Introduction

In studying the contribution of gut bacteria to human health and disease, murine models of the gut microbiome are still considered essential due to limitations in human research. Given the fact that mice can be raised in germ-free (GF) conditions, murine models give the possibility for the cultured microbiome from a human or mouse donor to be inoculated for study purposes. Such models are useful to investigate the effects of diet, drug uptake, at the interplay between host and microbiota. In recent studies, Bidelhier showed that the variance in mice gut microbiota can be explained, to varying degree, by host’s genotype, its cage microenvironment, and interindividual variation. However, model-based analysis built to understand these various confounding factors in murine disease models are not always straightforward. Several constraints include the special distributional properties of microbiome data, such as zero inflation, overdispersion, and methodology to correct for library size. Additionally, zero abundance does not necessarily mean OTU (Operational Taxonomic Unit) is not present at a certain time point.

Methodology

To assess the cage microenvironment effect on the mouse gut microbiome, we analyzed fecal samples from 48 (n = 15 per cage) GF and 82 (n = 26 per cage) zero inflammation, overdispersion, and methodology to correct for library size. Additionally, zero abundance does not necessarily mean OTU (Operational Taxonomic Unit) is not present at a certain time point.

Mixed effect models

Cage effect on richness estimates

Richness estimates (also known as alpha diversity) is defined as the number of distinct OTUs in each sample. Models for richness estimates (observed and Chao1)

- $M_0$: $Y_{ijkl} = g + T_i + c_j + e_{ijkl}$
- $M_1$: $Y_{ijkl} = g + T_i + c_j + e_{ijkl} + g_{ijkl}$
- $M_2$: $Y_{ijkl} = g + T_i + c_j + e_{ijkl} + g_{ijkl} + c_{ijkl}$
- $M_3$: $Y_{ijkl} = g + T_i + c_j + e_{ijkl} + g_{ijkl} + c_{ijkl} + a_{ijkl}$

where:
- $Y_{ijkl}$ represents the richness (number of active OTUs) value for treatment $i$, cage $j$, day $k$, and replicate $l$.
- $g$ is the effect of the $g$-th treatment.
- $T_i$ is the effect from the $i$-th time point.
- $c_j$ is the random effect from the $j$-th cage.
- $e_{ijkl}$ is the random effect noted within the $ijkl$ cage.

In this analysis, we did not add a term for purely marginal effects. This is due to the fact that the main cages from the same mother tend to be located in the same cage which means the mother and cage effect will highly correlated to each other and we would not be able to observe the cage effect on richness estimate anymore once the maternal effect is placed in the model.

Differential statistical assumptions were compared: normal, Poisson, and Negative Binomial (NB).

Cage effect on OTU appearance (control group only)

For this analysis, $Y_{ijkl}$ is defined as the OTU appearance on mouse at the time point $k$ in cage $j$. It is Bernoulli random variable with a “success” ($Y_{ijkl} = 1$) being the OTU is observed and a “failure” ($Y_{ijkl} = 0$) being the OTU is not observed. From here, we define our response variable $Z_{ijkl}$ as

$$Z_{ijkl} = \begin{cases} 1, & Y_{ijkl} = 0 \\ 0, & Y_{ijkl} = 1 \end{cases}$$

$M_0$: $G_{ijkl} = \lambda_{ijkl} + \epsilon_{ijkl}$

$M_1$: $G_{ijkl} = \lambda_{ijkl} + \epsilon_{ijkl} + a_{ijkl}$

$M_2$: $G_{ijkl} = \lambda_{ijkl} + \epsilon_{ijkl} + a_{ijkl} + p_{ijkl}$

$M_3$: $G_{ijkl} = \lambda_{ijkl} + \epsilon_{ijkl} + a_{ijkl} + p_{ijkl}$

These models were run on OTU, genus, and family level.

Modeling the cage effect on OTU appearance

In both absolute richness as well as Chao1 estimates, the Poisson models have the highest AIC and BIC as compared to models with normal and negative binomial distribution assumption. This may be due to the overdispersion issue when the richness estimates are modeled under the Poisson distribution assumption. The overdispersion measure (deviance/df) for Chao1 richness estimate for M1, M2, and M3 respectively are 6.55, 4.94, 4.92. And for the absolute richness estimates, the overdispersion measure for Chao1 richness estimate for M1, M2, and M3 respectively are 3.38, 2.59, 2.32.

Modeling the cage effect on OTU appearance

In both absolute richness as well as Chao1 estimates, the Poisson models have the highest AIC and BIC as compared to models with normal and negative binomial distribution assumption. This may be due to the overdispersion issue when the richness estimates are modeled under the Poisson distribution assumption. The overdispersion measure (deviance/df) for Chao1 richness estimate for M1, M2, and M3 respectively are 6.55, 4.94, 4.92. And for the absolute richness estimates, the overdispersion measure for M1, M2, and M3 respectively are 3.38, 2.59, 2.32.

Table 1: Table of likelihood ratio test on random effect (cage and mouse)

<table>
<thead>
<tr>
<th>Cage effect on richness estimates</th>
<th>Distribution</th>
<th>Statistics</th>
<th>AIC</th>
<th>BIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Richness estimates</td>
<td>Normal</td>
<td>0.0006</td>
<td>0.0001</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Poisson</td>
<td>0.0006</td>
<td>0.0001</td>
<td></td>
</tr>
<tr>
<td></td>
<td>NB</td>
<td>0.0006</td>
<td>0.0001</td>
<td></td>
</tr>
</tbody>
</table>

Model building

### Linear mixed effect model

$Y_{ijkl} = \alpha + \beta X_{ijkl} + \gamma Z_{ijkl} + \epsilon_{ijkl}$

### Fisher GLMM

$Y_{ijkl} \sim \text{Binomial}(\theta_{ijkl})$

### GLMM

$Y_{ijkl} \sim \text{Poisson}(\lambda_{ijkl})$

Check the random effects

Check if the model with random effect and all in the model without the random effect variances, all.

Cage effect on the OTU appearance

### Statistic

#### F test

$F = \frac{\text{variance in model with random effect}}{\text{variance in model without the random effect}}$

### Model with random effect

$Y_{ijkl} \sim \text{Normal}(

Modeling the cage effect on OTU appearance

Figure 5: The samples are the cage to which they belong. Their cages can be identified by the colors above each heatmap. Darker reds indicate a greater abundance of OTUs detected in the samples. If we look into the OTU activity by cage, we can see the abundance increases over the observed time period. For example, starting from only one animal on Day 21, to spreading to other animals in the same cage on subsequent periods as time progresses, we were interested in modeling the cage effect on OTU appearance.

### Conclusion

Only 6 out of 24 families are active with high abundances during the study period.

The cage has a strong effect on absolute as well as Chao richness estimates over time in both controlled and antibiotic group.

The subject specific variations, coming from each mouse, could explain the total variations in richness estimates when there was no cage effect put in the model. However, once the cage effect has been placed in the model, these variations are no longer meaningful in explaining the total variations.

In the control group only, there were 41 out of 346 OTUs that have a strong association with the cage, detected by both Fisher’s exact test and Generalized Linear Mixed Model (GLMM). This strong association was also observed on the genus as well as family level.

Figures

- Figure 1: Mouse and fecal samples collection scheme
- Figure 2: Modeling scheme to study the cage effect on OTU appearance
- Figure 3: The antibiotic group has relatively lower and less varied richness estimates compared to the control group. Each curve represents a mouse. Same color indicates common cage.
- Figure 4: The y axis indicates the number of active OTUs for each family in each mouse. The 346 OTUs were grouped into 24 different families. Only some families were active during observation period.
- Figure 5: The samples are the cage to which they belong. Their cages can be identified by the colors above each heatmap. Darker reds indicate a greater abundance of OTUs detected in the samples. If we look into the OTU activity by cage, we can see the abundance increases over the observed time period.
- Figure 6: Venn diagram of OTUs that are strongly associated with the cage, detected by Fisher’s exact test and GLMM.