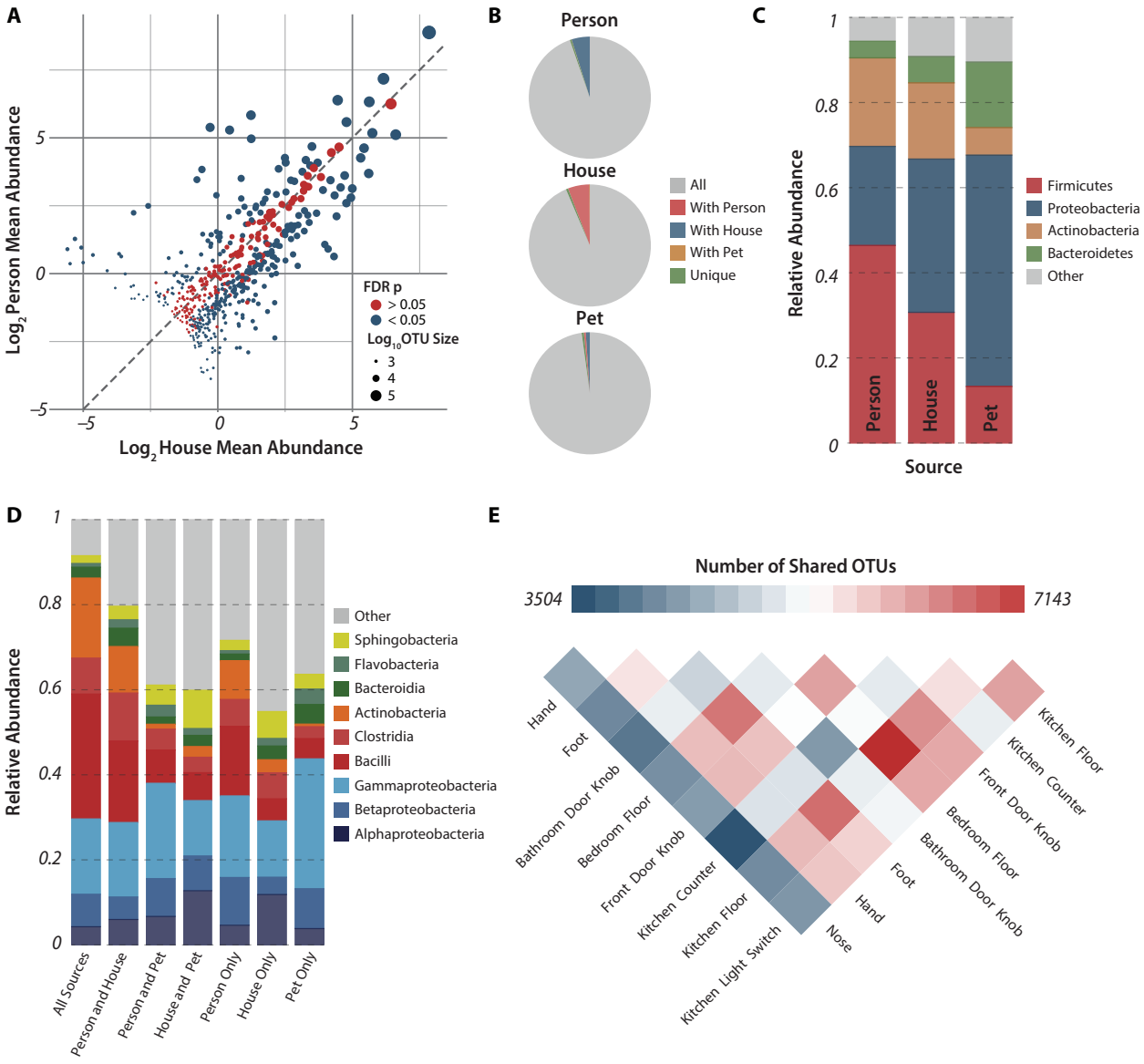

Figure 2: Widespread sharing of microbial taxa between human and home surfaces. (**A**) Plot of log2-transformed average relative abundances in the human and home environments for all OTUs in the study with greater than 500 reads. OTUs are colored by whether their average relative abundance is significantly different between the home and person environments based on the FDR-corrected *P* value from a nonparametric *t* test run with 1000 permutations and are sized according to their log10-transformed number of reads. The dashed line is $y = x$, indicating an equal average relative abundance. (**B**) Fraction of all reads from within a source belonging to OTUs shared with other sources, demonstrating the ubiquitous sharing of OTUs between homes and the humans and pets that occupy them. The percent of reads that cluster within source-specific OTUs is less that 0.6% for all three sources. (**C**) Taxonomic summary of observed relative abundance of abundant phyla across all samples divided by source. (**D**) Taxonomic summary of observed relative abundance of taxa at class level for all reads in the study by source-specific OTU overlap. (**E**) Shared phylotypes heatmap for individual surfaces after consolidation of samples taken from the same surface type across temporal sampling series and homes. Pooled samples were rarified to an even depth of 277,500 reads.

OTU sharing between surfaces across all homes revealed the greatest phylotype overlap between the two floor environments, with the nose sharing the least OTUs with other surfaces (**Figure 2E**). The number of OTUs shared by the surfaces with the greatest overlap and by the surfaces with the least overlap differed only by a factor of two.

We tested whether microbial community profiles could identify the house or surface a sample originated from by using random forest classifiers (**Table 1**). Floor samples were highly diagnostic of the family associated with that sample (ratios of random error to model error of 53.62 and 40.17 for kitchen and bedroom floors, respectively), and even considering all home surface samples together, the family that a sample was taken from was easily predicted (error ratio of 19.91). Models trained to predict the surface type from which a sample was taken were comparatively unsuccessful (error ratio of 3.29), with less predictive accuracy than that of those trained to predict family origin using broader taxonomic groupings. Families 5, 6, and 7 showed no significant difference between pre- and post-move homes, with error ratios of <1.75 for each model ($P > 0.05$). The relative success of predicting family of origin, even when models are trained on broader taxonomic levels, suggests that even error-prone reads from degraded DNA might still be a strong signal of an individual family's microbiota. Rarefaction to lower sequencing depth resulted in a steep decline in the models' ability to classify the home a sample was taken from, suggesting that greater sequencing depth than used in this study might substantially strengthen the models' predictive ability.

We matched human-associated microbial communities to home surfaces using a Bayesian technique known as SourceTracker (**Figure 3**) (Knights *et al.*, 2011). Hand samples were pooled by family and considered "vectors" to the bathroom doorknob, front doorknob, and kitchen light switch "recipient" communities. Bare-foot samples were pooled by family and treated as vectors for the bedroom and kitchen floor communities. On average, 76.7% of models successfully attributed

Sample Subset	Predicted Category	Taxonomic Level	N	Estimated Error ± SD	Baseline Error	Ratio
All Home Samples	Family	OTU	673	0.0386 ± 0.0245	0.7682	**19.91**
All Home Samples	Surface	OTU	673	0.2529 ± 0.0243	0.8321	**3.29**
Bathroom Door Knobs	Family	OTU	111	0.0471 ± 0.0738	0.7658	**16.24**
Bedroom Floors	Family	OTU	113	0.0192 ± 0.0427	0.7699	**40.17**
Front Door Knobs	Family	OTU	113	0.0562 ± 0.0666	0.7699	**13.71**
Kitchen Counters	Family	OTU	112	0.0602 ± 0.0751	0.7679	**12.76**
Kitchen Floors	Family	OTU	113	0.0144 ± 0.0304	0.7699	**53.62**
Kitchen Light Switches	Family	OTU	111	0.0593 ± 0.0834	0.7658	**12.92**
Family 1 Home Samples	Surface	OTU	84	0.1083 ± 0.1364	0.8333	**7.69**
Family 2 Home Samples	Surface	OTU	72	0.0833 ± 0.1178	0.8333	**10.00**
Family 3 Home Samples	Surface	OTU	82	0.2650 ± 0.1208	0.8293	**3.13**
Family 4 Home Samples	Surface	OTU	70	0.1367 ± 0.1543	0.8286	**6.06**
Family 5 Home Samples	Surface	OTU	156	0.1583 ± 0.0908	0.8333	**5.26**
Family 6 Home Samples	Surface	OTU	95	0.1083 ± 0.1115	0.8316	**7.68**
Family 7 Home Samples	Surface	OTU	114	0.2833 ± 0.1427	0.8333	**2.94**
All Home Samples	Family	Phylum	673	0.2795 ± 0.0414	0.7682	**2.75**
All Home Samples	Family	Class	673	0.2277 ± 0.0795	0.7682	**3.37**
All Home Samples	Family	Order	673	0.1389 ± 0.0475	0.7682	**5.53**
All Home Samples	Family	Family	673	0.1014 ± 0.0405	0.7682	**7.58**
All Home Samples	Family	Genus	673	0.0532 ± 0.0213	0.7682	**14.43**
Family 5 Home Samples	Pre/Post Move	OTU	156	0.0760 ± 0.0220	0.0769	**1.01**
Family 6 Home Samples	Pre/Post Move	OTU	95	0.1241 ± 0.0306	0.1263	**1.02**
Family 7 Home Samples	Pre/Post Move	OTU	114	0.1923 ± 0.0962	0.3158	**1.64**

Table 1: Summary of predictive accuracy of random forest supervised learning models. Tenfold cross validation models were constructed with 1,000 trees using OTUs from evenly rarified samples as predictors of sample origin.

the recipient community to the correct vector (68.6% of hand samples were identified as vectors to the correct home's kitchen light switch; 82.9% of pooled family foot samples were identified as vectors to the correct home's bedroom floor). We also estimated the contribution of individual occupants to their home's surface communities, which appears to be highly variable between surfaces, between homes, and over time. The effect of an individual leaving his or her residence for three sampling days, as occurred in homes 1 and 4, resulted in a decline in that individual's predicted contribution to a number of the home surfaces, which varied between homes, during their absence (**Figure 4**). This suggests that the human microbiome signature on home surfaces (such as the bathroom, front doorknob, and kitchen counter) decays or is replaced rapidly. Because different surfaces respond differently to a human leaving, careful sampling of each surface could provide a metric for assessing the time course of events related to that house and those persons.

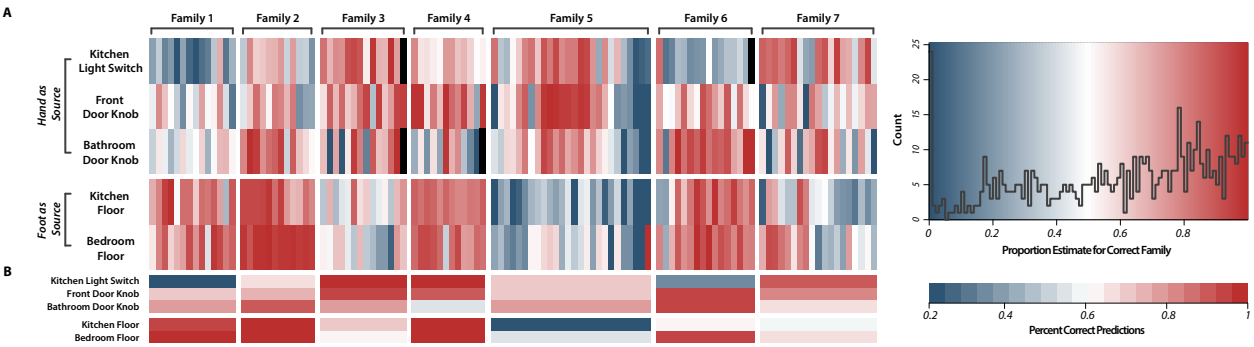

Figure 3: Summary of predictive accuracy of SourceTracker models. (**A**) Percent composition estimate for the correct source for each home surface sample in the study. Samples within each block are ordered by collection date, and black boxes occur where a sample is missing because it did not pass quality filtering standards. Across all surfaces, the models averaged a 59% prediction for the correct source. (**B**) Heatmap of model success across individual surface time series. The model was considered to be successful when the proportion of the sink community attributed to the correct source was greater than that attributed to any other source.

We tested the direction of microbial transfer among surfaces in the four homes where the subjects did not move houses using dynamic Bayesian networks (**Figure 5**). Humans were more likely to be sources of OTUs than were physical surfaces, with an average of 26 taxonomic edges leaving a human skin surface and arriving at a physical surface, versus eight edges in the opposite direction ($P < 0.001$). In contrast, human and home surfaces were equally likely to be recipients (human = 20.6 taxonomic edges; home = 19.3, nonsignificant). OTUs sourced from humans were mainly Actinobacteria and Proteobacteria, which are major components of the human skin microbial community (Segre *et al.*, 2009).

To assess whether personal relationships affect sharing of microbial taxa, we focused on home 4, where none of the residents were genetically related. The two occupants who were in a relationship shared more of their microbiota with each other than with the third occupant, who resided in a separate part of the house (**Figure 6**). This differentiation was observed across all surfaces, being greatest in nose samples ($R = 0.690$) and smallest in hand samples ($R = 0.300$) (all $P < 0.0001$). In contrast, only weak or insignificant differentiation was observed between married couples and their young children.

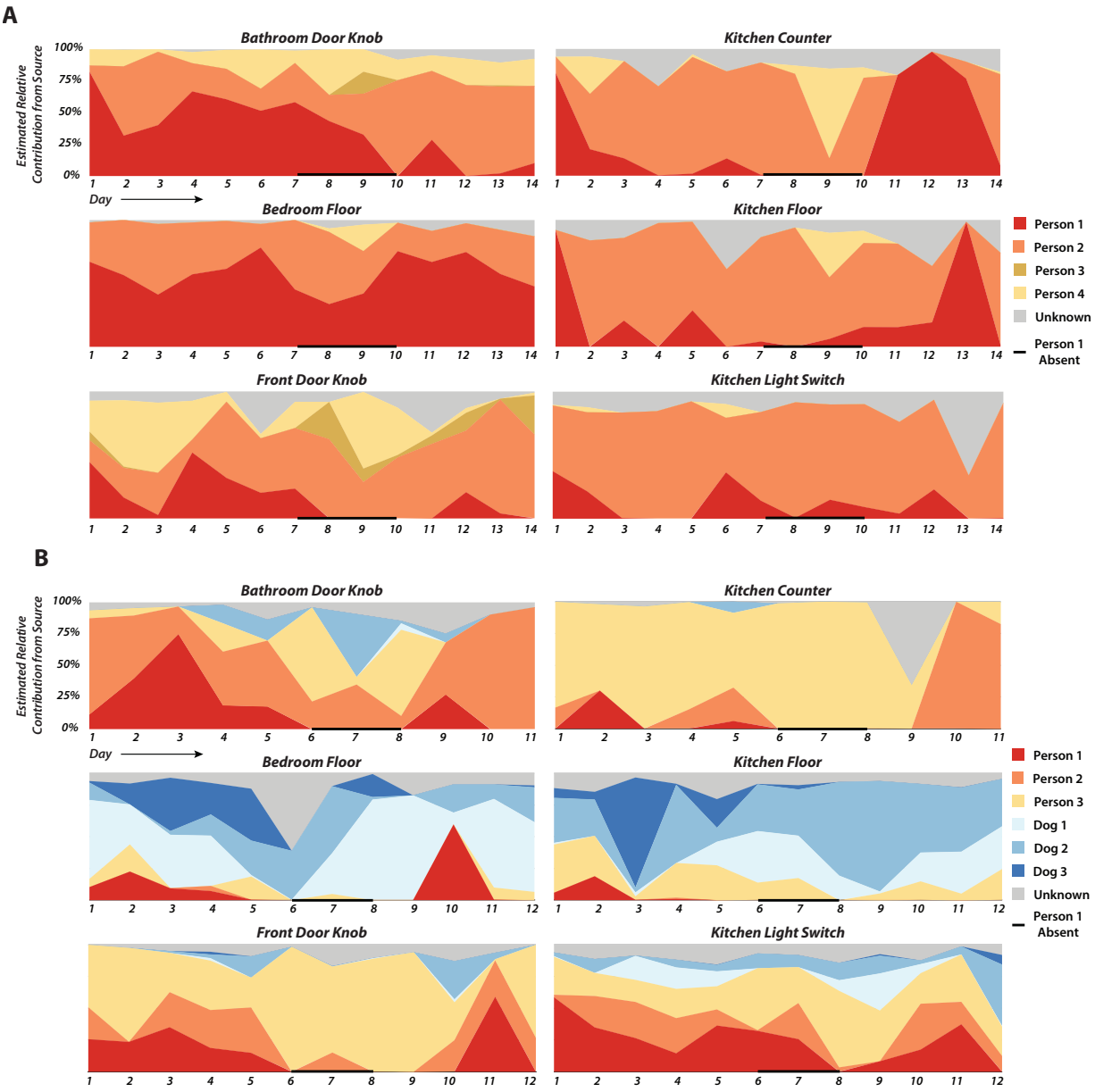

Figure 4: Time series of SourceTracker source contribution estimates for the 6 home surfaces in houses 1 (A) and 4 (B). All samples taken from the occupants of each home were consolidated and treated as sources impacting the home surface sinks. Data is visualized as a stacked area chart of the proportion of each sink sample estimated to originate from each source, with samples ordered by sampling date along the x-axis. In both houses, Person 1 was traveling and did not interact with the house for the time indicated by the black bar below each time series.

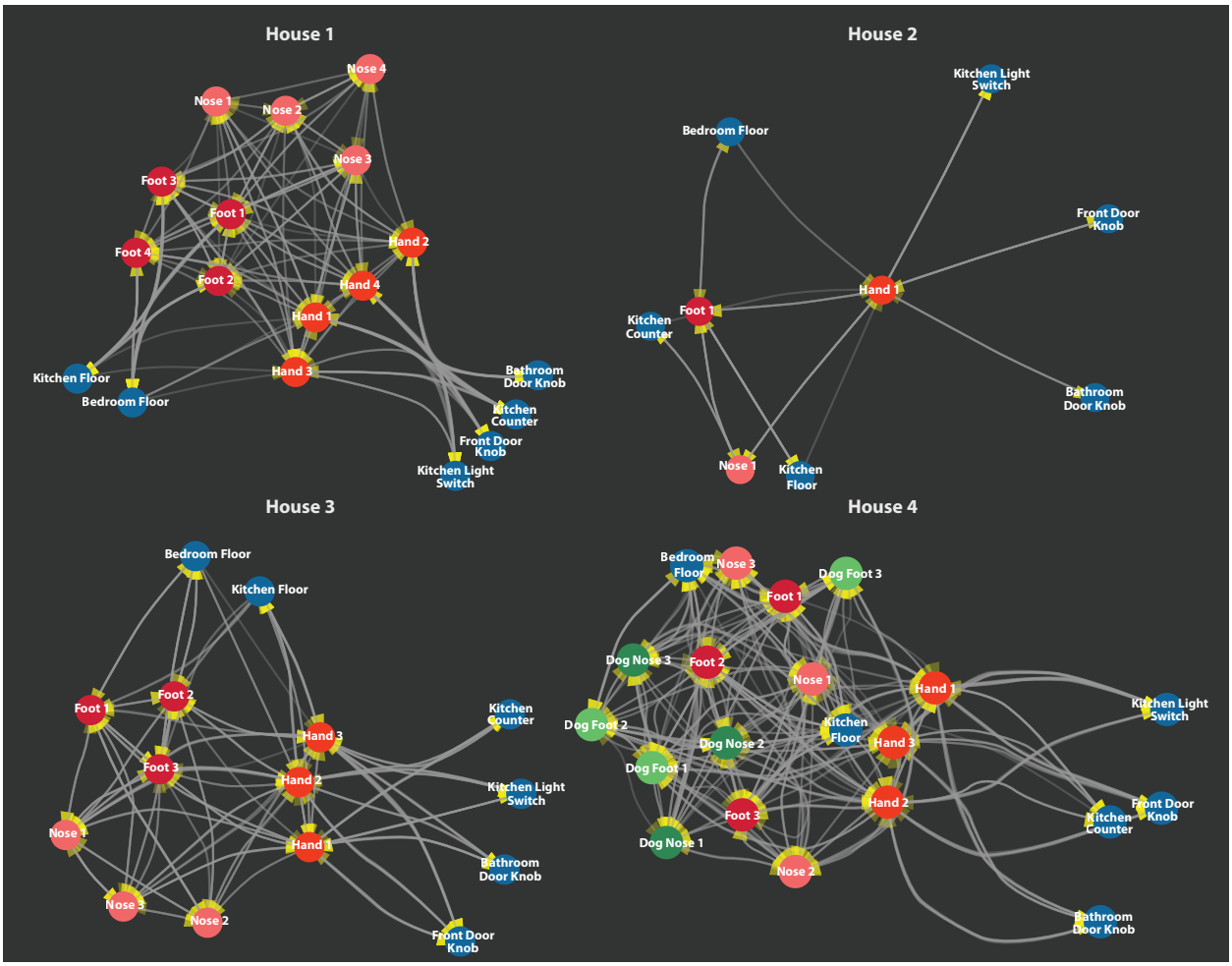

Figure 5: Dynamic Bayesian networks depicting interactions between surfaces in homes 1 though 4. Population data was considered at the phylum level and statistically significant Bayesian co-dependent relationships were defined between surfaces (nodes) in each family and home by phyla level taxonomic associations. Nodes are distributed using a spring-embedded layout that clusters highly-linked surfaces together. The direction of transfer is indicated by yellow arrowheads pointing from source to sink. Edges connecting the same source and sink are binned together and appear darker the more edges they encompass.

Overall, there were significant differences in the volatility of microbial communities associated with each surface type (Kruskal-Wallis $\chi 2 = 21.6$, $P = 0.0057$) (**Figure 7**). However, after a pairwise Wilcoxon test and FDR correction, the only significant differences were between hand and bedroom floor, hand and foot, and hand and nose. We can consequently conclude that the hand is especially variable over time relative to other body habitats and surfaces, presumably reflecting

high inputs from the various surfaces with which it comes in contact and/or more frequent disruption due to washing.

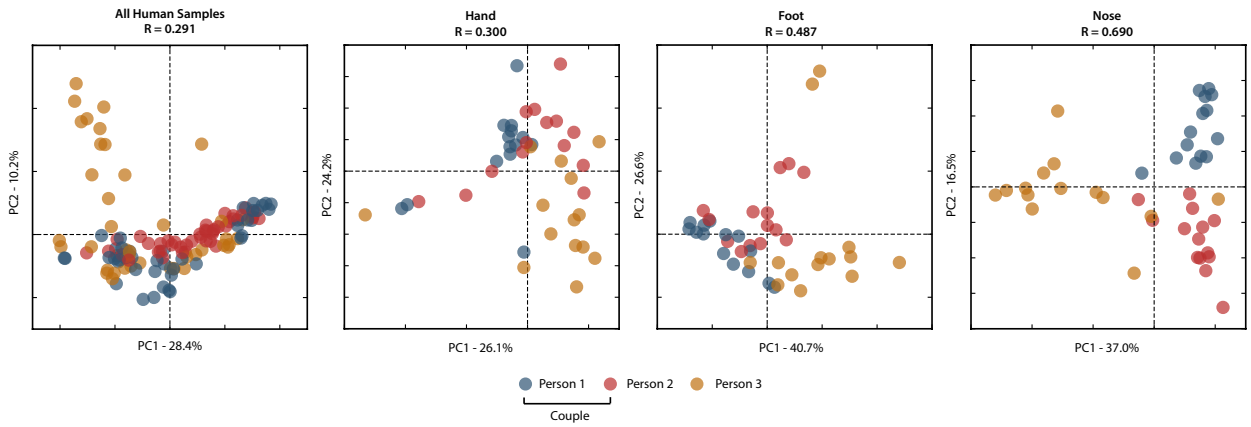

Figure 6: PCoA plots of human samples taken from home 3 based on weighted Unifrac distance. The left plot includes all human samples while the right three are subdivided by surface. R values are ANOSIM tests for significant differentiation between the two residents in a couple (individuals 1 and 2) and a third roommate (person 3). All p-values are less that 0.0001 based on 10,000 permutations of the randomized dataset.

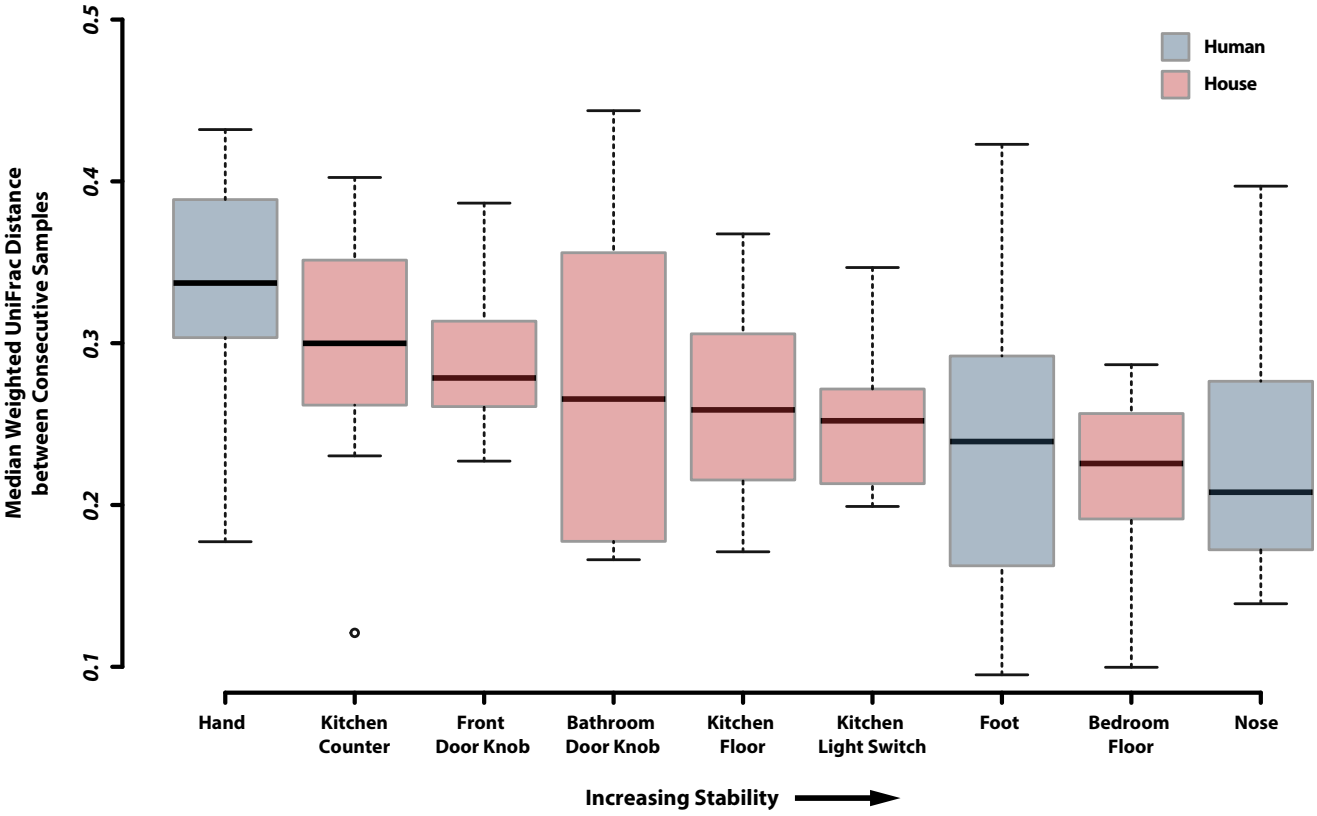

Figure 7: Temporal community volatility of individual surfaces, measured as the median weighted UniFrac distance between consecutive samples in a timeseries.

2.4 Conclusions

There is strikingly little research into relationships between microbial communities associated with home surfaces and their potential origins. Most studies explore fungal contamination of damp surfaces (Jaffal *et al.*, 1997; Hyvarinen *et al.*, 2002; Navalainen *et al.*, 2005; Lignell *et al.*, 2008), the role of hygiene in removing microbial communities (Grice & Segre, 2011; Bright *et al.*, 2010), and the length of time microbes can survive on surfaces (Kramer *et al.*, 2006; Moschandreas *et al.*, 2003). Here, we present an intensive longitudinal analysis of the microbial communities associated with the home environment and present evidence for substantial interaction among human, home, and pet microbiota. Such interactions could have considerable human and animal health implications. Further, we suggest that homes harbor a distinct microbial fingerprint that can be predicted by their occupants and that supersedes intersurface differentiation within the home. We further show the rapidity and extent to which a human population can influence the microbial diversity of the space they inhabit.

2.5 Methods

Sample collection

Seven households were recruited to this study, including three families who moved to a new house during the course of the study and four families who remained new or long term occupants of a single home for the duration of sampling. Six houses were located in Illinois, two houses were located in Washington State, and two houses were located in California. In total, 15 adults, 3 children, 3 dogs, and 1 cat contributed samples from their nasal cavity, hand, and foot heel to this investigation. These samples were collected by the participants by rubbing sterile swabs pre-moistened with 0.15M saline solution on each body site of interest every other day for six weeks. In addition, samples were collected in the same manner and frequency from the front doorknob, bathroom doorknob, bedroom floor, kitchen floor, kitchen counter, and kitchen light switch. At

each time point, participants recorded the day and time. Freezers were provided to each household to allow short-term storage of samples at -20°C pending shipment or transport by car to Argonne National Laboratory on dry ice. Temperature and relative humidity sensors were placed in each home's bedroom, bathroom, and kitchen, and measurements were recorded at an interval of five minutes throughout the six-week study period. All interaction with participants and data processing was conducted in accordance with University of Chicago IRB 12-0123.

Amplicon library preparation

All samples were processed using a modified version of the manufacturer's protocol of the Extract-N-Amp kit (Sigma-Aldrich). Swabbed tips were placed into 2ml 96-well Deep Well plates (Axygen). 200µl of Extract-N-Amp Extraction solution was added, vortexed for 5 seconds, and incubated at 90°C for 10 minutes. Samples were centrifuged a 2,500 x g for 1 minute. 200µl of Extract-N-Amp Dilution solution was added to each sample to obtain a 1:1 ratio of extraction to dilution solution. Genomic DNA was amplified using the Earth Microbiome Project barcoded primer set, adapted for Illumina HiSeq2000 and MiSeq by adding nine extra bases in the adapter region of the forward amplification primer that support paired-end sequencing. The V4 region of the 16S rRNA gene (515F-806R) was amplified with region-specific primers that included the Illumina flowcell adapter sequences. The reverse amplification primer also contained a twelve base barcode sequence that supports pooling of up to 2,167 different samples in each lane (Caporaso *et al.*, 2012). Each 20µl PCR reaction contains 5µl of MoBio PCR Water (Certified DNA-Free), 10µl of Extract-N-Amp Ready Mix, 1µl of Forward Primer (5uM concentration, 200pM final), 1µl Golay Barcode Tagged Reverse Primer (5µM concentration, 200pM final), and 4µl of template DNA. The conditions for PCR were as follows: 94°C for 3 minutes to denature the DNA, with 35 cycles at 94°C for 45s, 50°C for 60s, and 72°C for 90s; with a final extension of 10 minutes at 72°C to ensure complete amplification. PCR amplifications were completed in triplicate, and then pooled. Following pooling, amplicons were quantified using PicoGreen (Invitrogen) and a plate reader. Once

quantified, different volumes of each of the products were pooled into a single tube so that each amplicon was represented equally. This pool was then cleaned using the UltraClean® PCR Clean-Up Kit (MoBIO) and quantified using Qubit (Invitrogen). After quantification, the molarity of the pool was determined and diluted to 2nM, denatured, and then diluted to a final concentration of 4pM with a 30% PhiX spike for loading on the Illumina HiSeq2000 sequencer. Amplicons were then sequenced in a 151bp×12bp HiSeq2000 run using custom sequencing primers and procedures described in the supplementary methods of (Caporaso *et al.*, 2012).

Sequence processing and analysis

Unpaired 151 bp reads were screened for high quality using four criteria. Reads were required to have (1) an exact match to an expected multiplex barcode sequence, (2) zero ambiguous base calls, and (3) a minimum Phred score of 20 across the entire length of the read. In addition, (4) reads were discarded if they matched with 100% identity to a read from a negative control, that is, pre-moistened swabs returned unopened by the participants and processed alongside the other samples.

Downstream processing of sequence data utilized the QIIME v1.6.0-dev software suite (Caporaso *et al.*, 2010). Samples with more than 12,000 sequences were randomly subsampled to 12,000 sequences, then all 13,953,338 reads were clustered de novo using the QIIME script pick_otus.py at 97% identity. Clusters containing fewer than 10 sequences were discarded, reducing the number of clusters from 136,631 to 22,148. A representative sequence from each of these 22,148 clusters was phylogenetically classified using the QIIME script assign_taxonomy.py with the Greengenes 12.10 database. An OTU table constructed from these clusters was rarefied to a depth of 2,500 sequences. All sequence data was submitted to the EBI's database under accession number ERP005806.

Analysis of beta-diversity was performed by calculating the pairwise weighted UniFrac (Luzupone & Knight, 2005) distance between each pair of samples, and the resulting distance matrix

was used for all downstream statistical tests of sample similarity. The significance of sample groupings was assessed using ANOSIM (QIIME's compare_categories.py script) for categorical variables and Mantel tests (compare_distance_matricies.py) for continuous environmental metadata. In all cases, statistical significance was calculated by comparing the R statistic to a distribution generated by 10,000 permutations of the randomized dataset. Tests employing unweighted UniFrac distance produced uniformly higher ANOSIM results and similarly significant p-values, although weighted UniFrac distance is used throughout the study to minimize the effect the limited number of reads per sample may have on the detection of low abundance OTUs.

Random forest models

Random forest supervised learning models were used to determine the diagnostic power of microbial community profiles in predicting the house or surface a sample originated from. These models form "decision trees" using a subset of samples to identify patterns associated with a metadata category, and then test the accuracy of the tree on the remaining samples not used for training. Each model runs a number of independent trees, and reports the ratio of model error to random error as a metric for the predictive power of the category's microbial communities. A greater ratio of baseline to model error indicates a better ability to classify that grouping by microbial community alone. **Table 1** maintains values for Estimated Error (+/- SD); 1 minus the estimated error is the percentage of times the model is able to accurately predict the correct sample subset. The models were run using the supervised_learning.py command in QIIME, with 1,000 trees per model and 10-fold cross validation. All models were trained on OTU- level input tables except for the bottom five models of **Table 1**, which collapsed the OTU table to the taxonomic level indicated in the "Taxonomic Level" column.

SourceTracker models

To estimate the proportions of potential source microbial communities (e.g. hand, foot, etc) contributing to the composition of house surfaces, we used a Bayesian source tracking approach (Knights *et al.* 2011). Generally following QIIME tutorial guidelines (http://qiime.org/tutorials/source_tracking.html), we filtered rare taxa not present in less than 100 samples to get the table to a reasonable size for runtime. We used the hands from all the families as potential sources for sink surfaces that seemed likely to be touched by hands (Kitchen Light Switch, Front Doorknob, Bathroom Doorknob), and the feet from all the family members as potential sources for sink surfaces that seemed likely to be touched by feet (Kitchen Floor, Bedroom Floor). The percent confidence of sink and source matches for each timepoint is shown in **Figure 3A**. The percent correct family predictions of all feet and hands across all timepoints, for all sources is shown in **Figure 3B**. The predictions are strengthened by source tests showing the large majority of samples can be correctly identified as coming from the correct source.

Dynamic Bayesian networks

Dynamic Bayesian Network (DBN) Inference was used to generate interaction networks between surfaces in each home. Bayesian Networks (BN) are statistical models that identify the conditional dependencies of set of random variables. The structure of a BN is a directed acyclic graph (DAG). In a DBN, random variables are related across a time step, i.e. in this analysis, the bacterial population on a surface at time t can affect the population of another surface at time t+1.

Population data was prepared before use in DBN inference. Data was considered at the level of taxonomic Phylum and only taxa comprising the top 99.9% of population abundance in at least one home dataset were used in network determination. Taxa in the remaining <0.01% of population abundances were binned into the category 'Other'. There are 18 total taxa in the networks (Acidobacteria, Actinobacteria, AD3, Bacteroidetes, Crenarchaeota, Cyanobacteria, Firmicutes, Fusobacteria, Gemmatimonadetes, Planctomycetes, Proteobacteria, Spirochaetes, SR1, Tenericutes, Thermi, Verrucomicrobia, and WPS-2), although not all taxa are present in every home. Population

abundance was reported as percent of total population and log2 transformed. A separate DBN is generated for each taxa in every house (a total of 53 networks generated).

BANJO (http://www.cs.duke.edu/~amink/software/banjo/) was used to generate the networks. In BANJO, a discretization policy of 5 intervals, a maximum of 3 parents, a maximum and minimum Markov lag of 1 time step, random local move proposer, and greedy searcher were selected for network identification. In order to generate a network that also conforms to the topology of the homes and their occupants, the following restrictions were placed on the network. No house surfaces can directly interact with one another. All occupant surfaces can interact with one another. Hands interact with everything. The only surface feet, dog feet, and dog noses can interact with is the floor.

Separate networks were constructed for homes 1 through 4 and the topology of the resulting networks (e.g. the in- and out-degrees of surface nodes and the frequency with which directed edge for specific bacteria taxa indicate transfer of bacteria from occupant to home surface or home surface to occupant) were compared to identify potential underlying themes of bacterial transfer in the home microbiome.

Chapter 3

Forensic microbiology in built environments

3.1 Introduction

Humans are host to a vast array of microbial diversity on our skin, as well as in our urinary gastrointestinal and respiratory tracts. These microbes are readily dislodged to our surrounding environments, even as we pick up new microbes with each interaction we share with our environments or with other people, creating a constant microbial flux with our surroundings. In the built environment (e.g., homes, offices) where people in the developed world spend as much as 90% of their time (Custovic et al. 1994), the microbes we deposit account for the vast majority of observed microbial diversity, making it possible to link individuals to objects and spaces through microbial similarity (Lax et al. 2015). In this chapter, we review our current understanding of the factors shaping microbial diversity in built environments and discuss how microbial interaction between these environments and their occupants can be used for trace evidence in forensic science.

3.2 The human skin microbiome

We are born mostly sterile (Aagaard et al. 2014) and acquire our foundational microbiota from our mothers' vaginal canal or their skin, depending on delivery method (Dominguez-Bello et al. 2010). In early life, our microbiota is highly plastic as our limited neonatal microbial community increases in diversity, and may be radically altered by infection, exposure to antibiotics, exposure to healthcare environments or changes in diet. By 3-5 years old, our intestinal microbiota is mostly

stabilized (La Rosa et al. 2014), with major differences observed between modern urban settings and rural, less industrialized societies (Clemente et al. 2015; Tyakht et al. 2013; De Filippo et al. 2010).

Human skin is a complex ecosystem comprised of numerous folds, invaginations, and specialized niches that are colonized by a wide array of microorganisms including bacteria, fungi, viruses and mites (Grice & Segre 2011). We have numerous symbiotic relationships with these microorganisms, which may protect against invasion from pathogens or educate the billions of T cells found in our skin (Grice & Segre 2011).

The majority of our symbiotic bacteria, both on our skin and in inner mucosal surfaces, fall into four phyla: the Actinobacteria, Bacteroidetes, Firmicutes, and Proteobacteria. Although both skin and gut communities are of low diversity at the phylum level, they are characterized by very high diversity at the species level, which we are only beginning to delineate through culture-independent molecular approaches (Grice at al. 2009). Skin microbiome studies have found that the bacteria, which colonize a skin site are dependent on the physiology of that site, with different microbes adapted to moist, dry and sebaceous regions. *Propionibacterium* spp., for example, are especially abundant on sebaceous skin sites such as the forehead and back, while staphylococci and corynebacteria are most abundant on moist sites such as the axilla and foot sole. Dry areas such as the forearm and hand are the most diverse skin sites and are not usually dominated by a single genus (Grice at al. 2009).

As our primary interface with our surroundings, the skin ecosystem is subject to constant perturbations, both endogenous (such as selection driven by host genotype) and exogenous (showering, for example). Temporal variability of the skin microbiome varies by site, and molecular analyses have revealed that community membership and structure are most consistent in the ear, nose, and inguinal crease (Grice et al. 2009). There are numerous factors that influence an individual's skin microbiome, including host physiology (e.g., sex, age), immune responses (e.g., previous exposure to infective pathogens, inflammation), host genotype, host lifestyle (e.g.,

occupation and hygiene), and the climate in which the individual lives (reviewed in Grice & Segre 2011). Although skin microbial communities are volatile in their membership and structure, the interpersonal variation of symmetric skin sites that results from these myriad factors is generally greater than interpersonal variation over time (Gao et al. 2007; Costello et al. 2009; Bouslimani et al. 2015), making possible the forensic identification of individuals from their skin microbiota.

3.3 The microbiota of the built environment

Humans occupy an ecological niche unprecedented in evolutionary history: we have become an indoor species. In the developed world, we inhabit spaces that consist of surfaces and environmental systems that are specifically designed to reduce our exposure to microbial life. This biome of indoor spaces is expansive and constantly growing, with estimates of the extent of residential and commercial buildings ranging from 1.3% (Kitzes et al. 2007) to 6% (Hooke et al. 2012) of ice-free land area. As urbanization has accelerated over the last century, we have become increasingly bound to this ecosystem: from the hospital we are born in, to the house we grow up in, the microbial diversity of these environments plays a profound role in shaping our own microbiota and developing our immune system.

Recent studies have uncovered the extent to which humans influence the microbiota of the buildings they inhabit. From studies encompassing homes (Lax et al. 2014) classrooms (Meadow et al. 2014), residential kitchens (Flores et al. 2013), gyms (Wood et al. 2015), public restrooms (Flores et al. 2011; Gibbons et al. 2015), and hospitals (Hewitt et al. 2013), human skin and respiratory cavities are the primary source of microorganisms to built environments.

Human-home microbial dynamics

The strength of the microbial interaction between a home and its occupants was recently demonstrated by the Home Microbiome Project (Lax et al. 2014), which surveyed the microbial communities associated with six home surfaces and three human skin sites in seven homes over the

course of four to six weeks. These longitudinal surveys found very strong correlation between the abundance of microbial taxa on home surfaces and on the skin of home occupants, with an average of two-thirds of major operational taxonomic units (OTUs) (those with greater than 100 total observations) not significantly differentiated in abundance between the two surface types. Using a Bayesian source-tracking algorithm, the relative contributions of different occupants, including pets, to the microbial communities of home surfaces were modeled. When an occupant was traveling away from their home for multiple days, his/her microbial signatures either vanished or were substantially diminished, exemplifying how rapidly occupants can affect the microbiota of their surroundings. The study also used dynamic Bayesian networks to track microbial taxa as they moved between different surfaces, and found that transfer from human skin to a home surface was many times more common than transfer in the other direction, with transfer from one occupant's skin to another also shown to be widespread.

The strong association between human skin and the built environment microbiome allows for the forensic matching of spaces to their occupants. For example, samples taken from different surfaces in the same home have been shown to be significantly more similar than samples taken from the same surface in different homes, suggesting that occupants play a larger role in shaping home microbial communities than building materials or surface function (Lax et al. 2014). Models trained to predict which house a sample was taken from performed very well, correctly classifying which home a sample originated from 20 times better than expected by chance. When only floor samples were considered, the classifiers performed approximately 50 times better than expected by chance. By contrast, classifiers trained to determine which home surface a sample was taken from were only three times better than expected by chance. This is an important consideration for forensic studies, as trace evidence samples collected from different parts of a home environment could all be equally effective for identification of the people active in that space. However, this also belies potential differences in building architecture and how people interact with different surfaces.

Someone coming into a room for the first time may not interact with surfaces in the same way as someone who resides in that space. This difference in behavior could shape where effective trace evidence samples could be found or identified as a priority location for sample collection. Additionally, the design and layout of a space can affect how people interact with it, and as such these components need to be considered if trace evidence standards were to be developed. Source Tracker models have shown that the surface which is most easily matched to the skin of a home's occupants varies between houses (Lax et al. 2014). This difference must be considered when developing appropriate forensic tools.

Influence of pets

Pets can play an important role in altering the microbial communities of built environments. By serving as a vector that can bring in outdoor-associated microbiota, they can link the indoor and outdoor environment and increase the complexity of the indoor microbial community (Fujimura et al. 2014). Pets are likely to have the biggest effect on floor-associated microbial communities, where they have been predicted to be the source of most of the observed microbial taxa (Lax et al. 2014). Pets can also alter the microbiota on their owner's skin or even in their gut. For example, cohabiting adults share a greater degree of their microbiota with each other if they own a dog, although this appears to be driven largely by rare taxa (Song et al. 2013). The same study also found that humans were more microbially similar to their pet dogs than to dogs they did not own, suggesting that dogs can be a source of microbiota to human skin and hence will act as both a potential forensic marker, and a potential confounder of signals for trace evidence.

Influence of interpersonal relationships

Every contact between two people can transfer millions of microbial cells (Dawson et al. 2007), making it possible to elucidate personal interactions through microbial similarity, despite the possibility that this similarity is transient. It has been found that the microbiota of hands, feet, and noses are each significantly more similar between individuals inhabiting the same home than

between individuals living in different homes, with the effect especially pronounced in nasal microbial communities (Lax et al. 2014). Even within the same home, the microbial communities at those three sites were found to be significantly more similar between two individuals living as a physically interacting couple, than between those individuals and a roommate living in a different room (Lax et al. 2014).

Significant microbial transfer between individuals can occur even with only short-term contact. A study of roller derby players found that all players' skin microbiomes were more similar to one another after competition, suggesting a homogenization of skin-associated bacteria through contact sports (Meadow et al. 2013). Even personal effects such as shoes and cell phones can become more microbially similar when people temporarily inhabit the same space. A recent study in which attendees at three different scientific conferences sampled their shoes and personal cell phones revealed that samples taken from the same conference were significantly more similar than samples taken from different conferences (Lax et al. 2015).

3.4 Tools for the forensic classification of the built environment microbiome

Sampling and sequencing considerations

Culture-independent surveys have revolutionized the way in which researchers characterize microbial diversity. Rather than growing microbes in clonal cultures, these methods allow us to characterize the community holistically, providing a more comprehensive assessment compared to a small subset of taxa in isolation. Microbial communities are influenced not only by environmental factors but also by the myriad interactions between different taxa, making it imperative that we view diversity in an ecological context. We can now gain deep insights into community membership and function by directly sequencing a microbial sample, which in the case of human skin or the built environment, is often collected on a sterile, moistened cotton swab rubbed against the surface of interest.

There are a number of culture-independent molecular tools that can be used to characterize a microbial community, but they can generally be classified in one of two categories: amplicon sequencing and shotgun sequencing (Chapter 5). Amplicon sequencing relies on the PCR amplification of a small region of DNA, which in the case of microbial surveys is often the gene encoding the small subunit of the ribosomal RNA, known as 16S rRNA. Ribosomal genes are universal to all microbial life and are sufficiently conserved in some sequence regions to allow for the use of primers with the potential to catch a broad phylogenetic diversity; although recent work demonstrates that even 'state of the art' assessments are likely considerably biased against uncultured taxa with highly divergent ribosomal sequences (Brown et al. 2015; Luef et al. 2015). However, especially in human and built environment contexts, amplicon sequencing of the 16S rRNA gene with only a single primer set, enables the rapid and inexpensive characterization of the vast majority of the microbial communities. Although this method can scale to a very high resolution of community membership, inconsistencies can result from different DNA extraction methods or from primer bias that favors the amplification of certain taxa (Wesolowska-Andersen et al. 2014; Feinstein et al. 2009). Additionally, because gene content is so variable in bacteria, even between closely related phylotypes, 16S rRNA is a poor window into the functional capacity of the community. Finally, because 16S rRNA is so highly conserved, even distantly related phylotypes may have sufficiently similar sequences that they are indistinguishable in amplicon-based surveys. Although this can confound our ability to match microbial taxa to individuals in forensic studies, the ease and depth of ribosomal amplicon sequencing has made it an incredibly useful tool in forensic microbiology, and it is the method employed by many of the studies discussed in this chapter. 16S sequencing, in its benefits and shortcomings, is analogous to 18S amplicon sequencing in eukaryotes and internal transcribed spacer (ITS) sequencing in fungi.

An alternative method is to employ shotgun metagenomic analysis, which randomly sheers DNA extracted from a sample into small fragments that can be directly sequenced. These raw sequences encompass all genomic DNA extracted from the sample, so they can be used to assess the relative abundance of gene ontologies and infer the functional potential of a community. When the sample is sequenced deeply enough, it is often possible to assemble these short reads into full metagenomes that provide much greater insight into inter-sample differentiation than amplicon sequencing. As these methods will sequence DNA indiscriminately, human DNA will often account for a large percentage of reads in surveys of human skin or built environments with shed skin cells (Lax et al. 2014; Gibbons et al. 2015).

Machine learning and statistical classification

The last decade has seen an incredible rise in the throughput and capacity of DNA sequencing technologies, which has precipitated a vast decrease in the financial cost. Next generation sequencers like the Illumina HiSeq and MiSeq have made it possible to simultaneously sequence thousands of samples to a depth of many thousands of reads per sample. This has led to an unprecedented increase in the size and complexity of microbial datasets, which has required the development of new bioinformatics and statistical tools to enable appropriate interpretation.

Most studies store their data in an ecological data matrix, which has sample names on one axis and taxa names on the other, with the observation counts of each taxon in each sample populating the matrix, often measured as 16S amplicon read counts. In general, these matrices are incredibly sparse, with the vast majority of entries being zero because many taxa are found only in a small subset of samples. This is fundamentally different from other large biological datasets such as regulatory networks and microarray data, which are much denser.

Machine learning methods are particularly useful for recognizing patterns in these complex datasets (Knights et al. 2011a). Supervised classification is a form of machine learning that develops models based on a subset of the data that includes input features, such as person identity. The goal

of these classification methods is to train the models to determine taxa that are associated with a metadata criterion in a subset of the data, and to then test their predictive power on the data not used for training. For example, 16S gene sequences derived from human skin samples can be used to find which taxa discriminate best between the individuals within a study. These models attempt to optimize the expected prediction error of the model on future data, and provide importance scores to individual taxa that can be used to determine the probative value of individual species.

Numerous supervised learning models have been tested on microbial communities (Knights et al. 2011a), though random forest (RF) classifiers have emerged as a particularly well performing method. RFs are an ensemble learning method that operate by constructing a multitude of decision tress in training and then output the mode of the classes of the individual trees. The ratio of model error to the error expected from random classification is used as a metric for the predictive power of the microbial community in discriminating between sample categories. The decrease in decision tree accuracy when a taxon is not included in training is used as a metric for the feature importance (discriminatory power) of that taxon.

Sequence clustering

Determining which sorts of data optimize the success of machine learning algorithms remains an area of active research. Overcoming the high variability of microbial community composition between individuals is difficult, and it is important to collapse or cluster observed taxa or genes in order to overcome this complexity (Knights et al. 2011b). Canonically, amplicon sequences are clustered into operational taxonomic units (OTUs) based on a predetermined threshold of similarity (commonly 97% for species-level taxonomic resolution). The problem lies in making sure these clusters are not too specific or too broad so that they lose their discriminative signal. A study of six human-associated microbial communities found that optimal clustering thresholds ranged from 76% to 99% depending on the factor being classified, implying that predictive models are likely to benefit from a flexible approach to OTU clustering (Knights et al.

2011b). This study further suggested that information about nucleotide similarity and phylogenetic relationships between OTUs be supplied directly to machine learning classifiers in order to improve their accuracy. Novel sequence clustering methods, such as oligotyping (Eren et al. 2013) and minimum entropy decomposition (Eren et al. 2015) which use information theory approaches such as the Shannon entropy at different nucleotide positions to cluster sequence, may further enable us to fine tune how we process raw sequence data before analysis.

These tools provide us with the ability to refine bacterial communities associated with humans and human trace evidence down to the strain level. The gold standard for this analysis is the genotype, as bacterial organisms show strain level association with each person (Gilbert 2015). However, oligotyping can be used to identify strain level associations for 16S rRNA data that could be highly effective in creating a cost-effective and accurate mechanism to identify individuals with a sub-species level microbial fingerprint.

3.5 Applied forensic microbiology of the built environment

Tracking disease in the hospital environment

The ability to track disease-causing organisms in hospital environments represents a forensic epidemiology goal that, if realized, could significantly influence healthcare associated infection rates in hospital settings (Westwood et al. 2014). Using genomic sequencing approaches for cultured pathogenic strains, researchers have been able to forensically track microbial pathogens retrospectively to determine the risk factors in disease spread (Sebaihia et al. 2006; He et al. 2013). However, these retrospective studies do not facilitate rapid tracking of the spread of microbial infection. This is changing with the genotype detection and analysis in near real-time for cultures detected in disease outbreaks (Quick at al. 2015), but there is also the potential of being able to model the microbial ecosystem of a building in a predictive way. By monitoring the longitudinal dynamics of microbial communities in a hospital, and by tracking as many potentially influential

variables about the building and its occupants as possible, it is likely that a model could be constructed to determine how pathogens may survive and spread in that ecosystem. Using such a model it may be possible to test interventions *in silico*, to determine the most effective strategy for combating these infections. This is the current focus of studies such as the Hospital Microbiome Project (Lax *et al.*, 2017) that aims to build such a model by exploring the microbial communities that patients shed into hospital environments, and by routinely monitoring how these strains move around the hospital, through patients, equipment, air, water, and staff. In the future, such models could be highly effective in helping to design new hospital buildings, or in augmenting how these buildings are managed. New sequencing assays, such as propidium monoazide–quantitative polymerase chain reaction (PMA-qPCR) that allow for differentiation between live and dead bacteria in culture-independent sequencing, will further aid in the development of these models.

Tracking occupancy and activity in a built environment

Developing new forms of trace evidence is going to be extremely important for the next generation of forensic scientists. Genotype and amplicon sequencing technology could assist the investigator immediately when the case is at its most relevant. It is almost a scientific law, as well as an investigative law, that the most beneficial time to assess a case is soon after it occurs. The human skin and mucosal surface microbiome has a unique signature that represents a composite of all the physical and environmental interactions a human has throughout their life (Flores et al. 2014). This microbial signature is preserved when it is transferred to a new environment for a defined period of time (Lax et al. 2014). The unique character of the human skin microbiome can provide clues as to that person's lifestyle, diet, ethnicity, type of work, and even whom they regularly interact with (Lax et al. 2014; Gibbons et al. 2015). That this signature is left behind in a space provides unique forensic potential to provide trace evidence that can be used in investigations. Once the microbiome stabilizes around age 3-5 years-old, a unique microbial composition and structure has formed, that can fluctuate, but retains a core identity (Bäckhed et al. 2015). Human microbial signatures

associated with the indoor and outdoor built environment, demonstrate the integrity of the signature we leave behind, and even show hour by hour, the way in which that signature changes on different surfaces. Using this information it will likely be possible to track time since last occupancy for a space and even capture information about the lifestyles, geographic origin, and personal relationships of a previous or current occupant.

3.6 Conclusions

Forensic microbiology is a nascent field, but one with extraordinary potential to transform not only criminal investigations, but also personalized medicine, and disease prevention. Investigations of homes, hospitals, restrooms, and other urban environments are fundamentally altering our ability to track the microbial interactions we have with the environment we have built for ourselves.

Chapter 4

Forensic analysis of the microbiome of phones and shoes

4.1 Abstract

Microbial interaction between human-associated objects and the environments we inhabit may have forensic implications, and the extent to which microbes are shared between individuals inhabiting the same space may be relevant to human health and disease transmission. In this study, two participants sampled the front and back of their cell phones, four different locations on the soles of their shoes, and the floor beneath them every waking hour over a 2-day period. A further 89 participants took individual samples of their shoes and phones at three different scientific conferences. Samples taken from different surface types maintained significantly different microbial community structures. The impact of the floor microbial community on that of the shoe environments was strong and immediate, as evidenced by Procrustes analysis of shoe replicates and significant correlation between shoe and floor samples taken at the same time point. Supervised learning was highly effective at determining which participant had taken a given shoe or phone sample, and a Bayesian source-tracking method was able to determine which participant had taken each shoe sample based entirely on its similarity to the floor samples. Both shoe and phone samples taken by conference participants clustered into distinct groups based on location, though much more so when an unweighted distance metric was used, suggesting sharing of low-abundance microbial taxa between individuals inhabiting the same space. This work suggests that correlations between microbial community sources and sinks can allow for inference of the interactions between humans and their environment.

4.2 Introduction

In recent years, research into the microbial interactions between humans and their surroundings has revolutionized our understanding of the microbial ecology of the built environment (Kelley & Gilbert, 2013). The dynamic relationship between the bacteria associated with human skin and the microbiome of indoor surfaces and of objects we interact with has demonstrated the degree to which the human microbiome can shape the microbial ecology of our homes, offices, hospitals, and cities (Lax *et al.*, 2014; Gibbons *et al.*, 2015; Brooks *et al.*, 2014; Meadow *et al.*, 2014; Song *et al., 2013)*. Characterizing this microbial dynamic is critical for many purposes, such as determining the rate and progression of microbial colonization of human infants exposed to the indoor microbiome (Groer *et al.*, 2014, Lax *et al.*, 2015 A). We therefore believe it is essential to determine how the microbial ecology of the built environment establishes and fluctuates over time.

Research on the microbial exchange between human and built environments has illuminated the forensic potential of the microbiome. In some cases, human microbial signatures have been used to match individuals to objects they have interacted with, including computer keyboards (Fierer *et al.*, 2010). Work on the microbiome of multiple home surfaces has shown that the microbial signature of a family can be highly predictive of the microbiome of that family's home and that individuals within a home can be differentiated (Lax *et al.*, 2014). Indeed, recent work on the microbial assemblages associated with smart phones has shown that individuals leave their skin microbiome on the surface of their phones (Meadow *et al.*, 2014). The rate at which these microbial communities change after they are deposited on a surface is also potentially valuable for forensic applications. Recent work has shown that postmortem, the microbiome of animal hosts changes dramatically, but in a predictable manner (Metcalf *et al.*, 2013). This predictability enables us to use microbial assemblages to help explore not just where someone is right now but also where they may have been

recently (Blaser, 2010). To explore the potential to determine the microbial fingerprint of individuals on personal items, we performed a detailed biogeographic and longitudinal characterization of the microbial communities on personal mobile phones. Additionally, we examined whether the microbial communities associated with an individual's shoes were determined by the floor microbiome associated with where they were walking.

4.3 Results & Discussion

Identifying signatures on shoe and phone samples

To determine the extent to which the microbial communities of samples were driven by surface type (that is, shoe, phone, or floor) and study participant, we employed a combination of ordination and supervised learning analyses. We found that microbial community structure was determined both by surface type and participant (PERMANOVA on weighted UniFrac; Pseudo-F = 19.7 and 22.7, respectively; $P < 0.0001$). The relative influence of surface type and interacting individual on microbial community structure was demonstrated by the weighted (**Figure 8A, B**) and unweighted (**Figure 8C, D**) UniFrac distance between samples. In both cases, the first principle coordinate clearly demarcated sample surface while the second principal coordinate demarcated study participant. UPGMA hierarchical clustering of samples pooled by individual and surface type (**Figure 8E, F**) further suggested surface type as the dominant influence on microbial community structure, with phone and shoe samples forming distinct groups, which were in turn subdivided individually. In both ordination analyses, floor samples clustered tightly with their longitudinally associated shoe samples.

The diagnostic power of microbial community profiles for predicting which of the two study participants a shoe or phone sample had been taken from was determined using random forest supervised learning. Random forest models were highly successful at determining which of the two

participants' shoes a sample was taken from, correctly classifying samples more than 50 times as effectively as one would expect by chance (**Table 2**), which indicates consistent differentiation in the shoe microbial communities of these two different people, even accounting for temporal variability. This is likely due to the presence of a 'core microbiome' on the shoes of individual study participants, which we assessed by looking at the abundances over time of the 100 taxa with the highest feature importance scores in the model. The majority of those 100 operational taxonomic units (OTUs) were consistently detected on the shoes of one participant over the course of the time series, but not on those of the other participant.

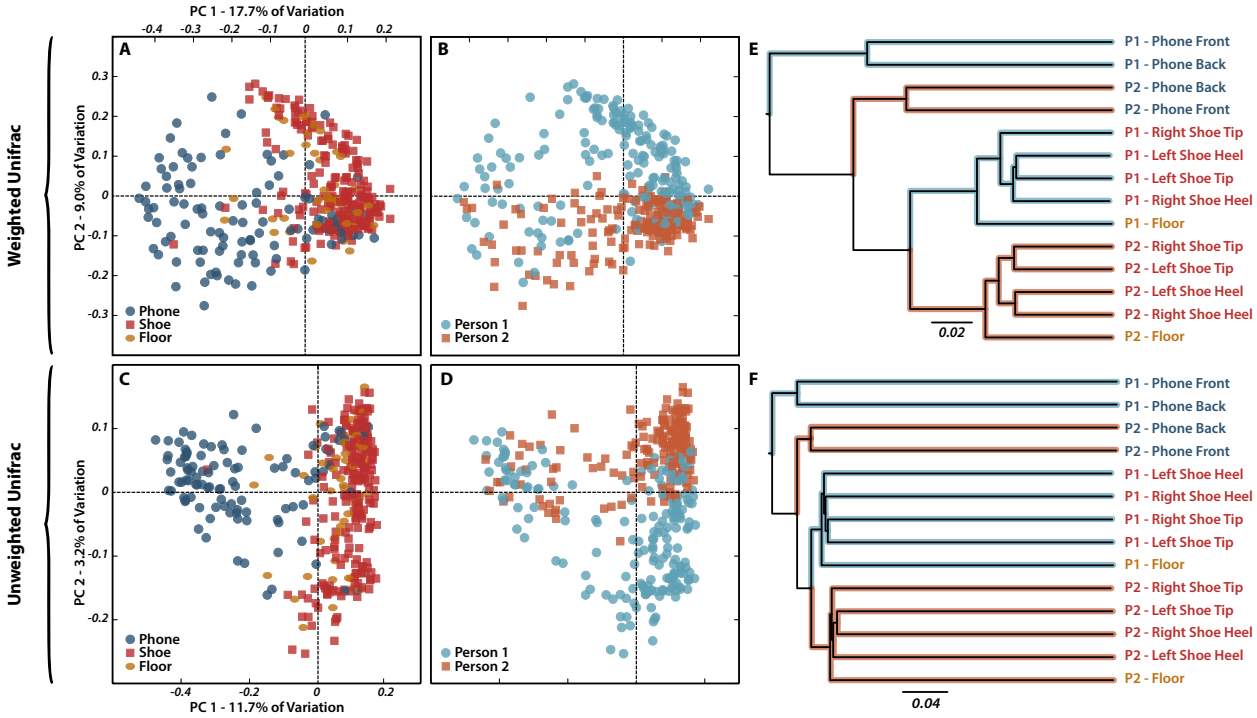

Figure 8: **Ordination of samples based on weighted and unweighted phylogenetic dissimilarity in community composition**. **(A, B)** depict principal coordinate (PCoA) plots for all samples in the study based on pairwise weighted UniFrac distance between samples, with sample points colored by surface and person, respectively. **(C, D)** are similarly colored by surface and person but are based on unweighted UniFrac distance. **(E, F)** depict UPGMA clustering of pooled and evenly rarified sample groupings based on weighted and unweighted UniFrac distance, respectively. Branches are highlighted to reflect person of origin (colors as in B and D) and group names at branch tips are colored by surface as in A and C.

In contrast to the high error ratio of models predicting study participant, the models did no better than expected by chance in determining which of the four shoe sites a sample had been taken from, even when models were segregated by study participant. We propose that this is due to the homogenization of communities across the shoe sole over time or to rapid changes in community structure at each sampling site. A similar pattern was observed in phone samples, with the models able to classify the participant a phone sample was taken from (error ratio of 13.6) but unable to determine whether the sample had been taken from the front or back of a given phone (**Table 2**).

Sample subset	Predicted category	*N*	Estimated error ± SD	Baseline error	Ratio
All phone samples	Person	104	0.037 ± 0.062	0.500	13.63
All shoe samples	Person	211	0.010 ± 0.020	0.479	50.26
P1 phone samples	Front/back	52	0.417 ± 0.206	0.481	1.15
P2 phone samples	Front/back	52	0.268 ± 0.180	0.481	1.79
P1 shoe samples	Shoe surface	110	0.705 ± 0.125	0.736	1.05
P2 shoe samples	Shoe surface	101	0.796 ± 0.090	0.732	0.92

Table 2: Summary of predictive accuracy of random forest supervised learning models. Tenfold cross validation models were constructed with 1,000 trees using OTUs from evenly rarified samples as predictors of sample origin. P1, person 1; P2, person 2, SD, standard deviation, N, number.

Random forest models were also used to assess which bacterial taxa were most associated with different surface types. Models were trained on a genus-level summary of the OTU table, and shoe and floor samples were merged into a single surface type based on their similarity in ordination analyses. When trained at the genus level, models were able to determine whether a sample was taken from a phone or a shoe/floor with an error ratio of 3.6. Skin-associated genera such as *Streptococcus*, *Propionibacterium,* and *Corynebacterium* were highly enriched in phone samples relative to shoe samples.

Longitudinal interaction between shoe and floor communities

To determine the extent to which the floor environments a shoe has interacted with influence the sole's microbial community and to assess whether individual shoe and floor time series

could be matched based on similarity, we employed a Bayesian source tracking approach (Knights *et al.,* 2011). These Bayesian models predicted a dominant influence from the correct source (**Figure 9**), which we believe shows the similarity between shoe and floor microbial community composition and may be used to infer where someone has recently walked. On average, the models predicted that a floor sample was the source of microbes for approximately three quarters of the microbial community associated with that shoe at that time point. Strikingly, floor samples had significant predictive power despite often being taken in areas the shoe did not directly touch (that is, proximate to where the participant had actually stepped), which suggests localized homogeneity of the floor microbial community. We also formulated individual SourceTracker models for each participant, in which the floor samples of individual locations were treated as sources to the shoe samples (**Figure 10**). These models demonstrated that bacterial taxa associated with the floor of a particular location often increased in abundance on the shoe soles of study participants while walking through that space.

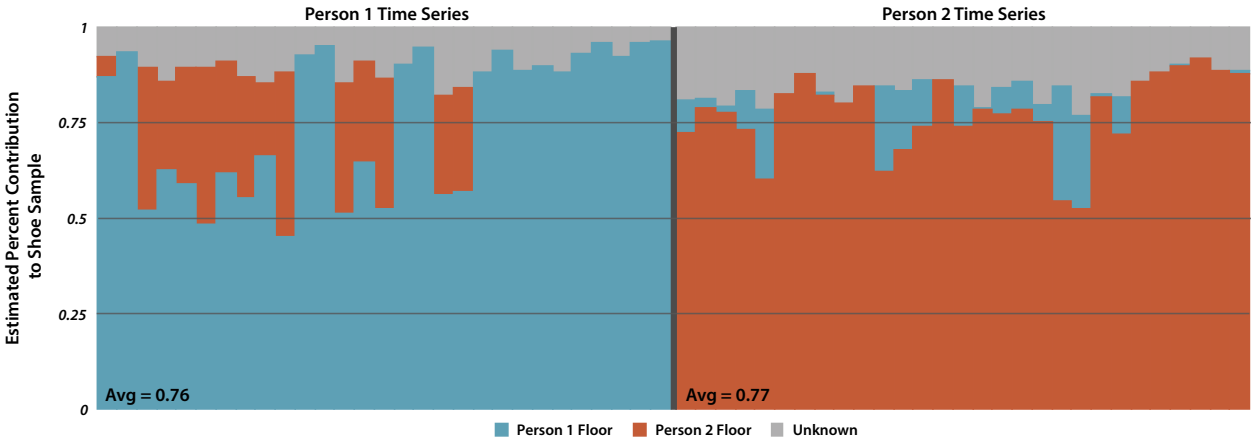

Figure 9: Summary of predictive accuracy of SourceTracker models in determining which of the two study participants a sample was taken from based only on the microbial communities of the floor samples those shoes had interacted with. For the models, all four shoe samples taken by each participant at a given time point were consolidated and treated as individual sinks (N = 29 and 27 for persons 1 and 2, respectively). All floor samples from the two participants' time series were collapsed and treated as the two possible sources to the shoe sink communities.

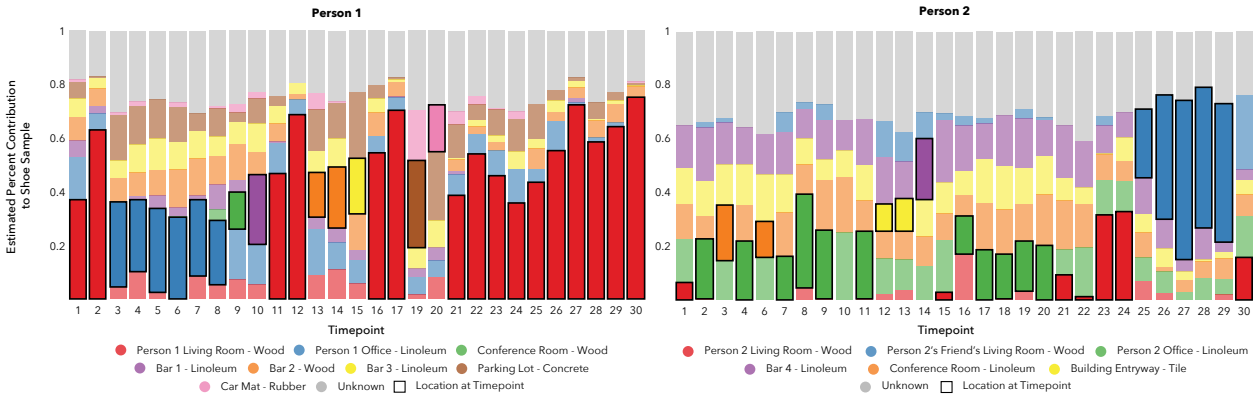

Figure 10: SourceTracker models for individual participants. All floor samples taken at each location were consolidated and treated as possible sources. All four shoe samples per time point were consolidated and treated as sinks. Bar height represents the mixing proportion estimate for each source in each sink sample, with the source environment where the participant was located at time of sampling indicated by a higher opacity and a black box. For person 2, time point 10, the participant was on lawn outdoors and the floor sample failed to produce enough reads to be included in the study.

To determine whether changes in the microbial community of the four shoe environments tended to be similar at each hourly sampling interval, we employed Procrustes analysis of the four sets of principal coordinates (**Figure 11**). All three pairwise comparisons for each study participant produced significant P values ($P < 0.005$), demonstrating that changes in the microbial communities of the four shoe environments resemble each other at each sampling interval, and thus suggesting a consistent impact from the floor microbial community. Procrustes analysis of the principal coordinates from the front and back of participants' phones did not produce significant P values, which we hypothesize is likely due to greater heterogeneity in community composition across the surface area of an individual phone at a given time point than would be observed across a shoe at a given time point due to lower overall biomass and high volatility in hand-associated microbial communities. It is also likely that microbes from the back of phones are likely to be sourced mostly from hands while the front may also be sourced from the face of the owner.

To assess the speed at which the floor environment influences the shoe sole microbial community, we looked at the relationships between shoe and floor samples taken from the same time point in principal coordinate (PC) space. For both study participants, PC1 values for the floor

and shoe samples at each time point were highly correlated for all four shoe environments (**Figure 12A**); we believe this is likely due to rapid contamination of the shoe sole by the floor microbial community. In all but one shoe environment (the right shoe heel of person 2), the correlation between shoe and floor PC1 values from the same time point was substantially higher than the correlation between samples taken one time step apart. In most cases, shoe microbial communities quickly converged on a PC space similar to that of the floor community (**Figure 12B**). These communities were largely segregated by the geographic location the sample was taken from and by the material of that location's floor (wood, linoleum, etc.), further supporting the possibility of rapid microbial transfer to the shoe sole.

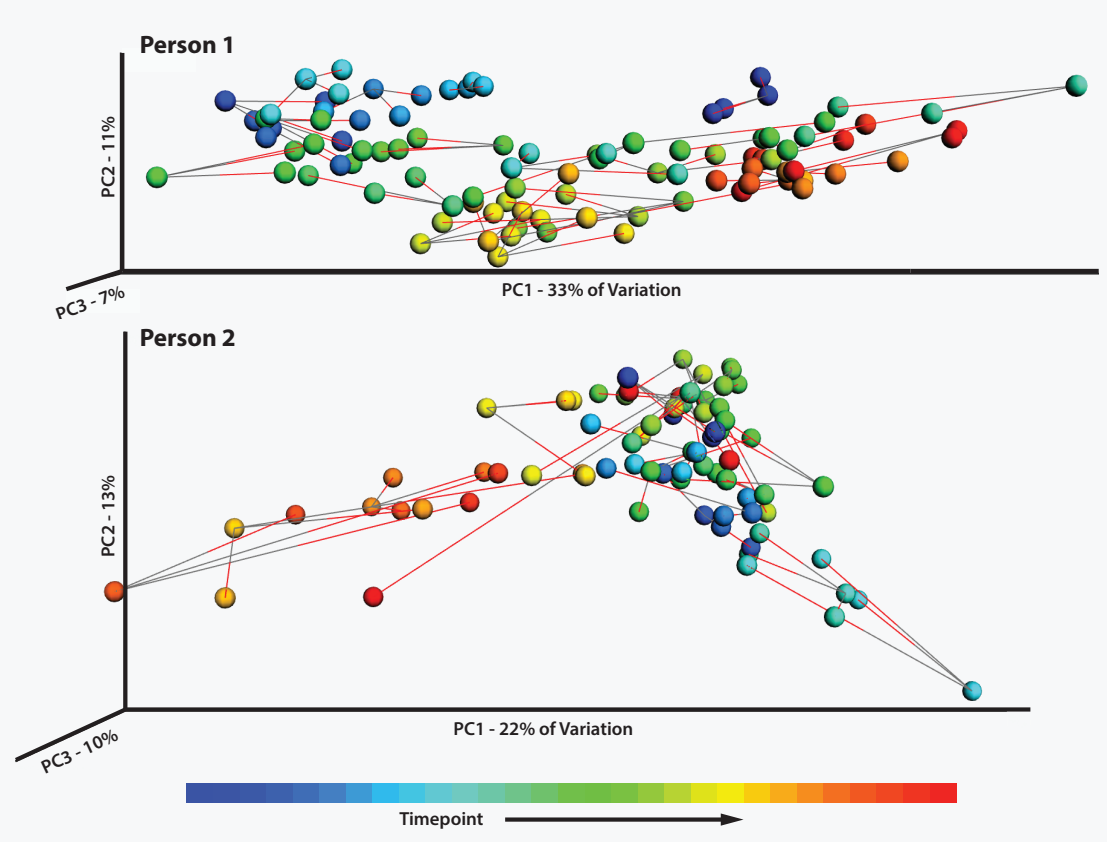

Figure 11: Procrustes analysis of shoe samples, demonstrating relatedness of community succession in the four shoe environments sampled in each person's time series. Samples in the PCoA plots are colored by the time point in which they were taken, and the four samples per time point (left heel, right heel, left tip, right tip) are connected by edges.

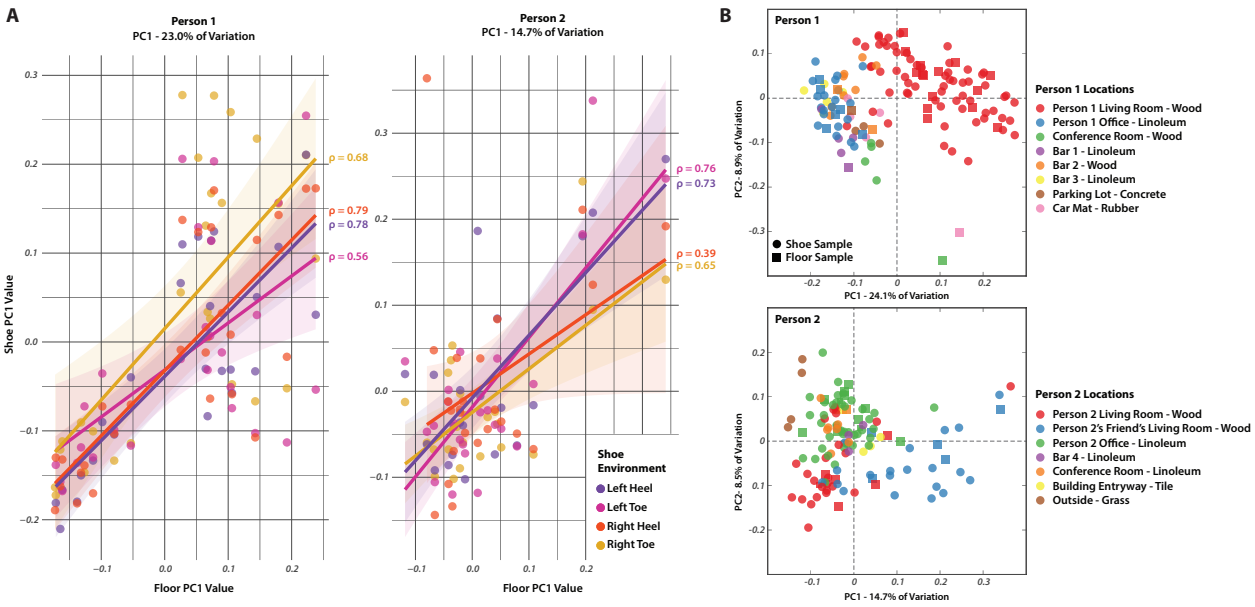

Figure 12: Immediate impact of floor microbial community on shoe microbial communities. (A) Correlation in the first principal coordinate values of shoe and floor samples taken at the same time point. **(B)** Principal coordinate plots of all shoe and floor samples, split by individual and colored by floor type and location at time of sampling.

Although our experimental design only allows us to assess the impact of the floor microbial community on that of the shoe sole, it is of course also true that shoes influence floor microbial communities by depositing microbes that have adhered to them. As participants walk, bacteria may adhere to shoes and be subsequently transferred back to the floor in a dynamic process of continual loading and unloading of microbes. A study of uptake and deposit of particles via indoor foot traffic showed that in many cases downplay of particles in the size range of bacteria from shoe to floor is greater than uptake by the shoe (Sippola *et al.*, 2014).

To assess the stability of microbial community structure across the 12 individual shoe and phone time series, we focused on weighted UniFrac distance between samples from consecutive time points and visualized community volatility as a density plot of those distances (**Figure 13**). Phone-associated microbial communities were observed to be both less stable (higher median distance) and more variable in their rate of change over time (broader distribution) than shoe-

associated communities. By contrast, little difference was observed between the four shoe environments or between the two phone environments. We hypothesize that the high volatility of phone-associated microbial communities is likely due to a small microbial biomass that would be prone to a rapid turnover in community composition and the very high volatility of hand-associated microbiota that has been observed in previous studies (Lax *et al.*, 2014).

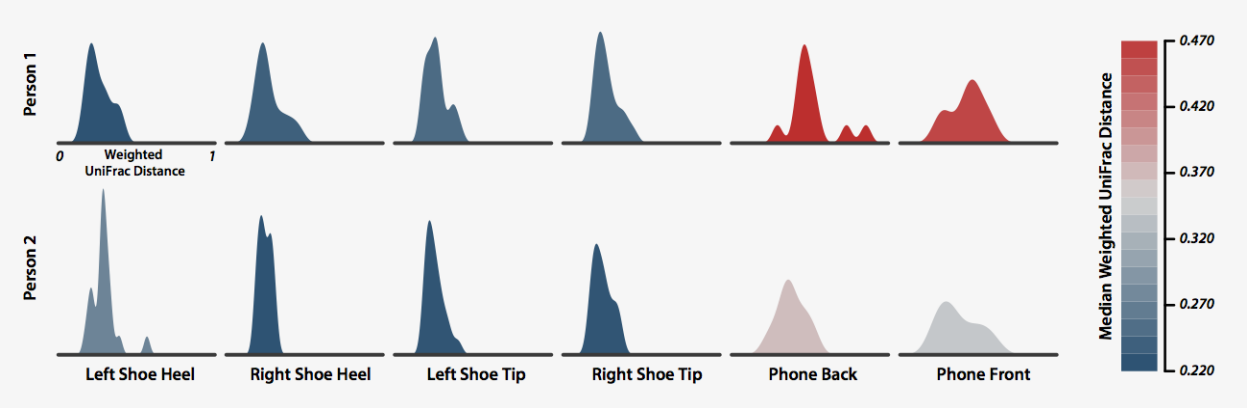

Figure 13: Volatility of individual surfaces across their time series, visualized as density plots of weighted UniFrac distances between samples from consecutive time points. Plots are colored by the median distance in those series of consecutive distances.

Biogeographic influence on community structure

In addition to the two time series participants, we also collected individual shoe and phone samples from volunteers at three academic conferences, one in Vancouver, BC (N = 29), one in Washington, D.C. (N = 26), and one in California (N = 34). California samples were taken from two different rooms at the same conference while Vancouver and Washington samples were all taken from the same room. We used these data both to corroborate the patterns of diversity observed in the time series with a larger number of participants and to assess the differentiation in community structure attributable to geographic segregation.

As in the time series analyses, phone and shoe microbial communities were significantly different (**Figure 14**; Pseudo-F = 38.2 for weighted UniFrac, $P < 0.0001$). The location at which samples were collected also played a significant role in shaping community similarity, especially in

shoe samples (Pseudo-F = 8.8, weighted UniFrac, P < 0.0001) though also significantly in phone samples (Pseudo-F = 4.9, weighted UniFrac, P < 0.0001). Random forest models were able to determine which of the three conferences a sample was taken from significantly better than expected by chance for both the shoe and phone environments (error ratio = 11.7 and 8.0, respectively). This suggests to us that, as seen in the time series data, different sites maintain a significantly different floor microbial community, which in turn shapes the microbial assemblage structure associated with the shoe samples.

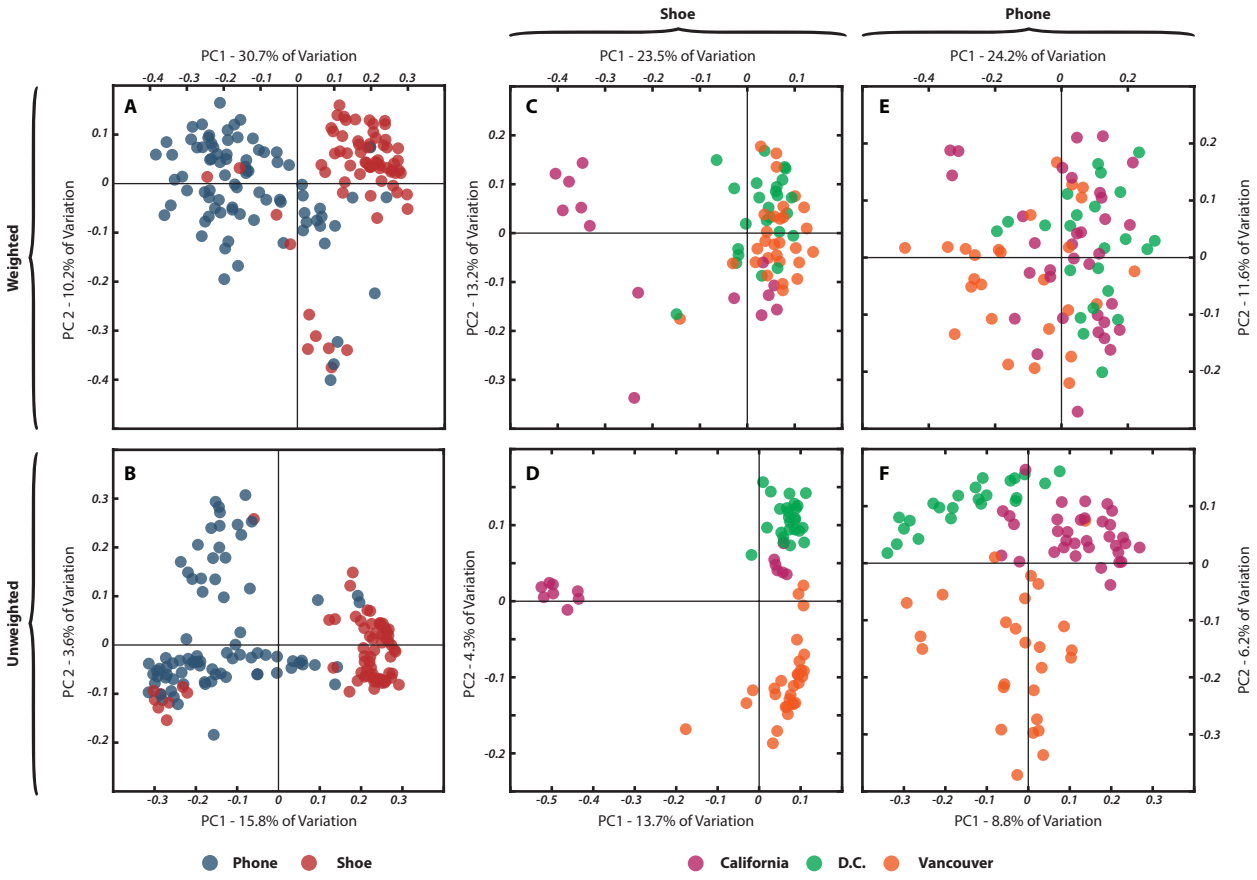

Figure 14: Ordination of biogeographic samples based on weighted and unweighted phylogenetic dissimilarity in community composition. Panels **A** and **B** depict principal coordinate (PCoA) plots for all biogeographic samples based on pairwaise weighted UniFrac distance between samples, with sample points colored by surface and location respectively. **C** and **D** depict ordinations of shoe samples, colored by sampling location, based on weighted and unweighted UniFrac distance, respectively. **E** and **F** depict ordinations of phone samples colored by sampling location and are based on weighted and unweighted UniFrac distance, respectively.

4.4 Conclusions

Microbial communities show unique structure and composition based on surface type, the identity of the person interacting with the surface, and geographic location. This has significant implications for a variety of applications. While we suggest that it is possible to infer individual identities based on the microbial community associated with their smart phone surface, it is less likely that this assemblage could be used to track where that person has been recently located in space due to the rapid turnover of the surface-associated microbial community. We believe that the personalized-nature of the human microbiome and the distinct community types associated with urban and built environments may play a significant role in future forensic investigations.

4.5 Methods

Sample collection

This chapter reports the results of two studies, one of which employed longitudinal sampling of shoe and phone microbial communities (time series study) and one of which collected individual shoe and phone samples from individuals attending three geographically disparate conferences (biogeographical study). For the time series study, two participants were recruited to sample their shoes and phones every hour over the course of two 12-hour time periods on consecutive days. Samples were collected by the participants by rubbing sterile swabs pre-moistened with 0.15 M saline solution on each site of interest. Floor samples were taken immediately adjacent from wherever the participant was standing at the time of shoe sampling; not necessarily in an area where they had recently stepped. All samples were immediately placed at −20°C, or on dry ice in cases where samples were collected while participants were away from home or office. At each sampling site, participants made note of their current environment and of all actions taken over the proceeding hour. Participant 1 wore flat-bottomed, rubber soled boots while participant 2 wore

sneakers with a more complex sole topography. Each participant wore the same pair of shoes on their 2 days of sampling, both with rubber soles.

For the biogeographical study: At three national and international conferences during 2012, samples were collected at random from participants' phones and shoes. Samples were collected by the participants by rubbing sterile swabs pre-moistened with 0.15 M saline solution on each site of interest. All samples were immediately placed on dry ice and shipped to Argonne National Laboratory, where they were stored at −80°C until processed.

Library preparation

Total DNA was extracted from swabs using the Extract-N-Amp plant PCR kit (Sigma, St. Louis, USA) following the manufacture's protocol with minor modifications. After extraction, DNA was quantified using PicoGreen (Invitrogen, Grand Island, USA) and a plate reader. DNA was then amplified using the Earth Microbiome Project barcoded primer set, adapted for the Illumina HiSeq2000 and MiSeq (Illumina, San Diego, USA) by adding nine extra bases in the adapter region of the forward amplification primer that support paired-end sequencing. The V4 region of the 16S rRNA gene (515 F-806R) was amplified with region-specific primers that included the Illumina flowcell adapter sequences and a 12-base barcode sequence (Caporaso *et al.*, 2010A; Caporaso *et al.*, 2012). Each 25 μl PCR reaction contained the following: 12 μl of MoBio PCR Water (Certified DNA-Free; MoBio, Carlsbad, USA), 10 μl of 5 Prime HotMasterMix (1×), 1 μl of forward primer (5 μM concentration, 200 pM final), 1 μl of Golay Barcode Tagged Reverse Primer (5 μM concentration, 200 pM final), and 1 μl of template DNA. The conditions for PCR were as follows: 94°C for 3 min to denature the DNA, with 35 cycles at 94°C for 45 s, 50°C for 60 s, and 72°C for 90 s, with a final extension of 10 min at 72°C to ensure complete amplification. Amplicons were quantified using PicoGreen (Invitrogen) and a plate reader. Once quantified, different volumes of each of the products are pooled into a single tube so that each amplicon is represented equally. This

pool is then cleaned up using UltraClean® PCR Clean-Up Kit (MoBio, Carlsbad, USA), and then quantified using Qubit (Invitrogen, Grand Island, USA). After quantification, the molarity of the pool is determined and diluted down to 2 nM, denatured, and then diluted to a final concentration of 4 pM with a 30% PhiX spike for loading on the Illumina HiSeq2000 sequencer (for the time series study), and a final concentration of 6.1 pM with a 30% PhiX spike for sequencing on the Illumina MiSeq (for the biogeographical study).

Sequence processing and analysis

Unpaired reads of length 151 bp for both the time series and biogeographic studies were clustered together at 97% identity using the Quantitative Insights Into Microbial Ecology (QIIME) script pick_open_reference_otus.py, with the May 2013 release of Greengenes (greengenes.lbl.gov) as the reference. OTUs comprising only a single read were discarded, and samples were rarified to an even depth of 1,000 reads. Analysis of beta-diversity was performed by calculating the pairwise weighted and unweighted UniFrac (Lozupone & Knight, 2005) distance between each pair of samples, and the resulting distance matrix was used for all downstream statistical tests of sample similarity. The significance of sample groupings was assessed using PERMANOVA (QIIME's compare_categories.py script) and statistical significance was calculated by comparing the Pseudo-F statistic to a distribution generated by 10,000 permutations of the randomized dataset.

Random forest models

Random forest supervised learning models were used to determine the diagnostic power of microbial community profiles in predicting the surface type or participant a sample originated from. These models form decision trees using a subset of samples to identify patterns associated with a metadata category and then test the accuracy of the tree on the remaining samples not used for training. Each model runs a number of independent trees and reports the ratio of model error to random error as a metric for the predictive power of the category's microbial communities. A greater ratio of baseline to model error indicates a better ability to classify that grouping by

microbial community alone. The models were run using the supervised_learning.py command in QIIME, with 1,000 trees per model and tenfold cross validation.

SourceTracker models

For the SourceTracker models, all four shoe samples taken by each participant at a given time point were consolidated and treated as individual sinks (N = 29 and 27 for participants 1 and 2, respectively). All floor samples from the two participants' time series were collapsed and treated as the two possible sources to the shoe sink community. Models were run following QIIME tutorial guidelines (http://qiime.org/tutorials/source_tracking.html).

Procrustes analysis

Procrustes analysis compares the shape of two PCoA plots by optimally rotating and scaling one plot to best fit the other, with the goodness of fit measured by the M2 statistic. P values are generated using a Monte Carlo simulation in which sample identifiers are shuffled (here 1,000 times) and the M2 statistic is compared to the distribution drawn from these permutations. The proportion of M2 values that are equal or lower than the actual M2 value is the Monte Carlo P value.

Only time points in which all four shoe samples passed quality filtering were considered (N = 24 for participant 1 and 19 for participant 2). For each participant, samples were divided by shoe environment and four different sets of principal coordinates were computed based on weighted UniFrac distance between samples. The QIIME script transform_coordinate_matricies.py was used for Procrustes analysis, with the left heel coordinates used as the reference and the other three coordinate matrices transformed to best fit the reference.

Availability of supporting data

All sequencing data as well as the OTU table and mapping file are available at http://figshare.com/articles/Forensic_analysis_of_the_microbiome_of_phones_and_shoes/1311743.

Chapter 5

Hospital associated microbiota and implications for nosocomial infections

5.1 Abstract

The rise of high-throughput sequencing technologies and culture-independent microbial surveys has the potential to revolutionize our understanding of how microbes colonize, move about, and evolve in hospital environments. Genome analysis of individual organisms, characterization of population dynamics, and microbial community ecology are facilitating the identification of novel pathogens, the tracking of disease outbreaks, and the study of the evolution of antibiotic resistance. Here we review the recent applications of these methods to microbial ecology studies in hospitals and discuss their potential to influence hospital management policy and practice and to reduce nosocomial infections and the spread of antibiotic resistance.

5.2 Introduction

As the global trend toward urbanization has accelerated over the past century, humans have become increasingly tethered to the built environment. From the hospital we are born in to the homes, apartments, and office buildings we live and work in, the indoor environment has become our most intimate ecosystem (Kelley & Gilbert, 2013), yet our ignorance of the microorganisms that share this habitat remains profound. The bacteria, fungi, and viruses that colonize these environments may help to shape our own microbiomes and can fundamentally alter the trajectory of our physiological, immunological, and neurological development. Designing our buildings and city spaces with the microbiome in mind may help to improve our health and mapping the microbial communities of our built environments may help us track biothreats and diseases, develop

sophisticated early warning systems, and understand how a changing climate and increasing population density will shape human health. The indoor ecosystem, and the urban environment in particular, are hotspots for reduced microbial diversity and this reduction may be having untold consequences for our health and well-being (Blaser & Falkow, 2009). This depauperate exposure to a complex microbiome that would normally train a healthy immune system has been linked to the rise in asthma and allergies (Blaser & Falkow, 2009; Fujimura *et al.*, 2014; Stefka *et al.*, 2014).

Our own microbiota is both shaped by and influences the microorganisms that inhabit built environments and recent studies have uncovered the extent to which humans influence the microbiota of the spaces they occupy. This research has furthered our understanding of how microbes interact with and survive and proliferate in built environments and in particular has demonstrated that the human skin microbiome comprises the major microbial source for indoor systems as varied as airplanes, kitchens, offices, and public restrooms (Lax *et al.*, 2014; Meadow *et al.*, 2014; Kembel *et al.*, 2014; Flores *et al.*, 2013; Gibbons *et al.*, 2015). The strength of this interaction is such that a built environment can be matched to its occupants based entirely on microbial similarity (Lax *et al.*, 2015) and that microbial signatures left on objects can be forensically matched to individuals (Fierer *et al.*, 2010).

The microbiology of the built environment has some of its most profound implications for health-care facilities, where hospital-acquired infections (HAIs) have long been among the leading causes of patient deaths (Anderson & Smith, 2005; Groseclose *et al.*, 2004; CDC 2009; Klevens *et al.*, 2007). Determining how microorganisms colonize, persist, and change in the hospital environment has the potential to elucidate the major sources of these HAIs, but the complexity of the microbial world confounds our attempts to focus on specific pathogens only. It is now essential that we understand the ecology of these frontline medical environments and elucidate the mechanisms by which the microbiology of a health-care setting can influence patient outcomes. Only then can we hope to engineer solutions to regain control over these outcomes. Traditional studies of health-care-

associated infections have relied heavily on cultured isolates, genotyping of known pathogens, and *post hoc* characterization of potential transmission routes. In this review, we discuss how novel sequencing and bioinformatics techniques are revolutionizing our understanding of hospital-associated microbiota, the origin and structure of hospital outbreaks, and the evolution of antibiotic resistance.

5.3 **Healthcare-Associated Infections**

Health-care-associated infections are an increasingly prevalent threat in the US health-care system. Patients may acquire a pathogenic infection after admittance to a health-care environment although it is often more complicated, as the patient's own microbiome may also harbor certain types of HAI (Olivas *et al.*, 2012). Exact determination of HAI prevalence is complicated by the lack of a single US surveillance system and the fact that most hospitals limit reporting of HAIs to device-associated and surgical infections, as well as those due to the pathogens *Clostridium difficile* or methicillin-resistant *Staphylococcus aureus* (MRSA). The vast increase in deaths attributed to these two hospital-associated pathogens (HAPs) over the past decade has been shocking, with more than 10 000 deaths per year in the UK attributed to these diseases (Pearson, 2009).

A 2011 survey by the Centers for Disease Control found that 4% of patients in acute-care hospitals had at least one health care-associated infection, more than half of which were not associated with devices or operative procedures (Magill *et al.*, 2014). They estimate that there were 648 000 patients with 721 800 HAIs in 2011, with a median interval of 6 days between hospital admission and the onset of HAI symptoms. Greater patient age, longer duration of hospital stay, larger hospital size, and the insertion of a central catheter were found to be the greatest risk factors in the contraction of HAIs. A wealth of prior studies have identified dominant HAPs and putative routes of transmission, including physicians' and nursing staff's clothing (Wiener-Well *et al.,* 2014; Lopez *et al.*, 2009; Treakle *et al.*, 2009), stethoscopes (Marinella *et al.*, 1997), personal phones (Brady

et al., 2009; Ulger *et al.*, 2009; Brady *et al.*, 2006; Akinyemi *et al.*, 2009), and computer keyboards (Bures *et al.,* 2000*)*. What these studies share is a specific focus on known pathogens and a reliance on microbial cultures.

5.4 **Metagenomic characterization of hospital microbial communites**

Historically, studies of hospital microbiota and infection control have relied on culture-dependent methods, taking a 'needle in a haystack' approach to select for specific pathogens rather than assessing the whole microbiome (Westwood *et al.*, 2014). Such methods are unable to effectively characterize the microbial diversity of abiotic hospital surfaces or the asymptomatic carriage of microbes by hospital staff (Westwood *et al.*, 2014). An assessment of the full microbial community, by contrast, allows inference of the factors structuring microbial assemblages in hospitals and the effects of building materials and design, cleaning regimens, and abiotic environmental factors on community diversity.

In many ways, hospitals represent an intriguing model system for the study of built-environment microbial communities, both because of their obvious connection to human welfare and because they allow almost total normalization of building materials, temperature, humidity, air source, and ventilation. This standardization allows us to disentangle microbial interactions between humans and the built environment from the myriad compounding abiotic factors of other systems. With a defined set of microbial sources, patients, staff, water, and air, it should be possible to elucidate patterns of microbial transfer within hospitals and determine how members of these microbial communities persist and grow in response to cleaning regimens and architectural choices.

Nowhere is the threat of HAI more concerning than in neonatal intensive care units (NICUs), where low-birth-weight infants are typically immunocompromised and susceptible to opportunistic pathogens (Couto *et al.*, 2007). Infants are born mostly sterile, with acquisition of the gut microbiota being first shaped by delivery method and then by diet and genotype (Dominguez-

Bello *et al.*, 2010; Palmer *et al.*, 2007; Trosvik *et al.*, 2010). In very-low-birth-weight (VLBW) infants, the development of this microbiome can become disrupted by the common use of broad-spectrum antibiotics, resulting in lower diversity, chaotic fluxes in community composition, and a large number of opportunistic pathogens (Morowitz *et al.*, 2011; Wang *et al.*, 2009; Morowitz *et al.*, 2010). A 2004 study found that 65% of VLBW infants contracted at least one HAI and 27% of infant deaths in NICUs included infection as a coded cause of death (Stoll *et al.*, 2004).

Some of the greatest efforts to characterize hospital-associated microbiota with high-throughput culture-independent techniques have focused on NICUs. The use of 16S rRNA gene sequencing enables the characterization of low-abundance and unculturable microorganisms. Although these methods cannot determine which microorganisms are viable and metabolically active, they can help track the spread of a microbial taxon through the hospital environment, which is of critical importance in determining the role of hospital surfaces and equipment in vectoring microorganisms.

A recent study of two San Diego NICUs found far more microbial diversity than revealed by earlier culture-based studies of NICU surfaces (Hewitt *et al*, 2013). Interestingly, comparison with other built environment surfaces, such as those in offices and restrooms, revealed similar taxonomic profiles dominated by the genera *Streptococcus*, *Staphylococcus*, *Pseudomonas*, *Enterobacter*, and *Neisseria*. Despite similarity in abundant genera, some NICU samples were differentiated by a large number of Enterobacteriaceae sequences, including the taxa *Escherichia coli*, *Klebsiella*, *Enterobacter*, and *Salmonella*, which commonly inhabit the digestive tract and are well known ICU pathogens that can easily proliferate in hospitals. Using a Bayesian source-tracking algorithm, the authors showed that the dominant source of NICU microorganisms was human skin.

A second NICU study used metagenomic analysis to determine the extent to which hospital surfaces are the source of colonizing microbes in the gastrointestinal (GI) tract of premature infants (Brooks *et al.*, 2014). The authors collected fecal samples from two infants every third day for the

first month of life, as well as samples from NICU surfaces. Dominant gut taxa were similar to those found in the nursing room, especially those on feeding and intubation tubing. Numerous antibiotic-resistance, biofilm-formation, and starvation-resistance genes were detected in the genomes of microbes in the fecal samples of both infants, potentially explaining how certain organisms are able to persist in such a regularly sterilized environment.

Few other studies have made use of the increasingly high-throughput sequencing pipelines that have reshaped the field of microbial ecology in recent years. The first ICU survey to make use of high-throughput sequencing characterized the microbial communities on inanimate surfaces of the ICU wards of a Spanish hospital (Poza *et al.*, 2012). The study also took samples from the entrance hall to the hospital, to test how the extent of cleaning and the number of people occupying a space influence hospital microbial communities. Although the study found substantially less microbial diversity in the ICU compared with the entrance hall, the 1145 taxa detected in the ICU demonstrate the surprisingly great diversity that can inhabit a space that is constantly being sanitized and treated with antibiotics. The advent of culture-independent microbial identification has upended our ideas about sterility in general. Even surveys of NASA clean rooms used for spacecraft assembly, which are rigidly sealed and sterilized, have revealed that these rooms harbor hundreds of microbial taxa (La Duc *et al.*, 2012). If it is indeed impossible to completely sterilize a built environment, that leaves open the question of which taxa are best able to survive intensive cleaning regimens and what potentially deleterious consequences may arise from this selective pressure placed on microorganisms.

5.5 **The spread of antimicrobial resistance in hospital environments**

The widespread use and accumulation of antibiotics in the environment over past decades has resulted in a worldwide crisis of antibiotic-resistant bacteria. This rapid rise in microbial resistance is largely driven by transfer of antimicrobial-resistance (AMR) genes between taxa

through lateral gene transfer (LGT), which represents one of the most dramatic and detrimental consequences of anthropogenic impacts on the evolution of other species (Gillings & Stokes, 2012). The saturation of the environment with antimicrobial compounds has placed strong selective pressure on the uptake of AMR genes, which are often contained on complex DNA vectors also carrying resistance to disinfectants and heavy metals (Baker-Austin *et al.*, 2006; Skippington & Ragan, 2011).

Most antibiotics were originally isolated from soil-dwelling bacteria that also carried AMR genes protecting them from their own metabolites. The AMR genes and DNA vectors found in modern pathogens can often be traced back to their environmental roots (Aminov, 2009; Gillings *et al.*, 2008; Martinez, 2008), suggesting that natural microbial ecosystems contain a vast reservoir of AMRs that can become acquired by pathogens. The human microbial ecosystem may be the most important of all for AMR transfer, which traditional clinical strategies have done much to exacerbate (Gillings & Stokes, 2012). Historically, antibiotics were designed to treat illness without the identification of a specific pathogen through broad-spectrum activity, which resulted in strong selection pressure on a correspondingly broad group of microbial taxa. This indiscriminate selective pressure can drive LGT events in groups of previously commensal organisms that were not the target of the antibiotics and can promote the growth of resistant pathogenic strains at the expense of commensal and beneficial ones (Gillings & Stokes, 2012).

There have been numerous studies documenting the LGT of AMR *in vivo* in hospital environments (Spanu *et al.*, 2012; Fluit *et al.*, 2013; Ruimy *et al.*, 2012; Otto, 2013; Cuisa *et al.*, 2012) . Many of these studies have focused on the acquisition of beta-lactam antibiotic resistance by *S. aureus* (MRSA). Methicillin resistance is conferred by the *mecA* gene, which is carried on a mobile genetic element called Staphylococcal Cassette Chromosome *mec* (SCC*mec*). SCC*mec* is widespread in coagulase-negative *Staphylococcus* (CoNS), a group not traditionally regarded as pathogenic but which shares the same ecological niche as *S. aureus* in the human anterior nares. There is evidence for

frequent horizontal transfer of SCC*mec* between CoNS and *S. aureus*, which will be of fundamental importance in combating and understanding MRSA epidemiology (Fluit *et al.*, 2013; Ruimy *et al.*, 2012). Phylogenetic studies of *S. aureus* isolates from the lungs of a chronically infected cystic fibrosis patient taking heavy doses of antibiotics found strong evidence for local adaptation of a single isolating strain that became heterogeneously insensitive to antibiotic treatment (McAdam *et al.*, 2011). These types of longitudinal studies are critical for documenting the emergence of AMR *in vivo* so that antibiotic-treatment failure can be more specifically explained (Hartfield *et al.*, 2014).

Although Gram-positive pathogens have dominated research into AMR evolution, *in vivo* AMR transfer has also been documented in Gram-negative strains such as *Klebsiella pneumoniae* and *E. coli* (Spanu *et al.*, 2012). It is extremely important that our understanding of nosocomial infections and antibiotic resistance in health-care environments is not limited to our experience but that we are able to reach beyond what we expect to find.

5.6 **Discovery and characterization of new pathogens**

Determining which microbial taxa are potentially disease causing can be difficult, especially because antibiotic-resistance genes are widespread and found even in the remotest of environments (Bhullar *et al.*, 2012). Advanced molecular analytical techniques, including whole-genome sequencing, allow the identification of nosocomial pathogens beyond the most studied diseases (e.g., MRSA, *C. difficile*) and into emerging threats such as Gram-negative multidrug-resistant bacteria.

Researchers were able to trace a 2011 outbreak of antibiotic-resistant *K. pneumoniae* at the US National Institutes of Health (NIH) Clinical Center through whole-genome sequencing of patient isolates, although only 41 nucleotides in a genome of 6 million base pairs were variable between isolates (Snitkin *et al.*, 2012). This very high level of resolution enabled the formation of a transmission network pointing to three independent transmission events from a single index case; these transmissions ultimately led to hospital-wide dissemination of the outbreak strain.

The *K. pneumonia* strains from the NIH study were carbapenem-resistant Enterobacteriaceae (CREs), which, in recent years, have become a particularly formidable threat, with some investigations reporting a mortality rate as high as 80% (Ben *et al.*, 2012). Carbapenems are an antibiotic class of last resort, making CREs a class of HAI that is almost impossible to effectively treat. They are easily transferred in health-care settings from patients and staff with asymptomatic colonization and have the potential to spread carbapenem resistance through plasmid transfer to other human gut microbiota (Snitkin *et al.*, 2012). Although such transfer is well documented in model organisms and laboratory strains, much less is understood about how antibiotic resistance may transfer in hospital settings (Ben *et al.*, 2012). This ability of microbial taxa to exchange genetic material requires us to think beyond whole-genome epidemiology and to think of pathogenic taxa within the context of the full microbial community.

Tracking pathogen movement with high-throughput sequencing has an impact beyond single hospitals and can be used to explore interhospital and even global routes of pathogen transmission. Using phylogenetic models, researchers were able to track the historical spread of a specific MRSA strain from London and Glasgow to smaller regional hospitals where local endemic strains began to circulate (McAdam *et al.*, 2012). On an even larger scale, phylogenetic models based on high-throughput sequencing have been used to compare globally distributed MRSA isolates and to track the worldwide spread of AMR over the course of four decades (Harris *et al.*, 2010).

The rapid advances in sequencing technology used to characterize the full genomes of cultured isolates also allow us to detect AMR genes across the full microbial community, including those in unculturable microorganisms. Although reducing our ability to characterize a specific strain, shotgun metagenomics enables us to embrace the full phylogenetic diversity of each microbial sample, including microorganisms of potential benefit.

5.7 **Effects of cleaning regimens and abiotic factors**

Despite the obvious public health interest in reducing nosocomial infection rates, there remains very little understood about the sources of most infections, including the extent of airborne transmission. Research on airborne microbes in built environments has looked at their relationship to airflow in hospital rooms and found that indoor air passed through mechanical ventilation was less diverse but more enriched in organisms closely related to human pathogens (Kembel *et al.,* 2012). By contrast, opening the windows in patient rooms has been found to significantly reduce the percentage of potentially pathogenic airborne bacteria (Escombe *et al.*, 2007), a realization that goes back as far as Florence Nightingale's demonstration that opening windows on wards of Crimean War casualties led to an improvement in patient outcomes.

Technological advances in our ability to control air circulation have lead to increased isolation between the indoor and outdoor environments, so that in many hotels, offices, and hospitals it is not possible to open windows. This serves two core purposes. First, it reduces the likelihood of negative air exchange with the outside, reducing the potential escape or ingress of pathogens or pollutants. Second, it makes control of the building environment's temperature and humidity significantly more efficient, increasing energy efficiency and reducing the cost of running these facilities. In many ways, almost inadvertently the built environment has been developed over the past 120 years to become increasingly inhospitable to microbial life. Environments are kept dry and surfaces are often covered with antimicrobial materials. While the built environment and control of its parameters has undoubtedly reduced the spread of communicable diseases, it has also changed our microbial relationship to the environment. In the hospital environment, where patients are often immunosuppressed or microbially dysbiotic, being exposed to an ecosystem with very limited microbial diversity may be a good thing. However, it is also possible that lack of exposure to a rich, diverse microbiome may exacerbate certain conditions and therefore negatively influence patient outcomes.

In the absence of a diverse microbial community, surface environments could play host to communicable pathogens that would otherwise be outcompeted by a diverse microbiota, as has recently been demonstrated in a study of post-cleaning microbial succession in public restrooms (Gibbons *et al.*, 2015). This study found that, after full decontamination with bleach, a late-stage successional community developed after 8 h that comprised mostly skin-associated taxa and that remained stable over the course of weeks. By contrast, early successional communities were unstable and more dominated by gut-associated taxa related to potential pathogens.

5.8 **Concluding remarks**

Built environments comprise chemical and physical habitats unprecedented in the natural world that may have untold consequences for the selection and growth of microorganisms. The hospital environment, despite the exquisite control imposed on its biological matrix, remains home to a bewildering diversity of microorganisms. Understanding the ecology of these complex communities will be likely to pay considerable dividends in the control of health-care associated infections and the spread of antibiotic resistance. However, we still have a considerable way to go in understanding and manipulating this environment so that we can control the ecological succession, structure, and pathogenicity of the indoor microbial world. Continued research is needed to quantify this environment, to explicitly examine how the metabolism of individual organisms and metabolic interactions between these organisms and with their environment structure the community dynamics that have been observed. If we can learn how to capture these processes in metabolic models, we may develop the potential to forecast how these communities will respond to changes in hospital management practice, patient treatment, and even architectural modifications. The complexity of these interactions is extraordinary and requires concerted, coordinated efforts with multidisciplinary teams to appropriately parameterize, quantify, and model these ecosystems. Instead of trying to

enforce our will on a great unknown, we will be able to tweak our environment to influence the microbial mechanism that has such a profound influence on our health and well-being.

Chapter 6

Bacterial colonization and succession in a newly opened hospital

6.1 Abstract

The microorganisms that inhabit hospitals may influence patient recovery and outcome, although the complexity and diversity of these bacterial communities can confound our ability to focus on potential pathogens in isolation. To develop a community-level understanding of how microorganisms colonize and move through the hospital environment, we characterized the bacterial dynamics among hospital surfaces, patients, and staff over the course of 1 year as a new hospital became operational. The bacteria in patient rooms, particularly on bedrails, consistently resembled the skin microbiota of the patient occupying the room. Bacterial communities on patients and room surfaces became increasingly similar over the course of a patient's stay. Temporal correlations in community structure demonstrated that patients initially acquired room-associated taxa that predated their stay but that their own microbial signatures began to influence the room community structure over time. The α- and β-diversity of patient skin samples were only weakly or nonsignificantly associated with clinical factors such as chemotherapy, antibiotic usage, and surgical recovery, and no factor except for ambulatory status affected microbial similarity between the microbiotas of a patient and their room.

6.2 Introduction

The indoor environment has become the most intimate ecosystem for most inhabitants of the developed world. A strong link has been observed between the microbial communities

associated with human skin and those recovered from buildings (Lax *et al.* 2014; Lax *et al.* 2015B; Adams *et al.*, 2015; Wood *et. al.*, 2015), and the reduced microbial diversity of these environments relative to the outside world may be linked to an increased incidence of immunological diseases such as asthma and allergies (Lax *et al.,* 2015A; Fujimura *et al.*, 2014; Stefka *et al.*, 2014; Blaser & Falkow, 2009). A potential link between hospital-associated microbial communities and hospital-acquired infections, a leading cause of patient death (CDC 2009; Anderson & Smiths, 2002; Groseclove *et al.*, 2004), needs further investigation (Lax *et al.*, 2015B). In particular, an understanding of bacterial community structure in hospital environments will be critical for mapping the dissemination of antimicrobial resistance genes (Pehrsson *et al.,* 2016). Culture-based analyses of pathogen genomes remain essential for inferring patterns of hospital-acquired infection transmission, and full genome sequencing may be required to establish virulence. Additionally, very subtle nucleotide variation can be used to establish transmission networks within hospitals (Snitkin *et al.*, 2012; Quick *et al.*, 2015) and around the globe (Harris *et al.*, 2010; McAfam *et al.*, 2012). Although of critical importance, whole-genome analyses necessarily restrict the focus of studies of hospital-associated bacteria to a small subset of clinically relevant taxa. Larger-scale analyses of commensal hospital microbiota can help to elucidate how microorganisms are vectored through the health care environment and the extent to which the skin microbiota of patients and staff influence, and are influenced by, these surroundings. Outside of a small number of studies limited to intensive care units and neonatal care rooms (Brooks *et al.*, 2014; Hewitt *et al.*, 2013; Poza *et al.*, 2012), no detailed longitudinal, culture-independent analyses of the hospital microbiota have yet been undertaken (Lax *et al.*, 2015B). Additionally, although a number of longitudinal studies have clarified the temporal dynamics of the human skin microbiota (Lax *et al.*, 2014; Oh *et al.*, 2016), little is known about how skin bacterial communities of patients respond to hospital stays and to clinical treatments such as chemotherapy and antibiotic administration.

Here, we present a yearlong survey of the bacterial diversity associated with the patients, staff, and surfaces of the newly constructed Center for Care and Discovery (University of Chicago), an inpatient hospital for medical and surgical patients. Sampling began 2 months before the hospital opening on 23 February 2013 and continued for a year thereafter. We collected 6523 microbial samples from multiple sites in 10 patient care rooms and two nurse stations split evenly across two hospital floors: the surgical subspecialty floor and the hematology and oncology floor (**Figure 15**). One patient room on each floor was sampled daily, whereas all other environments where sampled weekly. All rooms were noncritical care rooms that allowed for 24-hour visitation, and rooms were cleaned daily with a quaternary ammonium solution and at discharge with a 1:1000 bleach solution. The building's environmental and operational conditions, including temperature, relative humidity, illuminance, CO2 concentrations, and infrared doorway beam breaks, were continuously monitored (Ramos *et al.*, 2015; Dedesko *et al.*, 2015). At least 5000 high-quality 16*S* ribosomal RNA (rRNA) V4 amplicons were generated per sample using the protocols outlined by the Earth Microbiome Project.

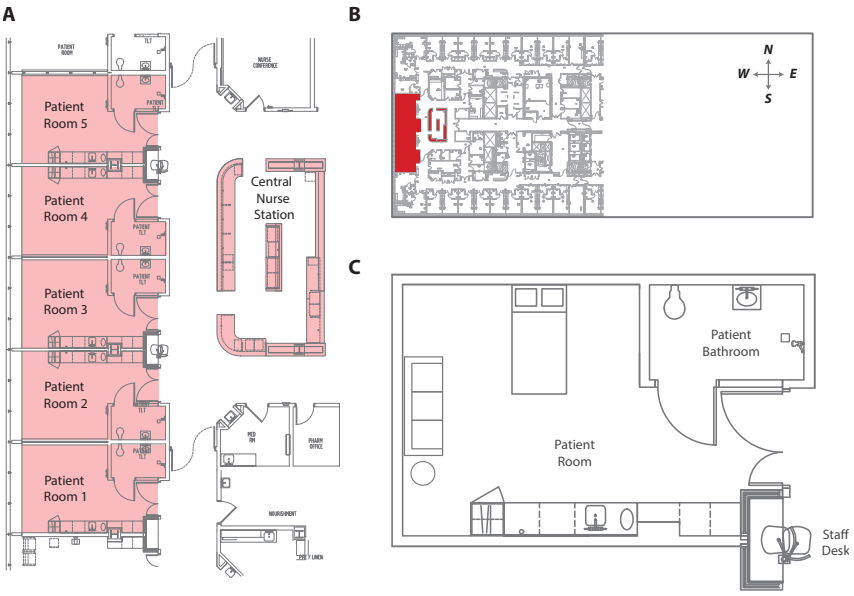

Figure 15: Floor plan of sampling locations. Floor plan for the 9th and 10th floors of the Center for Care and Discovery, University of Chicago. The floor plan is identical for the 9th and 10th floor, where all sampling for the study occurred. (**A**) Location of the 5 rooms sampled on the floor in the study in relation to the central nurse station, which was also sampled. (**B**) Location of the sampled rooms and nurse station within the hospital building, with geographic orientation. (**C**) Floor plan of an individual patient room.

6.3 Results & Discussion

Changes in hospital bacterial communities after hospital opening

As soon as the hospital became operational, the floor and nurse station surfaces demonstrated an increase in the relative abundance of the human skin–associated genera *Corynebacterium*, *Staphylococcus*, and *Streptococcus*, and a decrease in *Acinetobacter* and *Pseudomonas*, which dominated pre-opening (**Figure 16 A,B,C**). Network analysis revealed an almost complete shift in operational taxonomic unit (OTU)–level composition well beyond these dominant genera (**Figure 17**). Shannon diversity, which accounts for both the richness and evenness of observed OTUs, significantly increased in nurse station surfaces with which human skin commonly interacts ($P < 0.001$; **Figure 16 D**) but not in floor samples. Beyond the 5 genera discussed above, a further 15 genera were detected in the data set that satisfied at least one of the following criteria: an average abundance of 1% or greater across all sample types or an abundance of 5% or greater in at least one individual sample type. These included genera such as *Enterococcus*, other unclassifiable members of the family *Enterobacteriaceae*, *Finegoldia*, *Rothia*, *Prevotella*, and *Sphingomonas*, although the 16*S* amplicon–based methods used in this study were not suitable for establishing virulence. We noted that *Propionibacterium*, a dominant colonizer of human skin and built environments, was not well amplified by the primer set used in this study (Meisel *et al.*, 2016).

α- and β-diversity of sample types

We calculated the α-diversity of our samples using two different methods: the Shannon index, which is based on the abundance and evenness of the observed taxa, and the phylogenetic diversity of the samples, which is an unweighted measure of the branch length spanned by a phylogenetic tree of the observed taxa. Skin samples from patients and nurses were generally the least diverse of all sample types by both metrics, whereas sample sites most likely to interact with the outdoors, such as shoes, floors, and recirculated indoor air (which originated largely from high-

efficiency particulate air–filtered outdoor air) (Ramos *et al.*, 2015), were the most diverse (**Figure 18A**). We noted that these calculations were based on evenly rarefied data that provided insight only into the relative, rather than absolute, abundance of different community members.

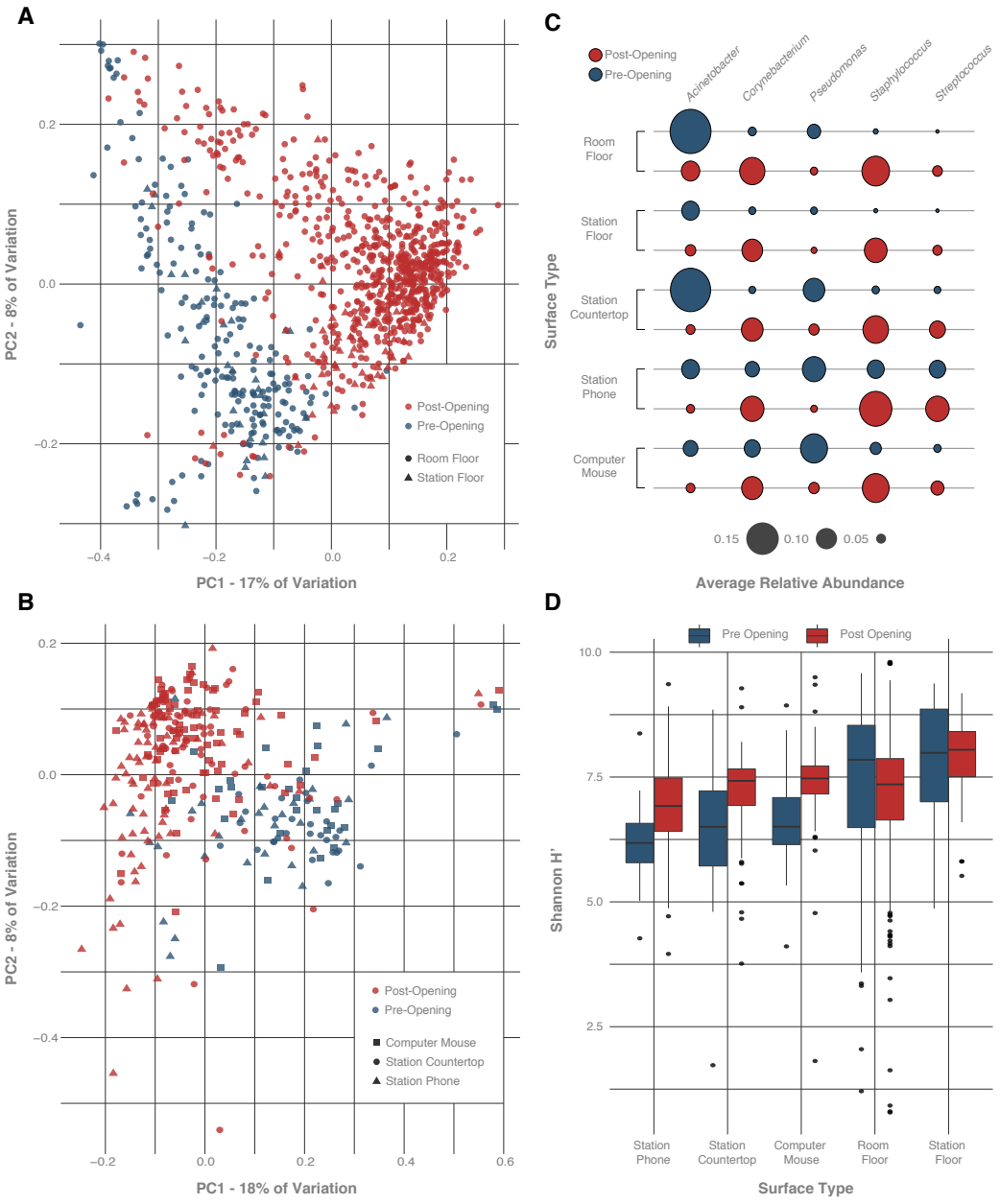

Figure 16: Change in microbial community structure after hospital opening. (**A**) Principal coordinate analysis (PCoA) of all floor samples based on weighted UniFrac distance and colored by whether they were taken before or after the hospital's opening. (**B**) PCoA of three nurse station surfaces colored as in (A). (**C**) Changes in the relative abundance of five key genera after hospital opening. (**D**) Box plot of changes in the Shannon diversity index of samples after hospital opening.

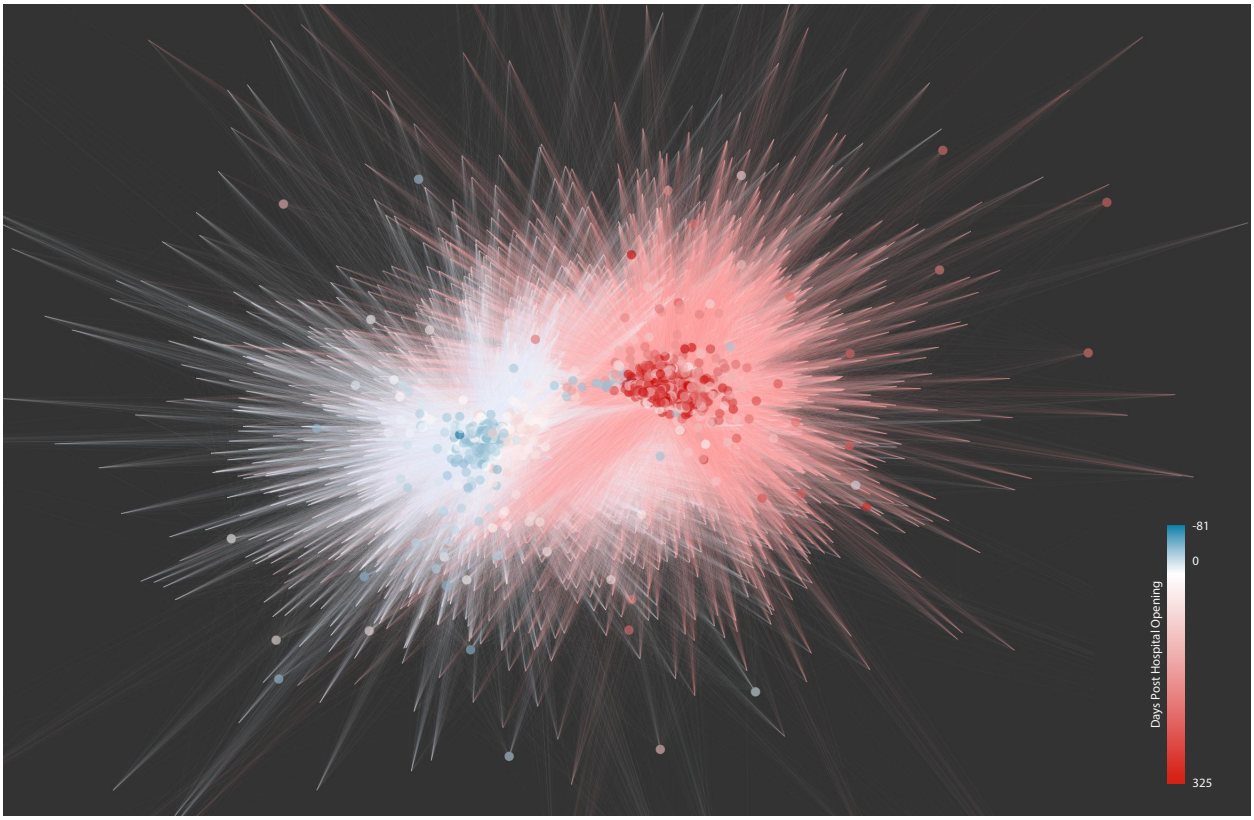

Figure 17: Bipartite OTU network of floor samples. Sample nodes (circles) are connected to each OTU (invisible at edge terminus) detected in that sample by edges, and the network is ordinated using a spring-embedded algorithm that clusters highly connected nodes together. Sample nodes and edges are colored by the number of days post-hospital opening that the sample was taken.

β-Diversity patterns grouped our sample types into three sets (**Figure 18B**). The first set, comprising patient skin samples, staff nose samples, and unused latex gloves, was dissimilar to all other sets and had high variance within sample types. The second set, comprising floor samples and most nurse station samples, was highly similar to other sample types within the set and had low variance within sample types. The final set, comprising staff clothing and personal effects such as shoes, pagers, and cell phones, was similar to the second set but not to other sample types within the group. Hand microbial communities of staff were more similar to the microbiota of hospital surfaces than were hand microbial communities of patients, likely as a result of the greater mobility of staff within the hospital. Supervised learning models could successfully differentiate both nose and hand sample microbial communities according to whether they were taken from staff members

or patients (error ratios of 2.5 and 3.9, respectively). Hand samples could even be differentiated using genus-level data (error ratio = 3.3), with the genera *Micrococcus* (staff-associated) and *Prevotella* (patient-associated) having the highest feature importance scores. In the preopening time period, room and nursing station floor samples had highly similar microbial communities but were dissimilar to other surfaces, whereas post-opening floor samples had a greater degree of similarity to all surfaces, which could be explained by higher foot traffic after opening.

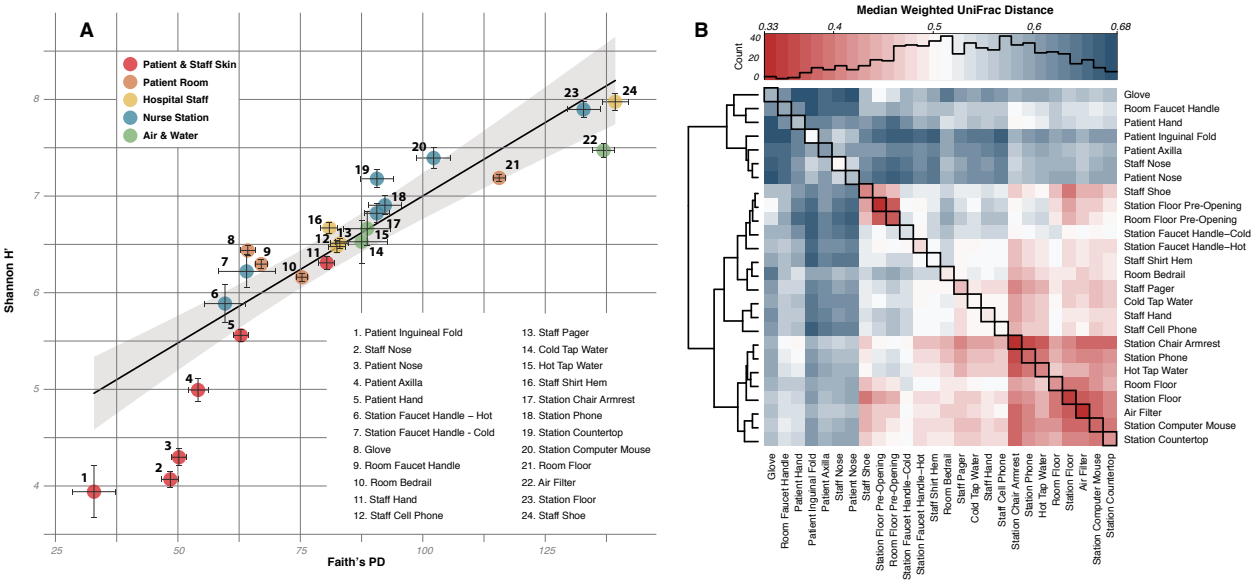

Figure 18: α- and β-diversity of hospital sample types. (**A**) Average α-diversity of sample types based on Faith's phylogenetic diversity (*x* axis) and the Shannon diversity index (*y* axis). Bars indicate SEM. (**B**) Heat map of β-diversity relationships between sample types based on the median weighted UniFrac distance between pairwise comparisons. Sample groups are clustered based on similarity in β-diversity patterns, and median distances within individual sample types are highlighted in black along the diagonal.

Temporal correlation of microbial communities across hospital surfaces

To determine the strength of microbial interactions with different hospital surfaces, we calculated the degree of resemblance between samples taken from two different surfaces on the same day and in the same patient room or nurse station. Here, we determined principal coordinate (PC) correlation, by calculating a weighted average of the correlations between samples taken from the same room and date along the dominant eigenvectors of the distance matrix (**Figure 19**). Patient hands and bedrails had a strong interaction (ϱ = 0.5), although all pairwise correlations

within patient rooms were significant, suggesting a degree of microbial homogenization on the same day within each room. The strongest observed correlations were between the hand microbiota of the hospital staff and their personal cell phones and pagers (ϱ = 0.48 and 0.52, respectively), which mirrors the findings in previous studies (Lax *et al.,* 2015C*,* Meadow *et al.*, 2014). Correlations within the nurse station environment were significant but comparatively weak, likely due to the diversity of people using these environments.

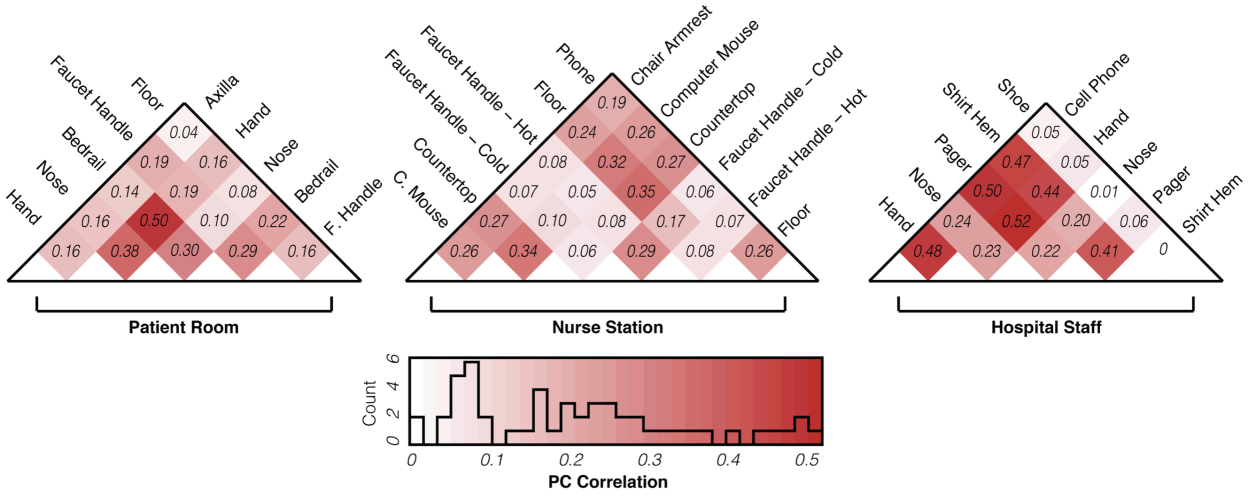

Figure 19: Heat map of principal coordinate space correlations between sample types. Values represent the average correlation between samples taken from the same location and date along the first 10 axes (eigenvectors) of the PCoA plot of all samples, weighted by the variance captured by each axis' eigenvalues.

We calculated the variability in the common microbiota (the set of OTUs found in at least a given percentage of samples) for five surface types as a function of the minimum threshold to be considered part of the common microbiota (**Figure 20A**). The average percent of 16*S* rRNA reads detected in every sample of a given room varied between 15 and 35%. This highlights the biogeographic variability of surfaces and the dynamic nature of the interaction between patient and room microbiotas. Variability was greatest for patient axilla (armpit) and lowest for patient noses; however, the trends for all five surfaces were markedly similar, suggesting that variability in the microbiotas of room bedrails and floors could be attributable to variation in patient skin microbiotas (**Figure 20A**). We used a Bayesian source-tracking approach (Knights *et al.*, 2011) to

evaluate predictive matching of microbial profiles of samples taken from the first and second days of a patient's stay for the 19 patients with such data (**Figure 21A**). For the same surface type (for example, hand to hand), the microbial community on day 2 was highly predictive of the day 1 microbial community of the corresponding patient (**Figure 21A**). Models using the microbial profile of day 1 hands to predict day 2 bedrails and vice versa were also highly accurate. In contrast, floor, nose, and axilla samples had much weaker, but still substantially better than random (1/19, 0.053), predictive accuracy with the exception of nose or floor predicting axilla.

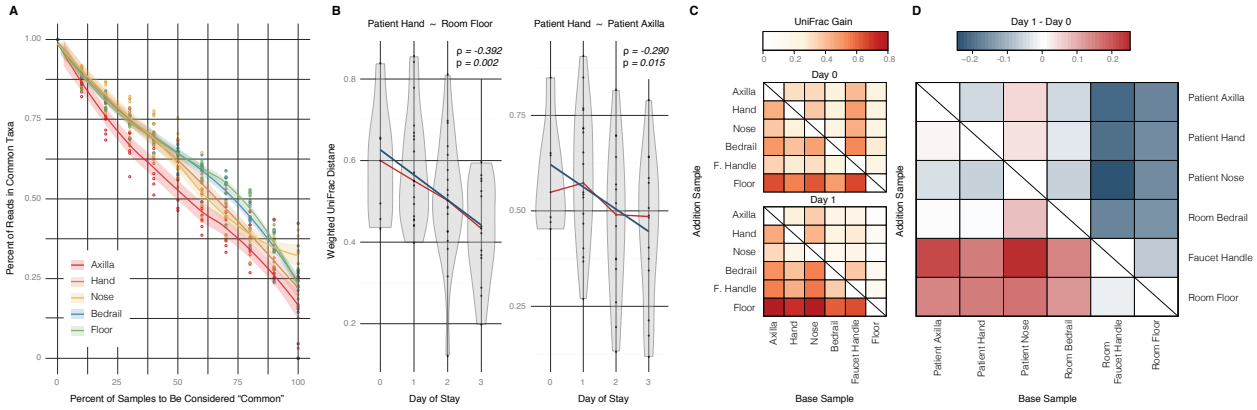

Figure 20: Interaction between patient skin and hospital room microbiota. (**A**) Scatter plot of the percent of 16*S* reads in the "common microbiome" for the eight rooms sampled weekly, with the common microbiome definition on the *x* axis and the percent of reads in that set on the *y* axis. Points represent the eight individual rooms, whereas the trend line is a moving average of the data. (**B**) Microbial similarity between surface types increases with day of stay. Red lines connect the medians of the box plots, and the blue lines are the best-fit linear regressions. ρ is the Spearman rank correlation, and the *P* value is calculated as the percent of 10,000 test statistics drawn from random permutations of the data set with more negative correlations than the one observed. (**C**) UniFrac gain between patient room surfaces illustrates directionality of microbial transfer. Values represent the proportional amount of branch length gained by the addition of one sample's community phylogeny (the "addition sample") to another's (the "base sample"). Heat maps are averages across seven patients on date of check-in (top) and after first night (bottom). (**D**) Change in UniFrac gain dynamics between day 0 and day 1 of hospitalization.

Patient skin and room surfaces became more microbially similar over the course of their stay, as evidenced by uniformly negative Spearman correlations between day of stay and community dissimilarity between surface types (**Figure 20B** and **Figure 21B**). The strength and significance of the correlation varied between comparisons but were strongest between patients' hands and their room floor ($\varrho = -0.39$, $P = 0.002$). Patients' hands and axillae became significantly more similar over

the course of their hospital stay ($\varrho = -0.29$, $P = 0.015$), perhaps because of reduced microbial exposure and homogenization resulting from bed confinement.

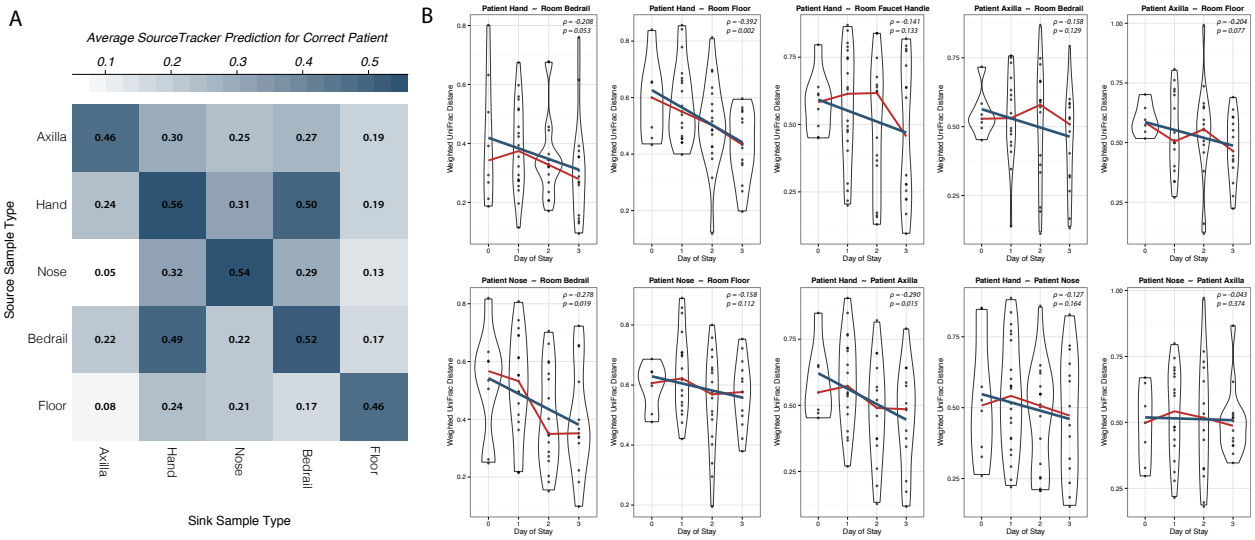

Figure 21: Interaction between patient skin and room samples. (**A**) Heat map of the predictive accuracy of SourceTracker models that used samples taken from the first day of a patient's stay (source sample) to predict which patient a day 2 sample was taken from (sink sample). (**B**) Microbial Similarity between Surface Types Increases with Day of Stay. Red lines connect the medians of the boxplots and the blue lines are the best-fit linear regressions. ρ is the Spearmen rank correlation and the p-value is calculated as the percent of 10,000 test-statistics drawn from random permutations of the dataset with more negative correlations that the one observed.

To assess directionality in microbial transfer, we turned to the seven patients whom we sampled both on their date of admission (day 0) and after they had spent their first night in the hospital (day 1). We calculated the UniFrac gain for each pair of samples, which is an asymmetric measure of how much phylogenetic diversity (branch length) is gained when one sample is added to another (**Figure 20C**). We found that hospital environment samples added more diversity to patient samples on day 1 than on day 0 and that patient samples added more diversity to hospital environment samples on day 0 than day 1 (**Figure 20D**). That is, taxa shared with the skin of the current patient were more abundant on room surfaces after the patient had spent a night in the room, whereas taxa shared with room surfaces were more abundant on patient skin when a patient

first entered the room. This asymmetry suggested that patients initially acquired room-associated taxa that predated their stay but that their own microbial signatures began to influence the room microbiota over time.

Influence of patient clinical factors

There are countless clinical factors that may influence the α- and β-diversity of patient skin, as well as the extent to which the bacterial communities of patients and their rooms resemble each other. To investigate how a patient's clinical history may inform the results of this study, we analyzed the medical records of the 49 patients who were sampled on multiple days. Although every patient had a unique medical history and reason for hospitalization, we coded this complexity into a number of binary variables that allowed for the statistical power to infer their influence on patient skin microbiota. Our analyses focused on the following clinical factors: admission through the emergency room, visit to the operating room before sampling, antibiotic use immediately before admission, antibiotic use at any time point during the hospital stay, antibiotic use at time of sampling, chemotherapy during hospital stay, and ambulatory status on admission. We further considered the patients' sex, age, and weight, as well as their length of stay in the hospital, the service they were admitted to, and the route of antibiotic administration if they were prescribed antibiotics during their hospital stay.

Canonical correspondence analysis (CCA) was used to infer relationships between the taxonomic composition of patient skin and bedrail samples and medical and environmental metadata. No significant effects were observed for any metadata criterion on the α- or β-diversity of room floor or room faucet handle samples. For patient hand and axilla samples, as well as bedrail samples, no factors were found that significantly constrained the variance in OTU-level composition, whereas chemotherapy treatment was the only significant constraint on nose bacterial community composition ($P = 0.02$; 1.3% of the variance). The effect of these factors was more evident in permutational multivariate analysis of variance (PERMANOVA) analyses of β-diversity

(**Table 3A**), with all seven factors significantly influencing the structure of hand microbial communities.

A

	Hand Samples		Axilla Samples		Nose Samples		Bedrail Samples	
	Pseudo-F	*p*	*Pseudo-F*	*p*	*Pseudo-F*	*p*	*Pseudo-F*	*p*
Antibiotics Pre-Admission	*4.506*	**0.0002***	*2.270*	**0.0264**	*2.205*	0.2140	*1.993*	**0.0189**
Antibiotics During Admission	*2.614*	**0.0052***	*1.047*	0.3585	*3.363*	**0.0010***	*2.264*	**0.0074***
Antibiotics At Sampling	*2.797*	**0.0031***	*0.878*	0.5178	*1.992*	**0.0383**	*1.742*	**0.0421**
Chemotherepy During Admission	*4.240*	**0.0002***	*3.608*	**0.0019***	*2.415*	**0.0102**	*2.778*	**0.0005***
Ambulatory on Admission	*1.897*	**0.0189**	*1.367*	0.1704	*2.863*	**0.0014***	*1.580*	0.0778
Samples Taken Post-Surgery	*5.238*	**<0.0001***	*1.684*	0.0842	*2.205*	**0.0243**	*6.176*	**<0.0001***
Emergency Room	*4.858*	**<0.0001***	*1.402*	0.1697	*1.920*	**0.0461**	*4.606*	**<0.0001***

B

Factor	Surface	N Yes	N No	Mean Yes	Mean No	NP T-Test p
Antibiotics Pre-Admission	Patient Hand	29	109	61.6	58.8	0.7184
	Patient Axilla	20	95	48.5	53.2	0.5944
	Patient Nose	24	114	50.0	46.3	0.5568
	Room Bedrail	24	108	73.2	68.3	0.4826
Antibiotics During Admission	Patient Hand	115	23	60.3	54.9	0.2626
	Patient Axilla	93	22	53.9	45.9	0.2206
	Patient Nose	113	25	48.8	38.5	0.0840
	Room Bedrail	109	23	69.4	68.2	0.8346
Antibiotics At Sampling	Patient Hand	40	98	62.7	58.1	0.4488
	Patient Axilla	33	82	45.1	55.3	0.1156
	Patient Nose	38	100	46.5	47.1	0.9040
	Room Bedrail	35	97	68.4	69.5	0.3828
Chemotherepy During Admission	Patient Hand	44	94	67.6	55.6	**0.0208**
	Patient Axilla	40	75	60.2	48.2	0.0610
	Patient Nose	42	96	54.7	43.5	**0.0390**
	Room Bedrail	40	92	78.1	65.3	**0.0154**
Ambulatory on Admission	Patient Hand	124	14	62.0	36.6	**0.0030***
	Patient Axilla	112	3	52.2	59.0	0.5180
	Patient Nose	125	13	48.6	31.3	**0.0220**
	Room Bedrail	120	12	71.9	42.4	**0.0002***
Samples Taken Post-Surgery	Patient Hand	59	79	63.2	56.5	0.1628
	Patient Axilla	53	62	52.3	52.4	0.9876
	Patient Nose	62	76	47.2	46.7	0.9304
	Room Bedrail	57	75	68.8	69.6	0.8626
Emergency Department	Patient Hand	22	116	36.0	63.8	**< 0.0001***
	Patient Axilla	12	103	34.3	54.5	**0.0310**
	Patient Nose	22	116	32.8	49.5	**0.0020***
	Room Bedrail	20	112	53.5	72.0	**0.0142**

Table 3: Effects of seven binary clinical factors on the α- and β-diversity of patient skin and bedrail bacterial communities. (**A**) PERMANOVA analyses of the effects of clinical metadata on observed β-diversity. Each test was based on the weighted UniFrac distance between samples, and significance was assessed through 10^5 permutations of the randomized data set. (**B**) Effects of clinical metadata on the α-diversity of skin and bedrail bacterial communities, based on Faith's phylogenetic diversity index. Significance was assessed through a two-sided nonparametric *t* test with 10^5 permutations. For both (A) and (B), significant test results are highlighted in bold, and those that are significant after a Bonferroni correction for multiple comparisons are indicated with an asterisk.

All factors, except antibiotics taken pre-admission, were significant for nose samples. All factors, except ambulatory status, were significant for bedrail samples. Axilla samples were least influenced by these variables, with only preadmission antibiotics and chemotherapy treatment showing significant effects (**Table 3A**). Despite the significance of these tests, the differences in the average ranked distances of within- and between-group comparisons were uniformly low (all analysis of similarity *R* statistics < 0.15), suggesting relatively weak effects of these factors on the variance of microbial communities between patients. Further, random forest supervised learning models were unable to successfully predict any of these patient factors based on OTU abundances (error ratio < 2), with the exception of weakly predicting whether a patient had visited an operating room based on hand (error ratio = 2.29) and nose (error ratio = 2.38) samples. The feature importance scores of individual OTUs for those two models were significantly correlated ($\varrho = 0.47$), suggesting that patients may either pick up certain taxa in the operating room (or en route to and from the operating room) or experience a consistent reduction in certain taxa due to presurgical preparation with antimicrobials such as chlorhexidine.

We also assessed the effects of these clinical factors on the α-diversity (Faith's phylogenetic diversity) (Faith, 1992) of patient skin and bedrail samples (**Table 3B**). The strongest observed effect was significantly lower diversity for all four sample types in patients who were admitted through the emergency department (**Table 3B**). We hypothesized that this may be due to prolonged duration in the health care environment because patients likely would have spent several hours in the emergency room before being transferred to the general care room where they were sampled. Patients undergoing chemotherapy had significantly lower diversity in nose and hand samples, as well as on bedrail samples, presumably due to indirect toxic effects of chemotherapeutic agents on the bacterial community or as a consequence of immune system changes related to the treatment. Patients who were not ambulatory had lower microbial diversity in nose, hand, and bedrail samples,

presumably because being confined to bed reduced their potential microbial exposure or due to a higher likelihood of microbial dysbiosis before admission. Ambulatory status was the only factor that significantly influenced the level of similarity between sample types, with the microbiota of non-ambulatory patients being less similar to that of hospital surfaces.

Antibiotics had no consistently observed effect on the α-diversity of the microbiota of patient skin (**Figure 22A**) or in ordination clustering of patient skin microbiota samples (**Figure 22B**), even when controlling for length of exposure and the route of delivery.

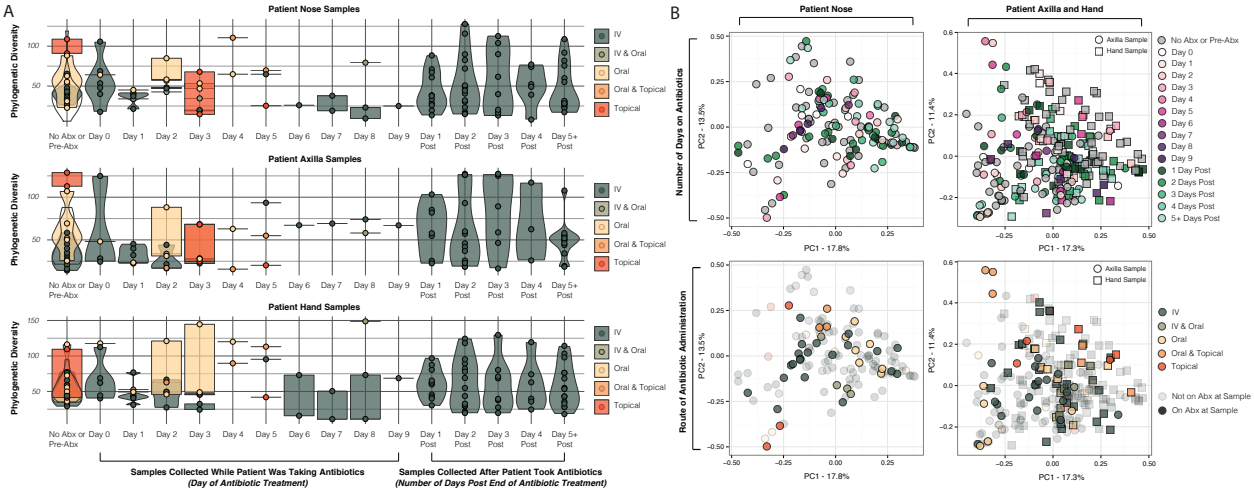

Figure 22: Effect of antibiotic treatment on the diversity of patient skin samples. (**A**) Effect of antibiotic (Abx) treatment on the phylogenetic α-diversity of patient skin samples. Antibiotic status is indicated on the *x* axis, and violin plots indicate the density of phylogenetic diversity for corresponding skin samples, segregated and colored by route of antibiotic administration. For the "no Abx" or "pre-Abx" values, distributions are segregated by the route of administration for later samples of that patient. The distribution for patients who never took antibiotics during their stay is indicated by a black violin plot with no fill color. (**B**) PCoA of patient skin samples based on weighted UniFrac distance, split by sample type. Plots at the top are colored by the antibiotic status of the patient, and the plots at the bottom are colored by the route of antibiotic administration if the patient was taking antibiotics at the time of sampling.

Most patients who received antibiotics were administered the drugs either intravenously or orally, suggesting that these routes of delivery may have negligible effects on the skin microbiota at the community level. A total of 26 different antibiotics were prescribed to the patients we analyzed over the course of their stay, and it is possible that the effects of individual drugs were masked by our grouping according to route of administration. Unfortunately, we do not have the statistical

power to assess the effects of individual drugs, although we do note that the four patients administered topical antibiotics (neosporin in all cases) saw decreases in skin microbial diversity after use of the antibiotic.

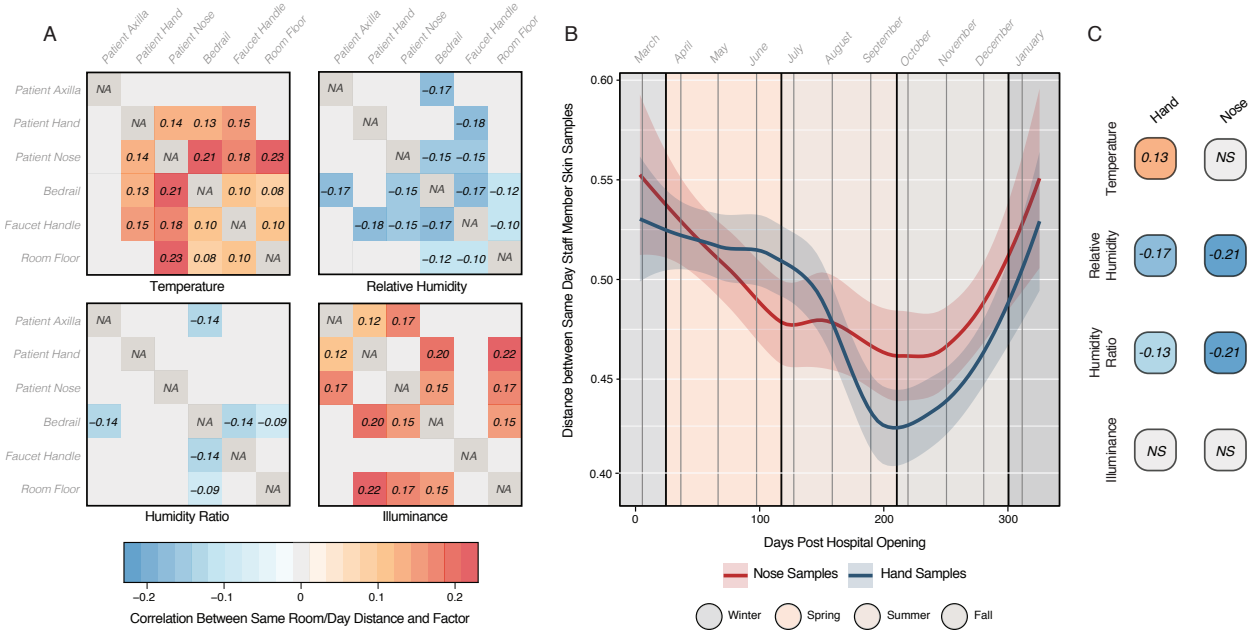

Figure 23: Effect of environmental factors on microbial transmission. (**A**) Heat maps of the correlation between environmental factors and the weighted UniFrac distance between samples taken from the same room on the same day. (**B**) Seasonal changes in the UniFrac distances between the hand and nose samples of nurses working on the same floor. Trend lines are a smoothed moving average of the data. (**C**) Correlations between environmental factors and the UniFrac distances of hand and nose samples of nurses working on the same floor on the same day. Color scheme is as in (A).

Effect of environmental conditions

As previously reported (Ramos *et al.*, 2015), temperature in patient rooms ranged from 20° to 26°C, with mean 23.5°C and SD 1.4°C. Relative humidity ranged from 14.2 to 48.5%, with mean 34.8% and SD 6.8%. Humidity ratio had mean 6.3 gW/kgda with SD 1.23 gW/kgda, and illuminance had mean 173 lx with SD 448 lx (illuminance includes both artificial and natural light) (Ramos *et al.*, 2015). Within a patient room on the same day, higher temperatures and higher illuminance were consistently associated with greater microbial dissimilarity between patient and surface microbial communities, whereas higher relative humidity and humidity ratio were consistently correlated with greater microbial similarity (**Figure 23A**). Microbial similarity among

staff members working on the same floor showed a seasonal trend in both hand and nose samples, with the greatest similarity in late summer/early fall and the least similarity in the winter (**Figure 23B**). Greater nose- and hand-associated microbial similarity among different staff members correlated with higher humidity; hand-associated microbial similarity also correlated with lower temperatures (**Figure 23C**).

6.4 Conclusions

When patients enter a hospital, they arrive with complex and dynamic microbial assemblages that will be shaped by the treatment they receive and by the interactions they have with staff and with the building itself. As the influence of human microbial ecology on patient care and recovery in the hospital environment becomes better understood, being able to reinforce beneficial microbial interactions and mitigate harmful ones throughout the course of hospitalization will become paramount. Here, we provide a descriptive interaction map of the hospital and human microbiomes. Although the study results represent an advance in our knowledge of hospital-associated microbial communities, our data are limited in their ability to provide immediate clinical impact because of the observational (rather than interventional) nature of the study. The use of 16*S* rather than shotgun metagenomic sequencing or culture-based methods also limits our ability to infer the transmission patterns of taxa with specific clinical relevance. In particular, our characterization of the clinical factors influencing patient skin microbiota is based only on community-level analyses without regard to potential virulence, antimicrobial resistance, or metabolic function. Further investigation is needed to clarify the specific effects of these factors on antibiotic-resistant and virulent taxa. However, this foundational knowledge demonstrates the extent to which the microbial ecology of patient skin and of hospital surfaces are intertwined and may provide context to future studies of the transmission of hospital-acquired infections.

6.5 **Methods**

Study design

The rationale for the study was to survey the microbial diversity in a newly opened hospital both before it opened to the public and for a full year thereafter. The surface types to be surveyed were chosen before sample collection began, and with the exception of a shift from sampling patient inguinal folds to patient axillae after the first month, collection sites remained unchanged over the course of the study. No power analyses were used to predetermine the sample size. The 10 patient rooms and the two nurse stations sampled in this study were chosen before the start of sample collection and were constant throughout the course of the study. Patients were consented and sampled based exclusively on whether they were residents of one of the predetermined rooms on a sampling day, and nursing staff were consented and sampled based entirely on whether they were assigned to the rooms chosen for sampling.

Sample collection

Samples were collected by trained staff at the Center for Care and Discovery at the medical center of the University of Chicago in compliance with IRB12-1508. With the exception of air samples, which were collected via ultraviolet-sterilized MERV 7 filter medium placed over the return air grilles in the patient rooms, all samples were collected by rubbing sterile swabs premoistened with 0.15 M saline solution on the site of interest. After collection, samples were immediately frozen at −20°C pending shipment to Argonne National Laboratory on dry ice. Environmental factors and proxies for human occupancy/activity were continuously collected as previously reported (Ramos *et al.*, 2015; Dedesko *et al.*, 2015), although neither of our methods could distinguish between hospital staff and nonstaff visitors. Hospital-acquired infection incidence among the 252 patients who participated in the study was assessed through analysis of ICD-9 (International Classification of Diseases, Ninth Revision) codes.

Amplicon sequencing

All samples were processed using a modified version of the manufacturer's protocol of the Extract-N-Amp kit (Sigma-Aldrich). Swabbed tips were placed into 2 ml of 96-Well Deep Well plates (Axygen). Extract-N-Amp Extraction solution (200 µl) was added, vortexed for 5 s, and incubated at 90°C for 10 min. Samples were centrifuged at 2500*g* for 1 min. Extract-N-Amp Dilution solution (200 µl) was added to each sample to obtain a 1:1 ratio of extraction to dilution solution. Genomic DNA was amplified using the Earth Microbiome Project barcoded primer set, adapted for Illumina HiSeq 2000 and MiSeq by adding nine extra bases in the adapter region of the forward amplification primer that support paired-end sequencing. The V4 region of the 16*S* rRNA gene (515F-806R) was amplified with region-specific primers that included the Illumina flowcell adapter sequences. The reverse amplification primer also contained a 12-base barcode sequence that supports pooling of up to 2167 different samples in each lane (Caporaso *et al.*, 2012). Each 20 µl of polymerase chain reaction (PCR) contains 5 µl of Mo Bio PCR Water (Certified DNA-Free), 10 µl of Extract-N-Amp Ready Mix, 1 µl of forward primer (5 µM concentration, 200 pM final), 1 µl of Golay Barcode Tagged Reverse Primer (5 µM concentration, 200 pM final), and 4 µl of template DNA. The conditions for PCR were as follows: 94°C for 3 min to denature the DNA, with 35 cycles at 94°C for 45 s, 50°C for 60 s, and 72°C for 90 s; with a final extension of 10 min at 72°C to ensure complete amplification. PCR amplifications were completed in triplicate and then pooled. After pooling, amplicons were quantified using PicoGreen (Invitrogen) and a plate reader. Once quantified, different volumes of each of the products were pooled into a single tube so that each amplicon was represented equally. This pool was then cleaned using the UltraClean PCR Clean-Up Kit (Mo Bio) and quantified using Qubit (Invitrogen). After quantification, the molarity of the pool was determined and diluted to 2 nM, denatured, and then diluted to a final concentration of 4 pM

with a 30% PhiX spike for loading on the Illumina HiSeq 2000 sequencer. Amplicons were then sequenced in two 151–base pair (bp) × 12-bp HiSeq 2000 runs and 3 MiSeq runs using the protocol outlined by the Earth Microbiome Project.

Quality control and sequence clustering

Forward reads were quality-trimmed and processed for OTU clustering using the open reference method implemented in the QIIME pipeline (Caporaso *et al.*, 2010A). The sequence identity cutoff was set at 97%, and taxonomy was assigned to the high-quality (<1% incorrect bases) candidate OTUs using the parallel_assign_taxonomy_rdp.py script of the QIIME software. Multiple sequence alignment and phylogenetic reconstruction were performed using PyNAST (Caporaso *et al.*, 2010B) and FastTree (Guindon *et al.,* 2010). OTUs containing less than 5 reads were discarded, and the OTU table was rarefied to an even depth of 5000 reads.

Oligotyping

We used the Oligotyping pipeline (Eren *et al.*, 2013) to identify sub-OTU level variation in four highly abundant genera: *Acinetobacter*, *Corynebacterium*, *Streptococcus*, and *Staphylococcus*. USEARCH (Edgar, 2010) was used to align reads back to OTUs based on a 97% identity cutoff, and mapped reads were quality-trimmed using the FASTX-Toolkit (http://hannonlab.cshl.edu/fastx_toolkit/). The minimum substantive abundance threshold for an oligotype (-M) was set to 500 reads, and the minimum number of samples (-s) and percent abundance cutoff (-a) were set to 1800 and 5%, respectively.

PC space correlation

We calculated the weighted UniFrac distance (Lozupone & Knight, 2005) between each pair of samples and then found the principal coordinates (eigenvectors) of the distance matrix. To reduce the complexity of the data and minimize noise, we focused only on the minimum set of eigenvectors whose eigenvalues summed to 50% of the variance, which, for our distance matrix, were the first 10. We calculated the PC correlation along these 10 eigenvectors (n) such that our PC

correlation (ϱ) was an average of the Pearson correlation along eigenvectors *i* weighted by their associated eigenvalues:

$$\rho = \sum_{i=1}^{n} \rho_i \frac{\lambda_i}{\sum_{j=1}^{n} \lambda_j}$$

Correlation along each eigenvector was checked for significance using the cor.test function in R (two-sided; confidence level, 95%), and all nonsignificant correlations ($P > 0.05$) were reduced to zero before averaging.

Correlations were determined between pairs of samples taken from within the same environment (patient room, nurse station, and hospital staff) on the same day. Glove and water samples were excluded from these analyses due to the small and unique subset of ordination space they occupied in the PCoA of all samples. Air filter samples were also excluded because each was collected over the course of a week rather than on a single day.

Supervised learning

Random forest supervised learning models were used to determine the diagnostic power of microbial community profiles in predicting whether hand and nose samples were taken from a hospital patient or staff member. The models were run using the supervised_learning.py command in QIIME, with 1000 trees per model and 10-fold cross-validation.

Canonical correspondence analysis

The OTU table was split by surface type into hand, axilla, and nose tables and filtered to include only samples that had metadata observations for all factors (n = 115 and 138 for axilla and for both nose and hand). To reduce the influence of widespread taxa, any OTUs detected in more than 80% of samples were filtered out, as were any OTUs detected in less than three samples or which comprised less than 0.01% of all reads within the OTU table. The ordistep command in the vegan R package was used to determine the optimal set of constraining variables. Patient age, patient

weight, number of days since hospital opening, room temperature, room humidity ratio, and room relative humidity were all standardized to mean 0 and variance 1. Measurements for relative humidity, temperature, and light were taken on the wall immediately next to the patient's bed. Chemotherapy treatment, antibiotic treatment before admission, antibiotic treatment during admission, surgery, admission through the emergency room, and ambulatory status were all treated as dummy variables (yes = 1; no = 0). We also attempted CCA on genus-level input data and on our oligotype data but were unable to find any significantly constraining variables for either data set.

Statistics

Statistical analyses were performed in R, except where noted. Changes in the similarity between patients and their rooms over time was assessed through Spearman correlation, and the *P* value was calculated as the percent of 10^5 test statistics drawn from random permutations of the data set with more negative correlations than the one observed. PERMANOVA analyses were performed in QIIME on the weighted UniFrac distance matrix between samples, and significance was assessed through 10^5 permutations of the randomized data set. Factors significantly affecting the α-diversity of skin and bedrail bacterial communities, based on Faith's phylogenetic diversity index, were assessed through a two-sided nonparametric *t* test with 10^5 permutations.

Acknowledgements

My dissertation83 relied on the collection of very large datasets that would not have been possible without the generosity and enthusiasm of many, many collaborators, to whom I could not be more grateful. I'd like to begin by thanking my advisor, Dr. Jack Gilbert, for always thinking big and encouraging others to do so as well. His passion for collaborative, interdisciplinary research has been inspiring, and I owe a great deal of my success as a graduate student to his encouragement and advice. I am also very grateful to Dr. Cathy Pfister and to Dr. Tim Wootton for giving me a home within the Darwinian Sciences cluster, and for their support these last five years. So much of what I've learned about ecology during my dissertation has come from attaining their lab meetings. I would also like to thank Dr. Sarah Cobey and Dr. Maureen Coleman for their fantastic advice and insights as committee members.

I am indebted to the citizen scientists who collected data for Chapter 2 and to the hospital patients and nursing staff who contributed samples for Chapter 6. These studies would not have been possible without their contributions. These projects also benefited tremendously from our wide array of collaborators, including Dr. John Alverdy, and the labs of Dr. Brent Stephems, Dr. Rob Knight, and Dr. Jeffrey Siegel. I am particularly grateful for the contributions of Dr. Kristina Guyton and Dr. Monika Krezalek, who combed through mountains of patient clinical data and turned it into something I could make sense of. Almost all of the work in this dissertation was funded by the Sloan Foundation, which not only supported the research itself but also brought together built environment scientists from around the globe and helped launch a wide array of collaborations in this developing field. It has been an honor to be a part of that community.

Finally, and most importantly, I want to thank the incredible group of fellow students in the Darwinian Sciences cluster that I've been honored to call my friends over the last five years. I've learned more from conversations with them than from any class, and they have made my time at The University of Chicago as joyous and memorable as earning a doctorate can be.

References

Aagaard, Kjersti, et al. "The placenta harbors a unique microbiome." *Science Translational Medicine* 6.237 (2014): 237ra65-237ra65.

Adams, Rachel I., et al. "Microbiota of the indoor environment: a meta-analysis." *Microbiome* 3.1 (2015): 1-18.

Akinyemi, Kabir O., et al. "The potential role of mobile phones in the spread of bacterial infections." *The Journal of Infection in Developing Countries* 3.08 (2009): 628-632.

Aminov, Rustam I. "The role of antibiotics and antibiotic resistance in nature." *Environmental Microbiology* 11.12 (2009): 2970-2988.

Anderson, Robert N., and Betty L. Smith. "Deaths: leading causes for 2002." *National Vital Statistics Reports* 53.17 (2005): 67-70.

Bäckhed, Fredrik, et al. "Dynamics and Stabilization of the Human Gut Microbiome during the First Year of Life." *Cell Host & Microbe* 17.5 (2015): 690-703.

Baker-Austin, Craig, et al. "Co-selection of antibiotic and metal resistance." *Trends in Microbiology* 14.4 (2006): 176-182.

Ben−David, D., et al. "Outcome of carbapenem resistant *Klebsiella pneumoniae* bloodstream infections." *Clinical Microbiology and Infection* 18.1 (2012): 54-60.

Bhullar, Kirandeep, et al. "Antibiotic resistance is prevalent in an isolated cave microbiome." *PLoS One* 7.4 (2012): e34953.

Blaser, Martin J., and Stanley Falkow. "What are the consequences of the disappearing human microbiota?." *Nature Reviews Microbiology* 7.12 (2009): 887-894.

Blaser, Martin J. "Harnessing the power of the human microbiome." *Proceedings of the National Academy of Sciences* 107.14 (2010): 6125-6126.

Boone, Stephanie A., and Charles P. Gerba. "Significance of fomites in the spread of respiratory and enteric viral disease." *Applied and Environmental Microbiology* 73.6 (2007): 1687-1696.

Bouslimani, Amina, et al. "Molecular cartography of the human skin surface in 3D." *Proceedings of the National Academy of Sciences* 112.17 (2015): E2120-E2129.

Brady, R. R. W., et al. "Is your phone bugged? The incidence of bacteria known to cause nosocomial infection on healthcare workers' mobile phones." J*ournal of Hospital Infection* 62.1 (2006): 123-125.

Brady, R. R. W., et al. "Review of mobile communication devices as potential reservoirs of nosocomial pathogens." *Journal of Hospital Infection* 71.4 (2009): 295-300.

Bright, Kelly R., Stephanie A. Boone, and Charles P. Gerba. "Occurrence of bacteria and viruses on elementary classroom surfaces and the potential role of classroom hygiene in the spread of infectious diseases." *The Journal of School Nursing* 26.1 (2010): 33-41.

Brooks, Brandon, et al. "Microbes in the neonatal intensive care unit resemble those found in the gut of premature infants." *Microbiome* 2.1 (2014): 1.

Brown, Christopher T., et al. "Unusual biology across a group comprising more than 15% of domain Bacteria." *Nature* 523.7559 (2015): 208-211.

Bures, Sergio, et al. "Computer keyboards and faucet handles as reservoirs of nosocomial pathogens in the intensive care unit." *American Journal of Infection Control* 28.6 (2000): 465-471.

Caporaso, J. Gregory, et al. "QIIME allows analysis of high-throughput community sequencing data." *Nature Methods* 7.5 (2010): 335-336.

Caporaso, J. Gregory, et al. "PyNAST: a flexible tool for aligning sequences to a template alignment." *Bioinformatics* 26.2 (2010): 266-267.

Caporaso, J. Gregory, et al. "Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms." *The ISME journal* 6.8 (2012): 1621-1624.

Ciusa, Maria Laura, et al. "A novel resistance mechanism to triclosan that suggests horizontal gene transfer and demonstrates a potential selective pressure for reduced biocide susceptibility in clinical strains of *Staphylococcus aureus*." *International Journal of Antimicrobial Agents* 40.3 (2012): 210-220.

Clemente, Jose C., et al. "The microbiome of uncontacted Amerindians." *Science Advances* 1.3 (2015): e1500183.

Costello, Elizabeth K., et al. "Bacterial community variation in human body habitats across space and time." *Science* 326.5960 (2009): 1694-1697.

Couto, Renato C., et al. "A 10-year prospective surveillance of nosocomial infections in neonatal intensive care units." *American Journal of Infection Control* 35.3 (2007): 183-189.

Custovic, A., S. C. O. Taggart, and A. Woodcock. "House dust mite and cat allergen in different indoor environments." *Clinical & Experimental Allergy* 24.12 (1994): 1164-1168.

David, Lawrence A., et al. "Host lifestyle affects human microbiota on daily timescales." *Genome Biology* 15.7 (2014): R89.

Dawson, P., et al. "Residence time and food contact time effects on transfer of *Salmonella typhimurium* from tile, wood and carpet: testing the five−second rule." *Journal of Applied Microbiology* 102.4 (2007): 945-953.

De Filippo, Carlotta, et al. "Impact of diet in shaping gut microbiota revealed by a comparative study in children from Europe and rural Africa." *Proceedings of the National Academy of Sciences* 107.33 (2010): 14691-14696.

Dedesko, Sandra, et al. "Methods to assess human occupancy and occupant activity in hospital patient rooms." *Building and Environment* 90 (2015): 136-145.

Dekio, Itaru, et al. "Detection of potentially novel bacterial components of the human skin microbiota using culture-independent molecular profiling." *Journal of Medical Microbiology* 54.12 (2005): 1231-1238.

Department of Health and Human Services (2009) Summary of Notifiable Diseases – United States, 2007, Centers for Disease Control and Prevention

Dominguez-Bello, Maria G., et al. "Delivery mode shapes the acquisition and structure of the initial microbiota across multiple body habitats in newborns." *Proceedings of the National Academy of Sciences* 107.26 (2010): 11971-11975.

Edgar, Robert C. "Search and clustering orders of magnitude faster than BLAST." *Bioinformatics* 26.19 (2010): 2460-2461.

Eren, A. Murat, et al. "Oligotyping: differentiating between closely related microbial taxa using 16S rRNA gene data." *Methods in Ecology and Evolution* 4.12 (2013): 1111-1119.

Eren, A. Murat, et al. "Minimum entropy decomposition: unsupervised oligotyping for sensitive partitioning of high-throughput marker gene sequences." *The ISME Journal* (2014).

Escombe, A. Roderick, et al. "Natural ventilation for the prevention of airborne contagion." *PLoS Medicine* 4.2 (2007): e68.

Faith, Daniel P. "Conservation evaluation and phylogenetic diversity." *Biological Conservation* 61.1 (1992): 1-1029.

Feinstein, Larry M., Woo Jun Sul, and Christopher B. Blackwood. "Assessment of bias associated with incomplete extraction of microbial DNA from soil." *Applied and Environmental Microbiology* 75.16 (2009): 5428-5433.

Fierer, Noah, et al. "Forensic identification using skin bacterial communities." *Proceedings of the National Academy of Sciences* 107.14 (2010): 6477-6481.

Flores, Gilberto E., et al. "Microbial biogeography of public restroom surfaces." *PLoS One* 6.11 (2011): e28132.

Flores, Gilberto E., et al. "Diversity, distribution and sources of bacteria in residential kitchens." *Environmental Microbiology* 15.2 (2013): 588-596.

Fluit, Ad C., et al. "Shared reservoir of ccrB gene sequences between coagulase-negative staphylococci and methicillin-resistant *Staphylococcus aureus*." *Journal of Antimicrobial Chemotherapy* (2013): dkt121.

Fujimura, Kei E., et al. "House dust exposure mediates gut microbiome *Lactobacillus* enrichment and airway immune defense against allergens and virus infection." *Proceedings of the National Academy of Sciences* 111.2 (2014): 805-810.

Funkhouser, Lisa J., and Seth R. Bordenstein. "Mom knows best: the universality of maternal microbial transmission." *PLoS Biology* 11.8 (2013): e1001631.

Gao, Zhan, et al. "Molecular analysis of human forearm superficial skin bacterial biota." *Proceedings of the National Academy of Sciences* 104.8 (2007): 2927-2932.

Gibbons, Sean M., et al. "Ecological succession and viability of human-associated microbiota on restroom surfaces." *Applied and Environmental Microbiology* 81.2 (2015): 765-773.

Gilbert, Jack A. "Our unique microbial identity." *Genome Biology* 16.1 (2015): 97.

Grice, Elizabeth A., et al. "A diversity profile of the human skin microbiota." *Genome Research* 18.7 (2008): 1043-1050.

Grice, Elizabeth A., et al. "Topographical and temporal diversity of the human skin microbiome." *Science* 324.5931 (2009): 1190-1192.

Grice, Elizabeth A., and Julia A. Segre. "The skin microbiome." *Nature Reviews Microbiology* 9.4 (2011): 244-253.

Gillings, Michael, et al. "The evolution of class 1 integrons and the rise of antibiotic resistance." *Journal of Bacteriology* 190.14 (2008): 5095-5100.

Gillings, Michael R., and H. W. Stokes. "Are humans increasing bacterial evolvability?." *Trends in Ecology & Evolution* 27.6 (2012): 346-352.

Groer, Maureen W., et al. "Development of the preterm infant gut microbiome: a research priority." *Microbiome* 2.1 (2014): 38.

Groseclose, S. L., et al. "Summary of notifiable diseases--United States, 2002." MMWR. *Morbidity and Mortality Weekly Report* 51.53 (2004): 1.

Guindon, Stéphane, et al. "New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0." *Systematic Biology* 59.3 (2010): 307-321.

Harris, Simon R., et al. "Evolution of MRSA during hospital transmission and intercontinental spread." *Science* 327.5964 (2010): 469-474.

Hartfield, Matthew, Carmen Lía Murall, and Samuel Alizon. "Clinical applications of pathogen phylogenies." *Trends in Molecular Medicine* 20.7 (2014): 394-404.

He, Miao, et al. "Emergence and global spread of epidemic healthcare-associated *Clostridium difficile*." *Nature Genetics* 45.1 (2013): 109-113.

Hewitt, Krissi M., et al. "Bacterial diversity in two neonatal intensive care units (NICUs)." *PloS One* 8.1 (2013): e54703.

Hooke, Roger LeB, José F. Martín-Duque, and Javier Pedraza. "Land transformation by humans: a review." *GSA Today* 22.12 (2012): 4-10.

Höppe, Peter, and Ivo Martinac. "Indoor climate and air quality." *International Journal of Biometeorology* 42.1 (1998): 1-7.

Hyvärinen, Anne, et al. "Fungi and actinobacteria in moisture-damaged building materials—concentrations and diversity." *International Biodeterioration & Biodegradation* 49.1 (2002): 27-37.

Jaffal, A. A., et al. "Residential indoor airborne microbial populations in the United Arab Emirates." *Environment International* 23.4 (1997): 529-533.

Kawamoto, Shimpei, et al. "Foxp3+ T cells regulate immunoglobulin a selection and facilitate diversification of bacterial species responsible for immune homeostasis." *Immunity* 41.1 (2014): 152-165.

Kelley, Scott T., and Jack A. Gilbert. "Studying the microbiology of the indoor environment." *Genome Biology* 14.2 (2013): 202.

Kembel, Steven W., et al. "Architectural design drives the biogeography of indoor bacterial communities." *PloS One* 9.1 (2014): e87093.

Kitzes, Justin, et al. "Current methods for calculating national ecological footprint accounts." *Science for Environment & Sustainable Society* 4.1 (2007): 1-9.

Klevens, R. Monina, et al. "Estimating health care-associated infections and deaths in US hospitals, 2002." *Public Health Reports* 122.2 (2007): 160.

Knights, Dan, Elizabeth K. Costello, and Rob Knight. "Supervised classification of human microbiota." *FEMS Microbiology Reviews* 35.2 (2011A): 343-359.

Knights, Dan, et al. "Human-associated microbial signatures: examining their predictive value." *Cell Host & Microbe* 10.4 (2011B): 292-296.

Knights, Dan, et al. "Bayesian community-wide culture-independent microbial source tracking." *Nature Methods* 8.9 (2011C): 761-763.

Kramer, Axel, Ingeborg Schwebke, and Günter Kampf. "How long do nosocomial pathogens persist on inanimate surfaces? A systematic review." *BMC Infectious Diseases* 6.1 (2006): 130.

La Duc, Myron T., et al. "Pyrosequencing-derived bacterial, archaeal, and fungal diversity of spacecraft hardware destined for Mars." *Applied and Environmental Microbiology* (2012): AEM-01435.

La Rosa, Patricio S., et al. "Patterned progression of bacterial populations in the premature infant gut." *Proceedings of the National Academy of Sciences* 111.34 (2014): 12522-12527.

Lax, Simon, et al. "Longitudinal analysis of microbial interaction between humans and the indoor environment." *Science* 345.6200 (2014): 1048-1052.

Lax, Simon, et al. "Forensic analysis of the microbiome of phones and shoes." *Microbiome* 3.1 (2015): 21.

Lax, Simon, Cathryn R. Nagler, and Jack A. Gilbert. "Our interface with the built environment: immunity and the indoor microbiota." *Trends in immunology* 36.3 (2015): 121-123.

Lax, Simon, and Jack A. Gilbert. "Hospital-associated microbiota and implications for nosocomial infections." *Trends in Molecular Medicine* (2015).

Lignell, U., et al. "Evaluation of quantitative PCR and culture methods for detection of house dust fungi and streptomycetes in relation to moisture damage of the house." *Letters in Applied Microbiology* 47.4 (2008): 303-308.

Lopez, Pedro-Jose, et al. "Bacterial counts from hospital doctors' ties are higher than those from shirts." *American Journal of Infection Control* 37.1 (2009): 79-80.

Lozupone, Catherine, and Rob Knight. "UniFrac: a new phylogenetic method for comparing microbial communities." *Applied and Environmental Microbiology* 71.12 (2005): 8228-8235.

Luef, Birgit, et al. "Diverse uncultivated ultra-small bacterial cells in groundwater." *Nature Communications* 6 (2015).

Magill, Shelley S., et al. "Multistate point-prevalence survey of health care–associated infections." *New England Journal of Medicine* 370.13 (2014): 1198-1208.

Marinella, Mark A., Carl Pierson, and Carol Chenoweth. "The stethoscope: a potential source of nosocomial infection?." *Archives of Internal Medicine* 157.7 (1997): 786-790.

Martínez, José L. "Antibiotics and antibiotic resistance genes in natural environments." *Science* 321.5887 (2008): 365-367.

McAdam, Paul R., et al. "Adaptive evolution of *Staphylococcus aureus* during chronic endobronchial infection of a cystic fibrosis patient." *PLoS One* 6.9 (2011): e24301.

McAdam, Paul R., et al. "Molecular tracing of the emergence, adaptation, and transmission of hospital-associated methicillin-resistant *Staphylococcus aureus*." *Proceedings of the National Academy of Sciences* 109.23 (2012): 9107-9112.

Meadow, James F., et al. "Significant changes in the skin microbiome mediated by the sport of roller derby." *PeerJ* 1 (2013): e53.

Meadow, James F., et al. "Bacterial communities on classroom surfaces vary with human contact." *Microbiome* 2.7 (2014): 1-7.

Meadow, James F., et al. "Indoor airborne bacterial communities are influenced by ventilation, occupancy, and outdoor air source." *Indoor Air* 24.1 (2014): 41-48.

Meadow, James F., Adam E. Altrichter, and Jessica L. Green. "Mobile phones carry the personal microbiome of their owners." *PeerJ* 2 (2014): e447.

Meisel, Jacquelyn S., et al. "Skin microbiome surveys are strongly influenced by experimental design." *Journal of Investigative Dermatology* 136.5 (2016): 947-956.

Metcalf, Jessica L., et al. "A microbial clock provides an accurate estimate of the postmortem interval in a mouse model system." *Elife* 2 (2013): e01104.

Morowitz, Michael J., et al. "Redefining the role of intestinal microbes in the pathogenesis of necrotizing enterocolitis." *Pediatrics* 125.4 (2010): 777-785.

Morowitz, Michael J., et al. "Strain-resolved community genomic analysis of gut microbial colonization in a premature infant." *Proceedings of the National Academy of Sciences* 108.3 (2011): 1128-1133.

Moschandreas, D. J., K. R. Pagilla, and L. V. Storino. "Time and space uniformity of indoor bacteria concentrations in Chicago area residences." *Aerosol Science & Technology* 37.11 (2003): 899-906.

Nevalainen, Aino, and Markku Seuri. "Of microbes and men." *Indoor Air* 15.s9 (2005): 58-64.

Olivas, Andrea D., et al. "Intestinal tissues induce an SNP mutation in Pseudomonas aeruginosa that enhances its virulence: possible role in anastomotic leak." *PloS One* 7.8 (2012): e44326.

Oh, Julia, et al. "Temporal Stability of the Human Skin Microbiome." *Cell* 165.4 (2016): 854-866.

Otto, Michael. "Coagulase−negative staphylococci as reservoirs of genes facilitating MRSA infection." *Bioessays* 35.1 (2013): 4-11.

Palmer, Chana, et al. "Development of the human infant intestinal microbiota." *PLoS Biology* 5.7 (2007): e177.

Pearson, A. "Historical and changing epidemiology of healthcare-associated infections." *Journal of Hospital Infection* 73.4 (2009): 296-304.

Pehrsson, Erica C., et al. "Interconnected microbiomes and resistomes in low-income human habitats." *Nature* 533 (2016): 212-216.

Poza, Margarita, et al. "Exploring bacterial diversity in hospital environments by GS-FLX Titanium pyrosequencing." *PloS One* 7.8 (2012): e44105.

Quick, Joshua, et al. "Rapid draft sequencing and real-time nanopore sequencing in a hospital outbreak of *Salmonella*." *Genome Biology* (2015)

Ramos, Tiffanie, and Brent Stephens. "Tools to improve built environment data collection for indoor microbial ecology investigations." *Building and Environment* 81 (2014): 243-257.

Ramos, Tiffanie, et al. "Spatial and temporal variations in indoor environmental conditions, human occupancy, and operational characteristics in a new hospital building." *PLoS One* 10.3 (2015): e0118207.

Ruimy, Raymond, et al. "Nasal carriage of Methicillin-Resistant Coagulase-Negative Staphylococcocci: A Reservoir of mecA Gene for *Staphylococcus aureus*." *New Frontiers of Molecular Epidemiology of Infectious Diseases*. Springer Netherlands, 2012. 219-237.

Sebaihia, Mohammed, et al. "The multidrug-resistant human pathogen *Clostridium difficile* has a highly mobile, mosaic genome." *Nature Genetics* 38.7 (2006): 779-786.

Sippola, Mark R., Richard G. Sextro, and Tracy L. Thatcher. "Measurements and modeling of deposited particle transport by foot traffic indoors." *Environmental Science & Technology* 48.7 (2014): 3800-3807.

Skippington, Elizabeth, and Mark A. Ragan. "Lateral genetic transfer and the construction of genetic exchange communities." *FEMS Microbiology Reviews* 35.5 (2011): 707-735.

Snitkin, Evan S., et al. "Tracking a hospital outbreak of carbapenem-resistant *Klebsiella pneumoniae* with whole-genome sequencing." *Science Translational Medicine* 4.148 (2012): 148ra116-148ra116.

Song, Se Jin, et al. "Cohabiting family members share microbiota with one another and with their dogs." *ELife* 2 (2013): e00458.

Spanu, Teresa, et al. "In vivo emergence of tigecycline resistance in multidrug-resistant *Klebsiella pneumoniae* and *Escherichia coli*." *Antimicrobial Agents and Chemotherapy* 56.8 (2012): 4516-4518.

Stefka, Andrew T., et al. "Commensal bacteria protect against food allergen sensitization." *Proceedings of the National Academy of Sciences* 111.36 (2014): 13145-13150.

Stoll, Barbara J., et al. "Neurodevelopmental and growth impairment among extremely low-birth-weight infants with neonatal infection." *JAMA* 292.19 (2004): 2357-2365.

Summary of Notifiable Diseases--United States, 2007. Department of Health and Human Services, Centers for Disease Control and Prevention, 2009.

Treakle, Amy M., et al. "Bacterial contamination of health care workers' white coats." *American Journal of Infection Control* 37.2 (2009): 101-105.

Tringe, Susannah G., et al. "The airborne metagenome in an indoor urban environment." *PloS One* 3.4 (2008): e1862.

Trosvik, Pål, Nils Christian Stenseth, and Knut Rudi. "Convergent temporal dynamics of the human infant gut microbiota." *The ISME Journal* 4.2 (2010): 151-158.

Turnbaugh, Peter J., et al. "Organismal, genetic, and transcriptional variation in the deeply sequenced gut microbiomes of identical twins." *Proceedings of the National Academy of Sciences* 107.16 (2010): 7503-7508.

Tyakht, Alexander V., et al. "Human gut microbiota community structures in urban and rural populations in Russia." *Nature Communications* 4 (2013).

Ulger, Fatma, et al. "Are we aware how contaminated our mobile phones with nosocomial pathogens?." *Annals of Clinical Microbiology and Antimicrobials* 8.1 (2009): 7.

Wang, Yunwei, et al. "16S rRNA gene-based analysis of fecal microbiota from preterm infants with and without necrotizing enterocolitis." *The ISME Journal* 3.8 (2009): 944-954.

Wesolowska-Andersen, Agata, et al. "Choice of bacterial DNA extraction method from fecal material influences community structure as evaluated by metagenomic analysis." *Microbiome* 2.1 (2014): 19.

Westwood, Jack, et al. "The hospital microbiome project: meeting report for the UK science and innovation network UK-USA workshop 'beating the superbugs: hospital microbiome studies for tackling antimicrobial resistance', October 14th 2013." *Standards in Genomic Sciences* 9 (2014): 12.

Wiener-Well, Yonit, et al. "Nursing and physician attire as possible source of nosocomial infections." *American Journal of Infection Control* 39.7 (2011): 555-559.

Wood, Mariah, et al. "Athletic equipment microbiota are shaped by interactions with human skin."*Microbiome* 3:25 (2015).

Zaborin, Alexander, et al. "Membership and behavior of ultra-low-diversity pathogen communities present in the gut of humans during prolonged critical illness." *MBio* 5.5 (2014): e01361-14.