Longitudinal analysis of microbial interaction between humans and the indoor environment

Simon Lax,[1,2]* Daniel P. Smith,[1,2,3]* Jarrad Hampton-Marcell,[1,2] Sarah M. Owens,[2,4] Kim M. Handley,[1,2] Nicole M. Scott,[1,2] Sean M. Gibbons,[2,5] Peter Larsen,[6,7] Benjamin D. Shogan,[8] Sophie Weiss,[9,10] Jessica L. Metcalf,[9] Luke K. Ursell,[9,11] Yoshiki Vázquez-Baeza,[9,11,12] Will Van Treuren,[9] Nur A. Hasan,[13,14] Molly K. Gibson,[15,16,17] Rita Colwell,[13,14] Gautam Dantas,[15,16,17] Rob Knight,[9,11,18] Jack A. Gilbert[1,2,5]†

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. Draft genomes of potential human pathogens observed on a kitchen counter could be matched to the hands of occupants. 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.

The global trend toward urbanization has increasingly bound humanity, as a species, to the indoor environment (*1*, *2*). 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" (*3*–*7*), which should transfer to a new indoor space with skin shedding, respiratory activity, and skin-surface contact (*8*), the latter of which can transfer millions of microbial cells per event (*9*). The microbial diversity of the home likely affects immune defense (*10*) and disease transmission (*11*) 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 (table S1), 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.

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 (Fig. 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 ($P < 0.0001$) (Fig. 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 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.

Humans sharing a home were more microbially similar than those not sharing a home, with samples taken from the same individual having the greatest similarity (Fig. 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

[1]Department of Ecology and Evolution, University of Chicago, 1101 E 57th Street, Chicago, IL 60637, USA. [2]Institute for Genomic and Systems Biology, Argonne National Laboratory, 9700 South Cass Avenue, Argonne, IL 60439, USA. [3]Alkek Center for Metagenomics and Microbiome Research, Department of Molecular Virology and Microbiology, Baylor College of Medicine, Houston, TX 77030, USA. [4]Computation Institute, University of Chicago, Chicago, IL 60637, USA. [5]Graduate Program in Biophysical Sciences, University of Chicago, Chicago, IL 60637, USA. [6]Department of Bioscience, Argonne National Laboratory, 9700 South Cass Avenue, Argonne, IL 60439, USA. [7]Department of Bioengineering, University of Illinois at Chicago, Chicago, IL 60607, USA. [8]Department of Surgery, University of Chicago Medicine, 5841 South Maryland Avenue, Chicago, IL 60637, USA. [9]Biofrontiers Institute, University of Colorado, 3415 Colorado Avenue, Boulder, CO 80304, USA. [10]Department of Chemical and Biological Engineering, University of Colorado at Boulder, Boulder, CO 80304, USA. [11]Department of Chemistry and Biochemistry, University of Colorado at Boulder, Boulder, CO 80304, USA. [12]Department of Computer Science, University of Colorado at Boulder, Boulder, CO 80304, USA. [13]CosmosID, 387 Technology Drive, Suite 3119, College Park, MD 20742, USA. [14]Center for Bioinformatics and Computational Biology, University of Maryland Institute for Advanced Computer Studies, University of Maryland College Park, College Park, MD 20742, USA. [15]Center for Genome Sciences and Systems Biology, Washington University School of Medicine, St. Louis, MO 63108, USA. [16]Department of Pathology and Immunology, Washington University School of Medicine, St. Louis, MO 63110, USA. [17]Department of Biomedical Engineering, Washington University, St. Louis, MO 63130, USA. [18]Howard Hughes Medical Institute, Boulder, CO 80309, USA.
*These authors contributed equally to this work and are listed in alphabetical order. **†Corresponding author. E-mail: gilbertjack@anl.gov**

sample having little effect on the observed strength of differentiation (fig. S1A).

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] (Fig. 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) (fig. S2A). Although relative OTU abundances always correlated tightly between human and home environments, they varied in their correlations with pets (fig. S2B).

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 (Fig. 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 (Fig. 2C, split by sample in fig. S3). 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 $P > 0.05$) (Fig. 2D). Pairwise comparison of 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 (Fig. 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 S2). 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

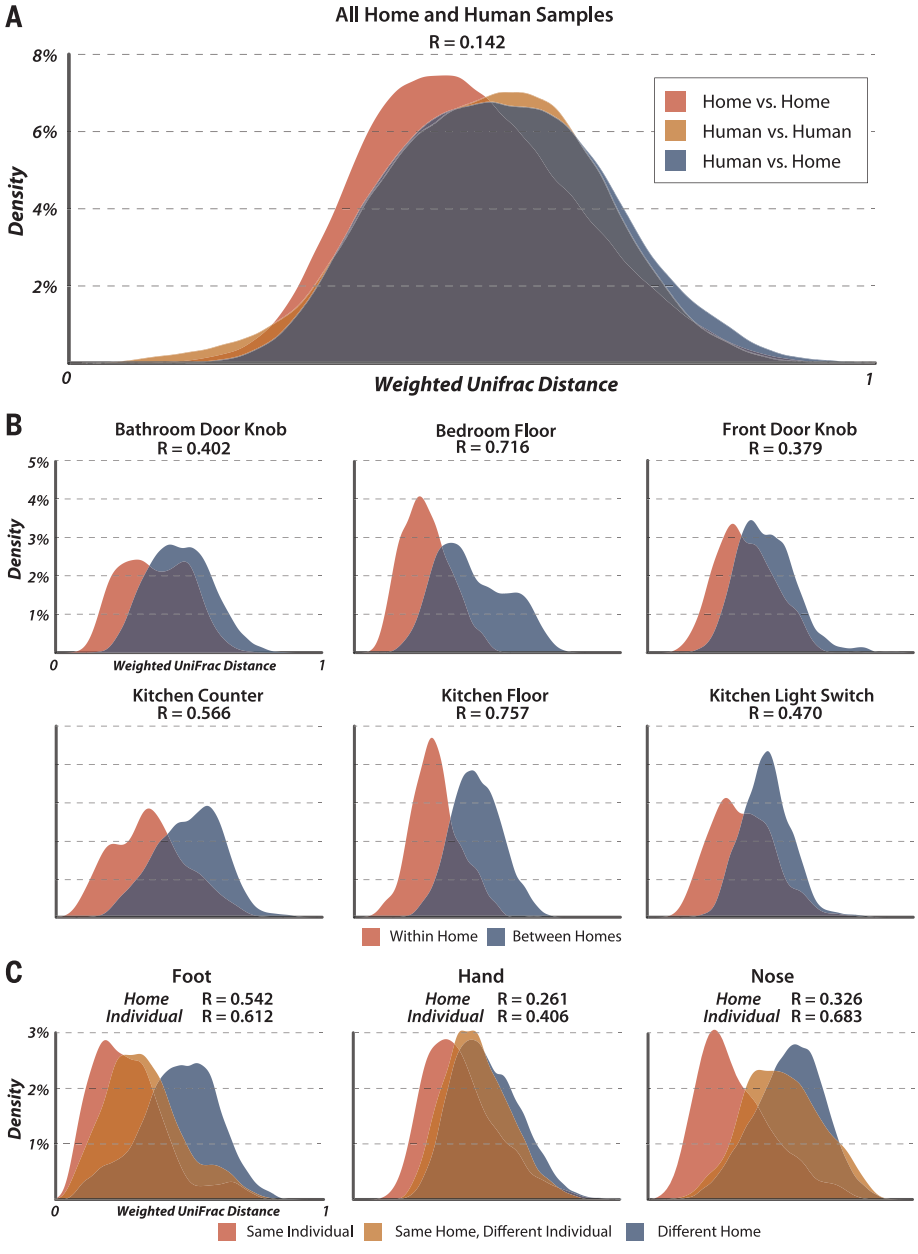

Fig. 1. Differentiation in microbial community structure between homes and individuals. Density plots comparing the distributions of weighted UniFrac distances calculated within and between various criteria with accompanying ANOSIM tests of differentiation (all *P* values are less than 0.0001 based on 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).

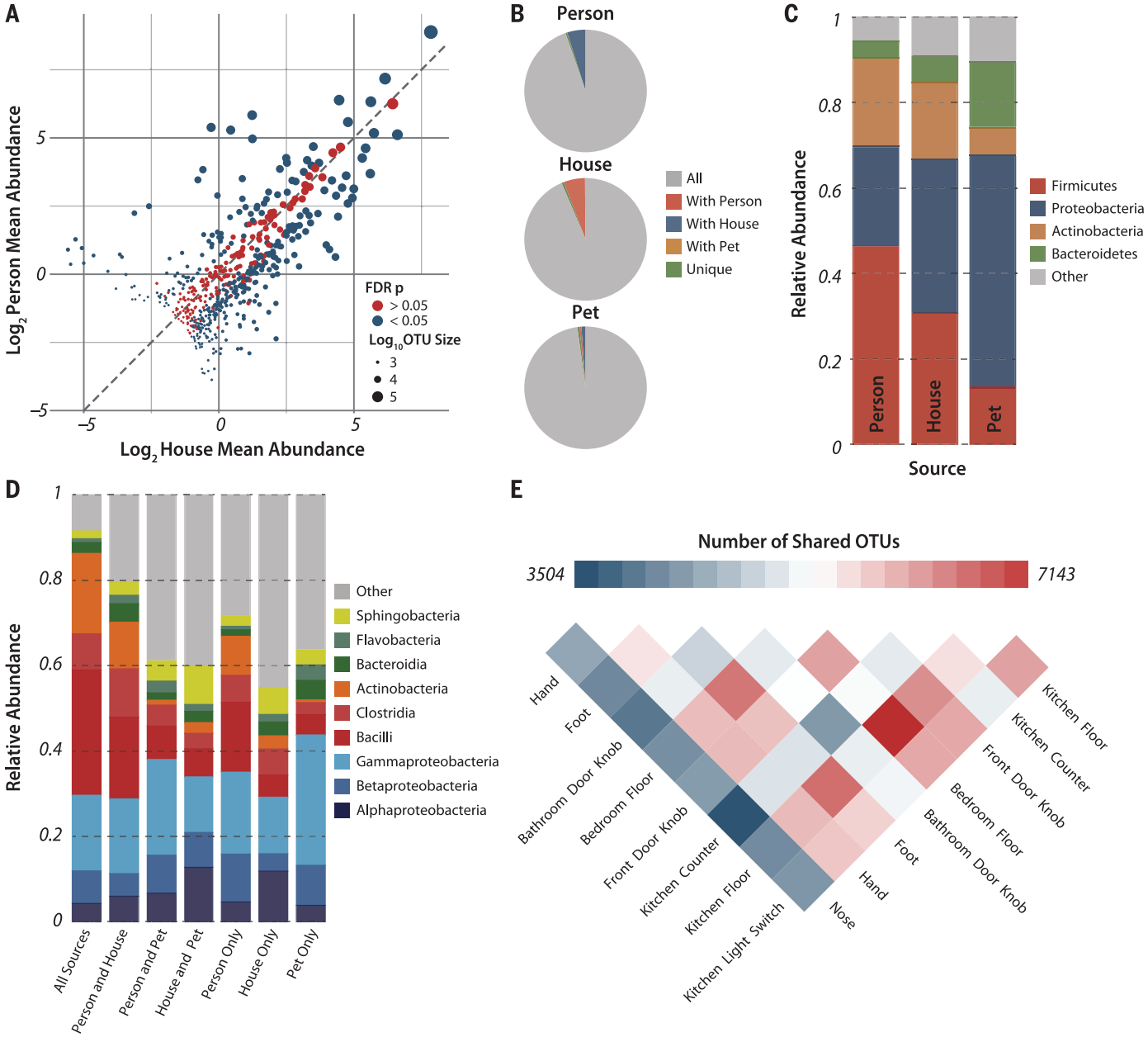

Fig. 2. Widespread sharing of microbial taxa between human and home surfaces. (**A**) Plot of $\log_2$-transformed average relative abundances in the human and home environments for all OTUs in the study with greater than 500 reads. OTUs are colored by whether their average relative abundance is significantly 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 $\log_{10}$-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.

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 (fig. S1B), 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 (Fig. 3) (*12*). 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 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

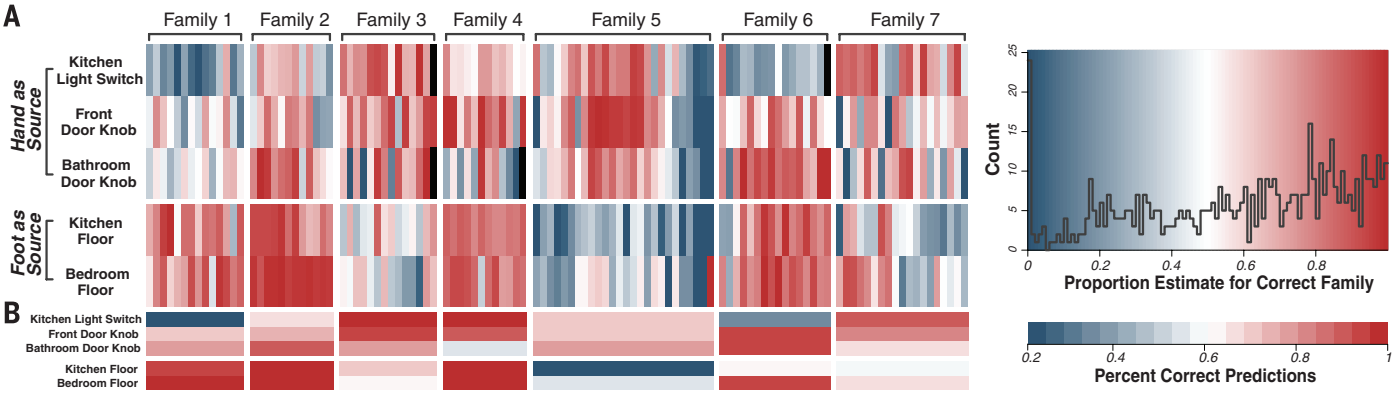

Fig. 3. Summary of predictive accuracy of SourceTracker models. (**A**) Percent composition estimate for the correct source for each home surface sample in the study. Samples within each block are ordered by collection date, and black boxes occur where a sample is missing because it did not pass quality 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.

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 (fig. S4). 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.

We tested the direction of microbial transfer among surfaces in the four homes where the subjects did not move houses using dynamic Bayesian networks (fig. S5 and table S3). 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 (table S3), which are major components of the human skin microbial community (*6*).

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 (fig. S6). 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.

Overall, there were significant differences in the volatility of microbial communities associated with each surface type (Kruskal-Wallis $\chi^2 = 21.6$, $P = 0.0057$) (fig. S7). 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.

To determine whether taxa transferred between surfaces and human occupants maintained genes associated with pathogenicity, we selected 56 samples from home 4 for longitudinal analysis via shallow shotgun metagenomic sequencing, including 18 home surface samples, 23 human samples, and 15 dog samples (fig. S8). Genes associated with phage and transposable elements were enriched in human samples. Taxonomic analysis of unassembled metagenomic reads revealed *Corynebacterium* on all human samples; *Enhydrobacter*-, *Corynebacterium*-, and *Streptococcus*-like bacteria on all bathroom doorknob samples; and *Enterobacter*-like bacteria on the kitchen counter. *Enterobacter*-like sequences were also identified on the hands of two occupants on days 2 and 4, further supporting the dynamic Bayesian network analysis above (in addition to genome reconstructions) that indicated a close link between these surfaces (fig. S4). After deeper metagenomic sequencing, multiple draft genomes were assembled from hand and kitchen counter samples, including uncultivated *Enterobacteraceae* and *Acinetobacter* genomes and associated bacteriophage. These latter genomes shared 99.7 and 99.9% reconstructed 16*S* rRNA gene sequence similarity with the respective opportunistic human pathogens *Pantoea agglomerans* and *Acinetobacter baumannii* and maintained genes associated with pathogenicity and antibiotic resistance. Representatives of these genomes sourced from both the kitchen counter and one household occupant's hand on day 2 shared >2400 genes with 100% protein sequence identity. When considering the whole bacterial community, and including phage sequence, a total of 84% (7671) of genes from the hand were shared with the countertop, suggesting a multiorganism transference event between these surfaces. A further 24 to 29% of the community genes (>3100) were also identical between the countertop across days (between days 2 and 4) and between the countertop on day 4 and the hand of one of the other occupants sampled on the same day.

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 (*13*–*16*), the role of hygiene in removing microbial communities (*7*, *17*), and the length of time microbes can survive on surfaces (*18*, *19*). 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.

REFERENCES AND NOTES

1. P. Höppe, I. Martinac, *Int. J. Biometeorol.* **42**, 1–7 (1998).
2. A. Custovic, S. C. O. Taggart, A. Woodcock, *Clin. Exp. Allergy* **24**, 1164–1168 (1994).
3. I. Dekio *et al.*, *J. Med. Microbiol.* **54**, 1231–1238 (2005).
4. Z. Gao, C. H. Tseng, Z. Pei, M. J. Blaser, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 2927–2932 (2007).
5. E. A. Grice *et al.*, *Genome Res.* **18**, 1043–1050 (2008).
6. E. A. Grice *et al.*, *Science* **324**, 1190–1192 (2009).
7. E. A. Grice, *Natl. Rev.* **9**, 244–253 (2011).
8. S. G. Tringe *et al.*, *PLOS One* **3**, e1862 (2008).
9. P. Dawson, I. Han, M. Cox, C. Black, L. Simmons, *J. Appl. Microbiol.* **102**, 945–953 (2007).
10. K. E. Fujimura *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **111**, 805–810 (2014).

11. S. A. Boone, C. P. Gerba, *Appl. Environ. Microbiol.* **73**, 1687–1696 (2007).
12. D. Knights *et al.*, *Nat. Methods* **8**, 761–763 (2011).
13. A. A. Jaffal *et al.*, *Environ. Int.* **23**, 529–533 (1997).
14. A. Hyvarinen, T. Meklin, A. Vepsalainen, A. Nevalainen, *Int. Biodeterior. Biodegradation* **49**, 27–37 (2002).
15. A. Nevalainen, M. Seuri, *Indoor Air* **15** (suppl. 9), 58–64 (2005).
16. U. Lignell *et al.*, *Lett. Appl. Microbiol.* **47**, 303–308 (2008).
17. K. R. Bright, S. A. Boone, C. P. Gerba, *J. Sch. Nurs.* **26**, 33–41 (2010).
18. A. Kramer, I. Schwebke, G. Kampf, *BMC Infect. Dis.* **6**, 130 (2006).
19. D. J. Moschandreas, K. R. Pagilla, L. V. Storino, *Aerosol Sci. Technol.* **37**, 899–906 (2003).

ACKNOWLEDGMENTS

This work was supported by the Alfred P. Sloan Foundation's Microbiology of the Built Environment Program and a NIH Director's New Innovator Award (DP2-DK-098089) to G.D. S.M.G was supported by a U.S. Environmental Protection Agency STAR Fellowship. M.K.G. is a NSF graduate research fellow (award DGE-11143954). This work was supported in part by the U.S. Department of Energy under contract DE-AC02-06CH11357. We acknowledge the University of Chicago Research Computing Center for support of this work. All sequence data were submitted to the European Bioinformatics Institute's database under accession no. ERP005806.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/345/6200/1048/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S9
Tables S1 to S3
References (*20*–*29*)

8 April 2014; accepted 18 July 2014
10.1126/science.1254529