



RESEARCH

Open Access



Microbiota of the indoor environment: a meta-analysis

Rachel I. Adams^{1*}, Ashley C. Bateman², Holly M. Bik^{3,4} and James F. Meadow²

Abstract

Background: As modern humans, we spend the majority of our time in indoor environments. Consequently, environmental exposure to microorganisms has important implications for human health, and a better understanding of the ecological drivers and processes that impact indoor microbial assemblages will be key for expanding our knowledge of the built environment. In the present investigation, we combined recent studies examining the microbiota of the built environment in order to identify unifying community patterns and the relative importance of indoor environmental factors. Ultimately, the present meta-analysis focused on studies of bacteria and archaea due to the limited number of high-throughput fungal studies from the indoor environment. We combined 16S ribosomal RNA (rRNA) gene datasets from 16 surveys of indoor environments conducted worldwide, additionally including 7 other studies representing putative environmental sources of microbial taxa (outdoor air, soil, and the human body).

Results: Combined analysis of subsets of studies that shared specific experimental protocols or indoor habitats revealed community patterns indicative of consistent source environments and environmental filtering. Additionally, we were able to identify several consistent sources for indoor microorganisms, particularly outdoor air and skin, mirroring what has been shown in individual studies. Technical variation across studies had a strong effect on comparisons of microbial community assemblages, with differences in experimental protocols limiting our ability to extensively explore the importance of, for example, sampling locality, building function and use, or environmental substrate in structuring indoor microbial communities.

Conclusions: We present a snapshot of an important scientific field in its early stages, where studies have tended to focus on heavy sampling in a few geographic areas. From the practical perspective, this endeavor reinforces the importance of negative “kit” controls in microbiome studies. From the perspective of understanding mechanistic processes in the built environment, this meta-analysis confirms that broad factors, such as geography and building type, structure indoor microbes. However, this exercise suggests that individual studies with common sampling techniques may be more appropriate to explore the relative importance of subtle indoor environmental factors on the indoor microbiome.

Keywords: Bacteria, Built environment, High-throughput sequencing, Microbiome, Negative controls, Source-sink dynamics

Background

The microorganisms in, on, and around our bodies comprise a large portion of the biodiversity we encounter in our lives. Our microbial associates impact our health, both positively and negatively. Understanding the processes that structure indoor microbial communities is an

important endeavor, since we spend the bulk of our time indoors and likely exchange many of our microbial passengers with various indoor habitats. In recent years, the scientific community has begun to recognize the importance of characterizing such human-associated habitats, with increasing numbers of studies seeking to determine the biodiversity, ecology, and public health implications of microbial assemblages present in the built environment. Investigations to date have included assessments of public restrooms [1, 2], hospitals [3–7], residences [8–18],

*Correspondence: adamsri@berkeley.edu

¹ Plant & Microbial Biology, University of California Berkeley, 94720 Berkeley, CA, USA

Full list of author information is available at the end of the article

university classrooms and office buildings [5, 19–22], artisan cheesemaking facilities [23], athletic facilities [24, 25], museums [5, 26], metropolitan subways [27–29], and even the evolutionary context of microbes indoors [30].

These individual studies typically target a specific location in order to elucidate the interrelationship between microbial communities and various environmental factors. Most recently, community assemblages have been described by targeting short fragments of ribosomal RNA (rRNA) genes and noncoding regions, such as the 16S rRNA gene in bacteria and archaea and the internal transcribed spacer (ITS) region in fungi. Gene fragments are amplified and sequenced from the pool of species present in environmental DNA samples, followed by data analyses that describe the community assemblage observed within a sample (α -diversity) as well as how much diversity is shared between different samples (β -diversity).

There is both promise and peril in combining separate studies into a meta-analysis. Individual efforts to characterize the microbes we encounter in buildings have shown that geography [10], building design and ventilation [31, 32], and occupant presence and activity [19, 20, 31, 32] can all contribute as drivers of indoor microbial communities. On the one hand, concurrent evaluation of these—and additional—studies has the potential to reveal large-scale biological trends and patterns in microbial community composition. On the other hand, there are several recognized limitations [33], with the dominant question being—can we detect true biological differences over the background of technical variation that exists across studies? In the case of microbial datasets, technical variation is defined as differences caused by experimental protocols, including but not limited to differences in the nature of the samples, DNA extraction methods, PCR amplification protocols, target genetic locus, sequencing primers, and sequencing platform. Differential experimental protocols can introduce significant bias and potentially obscure any meaningful biological patterns that may be observed across studies.

In this study, we conducted a meta-analysis of publicly available microbial datasets utilizing high-throughput sequencing (454, Illumina platforms) to investigate community patterns in the built environment. Additionally, supplementary datasets representing potential “source habitats” (human microbiome, outdoor environments) were analyzed alongside datasets from the indoor environment. Specifically, we aimed to address the following questions: (1) Are there consistent mechanisms evident across studies (biogeography, building operations, etc.) structuring microbial communities indoors? (2) Can we identify consistent source habitats for microbial communities in the built environment? (3) Are fungal and bacterial communities structured by the same processes across studies?

Results and discussion

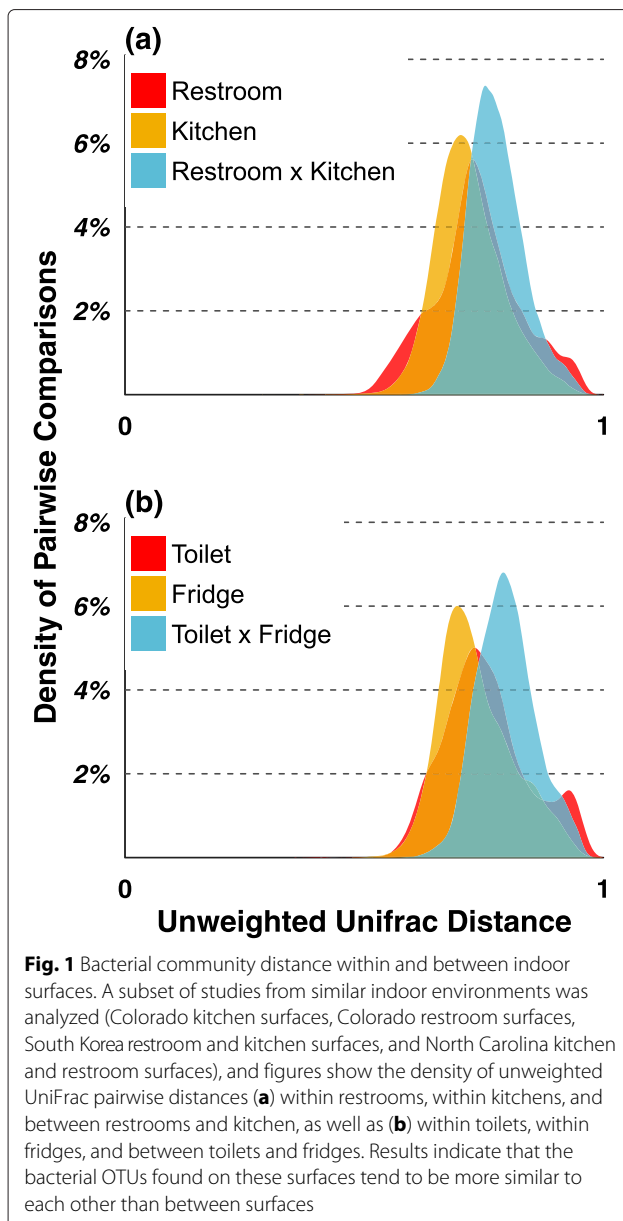
We compiled high-throughput 16S rRNA sequence data from 23 different studies, including 16 studies from built environments and seven from potential source environments, such as soil, the human microbiome, and outdoor air. We found that targeted biological comparisons were generally successful when using subsets of studies that shared a common experimental approach. Source tracking identified air and, to a lesser extent skin, as sources for indoor air, although for many samples, the sources were unidentified. In making comparisons across the entire set of studies, it was apparent that every dataset considered was unique in some aspect of habitat, sampling protocol, DNA extraction and amplification method, target 16S region, sequencing platform, or resulting sequencing depth and quality. Although our analysis approach mirrored that of other recent meta-analyses [33–35], unlike other studies, we found that technical variation across built environment studies overshadowed biological variation. This strong study-level effect was consistent across adjustments to the analysis pipeline. We conclude by discussing these findings in context of recommendations for future studies.

Biological insights

Studies with shared habitats

We looked for subsets of studies that examined identical habitats across broad geographic areas. The common indoor habitats were limited to bathrooms and kitchens, so for this analysis, we included four studies: South Korea homes [11], Colorado restroom surfaces [1], Colorado kitchen surfaces [12], and North Carolina homes [9].

For the human microbiome, some evidence exists that the same body sites across individuals are generally more similar to one another than different skin sites within a single individual [36]. We hypothesized that, much like the human body, kitchens and restrooms were the surfaces that were likely to reveal consistent microbial community patterns, since both rooms are among the most likely to accumulate moisture on surfaces, and they receive similar periodic inoculation from microbe-rich sources (such as humans and food). As with skin, site-level similarity has been noted across residential surfaces for bacteria [1, 9, 12, 14]. For the subset of studies considered here, bacterial communities generally did reflect their respective surface (Fig. 1). For instance, regardless of the study, we found that communities from toilets were more similar to other toilets than to other surfaces in kitchens or restrooms. Each surface or room type was dominated by bacterial OTUs consistent with the most likely human-influenced source. For example, plant chloroplasts, presumably food-based, were the dominant type in kitchens, representing 17 % of sequences, while in the bathrooms, chloroplasts represented only 3.6 % of



sequences. Conversely, skin-associated bacteria such as *Propionibacterium acnes*, *Corynebacterium*, and *Streptococcus* were dominant in the bathroom and less abundant than more environmental-associated bacteria *Acinetobacter* in kitchens, regardless of geographic location (South Korea, Colorado, and North Carolina).

Source tracking

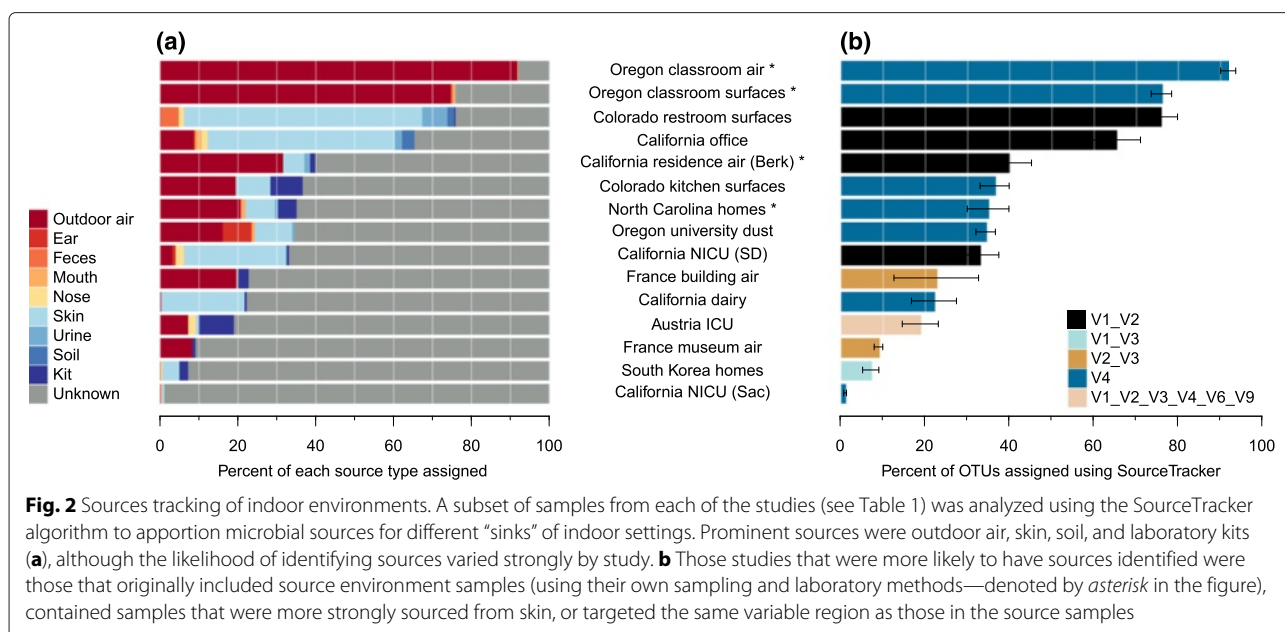
Source tracking is a Bayesian approach to estimate the proportion of a given “sink” community sample that is comprised of OTUs from a potential “source” sample [37]. For this study, sources were deemed to be outdoor air, soil, and human-associated samples (skin, feces, mouth, urine). Broadly, outdoor air and unidentified sources dominated

the sources for indoor air environments (Fig. 2a); outdoor air averaged a mean proportion of 0.52 (range 0.003–0.98) while unknown averaged 0.43 (range 0.016–0.99). Skin was the next most identified source with a mean proportion of 0.03 (range 0–0.25). Indoor surface environments, compared to airborne assemblages, tended to be more strongly sourced from human-associated taxa, with an average proportion of skin of 0.17 (range 0–0.96), and outdoor air contributing a similar amount (0.14; range 0–0.95). In looking within indoor surface types, individual sources became more important. For example, urine and feces were observed to be a more dominant source in bathrooms compared to other areas (Fig. 2a). Thus, from the biological perspective, source tracking results largely support the intuitive understanding of environment representing the most common source populations for microbial taxa that get dispersed indoors. These results also largely mirror what has been shown in individual studies (e.g., [9, 14, 17, 19, 32]).

From the perspective of combining studies in meta-analysis, our results suggest that site-specific sources may be particularly important for air environments (Fig. 2b). Although limited in number, two studies of bacteria in indoor air also had outdoor air samples [15, 32], and one study of settled dust was also accompanied by localized outdoor source samples representing air [9]. For these studies, outdoor air accounted for a mean proportion of 0.59 compared to 0.14 for those studies without study-specific designed outdoor source samples. Another study conducted in the same building [19] as a previous study that did include specific outdoor air samples [32] also showed a high proportion of outdoor air as the source. Thus, generic outdoor air sources were less informative than site-specific ones, indicating that bacteria in outdoor air can be highly localized [15, 32]. Moreover, we also observed differences in the power of generic sources to identify sources depending on the target variable region (Fig. 2b). Overall, this exercise suggests that processing even a few comparable outdoor samples alongside built environment samples may be much more effective for accurately identifying sources of indoor microbes versus analyses relying on a more extensive set of outdoor samples from another study.

Technical variation in indoor microbiome studies

When considering all studies together in principal coordinate analyses, a strong study effect is the most clearly discernible pattern, particularly when taxonomic-based metrics of ecological distance are used (Additional file 1: Figure S1) rather than a phylogenetically informed metric (Fig. 3a). While the results herein are presented based on UniFrac analysis, we discuss implications of this choice below. Clustering by study is perhaps unsurprising, since there are many opportunities for variations in sample



collection in the built environment depending on the study question being addressed (Table 1). For example, across studies, samples were collected from such different building types as homes, classrooms, hospitals, and industrial settings. Collection material included swabs of surfaces, vacuumed floor dust, and air subjected to filtration. For studies using swab sampling, collections were from such varied surfaces as toilet seats, kitchen counters, and door trims, and the material of the swab itself also varied across studies. All measured experimental parameters had a significant effect on community composition. The factors with the largest explanatory power for bacterial communities were individual study ($R^2 = 0.4$; Fig. 3a), geolocation, and specific sampling matrix (the physical sample type, e.g., dairy countertop versus toilet versus pillow, etc; $R^2 = 0.38$), as well as general sampling matrix (Fig. 3b; source of the sequenced material differentiating air, surfaces, dust, and water; $R^2 = 0.18$), and the use of the building (Fig. 3c; $R^2 = 0.17$). Individual studies were generally carried out in a single location with a single type of sampling method, so these top most explanatory variables were correlated with each other. For example, individual study were linked to geolocation ($R^2 = 0.20$) and building type ($R^2 = 0.49$). Thus, while we were able to reveal biological variation within the built environment, the inter-study variation hindered our ability to identify consistent mechanisms (e.g., biogeography and specific building operations).

To explore the individual taxa driving patterns observed in the PCoA, we used bi-plots, which display the subset of bacterial taxa exerting the most influence over community clustering and separation patterns. The top ten taxa

associated with indoor environments were recognizable as microbes associated with humans (e.g., *Corynebacterium*, *Streptococcus*, *Enterobacteriaceae*, *Staphylococcus*, *Propionibacterium*, *Lactococcus*) and outdoor habitats (e.g., *Streptophyta* [likely plant pollen], *Pseudomonas*, *Acinetobacter*, and *Sphingomonas*) (Fig. 3a).

Implications of analysis workflow

Given the large differences in approaches to sampling the indoor environment that magnified technical variation across studies, we wanted to explore how choices in the analysis workflow might have influenced this outcome. Specifically, we discuss database representation, ecological distance metrics, and OTU picking strategy.

Database representation

We used closed-reference workflows to generate operational taxonomic units (OTUs) using the Greengenes database, since it is not currently possible to conduct *de novo* OTU picking on datasets that include separate 16S gene regions. However, this choice potentially introduces a consequential bias since studies varied in the percentage of sequences that matched the reference database (Fig. 4). For instance, human microbiome studies tended to be well represented (greater than 90 % assigned), while soils were very poorly represented (most samples were less than 50 %; see Table 1). This variation is not entirely surprising, since different environments are unequally represented in the GreenGenes database; a large scientific effort has focused on human-associated microbes in recent years, while in contrast, the genetic diversity present in most natural ecosystems remains largely uncharacterized. Most



built environments in our meta-analysis ranged between these two extremes, which is to be expected since buildings include microbes sourced from both humans and outdoor environments to differing extents. While this study-level variation is interesting in itself, it may introduce a bias when using closed-reference OTU picking to compare across disparate studies and environments.

Choice of analysis metrics

Analysis of complex multivariate community data often necessitates the use of a distance or dissimilarity metric to compress many dimensions of variability into a single pairwise comparison. Four such metrics commonly used in sequence-based microbial community studies are Jaccard, Bray-Curtis, Canberra, and UniFrac

Table 1 Studies included in this meta-analysis

Study name	Total no. of samples per study	Target region	PCR primers	Sequencing platform	DNA extraction method	Avg. sequence length (bp)	Avg % assignment of sequences per sample	Sampling method	Geolocation	Sampling sites	Sourcetracker assignment	Citation	Data source
California residence air (Berk) ^a	59	V1–V2	8f_357r	454 GS FLX plus	Phenol:Chloroform MoBio PowerSoil DNA isolation kit	314	58.67 % (+/– 13.79 %)	Air (settle plate)	Albany, CA, USA	Rooms in apartment building (living room, balcony, bedroom, bathroom)	Sink/source	[14]	Contributed by author
California dairy ^b	86	V4	515f_806r	Illumina MiSeq	ZR96 Fecal DNA extraction kit	252	67.73 % (+/– 14.76 %)	Surface (cotton swab)	Davis, CA, USA	Rooms in industrial cheese processing building (wagon wheel room, scooping room, production room, brine_main room, bloomy_rind room, drying room, aging room, packaging room)	Sink	[23]	Qiime database
California NICU (Sac) ^b	147	V4	515f_806r	Illumina MiSeq	ZR96 Fecal DNA extraction kit	252	53.94 % (+/– 10.22 %)	Surface (cotton swab)	Sacramento, CA, USA	Hospital (neonatal intensive care units) surfaces and devices	Sink	[3]	Qiime database
Colorado kitchen surfaces ^{a,b,c}	285	V4	515f_806r	Illumina HiSeq2000	MoBio PowerSoil DNA isolation kit	89	67.9 % (+/– 27.1)	Surface (cotton swab)	Boulder, CO, USA	Kitchens in family homes	Sink	[12]	Qiime database
Colorado restroom surfaces ^{a,b,c}	126	V1–V2	27f_338r	454 GS Junior	MoBio PowerSoil DNA isolation kit	232	91.6 % (+/– 7.46 %)	Surface (cotton swab)	Boulder, CO, USA	Public restrooms	Sink	[1]	Qiime database
France building air	3	V2–V3	535f_789r	454 GS FLX Titanium	QiaAmp	228	92.23 % (+/– 1.86 %)	Air (filter)	Paris, France	Hospital, museum, office building	Sink	[5]	Contributed by author
South Korea homes ^c	28	V1–V3	9f_541r	454 GS Junior	FastDNA SPIN extraction kit	120	60.8 % (+/– 17.68 %)	Surfaces (Easy Swab Kit)	Seoul, Korea	Rooms in family homes (kitchen, bathroom)	Sink	[11]	Qiime database
California NICU (SD)	30	V1–V2	27f_338r	454 GS FLX	DNeasy Tissue Kit	221	94.22 % (+/– 4.67 %)	Surface (cotton swabs)	San Diego, CA, USA	Hospital (neonatal intensive care units)	Sink	[7]	Qiime database
California office	55	V1–V2	27f_338r	454 GS FLX	DNeasy Tissue Kit	227	93.95 % (+/– 4.07 %)	Surface (cotton swabs)	San Francisco, CA; New York, NY; Tucson, AZ	Office building objects	Sink	[22]	Qiime database

Table 1 Studies included in this meta-analysis *Continued*

Oregon university dust	173	V4	515f_806r	Illumina MiSeq	MO BIO PowerWater DNA Isolation Kit	130	85.67 % (+/- 7.83 %)	Dust (vacuum filter)	Eugene, OR, USA	University building (office, storage, classroom, food service)	Sink	[47]	Contributed by author
North Carolina homes ^c	384	V4	515f_806r	Illumina HiSeq 2000; Illumina MiSeq	directPCR	109	86.37 % (+/- 15.91)	Surface (Rayon-tipped swab)	Raleigh-Durham Area, NC, USA	Homes (condominium, low-rise apartment, family homes)	Sink/source	[9]	Qiime database
France museum air	3	V2-V3	Custom	454 GS FLX Titanium	Custom	263	87.42 % (+/- 0.58 %)	Air (cyclone sampler)	Paris, France	Art museum	Sink	[26]	Contributed by author
Oregon classroom air ^b	321	V4	515f_806r	Illumina MiSeq	MO BIO PowerWater DNA Isolation Kit	127	85.77 % (+/- 5.54 %)	Air (filter)	Eugene, OR, USA	University classroom	Sink/source	[19]	Contributed by author
Oregon classroom surfaces ^b	62	V4	515f_806r	Illumina MiSeq	MO BIO PowerWater DNA Isolation Kit	127	91.41 % (+/- 4.00 %)	Surfaces	Eugene, OR, USA	University classroom	Sink	[19]	Contributed by author
Austria ICU	24	V1-V2-V3-V4-V6-V9; V4	27f_1492r; 515f_927r	454 GS FLX plus	Custom XS PC	236	94.66 % (+/- 3.72 %)	Surfaces (BiSKits and nylon-flocked swabs)	Austria	Hopital (intensive care unit)	Sink	[6]	Contributed by author
Connecticut classroom air	15	V3-V4	331f_797r	454 GS FLX Titanium	MoBio PowerMax Soil DNA Extraction Kit	508	68.34 % (+/- 3.52 %)	Vacuum (dust and air), HVAC (air)	New Haven, CT, USA	University classroom	N/A	[20]	Contributed by author
Colorado mountaintop air	9	V1-V2	27f_338r	454 GS FLX Titanium	MoBio UltraClean PlantDNA Isolation Kit	225	83.38 % (+/- 20.64 %)	Air (filter)	Steamboat Springs, CO, USA	Aerial source of biota (atmosphere)	Source	[52]	Qiime database
Ireland elderly gut	168	V4	520f_802r	454 GS FLX Titanium	QIAGEN kit	221	93.83 % (+/- 3.16 %)	Surface (feces sample)	Cork, Ireland	Human gut (source)	Source	[53]	Qiime database
Colorado body sites	602	V1-V2	27f_338r	454 GS FLX Titanium	MoBio PowerSoil DNA isolation kit	229	92.76 % (+/- 4.56 %)	Surface (cotton swab)	Boulder, CO, USA	Human body sites (source)	Source	[36]	Qiime database
Europe/Africa children gut	29	V4-V5	784f_1061r	454 GS FLX Titanium	Custom proteinaseK phenolchloroform	247	93.24 % (+/- 3.17 %)	Surface (feces Sample)	Boulpon, Burkina Faso; Florence, Italy	Human gut (source)	Source	[54]	Qiime database

Table 1 Studies included in this meta-analysis *Continued*

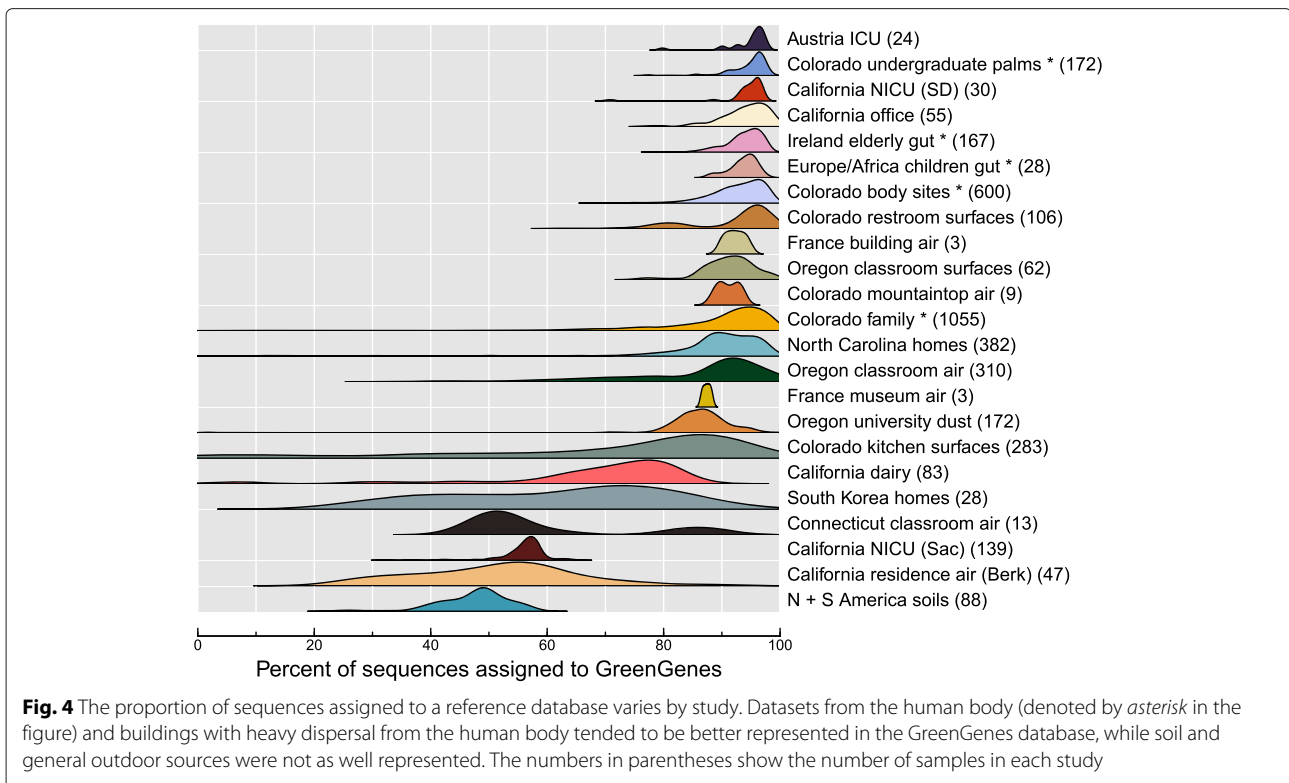
Colorado undergraduate palms ^a	182	V1–V2	27f_338r	454 GS FLX	MoBio UltraClean PlantDNA Isolation Kit	228	94.78 % (+/– 3.5 %)	Surface (cotton swab)	Boulder, CO, USA	Human palms (source)	Source	[43]	Qiime database
N + S America soils	89	V1–V2	27f_338r	454 GS FLX Titanium	MoBio PowerSoil DNA isolation kit	231	47.4 % (+/– 5.96 %)	Soil	USA, Puerto Rico, Peru, Argentina, Canada	Outdoor source of biota (soil)	Source	[55]	Qiime database
Colorado family	1071	V1–V2	27f_338r	Illumina GAIIx	MoBio PowerSoil DNA isolation kit	128	87.00 % (+/– 15.42 %)	Surface (cotton swab)	Boulder, CO, USA	Human and dog body sites (source)	Source	[56]	Qiime database

We combined high-throughput sequence data from 23 published, publicly available studies characterizing the built environment (sink) or studies that characterize potential sources of microbial dispersal into the built environment habitat (source). The Sourcetracker method requires a priori assignment of the samples in each of the studies as either an ecological source or sink, as indicated in the Sourcetracker Assignment column; three studies included source samples in their study design

^aIncluded kit controls

^bStudy also used in Open-reference OTU picking

^cShared habitats study



(including both the weighted and unweighted variations) [38]. The first three are classical taxonomic metrics that assume equal relationships among organisms [39], whereas UniFrac incorporates and emphasizes the phylogenetic relationships among OTUs.

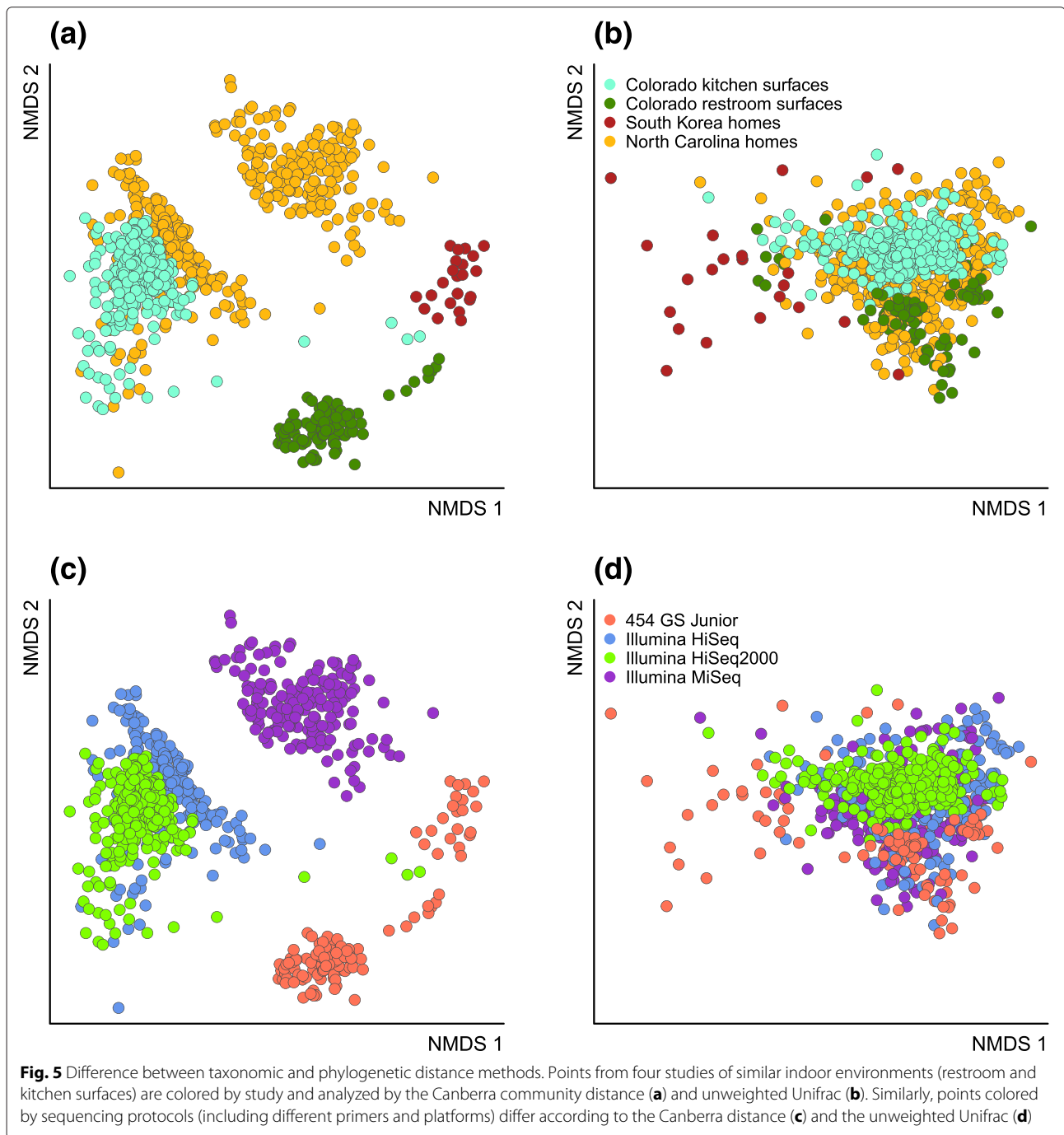
For many microbial community studies, taxonomic metrics can be desirable due to their ability to identify non-phylogenetic differences. For instance, the Jaccard metric can be evaluated as the simple proportion of taxa shared between two samples. The Bray-Curtis and Canberra metrics are both commonly used for their ability to prioritize shared distributions of the most abundant or rare taxa, respectively. The UniFrac metrics (weighted and unweighted), on the other hand, are heavily utilized in microbiome studies where broad-scale phylogenetic differences are important, such as when comparing the skin to gut microbiome where strong habitat differences favor entire functional groups or phylogenetic lineages. The trade-off is that a phylogenetic metric might miss out on subtle community differences when comparing very similar environments, while taxonomic metrics are inherently naive of functional community shifts.

Since the choice of distance/dissimilarity metric can be consequential, we compared the four metrics for their ability to overcome technical variation that results when disparate datasets are combined. Consistent with previous attempts [33–35], the unweighted UniFrac metric consistently succeeded in eliminating at least some of the

technical variation in our dataset, while study-to-study variation was overwhelmingly evident when employing any of the taxonomic metrics. Although there is no clear method to determine how well a metric reduces study-to-study variation, we evaluated metrics based on two characteristics: (1) the extent to which individual studies overlapped with other studies from similar environments, such as when comparing surface communities on toilets (Fig. 5), and (2) the distribution of dissimilarity values (Additional file 2: Figure S2). This second criterion is crucial when comparing disparate studies with potentially very few OTUs in common, which results in many pairwise observations at or near 1, as opposed to the relatively normal distribution when using UniFrac. Given that the unweighted UniFrac clearly masked some important sources of technical variation (such as sequencing protocols in Fig. 5), it yielded a normal distribution of pairwise comparisons. Since our questions in the present study were constrained to broad phylogenetic patterns, we felt confident using the UniFrac metric for all analyses.

OTU picking strategy

Qualitative conclusions based onto the two OTU-picking strategies were similar, in that patterns were generally consistent regardless of clustering method (Fig. 6). Moreover, the taxa indicative of each of the two environments overlapped between the two OTU-picking strategies. This exercise was useful for confirming that comparisons



between studies appear to be robust to slight differences in analysis parameters.

By selecting pairs of studies to explore with open-reference OTU picking (see the “Methods” section), we were able to ask whether certain taxa appeared to be associated with particular environments. The California dairy and California neonatal intensive care unit (NICU) did not show strong overlap in bacterial composition, and specific taxa seemed to be indicative of each dataset:

Lactococcus and the salt-associated *Pseudoalteromonas* were two taxa linked to dairy samples, while the NICU was dominated more by human *Enterobacteriaceae* and outdoor-associated bacteria *Acinetobacter*. Interestingly, an OTU of *Caulobacteraceae* was associated with the California NICU, and this bacterial family was also observed in a separate NICU study based in Pennsylvania [4]. Homes in North Carolina and kitchen surfaces in Colorado homes showed more overlap than the dairy

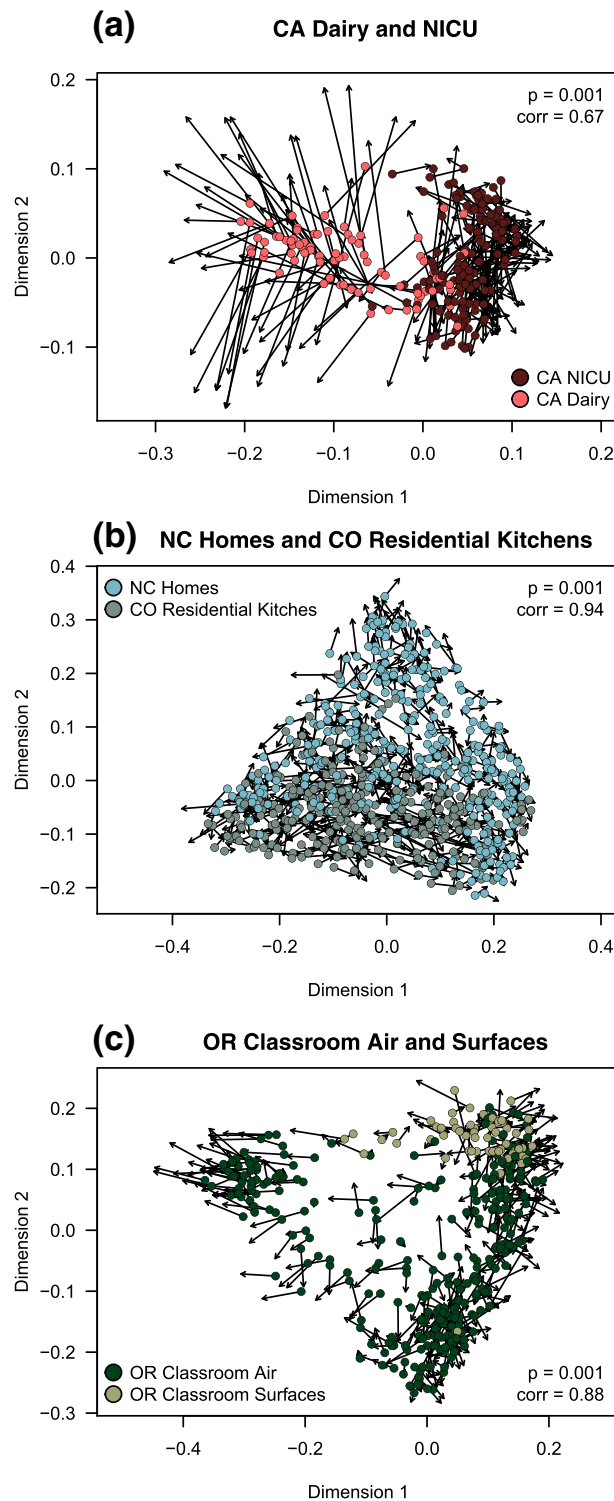


Fig. 6 Closed and open-reference OTU picking yielded similar β -diversity results. Three pairs of studies were separately analyzed using the two OTU picking methods and then compared using Procrustes analysis. Each *point* is the result of open-reference OTU picking, and each *arrowhead* is the same sample from closed-reference OTU picking. A significant Procrustes statistic indicates that the results from β -diversity analysis are strongly correlated. The same sample across the two methods are linked with an *arrow*. **a** California dairy and neonatal intensive care unit, both near Davis, California; **b** North Carolina homes in and near Raleigh, North Carolina, and Boulder, Colorado residential kitchen surfaces; **c** Oregon classroom air and surface samples from Eugene. Although the California Dairy study appears to be different between the two methods (**a**), the Dairy site was statistically distinct from the paired NICU study regardless of OTU picking method

and NICU. *Streptophyta*, *Enhydrobacter*, mitochondria, *Acinetobacter*, and *Pseudomonas* were more closely associated to kitchens while several human-associated groups including *Enterobacteriaceae*, *Streptococcus*, *Staphylococcus*, and *Corynebacterium* were found throughout homes. Like in homes, air and surfaces in an Oregon university classroom showed modest separation, with *Corynebacterium* and *Streptophyta* associated most strongly with surfaces while the air was more diverse, and linked to *Pedobacter*, *Microbacteriaceae*, *Sphingomonas*, *Oxalobacteraceae*, *Hymenobacter*, *Comamonadaceae*, and *Alicyclobacillus*.

Kit controls

While previously utilized in some HTS studies from the built environment [1, 9, 12–16, 29], the use of kit controls to check for the potential introduction of contaminant taxa when carrying out standard molecular protocols is an increasingly recognized issue for environmental sequencing studies [40–42]. Recently, Salter et al. [41] showed that laboratory reagents and commercial DNA extraction kits can harbor their own distinct microbial communities, with the composition and diversity of the “kit microbiome” varying across manufacturer. Contaminant taxa can have a particularly large impact on low biomass samples (which are common in built environment studies), with kit-derived microbes effectively introducing a sampling artifact that drives the resulting β -diversity patterns observed across samples. By using technical controls during sample processing (concurrently sequencing blank DNA extractions alongside all samples), studies can identify and remove kit-associated taxa, thus reducing technical artifacts and helping to elucidate the true biological differences among samples [41]. For the studies included in this meta-analysis, we found that most did not include technical controls to profile kit-associated taxa, or at least did not make any such blank samples publicly available. Only four studies made use of negative controls [9, 12, 14, 43], out of the 23 data sets examined in total. All of these studies relied on the recovery of material from swabs. As an example, we explored the profile of the kit controls using these handful of studies that diligently provided such data, but we stress that the presence of contaminant taxa are not unique to these studies.

We observed that kit controls appeared to have a distinct microbial profile compared to other samples, with some taxa, such as the bacterial phylum *Tenericutes*, exhibiting significant enrichment in kit controls versus environmental samples (Fig. 7). Further inspection revealed that most OTUs assigned to the *Tenericutes* fell within the bacterial class Mollicutes or the genus *Mycoplasma*. Both of these groups contain well-characterized taxa representing ubiquitous,

resilient laboratory contaminants of cell culture lines in particular [44, 45].

In contrast, the profile observed for other taxa such as Cyanobacteria (also potentially plant chloroplast sequences from pollen) showed an opposite trend, appearing abundant in dust samples while very few sequences from this taxon were found in the “kit microbiome” (Additional file 3: Figure S3). In the present meta-analysis, we used the results from technical controls to guide the downstream filtering of rRNA datasets and remove potential contaminating taxa, following the approach taken in [9]. While some entire taxonomic groups were easily removed due to their apparent contaminant status, such as *Tenericutes*, other taxonomic groups, such as *Corynebacterium* and *Staphylococcus*, were more difficult to remove since they are strongly human-associated (and could appear in kit controls due to “mistagging” of bar codes [46]), so removal could erroneously limit real biological insights. Notably, OTU picking strategy (open versus closed reference) generally did not increase the number of reads assigned to kit control samples (Additional file 4: Figure S4), suggesting contaminant taxa tend to be well represented in existing public databases and the kit microbiome can be sufficiently profiled even using closed reference workflows.

Qualitative comparison of included studies

Given the limitations we encountered with compiling the raw data across studies, we explored whether general biological conclusions across studies were shared. This approach was qualitative, as quantitative metrics, such as individual bacterial diversity estimates, are incomparable when experimental protocols vary. Indoor surfaces are highly influenced by the nature of human contact [1, 9, 11, 19] and predictable human behavior; for example, bacteria on kitchen surfaces appear to be influenced by the introduction of food [12], while surfaces most often touched by hands contain high proportions of skin-associated bacteria [1, 6, 12]. In non-residential settings, bacteria of outdoor origin may be a more influential source for surfaces, although human-associated microbes are still highly present [6, 19, 22]. Cleaning can reduce the human fingerprint [3, 9], and surfaces in industrial settings may show little influence of human contact [23]. The airborne bacteria found indoors also suggest a strong influence of human-associated bacteria as a source, in addition to outdoor-associated bacteria [14, 20, 26]. A longer sampling period of the air (on the order of weeks) demonstrates a shift towards human-associated bacteria in high-occupancy buildings [5, 20, 26], while shorter sampling periods (hours) with high outdoor ventilation suggest little effect of human occupancy on aerial bacterial composition indoors [32]. Features of building design, such as ventilation and connectedness of indoor spaces,

differing by building materials, maintenance, use, and location. Moreover, sampling within the built environment relies on different collection matrices (surface, air, dust) and materials (swabs, filters, wipes, etc.), as the particular research questions dictate the most appropriate sampling techniques. For instance, the focus of a study could be on time-resolved samples of microbes indoors, in which case vacuum filtration onto filters would be more appropriate strategies than collection of settled dust. The varied nature of the building microbiome complicates efforts to standardize methods. As a consequence of this spectrum of variation, statistical power is effectively reduced, making it much harder to detect overarching biological differences. Taken together, these issues increase the difficulty of conducting meta-analyses for built environment studies, more so than other environments such as the human body and soils. Combining data sets is an alluring prospect, whereby the power of combined data may be greater than individual studies in terms of inferring biological patterns. Unfortunately, the present investigation demonstrates that meta-analysis of combined data sets is not as straightforward as we would like. Importantly, we emphasize that this exercise does not question the findings of any one particular study, as technical variation typically does not exist to the same extent within an individual study.

For future studies of the built environment, there are a number of ways to move forward. Firstly, the inclusion of appropriate negative (“kit”) controls is of paramount importance for quality control of sequence data, so that studies can filter and remove well-characterized contaminant taxa and prevent erroneous biological inferences based on artifacts. While some study-level variation that we observed is likely due to differences in experimental protocols details above, it is also likely that lab-specific contaminants also cause communities to diverge, and including kit controls offer a way to identify these spurious taxa. Secondly, there are many opportunities for individual indoor environment studies—with a common sampling strategy—to expand in scope: increasing the number of samples collected, including sampling of potential source habitats for indoor microbes, increasing the number and type of buildings surveyed, expanding the geographic focus if appropriate to address the study questions, and including experimental manipulation of the built environment in order to test specific hypotheses. Lastly, as this field matures, standardized data collection and description methods of operational building characteristics (e.g., [48, 49]) will allow for more meaningful comparisons across disparate studies.

As the health implications of the indoor microbiome are continually tackled by the research community (e.g., [50, 51]), understanding the basic factors that govern the

potential pool of exposure is fundamental to modulating that environment. For example, it would be informative to be able to “rank” the mechanisms that structure microbial communities across different sampling types indoors. Specifically, how do biogeography, building function, ventilation type, and occupants and their activities interact to determine microbial composition, and how does that vary between airborne, dust, or surface microbes? Data presented here and elsewhere indicate that in the case of bathroom and kitchen surfaces, biogeography is less important than occupant activity, but source strengths are likely different for airborne microbes in those two residential locations. Moreover, it is unclear if the multitude of microbes identified in these different indoor environments are interacting (let alone alive), and how those interactions may affect its persistence in the built environment. Notable, fungi were excluded from full consideration in this meta-analysis due to limited study number, and it is unclear whether similar processes structure the two kinds of microbes. We are hopeful that as more built environment studies are conducted and made available, we will deepen our understanding of the global factors that structure and influence indoor microbial community assemblages.

Methods

Study inclusion criteria

We included studies if they met the following criteria: (1) published before May 15, 2014; (2) used high-throughput (HTS) amplicon sequencing to target 16S rRNA genes in bacteria/archaea or ITS rRNA in fungi; and (3) focused on built environments. We excluded several high-throughput studies due to low data quality compared to other studies (e.g., [13, 27, 31]), sequencing protocols that did not match the targeted single locus approach (e.g., [4]), and raw data that was not publicly deposited or otherwise obtainable (e.g., [8]). Table 1 describes some of the salient features of the 23 studies that met the above criteria and were thus included in the present meta-analysis [1, 3, 5–7, 9, 11, 12, 14, 19, 20, 22, 23, 26, 32, 36, 43, 47, 52–56].

Clone-based studies were not included in this study. In our search efforts for individual studies to include, we noted a large number of early built environment studies that relied on clone library construction followed by Sanger sequencing (e.g., [21, 57, 58]). While we acknowledge the potential contribution of such studies to a meta-analysis (providing an expanded set of geographical locations and a broadened survey of microbial diversity), preliminary exploration of these datasets indicated that clone libraries are unfeasible for inclusion with larger high-throughput datasets because of the orders-of-magnitude differences in sequence count and the fundamentally different laboratory protocols underlying such methods.

Preprocessing

Sequence datasets and sample metadata were either shared by the original authors, downloaded from public databases such as the NCBI Sequence Read Archive (SRA) or obtained from the QIIME online database, now superseded by the Qiita database (<http://qiita.ucsd.edu>).

Studies obtained from the QIIME database were downloaded as quality-filtered and demultiplexed datasets. For all other studies obtained from NCBI or the study authors, we implemented the same pre-processing quality filtering used for inclusion in the QIIME database, by setting parameters for the `split_libraries.py` (454 data) or `split_libraries_fastq.py` (Illumina data) scripts within QIIME version 1.8.0 [59]. Specifically, raw sequences from 454 Titanium and FLX platforms were excluded if not between 200 and 1000 nucleotides in length, had greater than six ambiguous bases, had homopolymer runs longer than 6 nucleotides, had mismatches in the primer, or could not be assigned to a sample using the barcode. For Illumina data, reads were truncated after more than three consecutive low-quality base calls and reads with <0.75 of the original read length remaining after truncation were subsequently discarded. Any reads containing ambiguous bases after quality trimming were also excluded. Because 454 Titanium chemistry yields longer read lengths, we trimmed all reads to the length achieved with standard FLX chemistry [33]. Pre-processed sequences from the QIIME database were then combined with sequences pre-processed in-house for downstream analysis.

Closed-reference OTU picking

Closed-reference operational taxonomic unit (OTU) picking followed the methods outlined in Lozupone et al. [33], using QIIME version 1.8. The closed-reference workflow is a database-dependent approach, using a pre-defined set of reference sequences with known taxonomy (the manually curated Greengenes database [60]) to cluster sequences into OTUs and assign taxonomy to environmental sequences. This approach is advantageous for comparing studies that target different 16S gene regions (e.g., V4 versus V9), since the underlying database is comprised of full-length gene sequences (≈ 1500 bp). However, in closed-reference OTU picking, taxonomic assignments are inherently constrained by the coverage of species and groups present in the reference database, and thus, results are limited to this subset of known microbial diversity. Closed-reference OTU picking was carried out using the `pick_closed_reference_otus.py` script with default parameters and the `-enable_rev_strand_match` flag. Identical sequences were first grouped using the prefix-suffix method implemented in QIIME, followed by clustering of representative sequences using UCLUST at 97 % sequence similarity against the reference sequences

in the Greengenes database (May 2013 release). Sequences that did not have at least 97 % identity to any reference sequences were discarded. The average percent of sequences assigned per study is described in Table 1. Finally, taxonomy was assigned to each representative OTU sequence using the corresponding Greengenes hit, generating an OTU table containing the sample taxonomy, sequence counts per sample, and metadata from the 23 individual studies. Downstream data exploration and β -diversity analyses were carried out using QIIME and R.

Few of the 23 studies that we evaluated included laboratory control samples, such as sequencing blanks from DNA extraction kits and reagents. This is increasingly recognized as a necessary component for low-biomass, high-throughput sequencing studies [41]. Putative contaminant taxa were removed from the OTU table when an abundant taxonomic group was identified as originating from laboratory contamination (that is, the taxa appeared only in studies originating from a single laboratory), and thus, their presence added to laboratory-centric technical variation. Specifically, control samples from two studies that provided control samples [1, 9] informed that OTUs in the phylum *Tenericutes* and orders *Oceanospirillales*, *Alteromonadales*, *EW055*, and *Tremblayales* were likely contaminants. Overall, our determination of contaminant taxa were constrained by the lack of laboratory control samples from the majority of studies. More robust analysis would likely result in the removal of hundreds of OTUs from this analysis and thus improve study-to-study comparison. Data exploration revealed that overall conclusions do not change without removal of these putative contaminant taxa.

To account for substantial variation in sequencing depth among studies, the OTU table was rarefied to 1000 sequences per sample, which is often sufficient to draw β -diversity conclusions in a variety of environments [61, 62]. For studies where metadata permitted, we removed samples that were clearly source environments (e.g., outdoor air), and used these in a SourceTracker analysis [37] to identify sources of indoor microbes. In addition to QIIME [59], the Phinch data visualization framework [63] was used to explore processed and rarefied OTU tables in order to explore biological patterns and compare microbial communities observed across studies.

Open-reference OTU picking

Since closed-reference OTU picking inherently constrains the number of individual OTUs recovered from environmental studies, and thus potentially limits β -diversity resolution, we ran additional analyses on a subset of studies in order to identify potential biases introduced through OTU picking. A subset of six studies representing the V4 region of the 16S rRNA gene (the most common

region analyzed across studies in this meta-analysis) were selected. We subdivided these studies into three pairs of datasets that utilized the same primer sequences and position on the 16S rRNA gene. The start and end positions of sequence reads within each dataset were confirmed by manually inspecting sequence alignments against the Greengenes 16S reference set. The final dataset pairs were (1) neonatal intensive care unit and dairy facility, both in California [3, 23], (2) two studies focused on the dust and air from an Oregon university classroom [19, 32], and (3) two studies of residential properties, one focused on North Carolina homes [9] and another focused on kitchens in Colorado [12]. For all dataset pairs, open-reference OTU picking was carried out using the `pick_open_reference_otus.py` script in QIIME, at 97 % sequence identity and 10 % subsampling of sequence reads not matching the Greengenes reference database. OTU picking was carried out using the `-enable_rev_strand_match` flag and default settings for all other parameters. Procrustes analysis was used to assess the influence of OTU-picking strategy on β -diversity inference.

Fungi

Fungal samples that relied on high-throughput sequencing in the built environment, at the time of analysis, were numbered at 469 across six studies (compared to nearly 4,000 samples across 23 studies for bacteria). Due to the limited number of samples and the limited overlap in community composition that was observed, fungi were not explored further (see Additional file 5: Text S1).

Availability of supporting data

Full documentation of all scripts, commands, and parameters used for data analysis in this study are available on GitHub (<https://github.com/jfmeadow/BEMAFinalAnalysis>). Final taxa tables (in .biom format) and mapping files are also available from the GitHub site.

Additional files

Additional file 1: Figure S1. Bray-Curtis ordination. Principal coordinate analysis (PCoA) of bacteria in the 16 “sink” studies using the Bray-Curtis taxonomic metric. Compare this to Fig. 3a in the main text. The top ten indicator taxa are shown as triangles. (PDF 105 kb)

Additional file 2: Figure S2. Pairwise distance observations. Comparison of the distributions of ecological distances among taxonomic and phylogenetic metrics. Y-axes on the histograms (along the diagonal) indicate distribution density, and all other axes are unitless and are bound by 0 and 1. The distribution of UniFrac distances (bottom panel) is nearly normal compared to the taxonomic distributions that are skewed toward 1. This suggests that this choice of unweighted UniFrac is best suited for ordination and β -diversity analysis. The scatter plots (off the diagonal) show the correlation of values between the pairs of metrics, indicating that

the different metrics are highly correlated with one another, and generally yield similar information. (PDF 3174 kb)

Additional file 3: Figure S3. *Cyanobacteria*. Example of taxonomic bias observed in technical control samples compared to environmental samples for taxa likely to be present in environmental samples. The composition of each of the pooled (a) dust and (b) kit controls is shown as a donut. The dust composition displayed higher abundances of *Cyanobacteria* (blue slice in donut chart indicated by arrows.) Per-sample abundance of *Cyanobacteria* is represented by blue bars displayed across all panels. Donut and bar charts were generated using the Pinch data visualization framework [63]. (PDF 146 kb)

Additional file 4: Figure S4. OTU picking strategy and kit controls. Comparison of OTU picking strategy on the sequences from the kit controls. Two studies were compared for the total number of sequence reads assigned in the kit controls, in closed-reference (a) versus open-reference (b) OTU picking workflows. One study, the North Carolina homes (24 control samples [9]), was remarkably similar in sequence reads assigned regardless of OTU picking method, while the other, Colorado kitchens (3 control samples [12]), was different. The relationship between the total number of assigned reads is shown in (c). (d and e) The taxonomic make-up of the samples is consistent at the phylum level between the two OTU picking strategies for one study but is different for another. (PDF 11 kb)

Additional file 5: Text S1. Fungal analysis. Description of the analysis steps for the fungal studies and basic findings, detailing why fungal studies were excluded from further analysis. (PDF 84 kb)

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

RIA, ACB, HMB, and JFM gathered the sequence data, performed the analyses, and wrote the paper. All authors read and approved the final manuscript.

Author details

¹Plant & Microbial Biology, University of California Berkeley, 94720 Berkeley, CA, USA. ²Biology and the Built Environment Center, Institute of Ecology and Evolution, University of Oregon, 97403 Eugene, OR, USA. ³UC Davis Genome Center, University of California, Davis, 95616 Davis, CA, USA. ⁴School of Biosciences, University of Birmingham, B15 2TT Birmingham, UK.

Acknowledgements

This manuscript was inspired by discussions at an “Evolution of the Indoor Biome” workshop sponsored by the National Evolutionary Synthesis Center and the Alfred P. Sloan Foundation (June 11–13, 2013 in Durham, NC). We would like to thank workshop organizers Craig McClain and Rob Dunn for leading this event and providing guidance and motivation for this manuscript. The authors gratefully acknowledge subsequent working group funding from NESCent and the Sloan Foundation which enabled in-person collaborative visits for data analysis. We also thank Jonathan Eisen, Brendan Bohannon and the Bohannon lab, Tom Bruns, and Jessica Green for their comments and constructive feedback on manuscript drafts.

Received: 27 April 2015 Accepted: 7 September 2015

Published online: 13 October 2015

References

- Flores GE, Bates ST, Knights D, Lauber CL, Stombaugh J, Knight R, et al. Microbial biogeography of public restroom surfaces. *PLoS One*. 2011;6(11):28132.
- Gibbons SM, Schwartz T, Fouquier J, Mitchell M, Sangwan N, Gilbert JA, et al. Ecological succession and viability of human-associated microbiota on restroom surfaces. *Appl Environ Microbiol*. 2014. doi:10.1128/AEM.03117-14.
- Bokulich NA, Mills DA, Underwood MA. Surface microbes in the neonatal intensive care unit: changes with routine cleaning and over time. *J Clin Microbiol*. 2013;51(8):2617–4. doi:10.1128/JCM.00898-13.
- Brooks B, Firek BA, Miller CS, Sharon I, Thomas BC, Baker R, et al. Microbes in the neonatal intensive care unit resemble those found in the gut of premature infants. *Microbiome*. 2014;2(1):1. doi:10.1186/2049-2618-2-1.

5. Gaüzère C, Godon JJ, Blanquart H, Ferreira S, Moularat S, Robine E, et al. 'Core species' in three sources of indoor air belonging to the human micro-environment to the exclusion of outdoor air. *Sci Total Environ*. 2014;485–486(0):508–17. doi:10.1016/j.scitotenv.2014.03.117.
6. Oberauner L, Zachow C, Lackner S, Högenauer C, Smolle KH, Berg G. The ignored diversity: complex bacterial communities in intensive care units revealed by 16s pyrosequencing. *Sci Rep*. 2013;3:1413. doi:10.1038/srep01413.
7. Hewitt KM, Mannino FL, Gonzalez A, Chase JH, Caporaso JG, Knight R, et al. Bacterial diversity in two neonatal intensive care units (NICUs). *PLoS ONE*. 2013;8(1):54703. doi:10.1371/journal.pone.0054703.
8. Nonnenmann MW, Coronado G, Thompson B, Griffith WC, Hanson JD, Vesper S, et al. Utilizing pyrosequencing and quantitative PCR to characterize fungal populations among house dust samples. *J Environ Monitor*. 2012;14(8):2038. doi:10.1039/c2em30229b.
9. Dunn RR, Fierer N, Henley JB, Leff JW, Menninger HL. Home life: factors structuring the bacterial diversity found within and between homes. *PLoS ONE*. 2013;8(5):64133. doi:10.1371/journal.pone.0064133.
10. Amend AS, Seifert KA, Samson R, Bruns TD. Indoor fungal composition is geographically patterned and more diverse in temperate zones than in the tropics. *Proc Natl Acad Sci*. 2010;107(31):13748–53. doi:10.1073/pnas.1000454107.
11. Jeon YS, Chun J, Kim BS. Identification of household bacterial community and analysis of species shared with human microbiome. *Curr Microbiol*. 2013;67(5):557–63.
12. Flores GE, Bates ST, Caporaso JG, Lauber CL, Leff JW, Knight R, Fierer N. Diversity, distribution and sources of bacteria in residential kitchens. *Environ Microbiol*. 2013;15(2):588–96.
13. Adams RI, Miletto M, Taylor JW, Bruns TD. Dispersal in microbes: fungi in indoor air are dominated by outdoor air and show dispersal limitation at short distances. *ISME J*. 2013;7(7):1262–73. doi:10.1038/ismej.2013.28.
14. Adams RI, Miletto M, Lindow SE, Taylor JW, Bruns TD. Airborne bacterial communities in residences: similarities and differences with fungi. *PLoS ONE*. 2014;9(3):91283. doi:10.1371/journal.pone.0091283.
15. Adams RI, Miletto M, Taylor JW, Bruns TD. The diversity and distribution of fungi on residential surfaces. *PLoS one*. 2013;8(11):78866.
16. Emerson JB, Keady PB, Brewer TE, Clements N, Morgan EE, Awerbuch J, et al. Impacts of flood damage on airborne bacteria and fungi in homes after the 2013 Colorado Front Range flood. *Environ Sci Technol*. doi:10.1021/es503845j. PMID: 25643125. <http://dx.doi.org/10.1021/es503845j>.
17. Lax S, Smith DP, Hampton-Marcell J, Owens SM, Handley KM, Scott NM, et al. Longitudinal analysis of microbial interaction between humans and the indoor environment. *Science*. 2014;345(6200):1048–52.
18. Wilkins D, Leung MH, Lee PK. Indoor air bacterial communities in hong kong households assemble independently of occupant skin microbiomes. *Environ Microbiol*. 2015. doi:10.1111/1462-2920.12889.
19. Meadow JF, Altrichter AE, Kembel SW, Moriyama M, O'Connor TK, Womack AM, et al. Bacterial communities on classroom surfaces vary with human contact. *Microbiome*. 2014;2(1):7.
20. Hospodsky D, Qian J, Nazaroff WW, Yamamoto N, Bibby K, Rismani-Yazdi H, Peccia J. Human occupancy as a source of indoor airborne bacteria. *PLoS One*. 2012;7(4):34867.
21. Pitkäranta M, Meklin T, Hyvärinen A, Nevalainen A, Paulin L, Auvinen P, et al. Molecular profiling of fungal communities in moisture damaged buildings before and after remediation—a comparison of culture-dependent and culture-independent methods. *BMC Microbiol*. 2011;11(1):235. doi:10.1186/1471-2180-11-235.
22. Hewitt KM, Gerba CP, Maxwell SL, Kelley ST. Office space bacterial abundance and diversity in three metropolitan areas. *PLoS ONE*. 2012;7(5):37849. doi:10.1371/journal.pone.0037849.
23. Bokulich NA, Mills DA. Facility-specific house microbiome drives microbial landscapes of artisan cheesemaking plants. *Appl Environ Microbiol*. 2013;79(17):5214–23. doi:10.1128/AEM.00934-13.
24. Bräuer SL, Vuono D, Carmichael MJ, Pepe-Ranney C, Strom A, Rabinowitz E, et al. Microbial sequencing analyses suggest the presence of a fecal veneer on indoor climbing wall holds. *Curr Microbiol*. 2014;69(5):681–9. doi:10.1007/s00284-014-0643-3.
25. Wood M, Gibbons SM, Lax S, Eshoo-Anton TW, Owens SM, Kennedy S, et al. Athletic equipment microbiota are shaped by interactions with human skin. *Microbiome*. 2015;3(1):25. doi:10.1186/s40168-015-0088-3.
26. Gaüzère C, Moletta-Denat M, Blanquart H, Ferreira S, Moularat S, Godon JJ, et al. Stability of airborne microbes in the Louvre Museum over time. *Indoor air*. 2014;24(1):29–40.
27. Robertson CE, Baumgartner LK, Harris JK, Peterson KL, Stevens MJ, Frank DN, et al. Culture-Independent Analysis of Aerosol Microbiology in a Metropolitan Subway System. *Appl Environ Microbiol*. 2013;79(11):3485–493. doi:10.1128/aem.00331-13.
28. Afshinnekoo E, Meydan C, Chowdhury S, Jaroudi D, Boyer C, Bernstein N, et al. Geospatial resolution of human and bacterial diversity with city-scale metagenomics. *Cell Syst*. 2015. doi:10.1016/j.cels.2015.01.001.
29. Leung MHY, Wilkins D, Li EKT, Kong FKF, Lee PKH. Indoor-air microbiome in an urban subway network: diversity and dynamics. *Appl Environ Microbiol*. 2014;80(21):6760–770. doi:10.1128/AEM.02244-14.
30. Martin LJ, Adams RI, Bateman A, Bik HM, Hawks J, Hird SM, et al. Evolution of the indoor biome. *Trends Ecol Evol*. 2015. doi:10.1016/j.tree.2015.02.001.
31. Kembel SW, Jones E, Kline J, Northcutt D, Stenson J, Womack AM, et al. Architectural design influences the diversity and structure of the built environment microbiome. *ISME J*. 2012;6(8):1469–79. doi:10.1038/ismej.2011.211.
32. Meadow J, Altrichter A, Kembel S, Kline J, Mhuireach G, Moriyama M, et al. Indoor airborne bacterial communities are influenced by ventilation, occupancy, and outdoor air source. *Indoor Air*. 2014;24(1):41–8.
33. Lozupone CA, Stombaugh J, Gonzalez A, Ackermann G, Wendel D, Vazquez-Baeza Y, et al. Meta-analyses of studies of the human microbiota. *Genome Res*. 2013;23(10):1704–14. doi:10.1101/gr.151803.112.
34. Ley RE, Lozupone CA, Hamady M, Knight R, Gordon JL. Worlds within worlds: evolution of the vertebrate gut microbiota. *Nat Rev Micro*. 2008;6(10):776–88. doi:10.1038/nrmicro1978.
35. Shade A, Gregory Caporaso J, Handelsman J, Knight R, Fierer N. A meta-analysis of changes in bacterial and archaeal communities with time. *ISME J*. 2013;7(8):1493–506.
36. Costello EK, Lauber CL, Hamady M, Fierer N, Gordon JL, Knight R. Bacterial community variation in human body habitats across space and time. *Science*. 2009;326(5960):1694–7. doi:10.1126/science.1177486.
37. Knights D, Kuczynski J, Charlson ES, Zaneveld J, Mozer MC, Collman RG, et al. Bayesian community-wide culture-independent microbial source tracking. *Nat Meth*. 2011;8(9):761–3.
38. Lozupone C, Lladser ME, Knights D, Stombaugh J, Knight R. UniFrac: an effective distance metric for microbial community comparison. *ISME J*. 2010;5(2):169–72. doi:10.1038/ismej.2010.133.
39. Legendre P, Legendre LF. *Numerical Ecology* vol 24. Oxford, UK: Elsevier; 2012.
40. Naccache SN, Greninger AL, Lee D, Coffey LL, Phan T, Rein-Weston A, et al. The perils of pathogen discovery: origin of a novel parvovirus-like hybrid genome traced to nucleic acid extraction spin columns. *J Virol*. 2013;87(22):11966–77. doi:10.1128/JVI.02323-13.
41. Salter SJ, Cox MJ, Turek EM, Calus ST, Cookson WO, Moffatt MF, et al. Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. *BMC Biol*. 2014;12(1). doi:10.1186/s12915-014-0087-z.
42. Lusk RW. Diverse and widespread contamination evident in the unmapped depths of high throughput sequencing data. *PLoS ONE*. 2014;9(10):110808. doi:10.1371/journal.pone.0110808.
43. Fierer N, Hamady M, Lauber CL, Knight R. The influence of sex, handedness, and washing on the diversity of hand surface bacteria. *Proc Natl Acad Sci*. 2008;105(46):17994–9.
44. Hwang JM, Lee JH, Yeh JY. A multi-laboratory profile of mycoplasma contamination in *Lawsonia intracellularis* cultures. *BMC Res Notes*. 2012;5(1):78.
45. Wang H, Kong F, Jelfs P, James G, Gilbert GL. Simultaneous detection and identification of common cell culture contaminant and pathogenic mollicutes strains by reverse line blot hybridization. *Appl Environ Microbiol*. 2004;70(3):1483–6.
46. Philippe E, Franck L, Jan P. Accurate multiplexing and filtering for high-throughput amplicon-sequencing. *Nucleic Acids Res*. 2015. doi:10.1093/nar/gkv107.
47. Kembel SW, Meadow JF, O'Connor TK, Mhuireach G, Northcutt D, Kline J, et al. Architectural design drives the biogeography of indoor bacterial communities. *PLoS ONE*. 2014;9(1):87093. doi:10.1371/journal.pone.0087093.

48. Glass EM, Dribinsky Y, Yilmaz P, Levin H, Van Pelt R, Wendel D, et al. MlxS-BE: a MlxS extension defining a minimum information standard for sequence data from the built environment. *ISME J.* 2014;8(1):1–3. doi:10.1038/ismej.2013.176.
49. Ramos T, Dedesko S, Siegel JA, Gilbert JA, Stephens B. Spatial and temporal variations in indoor environmental conditions, human occupancy, and operational characteristics in a new hospital building. *PLoS ONE.* 2015;10(3):0118207. doi:10.1371/journal.pone.0118207.
50. Nevalainen A, Täubel M, Hyvärinen A. Indoor fungi: companions and contaminants. *Indoor Air.* 2015;25(2):125–56. doi:10.1111/ina.12182.
51. Fujimura KE, Demoor T, Rauch M, Faruqi AA, Jang S, Johnson CC, et al. House dust exposure mediates gut microbiome lactobacillus enrichment and airway immune defense against allergens and virus infection. *Proc Natl Acad Sci.* 2014;111(2):805–10. doi:10.1073/pnas.1310750111.
52. Bowers RM, Lauber CL, Wiedinmyer C, Hamady M, Hallar AG, Fall R, et al. Characterization of airborne microbial communities at a high-elevation site and their potential to act as atmospheric ice nuclei. *Appl Environ Microbiol.* 2009;75(15):5121–30. doi:10.1128/AEM.00447-09.
53. Claesson MJ, Jeffery IB, Conde S, Power SE, O'Connor EM, Cusack S, et al. Gut microbiota composition correlates with diet and health in the elderly. *Nature.* 2012;488(7410):178–84.
54. De Filippo C, Cavalieri D, Di Paola M, Ramazzotti M, Poullet JB, Massart S, et al. Impact of diet in shaping gut microbiota revealed by a comparative study in children from Europe and rural Africa. *Proc Natl Acad Sci.* 2010;107(33):14691–14696. doi:10.1073/pnas.1005963107.
55. Lauber CL, Hamady M, Knight R, Fierer N. Pyrosequencing-based assessment of soil pH as a predictor of soil bacterial community structure at the continental scale. *Appl Environ Microbiol.* 2009;75(15):5111–120. doi:10.1128/AEM.00335-09.
56. Song SJ, Lauber C, Costello EK, Lozupone CA, Humphrey G, Berg-Lyons D, et al. Cohabiting family members share microbiota with one another and with their dogs. *eLife.* 2013;2:00458. doi:10.7554/eLife.00458.
57. Täubel M, Rintala H, Pitkäranta M, Paulin L, Laitinen S, Pekkanen J, et al. The occupant as a source of house dust bacteria. *J Allergy Clin Immunol.* 2009;124(4):834–40.
58. Kelley ST, Theisen U, Angenent LT, St Amand A, Pace NR. Molecular analysis of shower curtain biofilm microbes. *Appl Environ Microbiol.* 2004;70(7):4187–92. doi:10.1128/AEM.70.7.4187-4192.2004.
59. Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, et al. QIIME allows analysis of high-throughput community sequencing data. *Nat Methods.* 2010;7:335–6.
60. DeSantis TZ, Hugenholtz P, Larsen N, Rojas M, Brodie EL, Keller K, et al. Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Appl Environ Microbiol.* 2006;72(7):5069–72.
61. Kuczynski J, Liu Z, Lozupone C, McDonald D, Fierer N, Knight R. Microbial community resemblance methods differ in their ability to detect biologically relevant patterns. *Nat Methods.* 2010;7(10):813–9. doi:10.1038/nmeth.1499.
62. Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Huntley J, Fierer N, et al. Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *ISME J.* 2012;6(8):1621–1624. doi:10.1038/ismej.2012.8.
63. Bik HM. Phinch: an interactive, exploratory data visualization framework for -Omic datasets. *bioRxiv*, 009944 (2014) doi:10.1101/009944.

**Submit your next manuscript to BioMed Central
and take full advantage of:**

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit

