

MICROBIOTA

Bacterial colonization and succession in a newly opened hospital

Simon Lax,^{1,2} Naseer Sangwan,^{2,3} Daniel Smith,⁴ Peter Larsen,² Kim M. Handley,^{1*} Miles Richardson,¹ Kristina Guyton,³ Monika Krezalek,³ Benjamin D. Shogan,^{3†} Jennifer Defazio,³ Irma Flemming,^{3‡} Baddr Shakhsheer,³ Stephen Weber,⁵ Emily Landon,⁵ Sylvia Garcia-Houchins,⁵ Jeffrey Siegel,^{6,7} John Alverdy,³ Rob Knight,^{8,9} Brent Stephens,¹⁰ Jack A. Gilbert^{1,2,3§}

2017 © The Authors, some rights reserved; exclusive licensee American Association for the Advancement of Science.

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. Metagenomic analyses revealed that genes conferring antimicrobial resistance were consistently more abundant on room surfaces than on the skin of the patients inhabiting those rooms. In addition, persistent unique genotypes of *Staphylococcus* and *Propionibacterium* were identified. Dynamic Bayesian network analysis suggested that hospital staff were more likely to be a source of bacteria on the skin of patients than the reverse but that there were no universal patterns of transmission across patient rooms.

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 (1–4), 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 (5–8). A potential link between hospital-associated microbial communities and hospital-acquired infections, a leading cause of patient death (9–11), needs further investigation (1). In particular, an understanding of bacterial community structure in hospital environments will be critical for mapping the dissemination of antimicrobial resistance genes (12). 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 (13, 14) and around the globe (15, 16). 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 (17–19), no detailed longitudinal, culture-independent analyses of the hospital microbiota have yet been undertaken (1). Additionally, although a number of longitudinal studies have clarified the temporal dynamics of the human skin microbiota (2, 20), 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 (table S1 and fig. S1) 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. One patient room on each floor was sampled daily, whereas all other environments were 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,

¹Department of Ecology and Evolution, University of Chicago, 1101 East 57th Street, Chicago, IL 60637, USA. ²Division of Biosciences, Argonne National Laboratory, 9700 South Cass Avenue, Argonne, IL 60439, USA. ³Microbiome Center, Department of Surgery, University of Chicago, A27 South Maryland Avenue, Chicago, IL 60637, USA. ⁴Alkek Center for Metagenomics and Microbiome Research, Department of Molecular Virology and Microbiology, Baylor College of Medicine, Houston, TX 77030, USA. ⁵Department of Medicine, University of Chicago, 5841 South Maryland Avenue, Chicago, IL 60637, USA. ⁶Department of Civil Engineering, University of Toronto, 35 St. George Street, Toronto, Ontario M5S 1A4, Canada. ⁷Dalla Lana School of Public Health, University of Toronto, 223 College Street, Toronto, Ontario M5T 1R4, Canada. ⁸Department of Pediatrics, University of California, San Diego, San Diego, CA 92037, USA. ⁹Department of Computer Science and Engineering, University of California, San Diego, San Diego, CA 92037, USA. ¹⁰Department of Civil, Architectural and Environmental Engineering, Illinois Institute of Technology, 3201 South Dearborn Street, Chicago, IL 60616, USA.

*Present address: School of Biological Sciences, University of Auckland, 3A Symonds Street, Auckland 1010, New Zealand.

†Present address: Division of Colon and Rectal Surgery, Mayo Clinic, Rochester, MN 55905, USA.

‡Present address: Division of Trauma and Critical Care, Vanderbilt University, Nashville, TN 37212, USA.

§Corresponding author. Email: gilbertjack@gmail.com

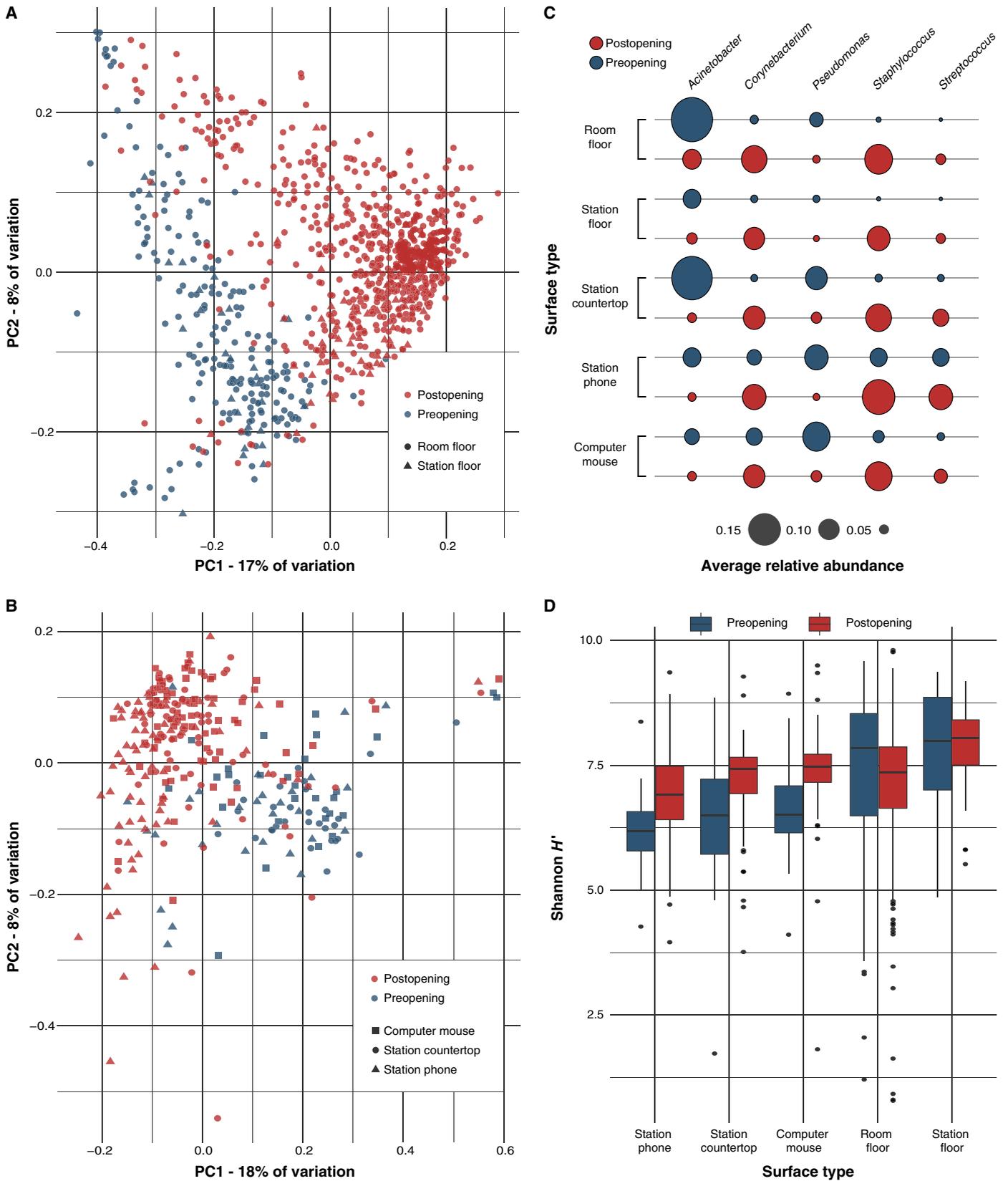


Fig. 1. Change in microbial community structure after hospital opening. (A) Principal coordinate analysis (PCoA) of all floor samples based on weighted UniFrac distance and colored by whether they were taken before or after the hospital's opening. (B) PCoA of three nurse station surfaces colored as in (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.

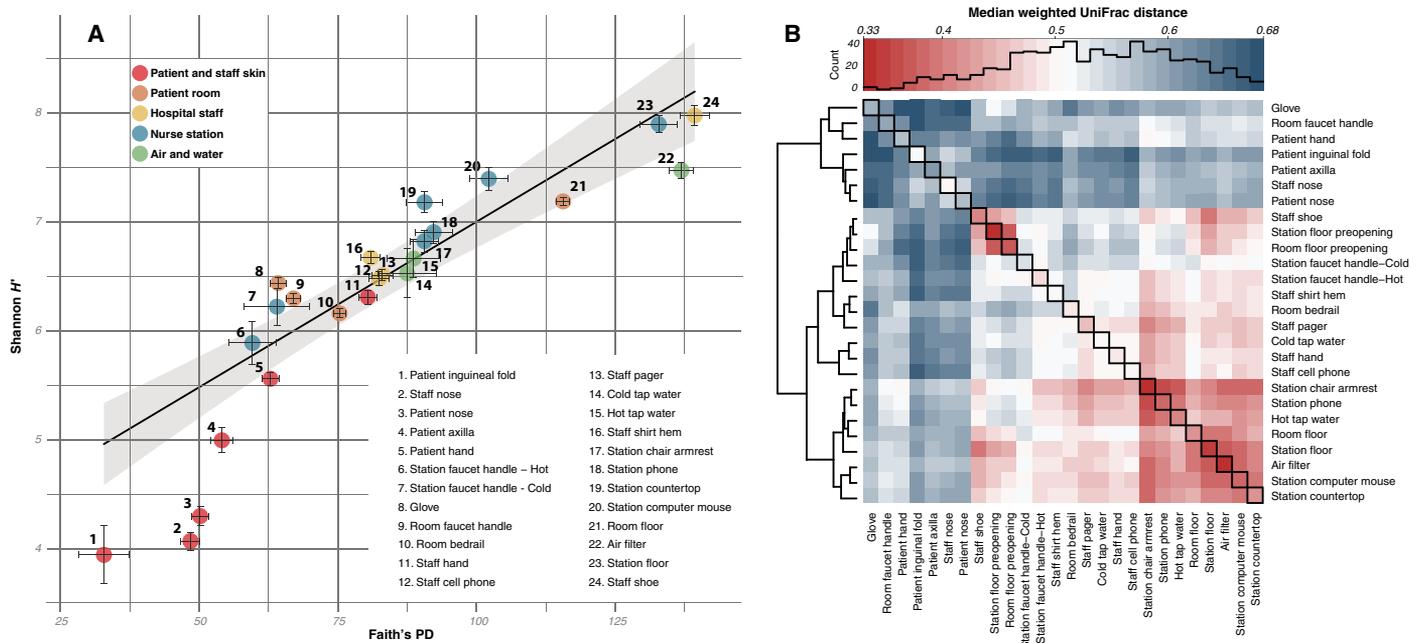


Fig. 2. α - and β -diversity of hospital sample types. (A) Average α -diversity of sample types based on Faith's phylogenetic diversity (x axis) and the Shannon diversity index (y axis). Bars indicate SEM. (B) Heat map of β -diversity relationships between sample types based on the median weighted UniFrac 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.

CO₂ concentrations, and infrared doorway beam breaks, were continuously monitored (21, 22). At least 5000 high-quality 16S ribosomal RNA (rRNA) V4 amplicons were generated per sample using the protocols outlined by the Earth Microbiome Project.

RESULTS

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 samples (Fig. 1, A to C). Network analysis revealed an almost complete shift in operational taxonomic unit (OTU)-level composition well beyond these dominant genera (fig. S2). 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 ($P < 0.001$; Fig. 1D) 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 (fig. S3). These included genera such as *Enterococcus*, other unclassifiable members of the family *Enterobacteriaceae*, *Finexgordia*, *Rothia*, *Prevotella*, and *Sphingomonas*, although the 16S 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 (23).

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. Rarefaction curves demonstrated that diversity calculations converged at sampling depths well below the 5000 reads per sample analyzed in this study (fig. S4). 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) (21), were the most diverse (Fig. 2A). 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.

β -Diversity patterns grouped our sample types into three sets (Fig. 2B). 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 (Fig. 2B), 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

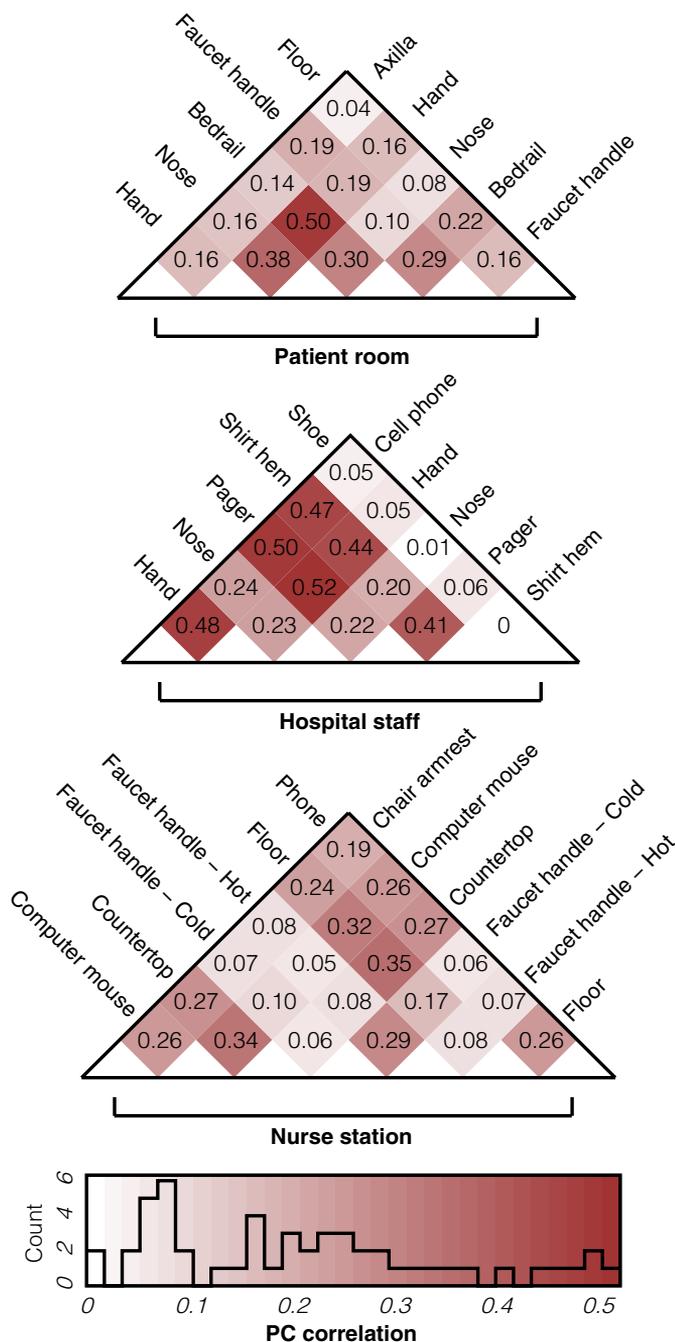


Fig. 3. Heat map of principal coordinate space correlations between sample types. Values represent the average correlation between samples taken from the same location and date along the first 10 axes (eigenvectors) of the PCoA plot of all samples, weighted by the variance captured by each axis' eigenvalues.

a greater degree of similarity to all surfaces, which could be explained by higher foot traffic after opening (Fig. 2B).

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 (Fig. 3). Patient hands and bedrails had a strong interaction ($\rho = 0.5$), 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 ($\rho = 0.48$ and 0.52 , respectively), which mirrors the findings in previous studies (24, 25). Correlations within the nurse station environment were significant but comparatively weak, likely due to the diversity of people using these environments.

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 (Fig. 4A). The average percent of 16S 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 (Fig. 4A). We used a Bayesian source-tracking approach (26) 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 (fig. S5A). 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 (fig. S5A). 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.

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 (Fig. 4B and fig. S5B). The strength and significance of the correlation varied between comparisons but were strongest between patients' hands and their room floor ($\rho = -0.39, P = 0.002$). Patients' hands and axillae became significantly more similar over the course of their hospital stay ($\rho = -0.29, P = 0.015$), perhaps because of reduced microbial exposure and homogenization resulting from bed confinement.

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 (Fig. 4C). 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 (Fig. 4D). 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 (Fig. 4D). 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.

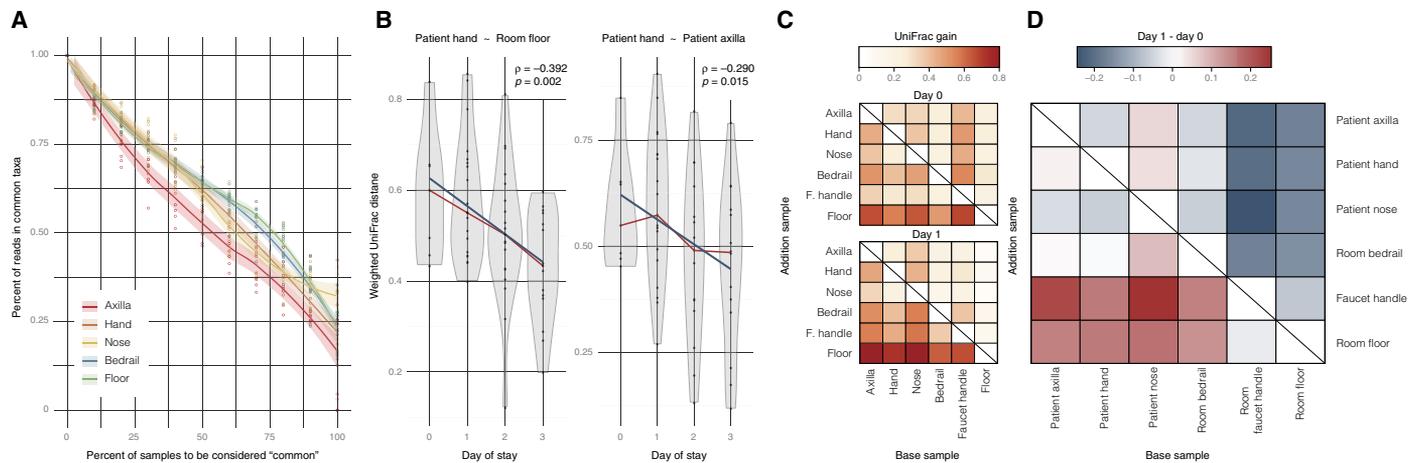


Fig. 4. Interaction between patient skin and hospital room microbiota. (A) Scatter plot of the percent of 16S reads in the “common microbiome” for the eight rooms sampled weekly, with the common microbiome definition on the x axis and the percent of reads in that set on the y axis. Points represent the 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.

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. A full summary of these factors is available in table S2, along with the correlations between these factors.

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 constrained). The effect of these factors was more evident in permutational multivariate analysis of variance (PERMANOVA) analyses of β -diversity (Table 1A), with all seven binary factors significantly influencing the structure of hand microbial communities. All factors, except antibiotics taken before 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 1A). 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 (table S3) 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 ($\rho = 0.47$), 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) (27) of patient skin and bedrail samples (Table 1B). The strongest observed effect was significantly lower diversity for all four sample types in patients who were admitted through the emergency department (Table 1B). 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 (Table 1B), 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 (table S4).

Table 1. 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.

	Hand samples		Axilla samples		Nose samples		Bedrail samples	
	Pseudo-F	P	Pseudo-F	P	Pseudo-F	P	Pseudo-F	P
Antibiotics preadmission	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
Chemotherapy 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 postsurgery	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*

Factor	Surface	n		Mean		NP t-test P
		yes	no	yes	no	
Antibiotics preadmission	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
Chemotherapy 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 postsurgery	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 room	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

Antibiotics had no consistently observed effect on the α -diversity of the microbiota of patient skin (Fig. 5A) or in ordination clustering of patient skin microbiota samples (Fig. 5B), even when controlling for length of exposure and the route of delivery. 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 (neospo- rin in all cases) saw decreases in skin microbial diversity after use of the antibiotic.

Effect of environmental conditions

As previously reported (21), 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 g_w/kg_{da} with SD 1.23 g_w/kg_{da}, and illuminance had mean 173 lx with SD 448 lx (illuminance includes both artificial and natural light) (21). 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 (Fig. 6A). 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 (Fig. 6B). Greater

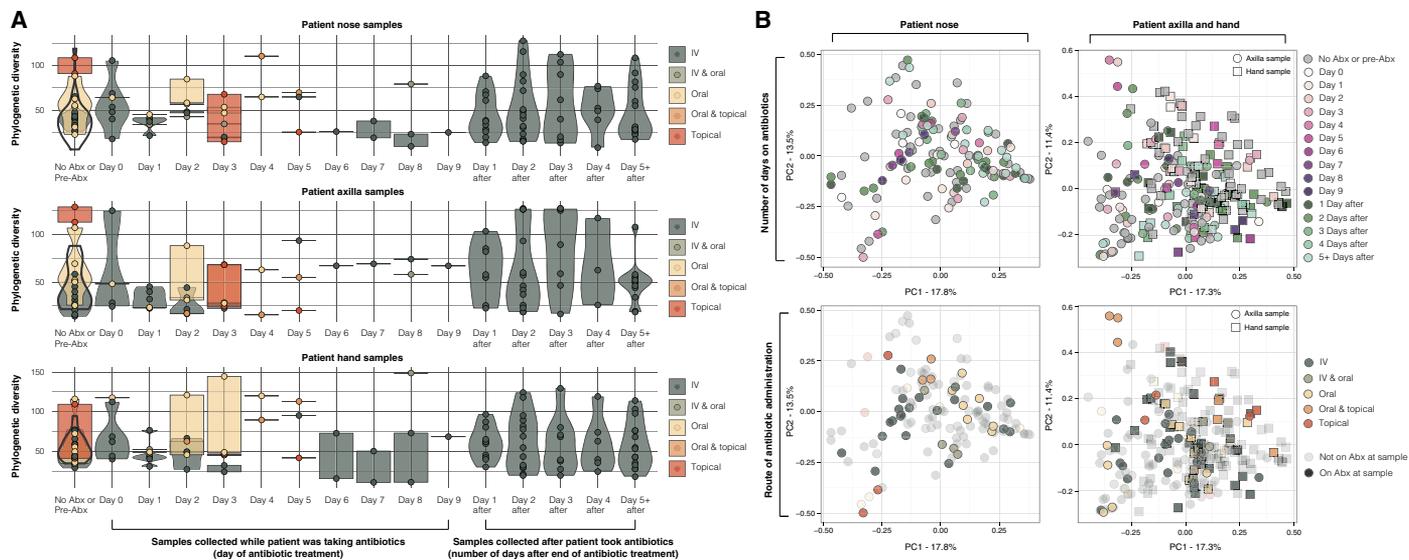


Fig. 5. Effect of antibiotic treatment on the diversity of patient skin samples. (A) Effect of antibiotic (Abx) treatment on the phylogenetic α -diversity of patient skin samples. Antibiotic status is indicated on the x axis, and violin plots indicate the density of phylogenetic diversity for corresponding 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.

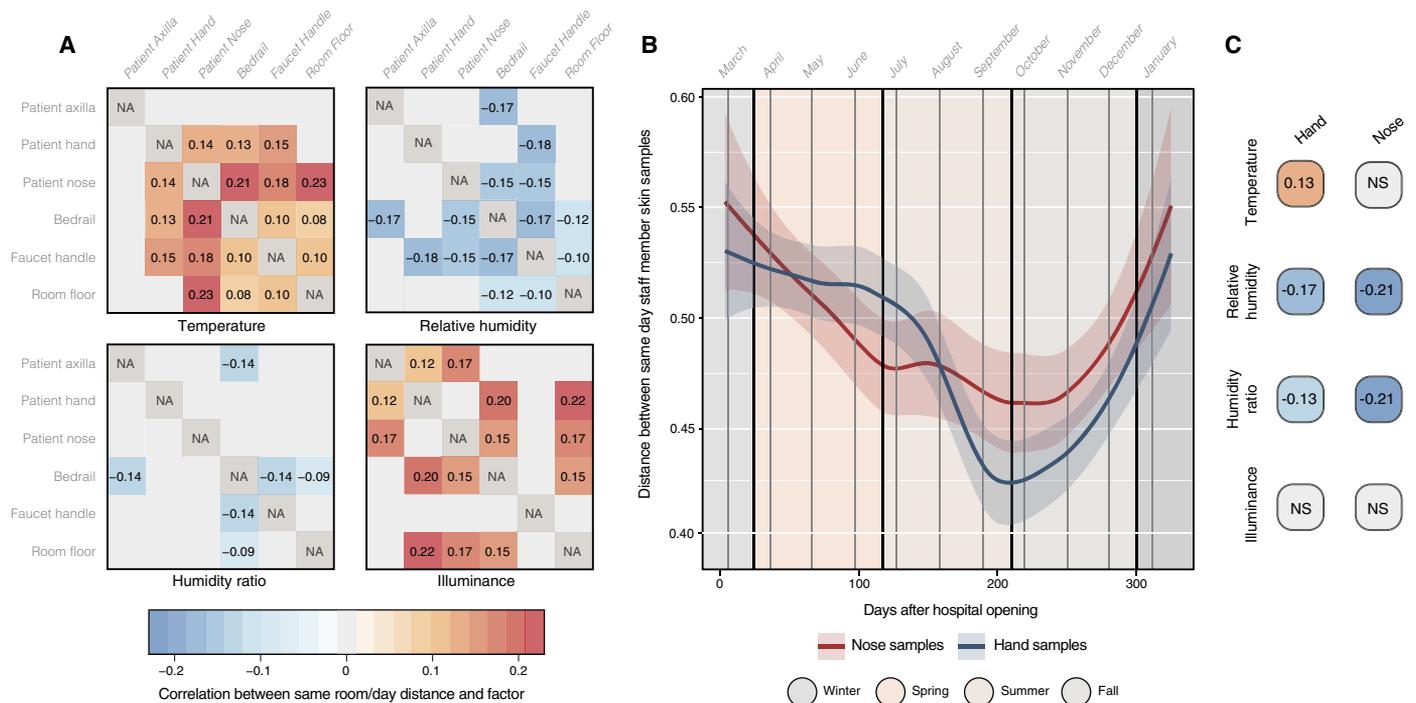


Fig. 6. Effect of environmental factors on microbial transmission. (A) Heat maps of the correlation between environmental factors and the weighted UniFrac distance between samples taken from the same room on the same day. (B) Seasonal changes in the UniFrac distances between the hand and nose samples 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).

nose- and hand-associated microbial similarity among different staff members correlated with higher humidity; hand-associated microbial similarity also correlated with lower temperatures (Fig. 6C). Hand, nose, and floor samples taken in the same week were generally more similar to each other than to samples taken during a different week (fig. S6).

Network analyses

To gain further insight into the bacterial diversity of our samples, we ran oligotyping (28) on reads assigned to four dominant genera: *Staphylococcus*, *Streptococcus*, *Corynebacterium*, and *Acinetobacter*. After filtering for oligotypes with >500 reads, these genera together comprised 303 unique

clusters, with 116 *Staphylococcus*, 83 *Streptococcus*, 77 *Corynebacterium*, and 27 *Acinetobacter* oligotypes (fig. S7). When we applied Bayesian source tracking to the two patient rooms that were sampled daily, using the oligotype-level profiles on hand or axilla samples from each patient as a source, it was often possible to accurately predict the correct patient ID based on the microbial profiles on the bedrails and, to a lesser extent, on floors and patient room faucet handles (figs. S8 and S9). This suggested that individual patients can harbor a unique strain-level skin microbial signature that might be obscured through traditional OTU clustering, although the fact that many oligotypes were found in a large number of samples (fig. S7) suggests a certain degree of ubiquity that limits predictive accuracy.

We used dynamic Bayesian network (DBN) analysis to infer patterns of oligotype interaction and transmission in the two rooms sampled daily. A DBN is a directed acyclic network of conditional dependencies in a set of random variables using probabilistic relationships. This conditional dependency permits the opportunity to predict the directionality of interactions between oligotypes. The networks generated for these two rooms, which both contained ~1000 edges, did not significantly resemble one another based on permutation analysis, suggesting an absence of universal transmission patterns across rooms. However, staff skin was more likely to be a source of oligotypes than a sink, whereas bedrails were more likely to be a source of oligotypes to patient skin than vice versa.

Metagenomic analyses

We selected 92 samples for metagenomic sequencing and pooled samples by type (skin or surface) and room day. Samples were chosen so that we could focus on different patients staying in the same hospital room over the course of multiple months. Apart from *Propionibacterium*, all major genera showed highly similar relative abundances when compared to the 16S rRNA data. Antibiotic resistance gene (ARG) abundance on a given day was almost always greater for room surface samples compared to skin samples (fig. S10A). However, tetracycline resistance genes were significantly more abundant in skin samples ($5.99 \pm 6.7\%$) compared to hospital surfaces ($2.099 \pm 3.6\%$; two-group *t* test, $P < 0.05$; Bonferroni correction), whereas multidrug efflux proteins were more abundant in surface samples compared to skin samples ($4.99 \pm 2.1\%$ and $9.099 \pm 7.6\%$, respectively) (fig. S10B). Taxonomic binning of ARG contigs assembled across skin samples revealed high abundance of *Staphylococcus aureus* ($29.4 \pm 11.6\%$), *Staphylococcus epidermidis* ($12.1 \pm 3.6\%$), and *Corynebacterium striatum* ($5.06 \pm 6.2\%$) genotypes. *Escherichia coli* ($11.2 \pm 6.3\%$) and *Pseudomonas aeruginosa* ($5.6 \pm 6.7\%$) showed high antibiotic resistance potential across surface samples (two-group *t* test, $P < 0.05$; Bonferroni-corrected).

Metagenomic contigs were assigned to 65 genome bins (table S5). Within a single patient room, we observed separate genotypes of *S. epidermidis*, *Propionibacterium acnes*, *Anaerococcus* sp., and *Corynebacterium* sp. that were >97% identical (gene complement) but sampled >71 days apart (99 and 170 days after hospital opening). This suggested either ubiquitous skin-associated microbial strains seeded by sequential room occupant or staff or the presence of persistent bacteria in the environment despite stringent room cleaning protocols. Genotypes assembled from day 170 had a consistently greater number of ARGs compared to those assembled on day 99 (Resfam; *S. epidermidis*: day 99 = 5 and day 170 = 9; *Anaerococcus*: day 99 = 4 and day 170 = 7). Genotypes with >99% average nucleotide identity were found on patient skin and room surfaces on the same day, and some of these genotypes were also persistent. For example, genotypes of *P. acnes* and *S. epidermidis* were found in both

skin and surface samples on days 97, 99, and 170 but showed correlations between synonymous versus nonsynonymous substitution ratio and codon usage bias (fig. S11), suggestive of continual selection pressure (29). For example, *S. epidermidis* genotypes were positively selected for different “invasion-specific” proteins (cell surface protein, SE2251; acetate-coenzyme A ligase, SE2161; and conserved hypothetical protein, SE0692) on different days (table S5) (30).

DISCUSSION

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 16S 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.

MATERIALS AND 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 (21, 22), 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 μ l) was added, vortexed for 5 s, and incubated at 90°C for 10 min. Samples were centrifuged at 2500g for 1 min. Extract-N-Amp Dilution solution (200 μ l) 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 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 12-base barcode sequence that supports pooling of up to 2167 different samples in each lane (31). Each 20 μ l of polymerase chain reaction (PCR) contains 5 μ l of Mo Bio PCR Water (Certified DNA-Free), 10 μ l of Extract-N-Amp Ready Mix, 1 μ l of forward primer (5 μ M concentration, 200 pM final), 1 μ l of Golay Barcode Tagged Reverse Primer (5 μ M concentration, 200 pM final), and 4 μ l 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) \times 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 (32). 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 (33) and FastTree (34). 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 (28) to identify sub-OTU level variation in four highly abundant genera: *Acinetobacter*, *Corynebacterium*, *Streptococcus*, and *Staphylococcus*. USEARCH (35) 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 (36) 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 (example calculation in fig. S12):

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

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 (fig. S13). 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 (21). 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.

Bayesian networks

DBNs were generated using BANJO (v2.2) with the following parameters: a maximum of five parents, a minimum Markov lag of 0, a maximum Markov lag of 1, discretization into five intervals, greedy searching, and all local moves. To reduce complexity, all oligotypes representing less than 0.05% of the total population were removed from analysis, resulting in data sets of 76 and 73 oligotypes for the two rooms sampled daily. All possible surfaces and oligotypes were allowed to interact with each other, and empty entries were replaced by average values from the nearest adjacent measurements. Data was \log_2 -transformed and then normalized within oligotypes across all samples.

Metagenomic analyses

A total of 92 samples were selected for shotgun metagenomic sequencing. Libraries were generated using 1 ng of genomic DNA and the Nextera XT protocol according to the manufacturer's instructions (Illumina). Sequencing was performed on the Illumina HiSeq platform (150 bp \times 2, six samples per lane; insert size range of 300 to 1200 \pm 100 bp). Raw metagenome reads were quality-trimmed using the nsoni pipeline (<http://vicbioinformatics.com/nesoni.shtml>) with the following parameters: minimum length, 75; quality cutoff, 30; adapter trimming, yes; and ambiguous bases, 0. Taxonomic information was assigned to the individual metagenome reads using MetaPhlAn (37). Quality-trimmed metagenome reads were assembled into contigs using IDBA_UD (38) using a k -mer length ranging from 31 to 41. Metagenome contigs with lengths of <300 bp were excluded from further analysis. To understand the strain-level population level dynamics of in situ genotypes across the surface and skin environments, we focused our further assembly efforts on binning population genomes using MetaBAT (39). Single-copy marker gene-based copy number variation analysis (40) was used to estimate the percentage completion and intraspecies contamination across each assembled genome. De novo assembled genotypes (same species with >80% completion) were grouped into strain-level bins using a whole genome-level average nucleotide identity cutoff of 99%. Genotypes assigned to *S. epidermidis*, *P. acnes*, *Corynebacterium*, and *Peptoniphilus* taxons were finalized for interhabitat (skin versus surface) strain-level evolutionary analysis. Using the reciprocal smallest distance algorithm (41), orthologous genes were identified across skin and surface genotypes (same strains assembled across a patient room on the same day). Pairwise selected orthologous protein coding genes were aligned using ClustalW (42). Multiple codon alignments were constructed from the corresponding aligned protein sequences using the pal2nal script (43). Final alignments (with stop codons removed) were processed for d_N/d_S analysis using PAML (44). To further validate the influence of in situ functional constraints on the observed selection patterns, we processed the orthologous gene pairs using codon bias variation. Using methods explained by Zhang *et al.* (45), codon deviation coefficient (CDC) was used as the measure of codon bias across orthologous gene pairs predicted across genotype. Mean value (CDC) of two orthologous genes was used for the correlation analysis against d_N/d_S values. Reconstructed genomes were submitted to RAST database (46) for automated genome annotation. ARGs were identified using Resfams (47).

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 10,000 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.

SUPPLEMENTARY MATERIALS

www.sciencetranslationalmedicine.org/cgi/content/full/9/391/eaah6500/DC1

- Fig. S1. Floor plan of sampling locations.
 Fig. S2. Bipartite OTU network of floor samples.
 Fig. S3. Relative abundances of common genera by surface type.
 Fig. S4. Rarefaction curves demonstrate convergence of α -diversity calculations.
 Fig. S5. Interaction between patient skin and room samples.
 Fig. S6. Differences in community dissimilarity between intra- and interweek samples.
 Fig. S7. Overview of oligotyping data.
 Fig. S8. Predictive accuracy of SourceTracker models using hand samples as source.
 Fig. S9. Predictive accuracy of SourceTracker models using axilla samples as source.
 Fig. S10. Patterns of antimicrobial resistance gene abundance.
 Fig. S11. Coupling between selection and codon usage bias shows differential impact of in situ functional constraints on the same strain of *P. acnes*.
 Fig. S12. Example PC space correlation calculation.
 Fig. S13. PCoA of all samples.
 Table S1. Summary of the 6523 samples included in the study, grouped by surface type.
 Table S2. Summary of clinical metadata for 49 patients sampled on multiple days.
 Table S3. Predictive accuracy of random forest supervised learning models predicting seven binary clinical factors.
 Table S4. Effect of binary clinical factors on the similarity between sample types taken from the same room on the same day.
 Table S5. Summary of the 65 genome bins assembled from metagenome contigs.

REFERENCES AND NOTES

1. S. Lax, J. A. Gilbert, Hospital-associated microbiota and implications for nosocomial infections. *Trends Mol. Med.* **21**, 427–432 (2015).
2. S. Lax, D. P. Smith, J. Hampton-Marcell, S. M. Owens, K. M. Handley, N. M. Scott, S. M. Gibbons, P. Larsen, B. D. Shogan, S. Weiss, J. L. Metcalf, L. K. Ursell, Y. Vázquez-Baeza, W. Van Treuren, N. A. Hasan, M. K. Gibson, R. Colwell, G. Dantas, R. Knight, J. A. Gilbert, Longitudinal analysis of microbial interaction between humans and the indoor environment. *Science* **345**, 1048–1052 (2014).
3. R. I. Adams, A. C. Bateman, H. M. Bik, J. F. Meadow, Microbiota of the indoor environment: A meta-analysis. *Microbiome* **3**, 1–18 (2015).
4. M. Wood, S. M. Gibbons, S. Lax, T. W. Shoo-Anton, S. M. Owens, S. Kennedy, J. A. Gilbert, J. T. Hampton-Marcell, Athletic equipment microbiota are shaped by interactions with human skin. *Microbiome* **3**, 25 (2015).
5. S. Lax, C. R. Nagler, J. A. Gilbert, Our interface with the built environment: Immunity and the indoor microbiota. *Trends Immunol.* **36**, 121–123 (2015).
6. K. E. Fujimura, T. Demoore, M. Rauch, A. A. Faruqi, S. Jang, C. C. Johnson, H. A. Boushey, E. Zoratti, D. Ownby, N. W. Lukacs, S. V. Lynch, House dust exposure mediates gut microbiome *Lactobacillus* enrichment and airway immune defense against allergens and virus infection. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 805–810 (2014).
7. A. T. Stefka, T. Feehley, P. Tripathi, J. Qiu, K. McCoy, S. K. Mazmanian, M. Y. Tjota, G.-Y. Seo, S. Cao, B. R. Theriault, D. A. Antonopoulos, L. Zhou, E. B. Chang, Y.-X. Fu, C. R. Nagler, Commensal bacteria protect against food allergen sensitization. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 13145–13150 (2014).
8. M. J. Blaser, S. Falkow, What are the consequences of the disappearing human microbiota? *Nat. Rev. Microbiol.* **7**, 887–894 (2009).
9. Department of Health and Human Services, *Summary of Notifiable Diseases—United States, 2007* (Centers for Disease Control and Prevention, 2009).
10. R. N. Anderson, B. L. Smith, Deaths: Leading causes for 2002. *Natl. Vital Stat. Rep.* **53**, 67–70 (2005).
11. S. L. Groseclose, W. S. Brathwaite, P. A. Hall, D. A. Adams, F. J. Connor, P. Sharp, W. J. Anderson, R. F. Fagan, J. Javier Aponte, G. F. Jones, D. A. Nitschke, C. A. Worsham, N. Adekoya, M.-H. Chang, T. Doyle, R. Dhara, R. Ann Jajosky, Summary of notifiable diseases—United States, 2002. *MMWR Morb. Mortal. Wkly Rep.* **51**, 1–84 (2004).

12. E. C. Pehrsson, P. Tsukayama, S. Patel, M. Mejia-Bautista, G. Sosa-Soto, K. M. Navarrete, M. Calderon, L. Cabrera, W. Hoyos-Arango, M. Teresita Bertoli, D. E. Berg, R. H. Gilman, G. Dantas, Interconnected microbiomes and resistomes in low-income human habitats. *Nature* **533**, 212–216 (2016).
13. E. S. Snitkin, A. M. Zelazny, P. J. Thomas, F. Stock; NISC Comparative Sequencing Program, D. K. Henderson, T. N. Palmore, J. A. Segre, Tracking a hospital outbreak of carbapenem-resistant *Klebsiella pneumoniae* with whole-genome sequencing. *Sci. Transl. Med.* **4**, 148ra116 (2012).
14. J. Quick, P. Ashton, S. Calus, C. Chatt, S. Gossain, J. Hawker, S. Nair, K. Neal, K. Nye, T. Peters, E. De Pinna, E. Robinson, K. Struthers, M. Webber, A. Catto, T. J. Dallman, P. Hawkey, N. J. Loman, Rapid draft sequencing and real-time nanopore sequencing in a hospital outbreak of *Salmonella*. *Genome Biol.* **16**, 114 (2015).
15. S. R. Harris, E. J. Feil, M. T. G. Holden, M. A. Quail, E. K. Nickerson, N. Chantratita, S. Gardete, A. Tavares, N. Day, J. A. Lindsay, J. D. Edgeworth, H. de Lencastre, J. Parkhill, S. J. Peacock, S. D. Bentley, Evolution of MRSA during hospital transmission and intercontinental spread. *Science* **327**, 469–474 (2010).
16. P. R. McAdam, K. E. Templeton, G. F. Edwards, M. T. G. Holden, E. J. Feil, D. M. Aanensen, H. J. A. Bargawi, B. G. Spratt, S. D. Bentley, J. Parkhill, M. C. Enright, A. Holmes, E. K. Girvan, P. A. Godfrey, M. Feldgarden, A. M. Kearns, A. Rambaut, D. A. Robinson, J. R. Fitzgerald, Molecular tracing of the emergence, adaptation, and transmission of hospital-associated methicillin-resistant *Staphylococcus aureus*. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 9107–9112 (2012).
17. B. Brooks, B. A. Firek, C. S. Miller, I. Sharon, B. C. Thomas, R. Baker, M. J. Morowitz, J. F. Banfield, Microbes in the neonatal intensive care unit resemble those found in the gut of premature infants. *Microbiome* **2**, 1 (2014).
18. K. M. Hewitt, F. L. Mannino, A. Gonzalez, J. H. Chase, J. Gregory Caporaso, R. Knight, S. T. Kelley, Bacterial diversity in two neonatal intensive care units (NICUs). *PLOS ONE* **8**, e54703 (2013).
19. M. Poza, C. Gayoso, M. J. Gómez, S. Rumbo-Feal, M. Tomás, J. Aranda, A. Fernández, G. Bou, Exploring bacterial diversity in hospital environments by GS-FLX Titanium pyrosequencing. *PLOS ONE* **7**, e44105 (2012).
20. J. Oh, A. L. Byrd, M. Park; NISC Comparative Sequencing Program, H. H. Kong, J. A. Segre, Temporal stability of the human skin microbiome. *Cell* **165**, 854–866 (2016).
21. T. Ramos, S. Dedesko, J. A. Siegel, J. A. Gilbert, B. Stephens, Spatial and temporal variations in indoor environmental conditions, human occupancy, and operational characteristics in a new hospital building. *PLOS ONE* **10**, e0118207 (2015).
22. S. Dedesko, B. Stephens, J. A. Gilbert, J. A. Siegel, Methods to assess human occupancy and occupant activity in hospital patient rooms. *Build. Environ.* **90**, 136–145 (2015).
23. J. S. Meisel, G. D. Hannigan, A. S. Tyldsley, A. J. SanMiguel, B. P. Hodkinson, Q. Zheng, E. A. Grice, Skin microbiome surveys are strongly influenced by experimental design. *J. Invest. Dermatol.* **136**, 947–956 (2016).
24. S. Lax, J. T. Hampton-Marcell, S. M. Gibbons, G. B. Colares, D. Smith, J. A. Eisen, J. A. Gilbert, Forensic analysis of the microbiome of phones and shoes. *Microbiome* **3**, 21 (2015).
25. J. F. Meadow, A. E. Altrichter, J. L. Green, Mobile phones carry the personal microbiome of their owners. *Peer J.* **2**, e447 (2014).
26. D. Knights, J. Kuczynski, E. S. Charlson, J. Zaneveld, M. C. Mozer, R. G. Collman, F. D. Bushman, R. Knight, S. T. Kelley, Bayesian community-wide culture-independent microbial source tracking. *Nat. Methods* **8**, 761–763 (2011).
27. D. P. Faith, Conservation evaluation and phylogenetic diversity. *Biol. Conserv.* **61**, 1–10 (1992).
28. A. M. Eren, L. Maignien, W. J. Sul, L. G. Murphy, S. L. Grim, H. G. Morrison, M. L. Sogin, Oligotyping: Differentiating between closely related microbial taxa using 16S rRNA gene data. *Methods Ecol. Evol.* **4**, 1111–1119 (2013).
29. W. Ran, D. M. Kristensen, E. V. Koonin, Coupling between protein level selection and codon usage optimization in the evolution of bacteria and archaea. *MBio* **5**, e00956-14 (2014).
30. Y. Yao, D. E. Sturdevant, A. Villaruz, L. Xu, Q. Gao, M. Otto, Factors characterizing *Staphylococcus epidermidis* invasiveness determined by comparative genomics. *Infect. Immun.* **73**, 1856–1860 (2005).
31. J. G. Caporaso, C. L. Lauber, W. A. Walters, D. Berg-Lyons, J. Huntley, N. Fierer, S. M. Owens, J. Betley, L. Fraser, M. Bauer, N. Gormley, J. A. Gilbert, G. Smith, R. Knight, Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *ISME J.* **6**, 1621–1624 (2012).
32. J. G. Caporaso, J. Kuczynski, J. Stombaugh, K. Bittinger, F. D. Bushman, E. K. Costello, N. Fierer, A. G. Peña, J. K. Goodrich, J. I. Gordon, G. A. Huttley, S. T. Kelley, D. Knights, J. E. Koenig, R. E. Ley, C. A. Lozupone, D. McDonald, B. D. Muegge, M. Pirrung, J. Reeder, J. R. Sevinsky, P. J. Turnbaugh, W. A. Walters, J. Widmann, T. Yatsunenko, J. Zaneveld, R. Knight, QIIME allows analysis of high-throughput community sequencing data. *Nat. Methods* **7**, 335–336 (2010).
33. J. G. Caporaso, K. Bittinger, F. D. Bushman, T. Z. DeSantis, G. L. Andersen, R. Knight, PyNAST: A flexible tool for aligning sequences to a template alignment. *Bioinformatics* **26**, 266–267 (2010).
34. S. Guindon, J.-F. Dufayard, V. Lefort, M. Anisimova, W. Hordijk, O. Gascuel, New algorithms and methods to estimate maximum-likelihood phylogenies: Assessing the performance of PhyML 3.0. *Syst. Biol.* **59**, 307–321 (2010).
35. R. C. Edgar, Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26**, 2460–2461 (2010).
36. C. Lozupone, R. Knight, UniFrac: A new phylogenetic method for comparing microbial communities. *Appl. Environ. Microbiol.* **71**, 8228–8235 (2005).
37. N. Segata, L. Waldron, A. Ballarini, V. Narasimhan, O. Jousson, C. Huttenhower, Metagenomic microbial community profiling using unique clade-specific marker genes. *Nat. Methods* **9**, 811–814 (2012).
38. Y. Peng, H. C. M. Leung, S. M. Yiu, F. Y. L. Chin, IDBA-UD: A de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. *Bioinformatics* **28**, 1420–1428 (2012).
39. D. D. Kang, J. Froula, R. Egan, Z. Wang, MetaBAT, an efficient tool for accurately reconstructing single genomes from complex microbial communities. *Peer J.* **3**, e1165 (2015).
40. C. Kerepesi, D. Bánky, V. Grolmusz, AmphoraNet: The webserver implementation of the AMPHORA2 metagenomic workflow suite. *Gene* **533**, 538–540 (2014).
41. D. P. Wall, H. B. Fraser, A. E. Hirsh, Detecting putative orthologs. *Bioinformatics* **19**, 1710–1711 (2003).
42. M. A. Larkin, G. Blackshields, N. P. Brown, R. Chenna, P. A. McGettigan, H. McWilliam, F. Valentin, I. M. Wallace, A. Wilm, R. Lopez, J. D. Thompson, T. J. Gibson, D. G. Higgins, Clustal W and Clustal X version 2.0. *Bioinformatics* **23**, 2947–2948 (2007).
43. M. Suyama, D. Torrents, P. Bork, PAL2NAL: Robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Res.* **34**, W609–W612 (2006).
44. Z. Yang, PAML 4: Phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* **24**, 1586–1591 (2007).
45. Z. Zhang, J. Li, P. Cui, F. Ding, A. Li, J. P. Townsend, J. Yu, Codon deviation coefficient: A novel measure for estimating codon usage bias and its statistical significance. *BMC Bioinformatics* **13**, 43 (2012).
46. R. K. Aziz, D. Bartels, A. A. Best, M. DeJongh, T. Disz, R. A. Edwards, K. Formosa, S. Gerdes, E. M. Glass, M. Kubal, F. Meyer, G. J. Olsen, R. Olson, A. L. Osterman, R. A. Overbeek, L. K. McNeil, D. Paarmann, T. Paczian, B. Parrello, G. D. Pusch, C. Reich, R. Stevens, O. Vassieva, V. Vonstein, A. Wilke, O. Zagnitko, The RAST server: Rapid annotations using subsystems technology. *BMC Genomics* **9**, 75 (2008).
47. M. K. Gibson, K. J. Forsberg, G. Dantas, Improved annotation of antibiotic resistance determinants reveals microbial resistomes cluster by ecology. *ISME J.* **9**, 207–216 (2015).

Acknowledgments: We acknowledge the University of Chicago Research Computing Center, the Argonne National Laboratory Sample Processing and Sequencing Core, and the University of Chicago hospital staff for support of this work. We thank the many nurses and patients who volunteered their time to this study. **Funding:** We thank the Alfred P. Sloan Foundation Microbiology of the Built Environment Program for funding this study. This work was supported, in part, by the U.S. Department of Energy under contract DE-AC02-06CH11357. **Author contributions:** D.S., R.K., J.A., J.S., B. Stephens, and J.A.G. conceived the study. S.L., D.S., K.M.H., K.G., M.K., B.D.S., J.D., I.F., B. Shakhsher, S.G.-H., and E.L. collected data. S.L. analyzed all 16S data with the exception of the Bayesian network analysis. N.S. analyzed the shotgun metagenomic data. P.L. completed the Bayesian network analysis. M.R. assisted in the SourceTracker analyses. S.L. and J.A.G. wrote the paper. All authors edited and approved the manuscript. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** 16S sequence data were submitted to the European Bioinformatics Institute database under accession no. ERP016116. Metagenome sequence data were uploaded to the European Molecular Biology Laboratory under accession no. PRJEB13117. Correspondence and requests for materials should be addressed to J.A.G. (gilbertjack@uchicago.edu).

Submitted 26 July 2016
 Accepted 27 February 2017
 Published 24 May 2017
 10.1126/scitranslmed.aah6500

Citation: S. Lax, N. Sangwan, D. Smith, P. Larsen, K. M. Handley, M. Richardson, K. Guyton, M. Krezalek, B. D. Shogan, J. Defazio, I. Flemming, B. Shakhsher, S. Weber, E. Landon, S. Garcia-Houchins, J. Siegel, J. Alverdy, R. Knight, B. Stephens, J. A. Gilbert, Bacterial colonization and succession in a newly opened hospital. *Sci. Transl. Med.* **9**, eaah6500 (2017).