Detecting Personal Microbiota Signatures at Artificial Crime Scenes

Jarrad T. Hampton-Marcell[1,2,3], Peter Larsen[1], Tifani Anton[1], Lauren Cralle[3], Naseer Sangwan[4], Simon Lax[5], Neil Gottel[6], Mariana Salas-Garcia[6], Candace Young[8], George Duncan[7], Jose V. Lopez[7], Jack A. Gilbert[1,6]

[1]Biosciences Division, Argonne National Laboratory, Lemont, IL
[2]Department of Biological Sciences, University of Illinois at Chicago, Chicago, IL
[3]Department of Surgery, University of Chicago, Chicago, IL
[4]Department of Cardiovascular and Metabolic Sciences, Cleveland Clinic, Cleveland, OH
[5]Department of Physics, Massachusetts Institute of Technology, Cambridge, MA
[6]Department of Pediatrics and Scripps Institution of Oceanography, University of California San Diego, La Jolla, CA
[7]Department of Biological Sciences, Nova Southeastern University, Fort Lauderdale, FL
[8]Department of Chemistry, Physics and Engineering Studies, Chicago State University, Chicago, IL

CORRESPONDING AUTHOR
Jarrad Hampton-Marcell
Argonne National Laboratory
9700 S. Cass Ave.
Bldg. 446, Rm. A174
Lemont, IL 60439
United States
708-526-3966
jarrad.marcell@gmail.com

HIGHLIGHTS

- Humans harbor traceable skin microbiota that discriminatively shape the microbial community structures in their residence
- Humans harbor unique and stable microbial combinations (uESVs) that can be used to trace human interactions with surfaces within residences
- Utilizing uESVs, burglars can be sourced to surface interactions during mock burglaries leaving behind traceable skin microbiota

ABSTRACT

When mapped to the environments we interact with on a daily basis, the 36 million microbial cells per hour that humans emit leave a trail of evidence that can be leveraged for forensic analysis. We employed 16S rRNA amplicon sequencing to map unique microbial sequence variants between human skin and building surfaces in three experimental conditions: over time during controlled and uncontrolled incidental interactions with a door handle, and during multiple mock burglaries in ten real residences. We demonstrate that humans (n=30) leave behind microbial signatures that can be used to track interaction with various surfaces within a building, but the likelihood of accurately detecting the specific burglar for a given home was between 20-25%. Also, the human microbiome contains rare microbial taxa that can be combined to create a unique microbial profile, which when compared to 600 other individuals can improve our ability to link an individual 'burglar' to a residence. In total, 5,512 discriminating, non-singleton unique exact sequence variants (uESVs) were identified as unique to an individual, with a minimum of 1 and a maximum of 568, suggesting some people maintain a greater degree of unique taxa compared to our population of 600. Approximate 60-77% of the unique exact sequence variants originated from the hands of participants, and these microbial discriminators spanned 36 phyla but were dominated by the Proteobacteria (34%). A fitted regression generated to determine whether an intruder's uESVs found on door handles in an office decayed over time in the presence or absence of office workers, found no significant shift in proportion of uESVs over time irrespective of the presence of office workers. While it was possible to detect the correct burglars' microbiota as having contributed to the invaded space, the predictions were very weak in comparison to accepted forensic standards. This suggests that at this time 16S rRNA amplicon sequencing of the built environment microbiota cannot be used as a reliable trace evidence standard for criminal investigations.

KEYWORDS
Forensic microbiology, built-environment, host-microbe, trace evidence, human microbiome

INTRODUCTION

Utilizing biological markers as trace evidence generally relies on identifying discriminate variants that are stable relative to the suspect and persistent in the environment over time, as is the case for human DNA profiling [1]. Humans cultivate a highly individualized microbial population on their skin, and shed approximately 36 million of these bacterial cells per hour into their immediate environment [2], [3], and it is possible to detect this individualized signature on an environmental surface following a person's interaction with that surface [4]–[6]. For example, the microbial signature on a participant's finger tips to the keys on a computer keyboard [7], as well as an individual's personal devices such as a smart phone [8], [9].

This creates the intriguing possibility that the microbial signature of an individual could be reliably detected if they transiently interact with a surface, such as in a home environment [10]. Yet the proportions of each bacterial species in the human microbiome are not as stable as our genomic DNA. Even following the onset of an 'adult-like' signature, the gut microbiome is constantly adapting to a myriad of factors, such as diet and medicines, which creates shifting selection

pressures that promote one species over another, often leading to daily fluctuations in relative proportions [11]–[13]. However, in the absence of major disturbances such as antibiotic use, the skin microbiome appears to remains relatively stable, both in terms of composition and the proportion of each taxon [14], [15]. Is it therefore possible that the residual skin microbial profile left behind when someone transiently interacts with a physical surface in the built environment could be used to identify that person with a reasonable degree of accuracy?

Provided each person is unique in their ancestry, genetics and lifestyle, the composition of the microbial community that inhabits our bodies is predicted to be unique between those individuals. Of course, there remains the possibility that two individuals may have an compliment of species that are indistinguishable using 16S rRNA amplicon sequencing techniques, but even the microbiome of identical twins failed to identify such a phenomenon [16]. Recently, researchers have demonstrated that people emit a microbial 'cloud' that can be traced to the individual [17], and studies of different home environments have shown that the microbial communities deposited on high-touch surfaces are most similar to the occupants that frequently use that space and touch those surfaces, and that the use of a space could also be predicted based on microbial signatures [6], [10], [18]. Interestingly, continental scale biogeography has also been implicated in driving the differences in the home-associated microbial signatures e.g. [19], [20], which means this must be taken into consideration in any experimental design.

Here, we determine whether the residual skin microbiota left behind by transient interaction with a surface could serve as trace evidence of that individual's presence using two experimental systems. First, an individual physically interacted with the door handles of 12 offices on two separate days, and the door handles were then sampled sequentially to determine a decay rate for the ability to detect the residual microbiota signature of that individual in the context of both the presence (day 1) and absence (day 2) of office workers in the space. Second, a series of 'mock' home invasions were conducted in residential homes in two different cities (Naperville, Illinois and Fort Lauderdale, Florida, USA) to determine if a 'burglar's' residual skin microbiota could be detected and reliably used to identify the same individual's presence in that home. The microbiota of occupants, 'burglars' and surfaces were characterized using a relatively inexpensive but powerful molecular tool, 16S rRNA amplicon sequencing. This was followed with statistical tests to determine the likelihood that an individual's residual skin microbial signature could be used as trace evidence to infer their presence.

METHODS

Study Design. Two principle experiments were performed.

Experiment 1. A time series analysis was conducted to analyze the stability of unique microbial profiles over time on a physical surface i.e. door handle in two different contexts i.e. the absence versus presence of other individuals (**Table 1**). Door handles from offices (n = 12) found within the same area were interacted with by a prime individual who had not been in that building before. This experiment was performed on two consecutive days, first a Sunday during which no workers were present, and second a Monday morning during which all offices spaces (and hence door handles) were being utilized by the workers. Door handles were sanitized using a 70% bleach

solution approximately ten minutes prior to the first interaction. Each experimental period was ~60 mins, whereby a 5-minute interval occurred between each time the prime individual interacted with a different door handle. At the end of the experimental period, each door handle was sampled by swabbing two sterile cotton BD-Swube applicators over the whole door handle for 20 seconds (**Fig. 1**). Therefore, each door handle had a different time elapse between interaction with the prime individual's hand and sampling. We hypothesized that during the Sunday event, all door handles would remain mostly microorganism free prior to the interaction by the prime individual, while on the Monday event, each door handle would have a differential probability of having been interacted with by an office worker both prior and post the interaction with the prime individual. During the Monday event, tallies were taken by observers to record the number of times each door handle was touched by an office worker.

Experiment 2. A series of mock burglaries were conducted in three residential homes in Naperville, Illinois (USA) and five residential homes in Fort Lauderdale (FL), Florida (USA) in August 2016 (**Table 2**), and then repeated in March 2017. For each home, duplicate samples were collected by swabbing two sterile cotton BD-Swube applicators against various surfaces within each home for 20 seconds prior to and following burglarlies. The nares and hand of each occupant and of each mock burglar (n = 4) were sampled prior to the burglar's introduction to the home. For each location, a pair of individuals (n = 2) that were not residents to any of the homes served as burglars. The burglar pair had no previous interactions with the homes in August 2016 and had not interacted with the homes for more than 6 months in March 2017. During a burglary, occupants were asked to temporarily vacate the premises, and the burglar was asked to enter via the front door. They were allowed to interact with the home environment (as dictated by the owner, i.e. they were not allowed in rooms identified by the owner as private) for 10 mins, during which time they applied a piece of colored tape to each physical surface they interacted with which were limited to surfaces that had been sampled beforehand. This was done to limit the number of possible surfaces within each home, as well as to ensure a comparison of microbial communities pre- and post-burglary. At no time were any physical objects removed from the premises, and as such these events are considered failed burglaries. Following the mock burglary, the 'burglar' was sampled again, on their hands and in their nares (**Fig. 2**). Subsequently, an investigator entered the premises with a disposable face mask, hair covering, gloves and full clothing (to minimize their microbial impact on the space, as well as to reproduce how real crime scene investigators might interact with the space to minimize trace evidence disruption). The investigator proceeded to swab each surface identified by colored tape as having been interacted with. All samples were stored on wet ice, and then transferred to a -80C freezer within 3 hours for storage until processing.

Sample processing. DNA was extracted from each sample using a low biomass variation of the MO BIO Powersoil DNA extraction, and the 16S rRNA was amplified with the Earth Microbiome 16S Illumina Amplicon Protocol [21]. The V4 region of the 16S rRNA gene was targeted with the 515F-806RB primer pair and sequenced using a Illumina MiSeq sequencer [22], [23].

Sequence processing. Deblur. Exact Sequence Variants (ESVs) were assigned and clustered using Deblur [24]. Deblur circumvents the problems surrounding clustering of OTUs at an

arbitrarily threshold by obtaining single-nucleotide resolution ESVs after correcting for Illumina sequencing errors. This results in exact/amplicon sequence variants, also called ESVs differing from the traditional 97% clustering approach [25]. ESVs identified within sequencing blanks were filtered from the raw microbial table. The minimum reads-option was set to 0 to disable filtering inside Deblur and sequences were trimmed to 150bp. Following filter of samples with less than 2 reads, a table was constructed mapping the abundance of each ESV to each sample rarefied to 1000 sequences per sample to ensure that all samples contained the same number of sequences for downstream analysis. To determine unique ESV identification (uESVs), human-associated samples were grouped based on their originating source for hand- and nare-associated samples. ESVs that were solely expressed in one individual in comparison the entire population were identified as unique ESVs (uESVs). uESV sequences that did not occur at least twice were discarded prior to analysis.

Data Analysis. Diversity metrics were calculated using Bray-Curtis distances for ESVs including community similarity summarized by Principle Coordinate Analysis (PCoA). PCoA results were visualized using the R package 'phyloseq' [26], which allows for sample distances and natural clustering behavior to be visualized in the context of sample metadata. Unique ESV profiles were generated using QIIME v1.9.1 [27]. Sourcetracker was implemented to observe the probability of finding microbial assemblages associated with families (source) with various surfaces in the built environment (sink) [28]. Posterior probabilities were plotted using the R package 'ggplot2' [29]. Sequences were retrieved from longitudinal studies [10], [14] observing the microbial community structure in residential homes, as well as the American Gut Project [30]. Only hand- and nares-associated samples were utilized for downstream processing with sequence data obtained from our studies and produced a participant sample size of n = 1,884. Samples were grouped by the originating individual for all studies, including American Gut Initiative, and selected for ESVs that were only observed in a single individual. Unique ESV counts were converted into presence versus absence using the R package 'vegan' and heatmaps were generated using the 'gplots' package. Random forest models were generated to discern the importance of environmental features on microbial community composition. uESVs were mapped to surfaces in the built environment using 'ggplot2'. Using the R package 'microbiome', we computed core microbial communities using only ESVs. Densities of ESVs found in this study were compared against those of the other studies.

To identify if uESVs in a surface microbiome community can be significantly attributed to interactions with a specific human microbiome we derived a probability estimate based on comparison to a database of 662 human skin microbiota. This analysis was based on some underlying observations. There are 'unique' ESVs found in surface microbiota not found in Intruders or Residents. Not all Resident and Intruder uESVs are represented in surface microbiota. This suggests that the Home microbial composition is larger and is comprised of more possible interactions than can be accounted for by only resident-surface or intruder-surface interactions. Therefore, our observations here are only a thin slice of a much larger network of interactions. 'Unique' human ESVs aren't really unique to the home microbiota. Some uESVs are found on surfaces and humans that *did not* interact. Therefore, for this analysis we considered 'Unique' ESVs to be merely 'rare' in the meta-human microbiota and to occur at a frequency of 1

in 662 (which is the total number of microbial samples we analyzed here). Finally, not every rare ESV found on a human resident can be found on their home's surfaces. Therefore, *absence* of a rare ESV on a surface cannot be considered evidence of non-interaction or true absence. We took a 'surface-centric' approach for the calculation of probability; i.e. what is the probability that the population of ESVs on a particular surface is significantly enriched for a *specific* human's composition of uESVs? In this framework, we can imagine a microbiome on a subway handrail that has *every* uESV found in an individual's skin microbiota but isn't significantly *enriched* for any of their uESVs. Therefore, the population on the subway handrail is so diverse that it has the possibility of harboring the uESVs of virtually everyone on the subway. Conversely, an infrequently touched lockbox that has a very limited uESV composition, say as deriving from 1 person that interacts with it, can be significantly statistically enriched for another person's uESVs even if it that individual only leaves behind 2 detectable uESVs. To calculate this, we considered the enrichment of a surface microbiome for a specific human microbiome using a Cumulative Binomial Distribution (CBD):

$$enrichment = 1 - \sum_{y=0}^{k} \binom{n}{y} p^n (1-p)^{n-y}$$

Where *n* = number of uESVs present on *surface*
k = number of uESVs shared by the surface and human microbiome
p = frequency of uESVs in a defined human population (=1/662)

From this, we can calculate a p-value for surface microbiome enrichment. However, to determine if even a 1-in-million match is statistically inevitable given the vast number of possible ESV combinations in a million human microbiota we calculated a Family Wise Error Rate (FWER). We set a threshold for significance using the CBN of 0.05. We calculate the uESV enrichment for all known human microbiota (n=662 in this case). The FWER is calculated:

FWER = # human uESVs found to be enriched on surface / Total # human samples in database

In this study, FWER will always be some integer number over 662. For the FWER table provided, we converted that decimal to the nearest integer for "# expected enriched uESVs / 1000 microbial samples".

RESULTS

Skin-microbiota profiles both discriminate between individuals and leave a residual signature on building surfaces.

Following filtering and normalization, 1,884,000 sequence reads were generated comprising 39,346 Exact Sequence Variants (ESVs) for both experiments. Experiment 1 was conducted to understand the transfer and stability of traceable microbiota to a specified surface (door handle). This longitudinal analysis was conducted in an office space under 2 different scenarios: either the presence or absence of office workers co-interacting with the door handles. An intruder interacted with the door handles, which was then be sampled every 5 mins for 1 hour to determine if the intruder's signature decayed or remained stable under both scenarios. Regardless of occupancy, the microbial community structure showed similar trends over time, i.e. showing an initial

decrease in similarity followed by an increase, and finally a large decrease (**Fig. 3A**), suggesting that the presence of office workers co-interacting with the door handles had no significant impact on the longitudinal variance in microbial community structure (ANOVA, $p > 0.05$) (**Fig. 3B**). Furthermore, a fitted regression generated to determine whether the intruder's uESVs found on the door handles decayed over time under the different scenarios found no significant shift in proportion of uESVs over time (**Fig. 3C-D**). This suggests that for at least 1 hour after an interaction event, traceable microbiota left behind by an individual may remain detectable on a built-environment surface irrespective of other people's random interaction with that surface.

Skin-microbiota profiles both discriminate between individuals and leave a residual signature on building surfaces.

For experiment 2 (the mock burglaries conducted in Naperville (Illinois, USA) and Fort Lauderdale (Florida, USA)), Bray-Curtis distances were generated to quantify microbial community beta diversity dissimilarity among residents and burglars. Both nares-associated microbiota (**Fig. 4A**; PERMANOVA, $p < 0.001$, $R = 0.731$) and hand-associated (**Fig. 4B**; PERMANOVA, $p < 0.001$, $R = 0.387$) had significantly greater beta diversity between individuals than within individuals, with nares microbiota exhibiting more variation between individuals than the hand microbiota (**Fig. 4C**). As hand and nares samples were acquired from each of the mock burglars both prior to and following the burglary, temporal stability of their microbiota was assessed for each site. Hand microbiota displayed a significantly greater shift (ANOVA, $p < 0.05$) in microbial community structure post burglary in comparison to nares microbiota, suggesting greater temporal instability in the hand microbiota; this is potentially as a result of the physical interaction between hands and home surfaces. Additionally, burglar pairs shared more microbial ESVs and shared more ESVs with the home surfaces, post burglary, suggesting acquisition of microbes from the home environment (**Table 3**).

Microbial community structure for the surfaces in the residential homes was also analyzed. The beta diversity (Bray Curtis dissimilarity) of each home was significantly different, while no significant difference was found based on regional location (i.e. Illinois versus Florida; **Fig. 3D**). Following mock burglaries, ESVs belonging to 10 bacterial genera were enriched on home surfaces; specifically, *Acinetobacter*, *Bacteroides*, *Bifidobacterium*, *Chroococcidiopsis*, *Chryseobacterium*, *Comamonas*, *Corynebacterium*, *Lactobacillus*, *Sphingomonas*, and *Streptococcus*; (**Fig. 4E**). When comparing home surfaces prior to and following burglaries, microbial community structures were highly dissimilar with an average Bray-Curtis dissimilarity of 76.9% (**Fig. 4F**), suggesting microbial community structure within the residential homes were altered following mock burglaries. The different proportions of bacterial genera across homes (**Supp. Fig 1A**), suggests that each home had an individually distinguishable background microbiota, and that this was only slightly altered by a burglary. Random Forest models demonstrated that variance in the proportion of differentially enriched taxa could mostly be accounted for by differences in the microbiome of the residence (**Supp. Fig. 1B**; OOB, $R = 0.27$). Sourcetracker was used to determine whether the surface associated ESVs found in homes post burglary had a significantly greater probability of having originated from one of the burglars compared to one of the residents or other people associated with the available datasets. While the likelihood of accurately detecting the specific burglar for a given home was very low (an

average of 25% for Naperville and 20.8% for Fort Lauderdale), the likelihood of predicting a Fort Lauderdale burglar as having invaded a Naperville home, or vice versa, was significantly lower than correctly matching the burglar to their specific city (**Fig. 4G**; ANOVA, $p < 0.05$).

Rare ESVs as trace evidence in residential homes

We analyzed the ESV profiles of all participants in this study (n = 30) in comparison to individuals who provided skin samples to the American Gut Project (n = 577), and created a database of ESVs that were unique to each individual; specifically, an ESV was unique to an individual if it was not shared between any two individuals among the observed population. We hypothesized that these unique ESVs (uESVs) may provide discriminate variants to detect a burglar's presence. The phylogenetic resolution of 16S rRNA amplicon sequencing is such that some ESVs may represent sequences from taxa that are inherently more variable than others, and as such those would be likely to be able to discriminate between individuals more accurately. Following filtering, 5,512 non-singleton uESVs were identified across all the residents and burglars (n=29; **Supp. Fig. 2**). uESVs were detected with a minimum of 2, a maximum of 568 and a mean of 13; suggesting some people maintain a greater degree of unique ESVs compared to our population (**Supp. Fig. 3**). The number of uESVs detected increased with the size of the population observed, which suggests that for at least a randomized selection of ~400 individuals, each person will have unique discriminators (**Supp. Fig. 4**). In total, 77% and 60% of the uESVs originated from the hands of the burglars and residents, respectively (**Fig. 5A**). While uESVs spanned 36 phyla, >70% of all uESVs belonged to Proteobacteria (34%), Bacteroidetes (15%), Firmicutes (12%), and Actinobacteria (11%) (**Fig. 5B**).

We then asked the question, what is the probability that surfaces in Home X are enriched for uESVs found in Suspect Y post burglary? To answer this question, we applied a cumulative binomial distribution analysis to identify uESVs that were enriched on a surface, and then applied additional family-wise error correction approach that accounted for the number of human microbiota compared in the analysis (n=662; **Table 4**). In the Naperville, Illinois homes, surfaces were significantly enriched (FWER-corrected p-value < 0.01) for uESVs from persons 8 and 13 (Naperville burglars) post burglary. While, in the Fort Lauderdale, Florida homes, surfaces were significantly enriched for uESVs from persons 6 and 7 (Fort Lauderdale Burglars) post burglary. Hence, of all the ESVs unique to the burglars, only certain ones were found to be significantly enriched on surfaces in the relevant homes. We next applied Sourcetracker to determine if the uESVs from the burglars were a significant source of ESVs to the surfaces in each home, and whether these proportions varied between uESVs that were significantly enriched, compared to those that were not. However, the likelihood of uESVs being a source to the home surfaces post-burglary did not differ, irrespective of whether they were significantly enriched post burglary compared to pre-burglary (**Fig. 5C**); likely because the burglars uESVs that were enriched were very rare. Interestingly, floor samples were always the most significantly enriched for burglar-associated uESVs post burglary, when compared to tables, dressers, doorknobs or counters, which suggests that microbes shed from burglars during their movements throughout the home are more likely to be deposited on the floor than they are on surfaces touched by hands (**Fig. 5D**). Sourcetracker was used to determine the percentage of uESVs that could be traced to the burglars for each surface (**Fig. 5E**). While burglars' uESVs were observed across all surfaces in

most homes, in Naperville homes 1, 2, and 3 uESVs from burglars were differentially detected across surfaces, suggesting less uniform distribution. Together, these data suggest that identifiable and discriminatory uESVs exist, and can be detected on home surfaces albeit at very low proportional abundance and with variable detection success.

DISCUSSION

This study adds further evidence that the human microbiome embellishes the built environment in a way that could be used to identify individuals interacting with a space. We were able to observe significant shifts in microbial community structure of building surfaces following repeated mock burglaries of residential homes that mirrored changes in the skin microbiota of the burglars. We also demonstrated that this signal could be reliably detected for up to 1-hour post invasion, irrespective of incidental interaction with that surface by other occupants. The results suggest that the act of transient, elicit interaction with a built space does have an observable impact of the microbial community both on built surfaces and human skin. However, since microbial retention may not sustain for more than an hour, longitudinal data from participants would be needed to determine the stability of these impacts and whether they continue to be observable for a long duration, as well as to determine the temporal variability in signal detection on different surface materials (we only examined this on metal door handles). That is, would it be possible to detect a signature of a burglar hours to days after an event, or conversely would the influence of a built environment on a burglar be detectable hours to days after the intrusion?? We also demonstrated that Exact Sequence Variants of the 16S rRNA V4-5 hypervariable region that are unique to individuals (based on a comparison of skin-microbiota from 600 people) are left behind on surfaces by burglar's post invasion of a home, and that these can be used to correctly identified the intruder. More importantly, we demonstrated that skin microbiota can be used to track individuals to various surfaces, which was previously believed to be unlikely due to frequent microbial turnover within the built-environment . This study is novel in that it uses the human microbiota as trace evidence to connect potential burglars to a crime scene, as well as observing the likelihood of the number of contributors [31].

Despite being inside each home for no more than 10 minutes, we observed changes in hand- and nares-associated microbiota for the burglars, as well as noticeable shift among numerous built-environment surfaces. This supports previous findings that humans shape the immediate spaces they interact with and that the built-environment serves as a place for human microbiota to accumulate [32]. More importantly, our findings demonstrate that given a number of possible microbiomes, an individual's microbial contribution can be detected on surfaces within a residential home. Our study was designed to determine the likelihood of detecting human intrusion into a space using their microbiome, but we could not properly assess the longevity of their microbial signature as the homes were not continuously sampled following each burglary. As databases and the number of human microbiomes grow, so will our statistical power and its ability to be implemented as a forensic tool.

We identified unique ESVs from a moderately sized human population (~600 individuals), which when we randomized to 400 people demonstrated that each individual has at least 13 uESVs on average. Interestingly, the majority of uESVs were annotated to the phylum Proteobacteria, a

common human skin associated phylum [33]. This is a significant advance on most microbial forensic investigations, which could not differentiate on whether microbial markers were from human sources or a result of dispersion from the environmental exposure [7], [8], [34], [35]. By cataloguing potentially unique ESVs we were able to say with greater certainty that an individual had been associated with the space for a period of time, although as before, this may not be true across time. Longitudinal analysis of the stability of unique ESVs would be required, and as of yet there are too few dense longitudinal time series of skin microbiota to be able to adequately assess this phenomenon. However, it is likely that as most of the uESVs detected had very low proportional abundance that more comprehensive sequencing of people (i.e. sequencing more reads per sample) may eradicate some of this uniqueness. Furthermore, we do recognize that it is not definitively clear whether the uESVs detected in this preliminary study would remain forensic identifiers of each individual were the population size were to increase 10- or 100-fold. Also, it is also plausible that some of the lowly abundant uESVs may be products of background noise likely from environmental dispersion considering samples were taken from skin. While we did generate uESVs for approximately 600 individuals to serve as a comparative control in order to remove false positives, we do acknowledge that it would have been better to compare uESV profiles for mock burglars against burgled and non-burgled homes. It is also likely, and untested by this study, that people who frequently physically interact are more likely to share a much greater proportion of uESVs, which obviously is a significant flaw for forensic utility.

Importantly, the FWER calculation we employed to determine the probability that enriched uESVs can predict a burglar's interaction with a space will be skewed due to the lack of truly independent observations in the database; for example, out of the 662 microbiota included in the FWER analysis, a large proportion of those included people living together, who due to sharing of microbes tend to have substantially similar microbiota. Another weakness of the FWER analysis, is that our determination of the frequency of uESVs as 1/662 is only a rough estimate of true frequency in the meta-human microbiome, due to substantial variance in the detection threshold of the total compliment of ESVs in anyone sample. To improve this analysis, we would need substantially more samples, sequenced at significantly greater depth, across multiple time points. We would need to improve the likelihood of true negatives, i.e. if an uESV is not detected in an individual we are certain that the probability it was missed due to a lack of observational depth is virtually improbable. Due to the inherent complexity of the microbiota and the way in which microbial communities are sampled, this outcome seems highly unlikely. As it stands, we can think of no feasible way to achieve this.

We have only begun to explore the potential of the human microbiome as a tool for forensics, and while our initial analysis, and that of previous studies, tends to suggest that it lacks obvious utility, further research is needed to absolutely confirm this finding. However, this study demonstrates that humans do leave behind traceable microbial signatures that can be used to track interaction with various surfaces within a building. Also, we observed that humans maintain rare microbiota that can be combined to create a unique microbial profile when compared to 600 other individuals; whether this uniqueness is an artefact of the population size under comparison or of the lack of depth or duration of observation in individuals, remains to be seen. Finally, we demonstrated that the unique microbial fingerprint of an intruder was detectable on a surface for up to 1 hour with

no obvious decay, irrespective of incidental use of these surfaces by other occupants. These findings present tantalizing evidence that the microbiota may be able to provide valuable forensic information in conjunction with more traditional methods. However, it is clear that there is a need to build better models to source individuals based on their microbial community composition, and to more accurately determine if a microbial signature is truly unique and stable for an individual. Future studies would benefit from focusing on the stability of traceable microbial markers over time in both people and on surfaces. Additionally, while the 16S rRNA gene is a serviceable and cheap indicator of microbial composition and structure, analyzing microbial metagenomes to identify genotype level differences, while more expensive and time consuming, might be better suited for forensic analysis. This will be true only if bacterial strains remain stable and unique to individuals over time, which due to the great dispersal rate of microbes between people maybe highly improbable. In summary, we believe the human microbiome, while having some potential value as a trace evidence marker for forensic analysis, is currently under-developed and unable to provide the level of security, specificity and accuracy required for a forensic tool.

Authors' contributions

JHM, JAG and JVL designed the study. KR, TAW, and LC were involved in DNA extraction, library preparation, and processing sequencing data. JHM interpreted the data, created figures, and wrote the manuscript. JAG, JVL, and GD participated in interpreting the data and revising manuscript. All authors read and approved the final manuscript.

Acknowledgements

This work was sponsored by National Institutes of Justice award 2015-DN-BX-K430. We thank members of the Gilbert and Lopez Lab in assisting in collection and de-identification samples and their compositional data, as well as assisting in data analysis. We thank all volunteers who provided consent for mock burglaries at their residence. The human study reported here was approved under Institutional Review Board Approval Number IRB16-0129 at the University of Chicago.

Availability of data and materials

All sequence data and associated metadata has been made available through Figshare.

REFERENCES

[1] C. J. Frégeau and R. M. Fourney, "DNA typing with fluorescently tagged short tandem repeats: a sensitive and accurate approach to human identification," *BioTechniques*, vol. 15, no. 1, pp. 100–119, Jul. 1993.

[2] S. Bhangar *et al.*, “Chamber bioaerosol study: human emissions of size-resolved fluorescent biological aerosol particles,” *Indoor Air*, vol. 26, no. 2, pp. 193–206, Apr. 2016, doi: 10.1111/ina.12195.

[3] R. You and C. Chen, “Measuring the Short-Term Emission Rates of Particles in the ‘Personal Cloud’ with Different Clothes and Activity Intensities in a Sealed Chamber,” *Aerosol Air Qual. Res.*, vol. 13, Jan. 2013, doi: 10.4209/aaqr.2012.03.0061.

[4] J. T. Hampton-Marcell, J. V. Lopez, and J. A. Gilbert, “The human microbiome: an emerging tool in forensics,” *Microb. Biotechnol.*, vol. 10, no. 2, pp. 228–230, Feb. 2017, doi: 10.1111/1751-7915.12699.

[5] S. Lax and J. A. Gilbert, “Forensic microbiology in built environments,” in *Forensic Microbiology*, John Wiley & Sons, Ltd, 2017, pp. 328–338.

[6] M. Richardson, N. Gottel, J. A. Gilbert, and S. Lax, “Microbial Similarity between Students in a Common Dormitory Environment Reveals the Forensic Potential of Individual Microbial Signatures,” *mBio*, vol. 10, no. 4, pp. e01054-19, Aug. 2019, doi: 10.1128/mBio.01054-19.

[7] N. Fierer, C. L. Lauber, N. Zhou, D. McDonald, E. K. Costello, and R. Knight, “Forensic identification using skin bacterial communities,” *Proc. Natl. Acad. Sci.*, vol. 107, no. 14, pp. 6477–6481, Apr. 2010, doi: 10.1073/pnas.1000162107.

[8] S. Lax *et al.*, “Forensic analysis of the microbiome of phones and shoes,” *Microbiome*, vol. 3, p. 21, 2015, doi: 10.1186/s40168-015-0082-9.

[9] J. F. Meadow, A. E. Altrichter, and J. L. Green, “Mobile phones carry the personal microbiome of their owners,” *PeerJ*, vol. 2, p. e447, Jun. 2014, doi: 10.7717/peerj.447.

[10] S. Lax *et al.*, “Longitudinal analysis of microbial interaction between humans and the indoor environment,” *Science*, vol. 345, no. 6200, pp. 1048–1052, Aug. 2014, doi: 10.1126/science.1254529.

[11] L. A. David *et al.*, “Diet rapidly and reproducibly alters the human gut microbiome,” *Nature*, vol. 505, no. 7484, pp. 559–563, Jan. 2014, doi: 10.1038/nature12820.

[12] H. Watanabe *et al.*, “Minor taxa in human skin microbiome contribute to the personal identification,” *PLOS ONE*, vol. 13, no. 7, p. e0199947, Jul. 2018, doi: 10.1371/journal.pone.0199947.

[13] A. E. Woerner *et al.*, “Forensic human identification with targeted microbiome markers using nearest neighbor classification,” *Forensic Sci. Int. Genet.*, vol. 38, pp. 130–139, Jan. 2019, doi: 10.1016/j.fsigen.2018.10.003.

[14] J. G. Caporaso *et al.*, “Moving pictures of the human microbiome,” *Genome Biol.*, vol. 12, p. R50, 2011, doi: 10.1186/gb-2011-12-5-r50.

[15] J. Oh, A. L. Byrd, M. Park, H. H. Kong, and J. A. Segre, “Temporal Stability of the Human Skin Microbiome,” *Cell*, vol. 165, no. 4, pp. 854–866, May 2016, doi: 10.1016/j.cell.2016.04.008.

[16] J. K. Goodrich *et al.*, “Genetic Determinants of the Gut Microbiome in UK Twins,” *Cell Host Microbe*, vol. 19, no. 5, pp. 731–743, May 2016, doi: 10.1016/j.chom.2016.04.017.

[17] J. F. Meadow *et al.*, “Humans differ in their personal microbial cloud,” *PeerJ*, vol. 3, p. e1258, Sep. 2015, doi: 10.7717/peerj.1258.

[18] A. Sharma *et al.*, “Longitudinal homogenization of the microbiome between both occupants and the built environment in a cohort of United States Air Force Cadets,” *Microbiome*, vol. 7, no. 1, p. 70, May 2019, doi: 10.1186/s40168-019-0686-6.

[19] A. Barberán *et al.*, “Continental-scale distributions of dust-associated bacteria and fungi,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 112, no. 18, pp. 5756–5761, May 2015, doi: 10.1073/pnas.1420815112.

[20] R. R. Dunn, N. Fierer, J. B. Henley, J. W. Leff, and H. L. Menninger, “Home Life: Factors Structuring the Bacterial Diversity Found within and between Homes,” *PLOS ONE*, vol. 8, no. 5, p. e64133, May 2013, doi: 10.1371/journal.pone.0064133.

[21] L. R. Thompson *et al.*, “A communal catalogue reveals Earth’s multiscale microbial diversity,” *Nature*, vol. 551, no. 7681, pp. 457–463, Nov. 2017, doi: 10.1038/nature24621.

[22] J. J. Minich *et al.*, “High-Throughput Miniaturized 16S rRNA Amplicon Library Preparation Reduces Costs while Preserving Microbiome Integrity,” *mSystems*, vol. 3, no. 6, pp. e00166-18, Dec. 2018, doi: 10.1128/mSystems.00166-18.

[23] W. Walters *et al.*, “Improved Bacterial 16S rRNA Gene (V4 and V4-5) and Fungal Internal Transcribed Spacer Marker Gene Primers for Microbial Community Surveys,” *mSystems*, vol. 1, no. 1, pp. e00009-15, Feb. 2016, doi: 10.1128/mSystems.00009-15.

[24] A. Amir *et al.*, “Deblur Rapidly Resolves Single-Nucleotide Community Sequence Patterns,” *mSystems*, vol. 2, no. 2, Apr. 2017, doi: 10.1128/mSystems.00191-16.

[25] B. J. Callahan, P. J. McMurdie, and S. P. Holmes, “Exact sequence variants should replace operational taxonomic units in marker-gene data analysis,” *ISME J.*, vol. 11, no. 12, pp. 2639–2643, 2017, doi: 10.1038/ismej.2017.119.

[26] P. J. McMurdie and S. Holmes, “phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data,” *PLoS ONE*, vol. 8, no. 4, p. e61217, Apr. 2013, doi: 10.1371/journal.pone.0061217.

[27] J. G. Caporaso *et al.*, “QIIME allows analysis of high-throughput community sequencing data,” *Nat. Methods*, vol. 7, no. 5, pp. 335–336, May 2010, doi: 10.1038/nmeth.f.303.

[28] D. Knights *et al.*, “Bayesian community-wide culture-independent microbial source tracking,” *Nat. Methods*, vol. 8, no. 9, pp. 761–763, Sep. 2011, doi: 10.1038/nmeth.1650.

[29] H. Wickham, “A Layered Grammar of Graphics,” *J. Comput. Graph. Stat.*, vol. 19, no. 1, pp. 3–28, Jan. 2010, doi: 10.1198/jcgs.2009.07098.

[30] D. McDonald *et al.*, “American Gut: an Open Platform for Citizen Science Microbiome Research,” *mSystems*, vol. 3, no. 3, pp. e00031-18, Jun. 2018, doi: 10.1128/mSystems.00031-18.

[31] H. Haned, L. Pène, J. R. Lobry, A. B. Dufour, and D. Pontier, “Estimating the number of contributors to forensic DNA mixtures: does maximum likelihood perform better than maximum allele count?,” *J. Forensic Sci.*, vol. 56, no. 1, pp. 23–28, 2011, doi: 10.1111/j.1556-4029.2010.01550.x.

[32] J. Chase *et al.*, “Geography and Location Are the Primary Drivers of Office Microbiome Composition,” *mSystems*, vol. 1, no. 2, pp. e00022-16, Apr. 2016, doi: 10.1128/mSystems.00022-16.

[33] P. H. Bradley and K. S. Pollard, “Proteobacteria explain significant functional variability in the human gut microbiome,” *Microbiome*, vol. 5, no. 1, p. 36, Mar. 2017, doi: 10.1186/s40168-017-0244-z.

[34] S.-Y. Lee, S.-K. Woo, S.-M. Lee, and Y.-B. Eom, “Forensic analysis using microbial community between skin bacteria and fabrics,” *Toxicol. Environ. Health Sci.*, vol. 8, no. 3, pp. 263–270, Sep. 2016, doi: 10.1007/s13530-016-0284-y.

[35] S. E. Schmedes *et al.*, “Targeted sequencing of clade-specific markers from skin microbiomes for forensic human identification,” *Forensic Sci. Int. Genet.*, vol. 32, pp. 50–61, Jan. 2018, doi: 10.1016/j.fsigen.2017.10.004.

FIGURE LEGEND

Figure 1. A schematic for experiment 1 is depicted demonstrating the burglar's interactions with connected offices during the absence (Day 1) and presence (Day 2) of other individuals.

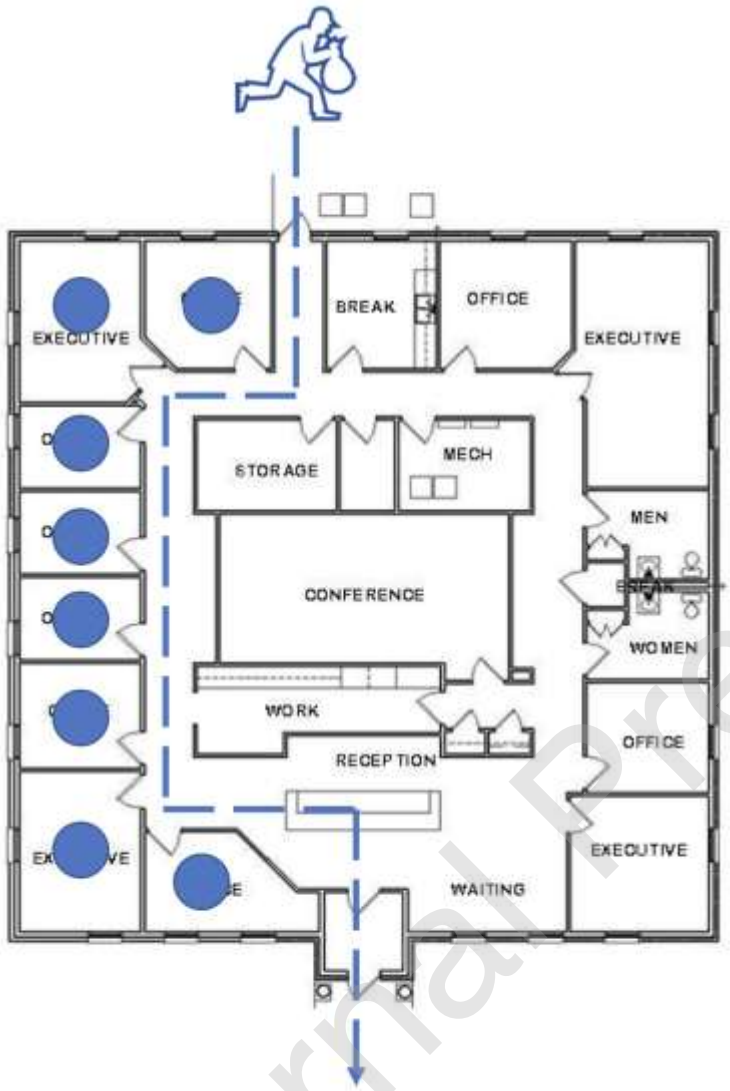

Figure 2. A schematic for experiment 2 is depicted demonstrating the sampling design for mock burglaries in Naperville, IL and Fort Lauderdale, FL. Swabbed samples were taken prior to (burglars, residents, and residential home) and following (burglars and residential home) mock burglaries.

Figure 3. (**A**) Temporal changes in microbial community structure were plotted for office door handle samples. (**B**) Overall change in microbial community structure were compared for Day 1 (absence of co-workers) and Day 2 (presence of coworkers). Means were compared to assess whether changes were significantly different between days. (**C**) Changes in burglar uESV abundances were tracked for both days. (**D**) Overall change in uESV abundances were compared and means were compared to asses significance.

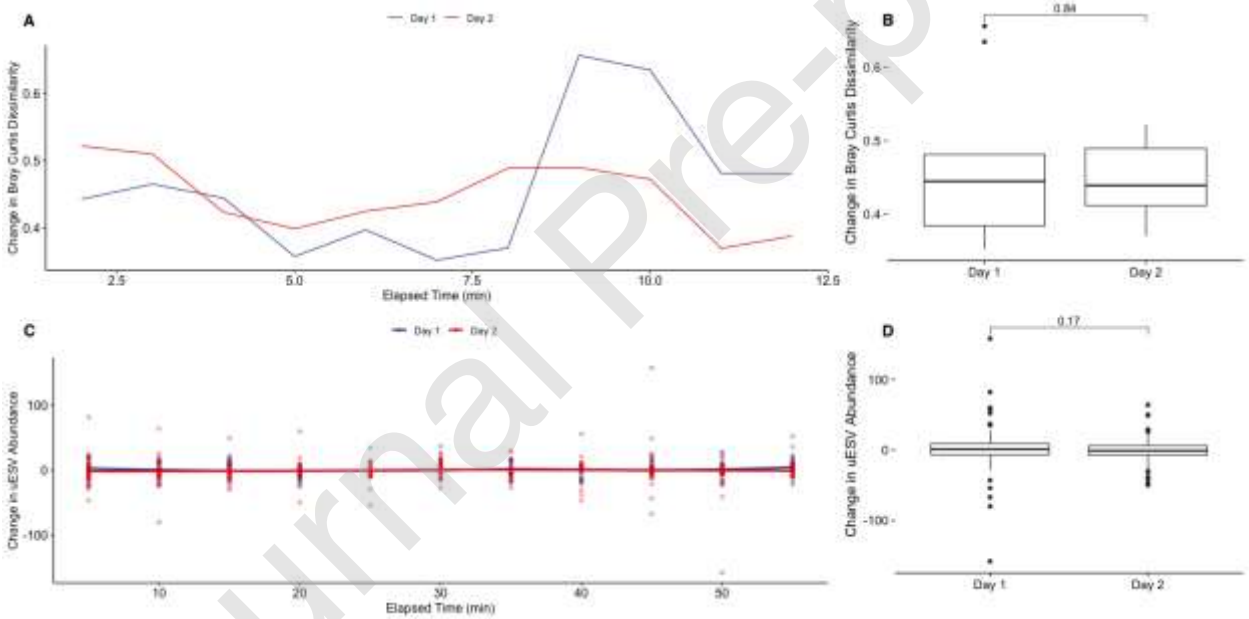

Figure 4. (A and B) Principle coordinate analysis (PCoA) were generated for nare and hand microbiota and colored by burglars in Naperville, IL and Fort Lauderdale, FL. (**C**) Densities were plotted to observe changes in microbial community structure for nare and hand microbiota. (**D**) Principle coordinate analysis (PCoA) was generated for residences in Naperville, IL and Fort Lauderdale, FL. (**E**) Log-fold change were detected using DESEQ2 to analyze differentially enriched taxa before versus after burglaries for surfaces within the residences. (**F**) Densities were plotted to observe changes in microbial community structure for residences. (**G**) Sourcetracker was utilized to report a burglar's likelihood of interacting with a residence.

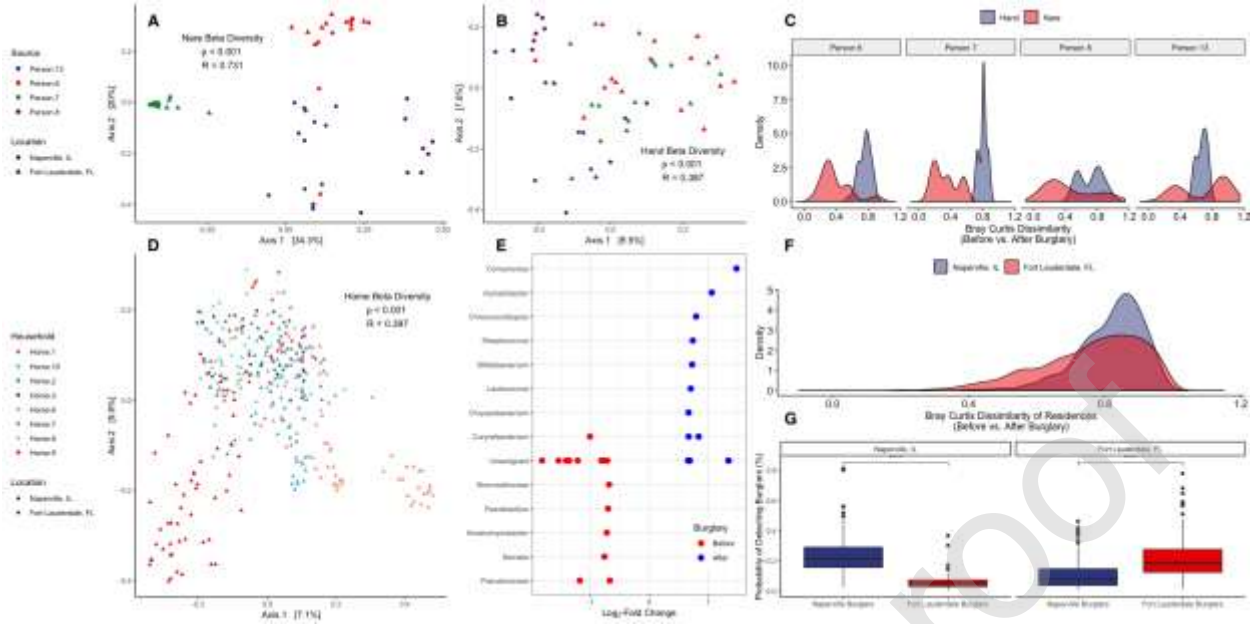

Figure 5. (**A**) Relative distribution of unique ESVs were plotted for hand and nare microbiota among burglars and residents. (**B**) Microbial composition of uESVs were plotted at the phylum for burglars and residents. (**C**) Probabilities were generated to discern between significant (ANOVA, $p < 0.05$) and non-significant (ANOVA, $p > 0.05$) and combined with Sourcetracker to plot the likelihood of detection. (**D**) Using Sourcetracker, uESV distribution was plotted among surfaces within residences in Naperville, IL and Fort Lauderdale, FL. (**E**) Using Sourcetracker, likelihood of burglar detection among various surfaces within residences were plotted.

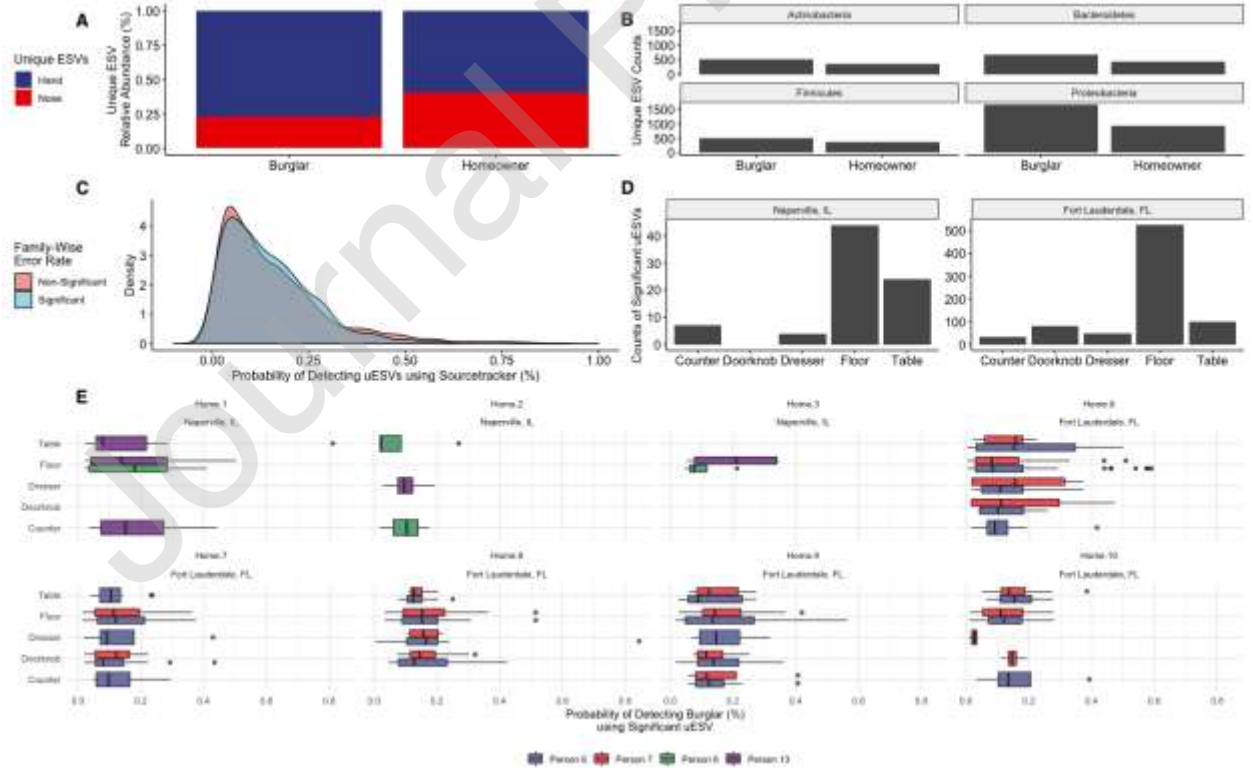

Table 1. For experiment 1, the elapsed time for single source interactions were recorded for both days, as well as the number of touches by resident office workers.

Date	Sample ID	Time Between Interaction and Sampling (min)	No. of Touches by Officer Worker
1/28/2018	FD009	20	0
	FD005	10	0
	FD011	25	0
	FD021	50	0
	FD015	35	0
	FD001	0	0
	FD003	5	0
	FD019	45	0
	FD017	40	0
	FD018	55	0
	FD019	30	0
	FD020	15	0
1/29/2018	FD021	40	3
	FD022	45	2
	FD023	10	1
	FD024	25	1
	FD025	35	5
	FD026	20	1
	FD027	15	2
	FD028	55	3
	FD029	0	2
	FD030	5	1
	FD031	30	2
	FD032	50	1

Table 2. For experiment 2, the number of occupants and their physical characteristics were detailed for each home including burglars interacting with residences during mock burglaries.

Residence	Location of home	Number of Occupants	Age ranges; gender of occupants	Age of burglars
Home 1	Naperville, IL	3	7-52; M,F	21-40

Home 2	Naperville, IL	4	4-39; M,F	21-40
Home 3	Naperville, IL	4	6-47; M,F	21-40
Home 6	Fort Lauderdale, FL	4	10 – 50; M,F	24-28
Home 7	Fort Lauderdale, FL	2	58-68; M, F	24-28
Home 8	Fort Lauderdale, FL	3	18 – 68; M,F	24-28
Home 9	Fort Lauderdale, FL	3	5 – 38; M, F	24-28
Home 10	Fort Lauderdale, FL	4	7-35: M, F	24-28

Table 3. Shared ESVs were analyzed prior to and following mock burglaries for burglar pairs, as well as the number of burglar ESVs observed among residences.

	Location	Shared ESVs	
		Before	After
Burglars	Naperville	157	821
	Fort Lauderdale	387	1562
Residences	Naperville	655	1177
	Fort Lauderdale	1103	2132

Table 4. Significance of enrichment for intruder microbiome signature by residence were calculated using FWER p-values, which indicate the probability that a given individual (out of 1,000) will match the uESVs of the indicated burglar.

Residence	Burglar ID	Enriched for uESVs FWER p-vlaue
Home 1	Person 8	0.002
	Person 13	0.000
Home 2	Person 8	0.006
	Person 13	0.008
Home 3	Person 8	0.002
	Person 13	0.006
Home 6	Person 6	0.000
	Person 7	0.002
Home 7	Person 6	0.000
	Person 7	0.003
Home 8	Person 6	0.000
	Person 7	0.002

Home 9	Person 6	0.000
	Person 7	0.003
Home 10	Person 6	0.000
	Person 7	0.002