

## ORIGINAL ARTICLE

# Stochastic changes over time and not founder effects drive cage effects in microbial community assembly in a mouse model

Jonathan McCafferty<sup>1,5</sup>, Marcus Mühlbauer<sup>2,5</sup>, Raad Z Gharaibeh<sup>3</sup>, Janelle C Arthur<sup>2</sup>, Ernesto Perez-Chanona<sup>2</sup>, Wei Sha<sup>3</sup>, Christian Jobin<sup>2,4,6</sup> and Anthony A Fodor<sup>1,6</sup>

<sup>1</sup>Department of Bioinformatics and Genomics, University of North Carolina at Charlotte, Charlotte, NC, USA;

<sup>2</sup>Department of Medicine, Pharmacology and Immunology-Microbiology, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA and <sup>3</sup>Bioinformatics Services Division, Department of Bioinformatics and Genomics, University of North Carolina at Charlotte, Kannapolis, NC, USA

**Maternal transmission and cage effects are powerful confounding factors in microbiome studies. To assess the consequences of cage microenvironment on the mouse gut microbiome, two groups of germ-free (GF) wild-type (WT) mice, one gavaged with a microbiota harvested from adult WT mice and another allowed to acquire the microbiome from the cage microenvironment, were monitored using Illumina 16S rRNA sequencing over a period of 8 weeks. Our results revealed that cage effects in WT mice moved from GF to specific pathogen free (SPF) conditions take several weeks to develop and are not eliminated by the initial gavage treatment. Initial gavage influenced, but did not eliminate a successional pattern in which Proteobacteria became less abundant over time. An analysis in which 16S rRNA sequences are mapped to the closest sequenced whole genome suggests that the functional potential of microbial genomes changes significantly over time shifting from an emphasis on pathogenesis and motility early in community assembly to metabolic processes at later time points. Functionally, mice allowed to naturally acquire a microbial community from their cage, but not mice gavaged with a common biome, exhibit a cage effect in Dextran Sulfate Sodium-induced inflammation. Our results argue that while there are long-term effects of the founding community, these effects are mitigated by cage microenvironment and successional community assembly over time, which must both be explicitly considered in the interpretation of microbiome mouse experiments.**

*The ISME Journal* (2013) 7, 2116–2125; doi:10.1038/ismej.2013.106; published online 4 July 2013

**Subject Category:** Microbial population and community ecology

**Keywords:** cage; time; succession; microbiome; mixed linear models

## Introduction

Since Darwin's formation of the theory of evolution based on observations made in the Galapagos Islands, island ecology has played a key role in our

understanding of how communities form and respond to environmental pressure. The mammalian gut can be thought of as an island inhabited by a complex assemblage of microbes. It has been demonstrated in humans that the initial microbiota in different body sites is undifferentiated and the community assembly is strongly influenced by mode of delivery (Turnbaugh *et al.*, 2007). Over time, selection pressure on the human microbiota sculpts different microbial communities in each body site, so that, for example, the adult oral microbiota is largely distinct from the adult gut microbiota (Ursell *et al.*, 2012).

Mouse models of the gut microbiome are essential tools for studying the contribution of gut bacteria to human health and disease (Ley *et al.*, 2005; Turnbaugh *et al.*, 2006; Mazmanian *et al.*, 2008; Arthur *et al.*, 2012). Because mice can be raised in germ-free (GF) conditions and then inoculated with any cultivated microbe or microbial community

Correspondence: C Jobin, Department of Infectious Diseases and Pathology-College of Medicine, Department of Medicine, Division of Gastroenterology, Hepatology and Nutrition, University of Florida, Gainesville, FL, USA.

E-mail: Christian.Jobin@medicine.ufl.edu

or A Fodor, Department of Bioinformatics and Genomics, University of North Carolina at Charlotte, 9201 University City Boulevard, Charlotte, NC 28223, USA.

E-mail: afodor@uncc.edu

<sup>4</sup>Current address: Department of Infectious Diseases & Pathology - College of Medicine, Department of Medicine, Division of Gastroenterology, Hepatology & Nutrition, University of Florida, Gainesville, FL, USA

<sup>5</sup>These authors are joint first authors.

<sup>6</sup>These authors are joint senior authors.

Received 12 February 2013; revised 23 May 2013; accepted 28 May 2013; published online 4 July 2013

harvested from a human or mouse donor, the mouse model allows for experimental manipulations not possible in humans. Despite the power inherent to this model, recent studies have demonstrated the existence of potential confounding variables which, if uncontrolled, can complicate the execution and interpretation of experiments. It has been argued that apparent differences in the bacterial community initially thought to be driven by host genotype can in fact be better explained by direct microbial maternal transmission (Ubeda *et al.*, 2012). The use of GF mice avoids the pitfall of microbial maternal transmission because the mice are born and maintained in microbe-free conditions. Transferring GF mice into specific pathogen free (SPF) conditions, a process called conventionalization, allows mice instead to acquire bacteria from their environment. Whether using unique biome gavage over a conventionalization approach eliminates microbial community differences among experimental cohorts and functionally shapes their response remains unclear.

In a recent study, Hildebrand *et al.* (2013) showed that the variance in mice gut microbiota can be explained, to varying degrees, by the host's genotype, its cage microenvironment and inter-individual variation. In a previous study comparing the microbiome of conventionalized wild-type (WT) (colitis resistant) and *Il10*<sup>-/-</sup> mice (colitis susceptible) at 20 weeks, we found that the Proteobacteria *E. coli* was greatly expanded in the week 20 *Il10*<sup>-/-</sup> mice (Arthur *et al.*, 2012). We noticed a strong cage effect in which animals within the same cage had similar microbial communities. To account for these cage effects, we reported median values of each cage and verified that phenotypic difference between WT and *Il10*<sup>-/-</sup> mice was not due to these cage effects. In the present study, we wished to explore the causes and functional consequences of these cage effects and also to develop a formal statistical model to describe the cage effects. We therefore used Illumina sequencing to characterize the 16S gene from fecal samples collected over time (1, 2, 4 and 8 weeks after removal from GF conditions) from WT mice that were either gavaged with a 'typical' mouse gut microbiota harvested from adult WT mice, or allowed to acquire the microbiota from the cage microenvironment. We found that cage effects required several weeks to become significant and that while gavage had long-lasting effects on the recipient animals that appeared to influence Dextran Sulfate Sodium (DSS)-induced intestinal inflammation, it did not eliminate either cage effects or the succession of the gut microbial community over time. Our results suggest that stochastic differences that occur over time within each cage, rather than the composition of the initial microbial community, drive the formation of cage effects. Our results also demonstrate that whether or not an initial biome gavage is used, experimental

design must explicitly account for successional patterns over time and cage microenvironment or risk misinterpretation that could lead to flawed conclusions.

## Materials and methods

### *Acquisition of microbiota*

GF WT 129/SvEv mice (24 mice) were transferred to SPF housing conditions and the same day, either 12 mice were inoculated by gavage from an amalgamation of 4 WT 129/SvEv donor fecal samples from mice ranging in age from 2 to 3 months ('gavage' group) or 12 mice were allowed to naturally acquire a microbiota from the cage microenvironment ('acquire' group). Stool samples were collected and processed for sequencing following the protocol outlined by Arthur *et al.* (2012) at 1, 2, 4 and 8 weeks post transfer to SPF conditions (Supplementary Table S5). Mice were housed in eight cages with 2–4 mice per cage (four gavage cages and four acquired cages).

### *Illumina sequence pipeline*

One lane of paired-end Illumina 16S rRNA sequencing of the V6 hypervariable region produced 15 467 365 reads ~75 bases in length (excluding the primer sequences). The paired-end reads were merged, following the protocol described in Arthur *et al.* (2012), into 14 931 625 sequences with an average length of  $74.46 \pm 1.18$  (mean  $\pm$  s.d.). Because our read length (100 bp) was longer than our amplicon size (75 bp), we had  $2 \times$  coverage on every read and could therefore remove sequences if the paired sequences were not in good concordance. Clustering those reads into Operational Taxonomic Units (OTUs) by AbundantOTU version 2.0 (<http://omics.informatics.indiana.edu/AbundantOTU/>) took 106 min and produced 873 OTUs using a 97% threshold incorporating 99.996% of all sequences and removing 50 863 singletons from downstream analysis. Chimera detection using UCHIME (<http://www.drive5.com/uchime/>) (Edgar *et al.*, 2011) and the Gold reference database identified five OTUs, which we then removed from downstream analysis. For taxonomic classification, the AbundantOTU consensus sequences were mapped to the Silva 108 database (<http://www.arb-silva.de/>) by BLASTn with an expectation value threshold of  $e-10$ . The top hits were selected and sent through RDP classifier version 2.1 (<http://sourceforge.net/projects/rdp-classifier/>) (Wang *et al.*, 2007) with an RDP confidence threshold of 80% or greater used for assignment. Reads were deposited in MG-RAST under project ID 4514986.

### *Statistical analysis*

OTU consensus sequences were collapsed into pivot table format where each row represents a sample and each column contains the raw counts for

each OTU consensus sequence. Raw counts were transformed using a log frequency calculation before use in downstream analysis:

$$\text{Log}_{10}\left(\frac{\text{RC}}{n} \times \frac{\sum x}{N} + 1\right)$$

where RC represents the number of raw counts in a column cell (OTU, phyla, and so on) for a sample,  $n$  is the number of sequences in a sample, the sum of  $x$  is the total number of counts in the table and  $N$  is the total number of samples.

Bray–Curtis dissimilarity matrixes were generated from normalized data and Principle Co-ordinate Analysis (PCoA) was conducted through the use of the software package mothur (Schloss *et al.*, 2009). Bray–Curtis dissimilarity has been shown to produce broadly similar results as other distance metrics (for example, see Supplementary Figure 1 in Claesson *et al.*, 2012). Cage effects were accounted for by incorporating mixed linear models utilizing SAS (Supplementary code Table S6) where cages were the random effects and genotype or treatment was fixed. Benjamini–Hochberg method for false discovery rate (FDR) correction was used for multiple testing correction. Richness is defined as the number of distinct OTUs in each sample. To correct for different numbers of sequences in each sample, we randomly subsampled (without replacement) 11 368 sequences in each sample (where 11 368 is the number of sequences in the sample with the smallest number of sequences) 1000 times and reported as richness the average number of OTUs seen across these 1000 re-samples. All statistical analyses were conducted through R (<http://www.r-project.org/>), custom Java code (available upon request) and SAS version 9.2 (SAS Institute Inc, Cary, NC).

Mixed linear models have many advantages including a solid theoretical base (Raudenbush and Bryk, 2002; Smyth, 2004), wide utilization in the literature (Brown *et al.*, 2011; Listgarten *et al.*; Ross *et al.*, 2012; Vilhjalmsson and Nordborg, 2013) and robust implementations in statistical packages such as R and SAS. However, mixed linear models impose an additional set of parametric assumptions over canonical linear models. In our case, they assume that the effects of cages are normally distributed with a mean of zero. These assumptions may be particularly inappropriate for metagenomics data, where it has been argued for the gut microbiome that only a few possible outcomes (or enterotypes) are likely (Arumugam *et al.*, 2011). While the enterotype hypothesis has been highly controversial (Huse *et al.*, 2012; Jeffery *et al.*, 2012; Segata *et al.*, 2012), it seems unlikely that the opposite assumption that there is no repeatable structure to the microbial community within cages is broadly true. A finite subset of possible cage outcomes might therefore violate the assumptions of mixed linear models. With this in mind, we compared our results with a simple model in which the median value for each cage was fed into a canonical two-way analysis of variance

(ANOVA) (data not shown). We saw a broadly similar pattern of  $P$ -values with this approach, although as we might expect this median-based linear models appeared to have substantially lower power than the full mixed linear model. Future research will undoubtedly pursue the question of the most appropriate model for cage effects that makes the fewest assumptions while preserving the most power, but the broad concurrency of the median and mixed linear models is encouraging in that it suggests that our results are not primarily driven by the additional parametric assumptions about cage distribution in the mixed linear model.

#### Functional prediction

Phylogenetic Investigation of Communities by Reconstruction of Unobserved States (PICRUSt) (<http://picrust.github.com/picrust/>) is a software package designed to infer metagenome functional content from 16S metagenomic data. The paired-end merged 16S sequences were used for closed-reference OTU picking using MacQIIME (Caporaso *et al.*, 2010) (<http://www.wernerlab.org/software/macqiime>) version 1.6.0, the resulting OTU table was then fed into PICRUSt version 0.9.1 and functional predictions were made according to the metagenome inference workflow described by the developers (<http://picrust.github.com/picrust/tutorials/quickstart.html#quickstart-guide>). PICRUSt results were normalized and log<sub>10</sub> transformed according to our normalization equation, and then analyzed using the mixed linear model where  $Y$  represents the log<sub>10</sub> abundance of each ortholog.  $P$ -values evaluating the null hypothesis that the ortholog was observed equally in each treatment or time point were generated for each ortholog. For every ortholog that was significantly different at an FDR adjusted  $P$ -value of  $<0.1$ , we used the KEGG pathways provided by PICRUSt to ask whether orthologs belonging to a particular functional category (for example, Cell Motility and Transcription) were likely to be significantly more abundant in one treatment/time point than another. For this purpose, we utilized Fisher's exact test to assess for each functional category the null hypothesis that there was an equal number of its orthologs present at higher or lower relative abundance between the treatments/time points.

#### DSS-induced acute injury and histological evaluation

Following the 8-week period, gavaged or acquired WT mice were given 2% DSS (MP Biomedicals, Aurora, OH, USA) in their drinking water for 13 days while control mice received water alone. Water consumption was comparable between the different experimental groups. Mice were monitored daily for weight loss and visible signs of rectal bleeding. Occult bleeding was evaluated 4 days after administration of DSS or water control (Hemocult;

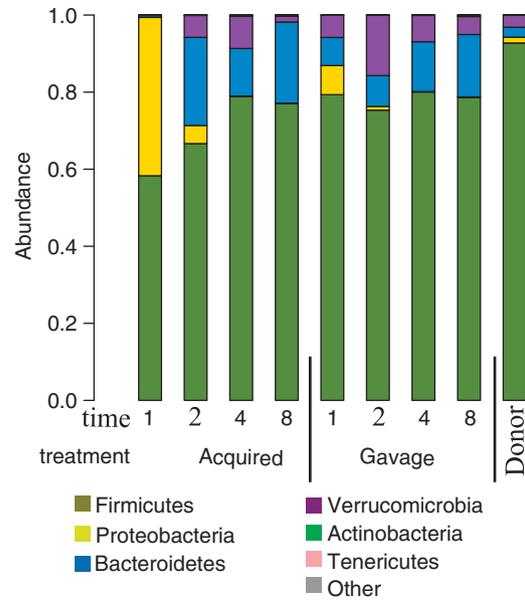
Beckmann Coulter Inc., Fullerton, CA, USA), in accordance with internal pilot studies that showed occult bleeding to be reliably observable in this model from day 4 onwards. Clinical score, assessing weight loss, occult blood and stool consistency were calculated as previously described (Goldsmith *et al.*, 2011). Mice were killed at the indicated time points by CO<sub>2</sub> asphyxiation followed by cervical separation. The colon was dissected and flushed with ice-cold phosphate-buffered saline, longitudinally splayed, swiss rolled, fixed in 10% formalin for 24 h, and then embedded in paraffin. Colitis severity was evaluated using Hematoxylin–Eosin-stained sections by a blinded investigator on a scale from 0 to 40, as described previously (Goldsmith *et al.*, 2011).

All animal experiments were approved by the Institutional Animal Care and Use Committee of the University of North Carolina at Chapel Hill.

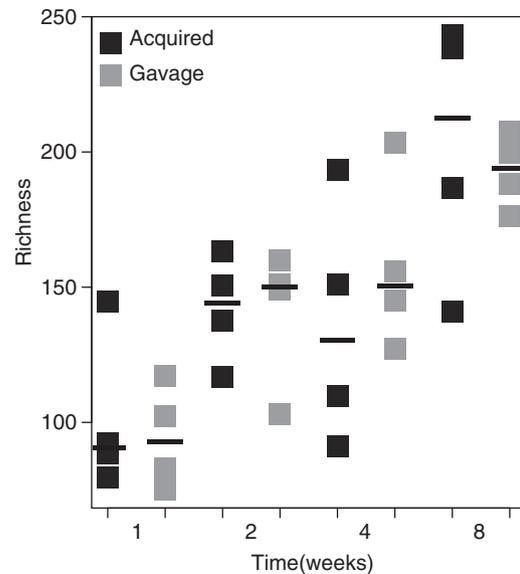
## Results

*Gavage can modulate the microbial community, but does not eliminate cage effects or succession over time*  
 In a previous study (Arthur *et al.*, 2012), we observed pronounced cage effects with distinct microbial communities present in different cages. To determine if such cage effects could be eliminated by an initial common biome gavage, GF WT mice were transferred to SPF conditions, where one cohort was immediately associated with a unique donor microbial community pooled from adult WT mice (hereafter referred to as the ‘gavage’ treatment) while another cohort was allowed to acquire their microbial community from the cage environment (hereafter referred to as the ‘acquired’ treatment). Fecal samples were collected at weeks 1, 2, 4 and 8 following removal from GF conditions and the microbial community was characterized by paired-end HiSeq Illumina sequencing targeting the V6 region of the 16S rRNA gene.

An examination of the results from all animals at the phyla level (Figure 1) demonstrates that at the 1-week time point, the gavage-treated animals appeared to have a microbial community that was in some ways a mixture of the donor community (Figure 1, rightmost bar) and the community in the ‘acquired’ group. The contribution of Proteobacteria (7.5%) in the gavage group was intermediate to the large fraction (41%) of Proteobacteria in the acquired cohort and the smaller fraction in the donor biota (1.5%). It appears, therefore, that the donor community influenced, but did not completely seed, the resulting microbial community at 1 week. Over time, the fraction of Proteobacteria decreased in both the acquired and gavage groups. By week 8, the phyla view of these two groups was very similar (Figure 1). Richness in both the gavage and treatment groups increased over time (Figure 2), suggesting that the initial seeding of the microbial community by gavage did not give the gavage group



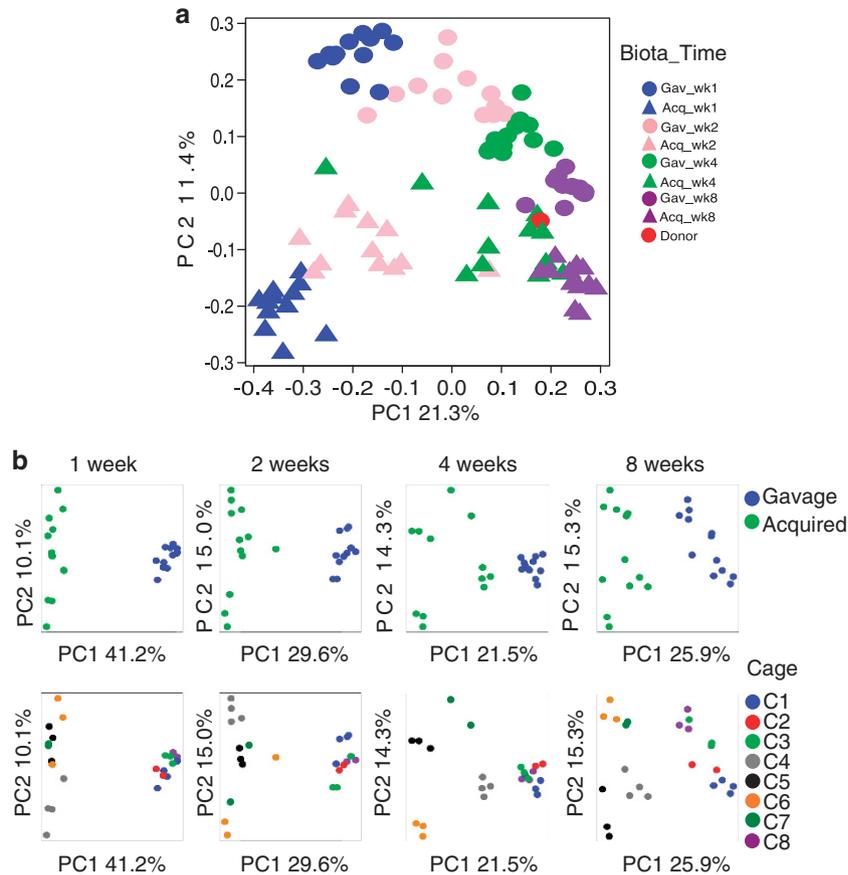
**Figure 1** Microbial community assembly over time after removal from GF conditions. Pie charts display the relative abundance of phyla at each time point. The donor microbiota is shown in the upper right hand corner.  $n=12$  at all time points for both the acquired and gavage groups.



**Figure 2** Richness as a function of time after removal from GF conditions for the acquired and gavage groups. The richness value for the donor biota was similar to week 1 values with a richness of 118.8. Values shown are the median of each cage.

a substantial ‘head start’ in forming a mature microbial community.

To perform inference on this data set and explicitly consider the effects of cage, treatment and time, we performed PCoA using Bray–Curtis dissimilarity at the OTU level (Figure 3a). We see that time is a dominant force in structuring the microbial



**Figure 3** (a) Bray–Curtis dissimilarity PCoA at the OTU level showing microbial community shifting over time. Gav: Gavage treatment; Acq: Acquired treatment. (b) Independent PCoA clusters were performed for each time point and are colored by treatment (top panel) and cage (bottom panel).

community with clear separation of samples at different time points. However, at each time point the gavage and acquired microbiota appear to be distinct (Figure 3b, top panel), although over the course of the experiment the differences between gavage and acquired microbiota become less pronounced.

If initial differences in the microbial community drive cage effects, we might expect to see a different pattern of cage effects in the gavage and acquired groups. Examination of the PCoA plots colored by cage (Figure 3b, bottom panel), however, revealed pronounced cage effects in both the acquired and gavage groups, especially at later time points. To quantify cage effects, we fit each treatment group at each time point with a one-way ANOVA with a fixed factor of cage. *P*-values generated from this model (Figure 4) show that the gavage and acquired groups have a similar pattern of cage effects. At the 1-week time point (Figure 4, black symbols), cage effects appear to be of marginal significance at best. At the 4- and 8-week time points, the cage effects have become much more pronounced in both the gavage and acquired groups, although the *P*-values are slightly larger in the gavage group.

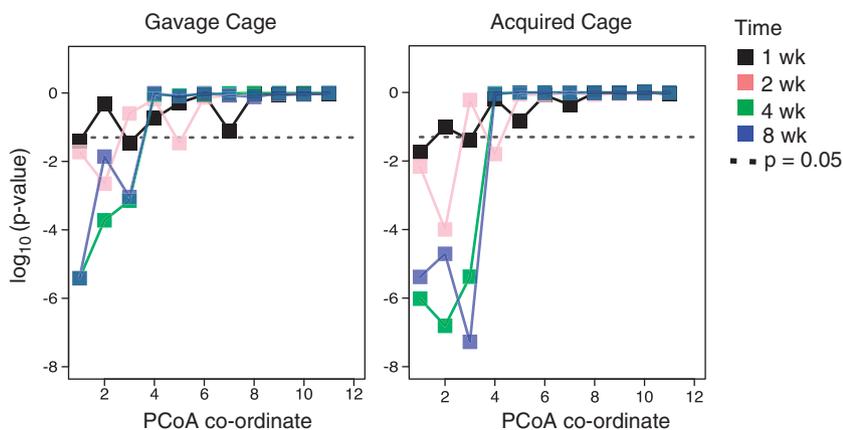
To account for the effects of cage, treatment and time, we evaluated a mixed linear model in which

treatment and time is a fixed effect and cage is a random effect (see Materials and methods). The model is formulated as:

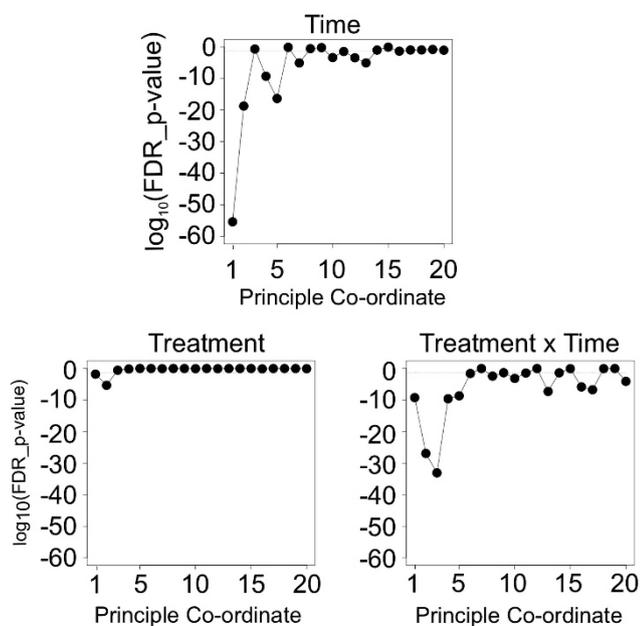
$$Y_{ijkl} = \mu + G_i + T_j + (GT)_{ij} + C_{k(i)} + \varepsilon_{ijkl}$$

where  $Y_{ijkl}$  represents PCoA axis value (Figure 5; Supplementary Table S1), phylum count (Supplementary Table S2), genus count (Supplementary Table S3) or richness value (Supplementary Table S4) for treatment  $i$ , time  $j$ , cage  $k$  and replicate  $l$ .  $G_i$  is the effect of the  $i$ th treatment. Treatment is set to one value for animals receiving gavage and another for animals allowed to acquire the microbial community from the cages.  $T_j$  is the effect from the  $j$ th time point.  $(GT)_{ij}$  is the interaction effect between treatment  $i$  and time  $j$ .  $C_{k(i)}$  is the effect from the  $k$ th cage that is nested within the  $i$ th treatment and  $\varepsilon_{ijkl}$  denotes the error associated with measuring  $Y_{ijkl}$ .

From this model, we conclude that time and time  $\times$  treatment interaction effects are generally more pronounced than the treatment effect, confirming that the implantation of a founder community through gavage did not eliminate the strong successional effect of time. Specifically, (i) richness changed over time but was not affected by treatment (Supplementary Table S4); (ii) at the phyla level



**Figure 4** Cage effects illustrated through the use of Bray–Curtis PCoA performed at the OTU level. Shown for the first 12 PCoA co-ordinates are the  $P$ -values from a one-way ANOVA with a fixed factor of cage evaluating the null hypothesis that cage had no effect on the distribution of the co-ordinate.



**Figure 5** For the first 20 co-ordinates from a PCoA at the OTU level,  $P$ -values from the mixed linear model evaluating the null hypothesis that the fixed effects of time, treatment (gavage vs acquired) and treatment  $\times$  time interactions had no effect on the co-ordinate shown on the x axis. Dotted line represents  $P=0.05$  significance level.

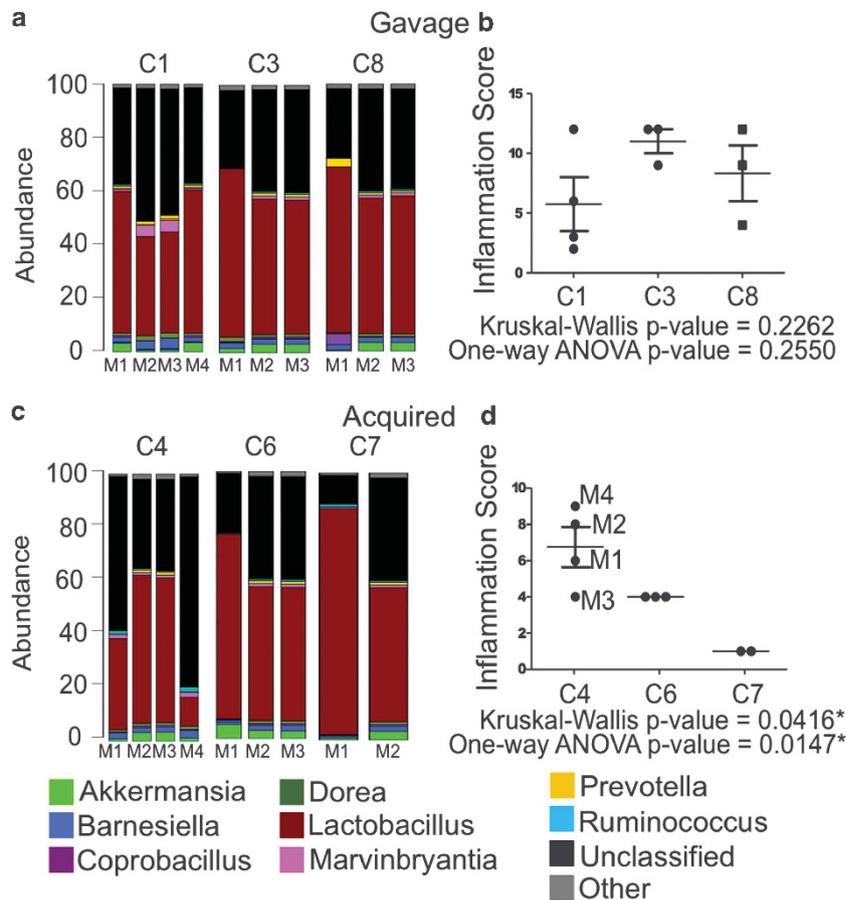
(Supplementary Table S2) and using a 10% FDR, time and treatment  $\times$  time interactions are significant for all evaluated phyla and (iii) at the genus level, time and treatment  $\times$  time interaction represents the first 49 most significant effects (Supplementary Table S3).

#### Gavage treatment protects from cage effects of inflammation

Even though gavage did not eliminate temporal effects, the abundance of several taxa was significantly different between gavage and acquired groups

independent of time. At a 10% FDR cutoff at the phyla level (Supplementary Table S2), the treatment variable for both Bacteroidetes and Firmicutes was significantly different between gavage and acquired groups independent of time. At the genus level, the 10% FDR cutoff yielded 20 genera whose abundance was significantly different between gavage and acquired groups independent of the time factor (Supplementary Table S3). This suggests that, despite the overall progress of acquired and gavage groups to become more similar to each other (Figure 1), the gavage treatment did have some long-lasting effects.

To study the functional consequences of cage effects, we treated acquired or gavaged mice after materials for sequencing were acquired at the 8-week time point with DSS for 13 days. At the end of DSS exposure, mice were killed and colonic inflammation was assessed using histological scoring. As a negative control, five mice ( $n=2$  for gavage;  $n=3$  for acquired) not exposed to DSS were euthanized at this same time point and scored for inflammation. All control mice showed absence of intestinal inflammation with histological scores of zero (data not shown). Inflammatory scores of mice exposed to DSS were highly divergent between cages of mice with an acquired biome compared with those that were gavaged (Figures 6b and d). Interestingly, *Lactobacillus*, a taxa that is generally considered to have anti-inflammatory properties (Servin, 2004; Santos Rocha *et al.*, 2012; von Schillde *et al.*, 2012), was found to be significantly associated with time, treatment  $\times$  time and treatment under the mixed linear model analysis of the gavage vs acquired data set (Figures 6a and c; Supplementary Table S3). With our small sample size, we cannot meaningfully speculate on whether these associations are robust and would be reproducible in future cohorts. Nonetheless, these data are intriguing in suggesting that long-term effects of an initial gavage may insulate an animal from environmentally induced susceptibility to cage effects in phenotypes of interest.



**Figure 6** Relative abundance of genera in (a) gavage and (c) acquired at the 8-week time point broken down by cages. Each bar represents an individual mouse's microbial community before treatment with DSS. Differences in inflammation scores between cages were not significant for the gavage mice (b) but were for the acquired mice (d) with both a parametric one-way ANOVA and a non-parametric Kruskal-Wallis (with the indicated *P*-values). Error bars represent the s.d. or dispersion from the mean in each set of samples.

*Successional differences over time can be associated with a shift in genome emphasis from motility and invasion to metabolic function*

Differences in community functional attributes were evaluated through the use of PICRUSt (<http://picrust.github.com/picrust/>), which provides an avenue for functional prediction from 16S sequences (see Materials and methods). To assess how these functional categories change over time, we again utilized our mixed linear model to compare PICRUSt predictions that differed between the earliest (1 week) and latest (8 weeks) time points independent of treatment. We used a Fisher's exact test to identify KEGG functional categories in which a significant number of orthologs differed between week 1 vs week 8 time points. At a 1% FDR cutoff, we observed a preferential selection for pathogenesis and motility at the early (1 week) time point. Of the orthologs that were significantly different between the week 1 and week 8 time points, 97% of orthologs with a KEGG functional category of 'Cell Motility' and 88% of the orthologs with a KEGG functional category of 'Infectious Diseases' showing a higher relative abundance at the early time point (Table 1). By contrast, functional

selection in the late (8 weeks) time point shifts to metabolic and cell maintenance (Table 1).

## Discussion

To explore the causes and consequences of cage effects, we here used Illumina sequencing to characterize the bacterial 16S gene repertoire from fecal samples collected over time for two cohorts (acquired and gavaged) of mice following removal from GF conditions. Time stands out as the primary influential factor in structuring microbial communities from an environment with low richness and a substantial fraction of Proteobacteria into a more 'adult' mammalian gut with increasing richness and greater domination by Bacteroidetes and Firmicutes. Interestingly, although Proteobacteria represent an important community member in both acquired and gavaged mice, this phylum is progressively phased out over time. This observation is in line with a previous study showing ecological succession in GF mice, with pioneering Proteobacteria being replaced by other community members (Gilliland *et al.*, 2012). Superimposed on these broad and reproducible successional patterns, however, was

**Table 1** Functional differences between early (1 week) and late (8 weeks) time points at 1% FDR identified with PICRUSt

Functional category	# of Orthologs	# of Non-significant orthologs	# of Orthologs up at week 1	# of Orthologs up at week 8	Ratio up
<i>Early</i>					
Cell motility	69	25	43	1	0.977272727
Infectious diseases	51	8	38	5	0.88372093
Signal transduction	135	22	93	20	0.82300885
Membrane transport	674	154	424	96	0.815384615
Cellular processes and signaling	490	106	294	90	0.765625
Metabolism of cofactors and vitamins	204	90	61	53	0.535087719
<i>Late</i>					
Translation	160	28	28	104	0.787878788
Replication and repair	146	30	48	68	0.586206897
Nucleotide metabolism	107	30	32	45	0.584415584

Ratio indicates the fraction of significant orthologs identified as having higher relative abundance at each time point.

substantial individual variation in microbial communities that could in large part be explained by the cage in which the animals were housed. Starting the microbial community with an initial gavage from a mature gut microbial community influenced, but did not eliminate, the dependency on time or cage effects.

A long-standing question in ecology is to what extent community structure is driven by selection vs stochastic events. In our data, we find evidence for both kinds of processes. Our data demonstrate that initial composition of the microbial community (gavage vs acquired), selection pressure over time common across all cages, and stochastic effects that develop differently over time in different cages all make important and measurable contributions in structuring the microbial community. Interestingly, while richness increased over time in our gavage vs acquired experiment, there was no significant difference from richness induced by the gavage treatment. This suggests that even though mice exposed to a 'bolus' of microbes (gavage) with a presumably higher number of bacteria than the acquired group, many of these microbes did not successfully colonize the gavaged group.

In addition to using 16S sequencing to characterize the microbial community, we studied the functional consequences of microbiota cage effects by inducing acute inflammation in WT mice with DSS. Inflammation scores established by histological examination of colon tissue revealed that animals that were allowed to acquire their microbiota from the cage environment displayed a more pronounced cage effect in inflammation severity than animals whose microbial community was seeded by gavage. In our analysis utilizing mixed linear models with cage modeled as a random variable, treatment (gavage vs acquired) effects on the structure of the microbial community were generally much less pronounced than effects due to time or interactions between time and treatment. There were, however, 2 phyla (Bacteroidetes and Firmicutes) and 20 genera (including *Lactobacillus*)

whose abundance was significantly different between gavage and acquired groups independently of time. These observations suggest that the initial founding community (that is, gavage) can have long-term effects on host health and disease, even in the face of robust community changes that occur over space and time.

There is considerable evidence for cage effects in the literature. In several disease models, microbial transfer of intestinal disease can be accomplished by housing healthy WT mice with colitic mice (Garrett *et al.*, 2007; Elinav *et al.*, 2011). Mice are coprophagic, and sharing the microbial community in this manner presumably has a substantial effect on maintenance of microbial community structure. We explicitly tested the hypothesis that cage effects are caused by initial differences in the microbial community within each cage. If these stochastic 'founder' effects drive cage effects, then we would expect (i) cage effects to be pronounced at early time points and (ii) gavage to substantially mitigate cage effects. Neither of these predictions were supported by our data. By contrast, cage effects clearly become more pronounced over time, moving from barely significant at week 1 to highly significant at weeks 4 and 8. This pattern is clearly seen in both the gavage and acquired groups, with only slightly less significant *P*-values in the gavage group (Figure 4). Therefore, our data argue that stochastic differences in community assembly occur within each individual cage microenvironment, and it is these changes, rather than founder effects, that are the primary drivers of cage effects. Attempts to eliminate cage effects by standardizing the initial microbial community within cages or with identical initial gavage to multiple animals are therefore likely to fail.

A recent paper (Ubeda *et al.*, 2012) has argued that family transmission, if not properly accounted for, can lead to confounded experimental design and incorrect inference regarding the effects of genotype differences on the microbial community. We note that since our animals were born in GF conditions, family transmission is not a variable that can be

considered in our experiments. However, by necessity, animals that have a similar path of family transmission have also shared cages. In animals born elsewhere than under GF conditions, cage and family transmission effects are therefore likely to be entwined and this has the potential to further confound and complicate experimental design. By co-housing gavaged and non-gavaged animals, future studies may provide further insight into how initial seeding effects interact with environmental constraints on the development of the gut microbiome.

Our analysis took advantage of PICRUSt (<http://picrust.github.com/picrust/>), a recently developed bioinformatics method that simulates whole genome metagenome sequencing based on 16S data. In a mouse model where there is a reproducible early selection for Proteobacteria, our PICRUSt data suggest that there is also an overrepresentation of gene function associated with community composition. Results from PICRUSt revealed a functional selection for motility and infection (Table 1). This is consistent with the hypothesis that Proteobacteria are optimized for conditions present in the early formation of the gut microbiome. These results, while biologically plausible, will need to be directly confirmed in future studies by whole-genome methods, such as whole community shotgun sequencing and RNA-seq, performed longitudinally over succession.

There are numerous potentially confounding variables to consider when planning microbiome studies, including the location of sampling (Gilliland *et al.*, 2012), maternal transmission (Ubeda *et al.*, 2012), the time after gavage experiments are performed (Gilliland *et al.*, 2012) and husbandry conditions (Ma *et al.*, 2012). To this growing list of potentially confounding variables, we here demonstrate that cage microenvironment is a powerful driver of individual differences in community structure. Despite the complexity inherent to so many potential confounding variables, we find that simple mixed linear models are adequately powered to produce significant results, even for the modest numbers of animals and cages that made up our experiment. This demonstrates that if experiments are planned with consideration given to the number of cages as well as the number of animals, successful microbiome experiments with interpretable outcomes can be achieved despite substantial variability induced by space and time.

## Conflict of Interest

Anthony Fodor has consulted for the Salix corporation.

## Acknowledgements

We thank the National Gnotobiotic Rodent Resource Center at University of North Carolina for assistance

with germ-free mice (NIH P40 R018603). Histology was performed at the CGIBD histology core (P30 DK034987). This work was supported by funding from R01 DK73338 (CJ), R01 DK47700 (CJ), UNC University Cancer Research Fund (CJ) and a Crohn's and Colitis Foundation of America fellowship to M Mühlbauer.

## References

- Arthur JC, Perez-Chanona E, Mühlbauer M, Tomkovich S, Uronis JM, Fan TJ *et al.* (2012). Intestinal inflammation targets cancer-inducing activity of the microbiota. *Science* **338**: 120–123.
- Arumugam M, Raes J, Pelletier E, Le Paslier D, Yamada T, Mende DR *et al.* (2011). Enterotypes of the human gut microbiome. *Nature* **473**: 174–180.
- Brown CT, Davis-Richardson AG, Giongo A, Gano KA, Crabb DB, Mukherjee N *et al.* (2011). Gut microbiome metagenomics analysis suggests a functional model for the development of autoimmunity for type 1 diabetes. *PLoS one* **6**: e25792.
- Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK *et al.* (2010). QIIME allows analysis of high-throughput community sequencing data. *Nat methods* **7**: 335–336.
- Claesson MJ, Jeffery IB, Conde S, Power SE, O'Connor EM, Cusack S *et al.* (2012). Gut microbiota composition correlates with diet and health in the elderly. *Nature* **488**: 178–184.
- Edgar RC, Haas BJ, Clemente JC, Quince C, Knight R. (2011). UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics* **27**: 2194–2200.
- Elinav E, Strowig T, Kau AL, Henao-Mejia J, Thaiss CA, Booth CJ *et al.* (2011). NLRP6 inflammasome regulates colonic microbial ecology and risk for colitis. *Cell* **145**: 745–757.
- Garrett WS, Lord GM, Punit S, Lugo-Villarino G, Mazmanian SK, Ito S *et al.* (2007). Communicable ulcerative colitis induced by T-bet deficiency in the innate immune system. *Cell* **131**: 33–45.
- Gilliland MG 3rd, Erb-Downward JR, Bassis CM, Shen MC, Toews GB, Young VB *et al.* (2012). Ecological succession of bacterial communities during conventionalization of germ-free mice. *Appl Environ Microbiol* **78**: 2359–2366.
- Goldsmith JR, Uronis JM, Jobin C. (2011). Mu opioid signaling protects against acute murine intestinal injury in a manner involving Stat3 signaling. *Am J Pathol* **179**: 673–683.
- Hildebrand F, Nguyen AT, Brinkman B, Yunta RG, Cauwe B, Vandenabeele P *et al.* (2013). Inflammation-associated enterotypes, host genotype, cage and inter-individual effects drive gut microbiota variation in common laboratory mice. *Genome Biol* **14**: R4.
- Huse SM, Ye Y, Zhou Y, Fodor AA. (2012). A core human microbiome as viewed through 16S rRNA sequence clusters. *PLoS one* **7**: e34242.
- Jeffery IB, Claesson MJ, O'Toole PW, Shanahan F. (2012). Categorization of the gut microbiota: enterotypes or gradients? *Nat Rev Microbiol* **10**: 591–592.
- Ley RE, Backhed F, Turnbaugh P, Lozupone CA, Knight RD, Gordon JI. (2005). Obesity alters gut microbial ecology. *Proc Natl Acad Sci USA* **102**: 11070–11075.

- Listgarten J, Lippert C, Kadie CM, Davidson RI, Eskin E, Heckerman D. (2012). Improved linear mixed models for genome-wide association studies. *Nat Methods* **9**: 525–526.
- Ma BW, Bokulich NA, Castillo PA, Kananurak A, Underwood MA, Mills DA *et al.* (2012). Routine habitat change: a source of unrecognized transient alteration of intestinal microbiota in laboratory mice. *PLoS one* **7**: e47416.
- Mazmanian SK, Round JL, Kasper DL. (2008). A microbial symbiosis factor prevents intestinal inflammatory disease. *Nature* **453**: 620–625.
- Raudenbush SW, Bryk AS. (2002). *Hierarchical linear models: applications and data analysis methods*, 2nd edn. Sage Publications: Thousand Oaks, xxiv, p 485.
- Ross EM, Moate PJ, Bath CR, Davidson SE, Sawbridge TI, Guthridge KM *et al.* (2012). High throughput whole rumen metagenome profiling using untargeted massively parallel sequencing. *BMC Genet* **13**: 53.
- Santos Rocha C, Lakhdari O, Blottiere HM, Blugeon S, Sokol H, Bermudez-Humaran LG *et al.* (2012). Anti-inflammatory properties of dairy lactobacilli. *Inflamm Bowel Dis* **18**: 657–666.
- Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB *et al.* (2009). Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl Environ Microbiol* **75**: 7537–7541.
- Segata N, Waldron L, Ballarini A, Narasimhan V, Jousson O, Huttenhower C. (2012). Metagenomic microbial community profiling using unique clade-specific marker genes. *Nat Methods* **9**: 811–814.
- Servin AL. (2004). Antagonistic activities of lactobacilli and bifidobacteria against microbial pathogens. *FEMS Microbiol Rev* **28**: 405–440.
- Smyth GK. (2004). Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Stat Appl Genet Mol Biol* **3**: Article3.
- Turnbaugh PJ, Ley RE, Hamady M, Fraser-Liggett CM, Knight R, Gordon JL. (2007). The human microbiome project. *Nature* **449**: 804–810.
- Turnbaugh PJ, Ley RE, Mahowald MA, Magrini V, Mardis ER, Gordon JL. (2006). An obesity-associated gut microbiome with increased capacity for energy harvest. *Nature* **444**: 1027–1031.
- Ubeda C, Lipuma L, Gobourne A, Viale A, Leiner I, Equinda M *et al.* (2012). Familial transmission rather than defective innate immunity shapes the distinct intestinal microbiota of TLR-deficient mice. *J Exp Med* **209**: 1445–1456.
- Ursell LK, Clemente JC, Rideout JR, Gevers D, Caporaso JG, Knight R. (2012). The interpersonal and intrapersonal diversity of human-associated microbiota in key body sites. *J Allergy Clin Immunol* **129**: 1204–1208.
- Vilhjalmsson BJ, Nordborg M. (2013). The nature of confounding in genome-wide association studies. *Nat Rev Genet* **14**: 1–2.
- von Schillde MA, Hormannsperger G, Weiher M, Alpert CA, Hahne H, Bauerl C *et al.* (2012). Lactocepin secreted by *Lactobacillus* exerts anti-inflammatory effects by selectively degrading proinflammatory chemokines. *Cell Host Microbe* **11**: 387–396.
- Wang Q, Garrity GM, Tiedje JM, Cole JR. (2007). Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl Environ Microbiol* **73**: 5261–5267.

Supplementary Information accompanies this paper on The ISME Journal website (<http://www.nature.com/ismej>)