Modeling cage effects in microbial community in controlled population mouse model using mixed effect models in R

Dea Rynanda Putri[1,2], Ziv Skhedy[2], Olivier Thas[3], Nolen Joy Perualila[2], Thierry van Effelterre[1], Xuesong Zhang[4], Marcus Rauch[5], Martin J. Blaser[4], Luc Bijnens[1]

[1]Janssen Research and Development, Beerse, Belgium; [2]Hasselt University, Diepenbeek, Belgium; [3]Ghent University, Ghent, Belgium; [4]New York University Langone Medical Center, New York, USA, [5]Janssen Prevention Center, London, UK

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 cultivated microbes from a human or mouse donor to be inoculated for study purposes. Such models are useful to investigate the effects of diets, drug uptake, or the interplay between host and microbiota.

In recent studies, Hildebrand showed that the variance in mice gut microbiota can be explained, to varying degrees, 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 in a subject (mouse). With such limitations, we try to explain the importance of the aforementioned factors in shaping the gut microbiome composition.

Methodology

To assess the cage microenvironment effect on the mouse gut microbiome, we analyzed fecal samples from 165 (n control = 83, n antibiotic = 82) same breed mice for 49 days (bi-monthly fecal samples collection). The mice were grouped into control and antibiotic groups.

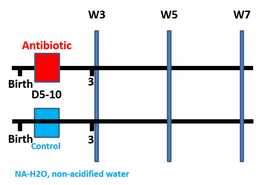


Figure 1: Mouse and fecal samples collection scheme

In modeling the cage effect on the richness estimates, the two treatment groups are considered in the model, while in modeling the cage effect on the OTU appearance, only the control group is considered.

Multiple tests corrections: Multiple tests corrections were applied for the analysis run on the cage effect on OTU appearance using the method described by Benjamini-Hochberg.

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: $Y_{ijkl} = \mu + G_i + T_j + (GT)_{ij} + \varepsilon_{ijkl}$
M_1: $Y_{ijkl} = \mu + G_i + T_j + (GT)_{ij} + c_k + \varepsilon_{ijkl}$
M_2: $Y_{ijkl} = \mu + G_i + T_j + (GT)_{ij} + c_k + m_{l(k)} + \varepsilon_{ijkl}$
M_3: $Y_{ijkl} = \mu + G_i + T_j + (GT)_{ij} + m_{l(k)} + \varepsilon_{ijkl}$

where:

- Y_{ijkl} represents the richness (number of active OTUs) value for treatment i, time j, cage k, and replicate l.
- G_i is the effect of the i-th treatment.
- T_j is the effect from the j-th time point.
- $(GT)_{ij}$ is the interactions between treatment i with time j.
- c_k is the random effect from the k-th cage.
- $m_{l(k)}$ is the random effect from l-th mouse nested within the k-th cage.

In this analysis, we did not add a term for purely maternal effects. This is due to the fact that the mice coming 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.

Different distributional assumptions were compared: normal, Poisson, and Negative Binomial (NB).

Cage effect on OTU appearance (control group only)

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

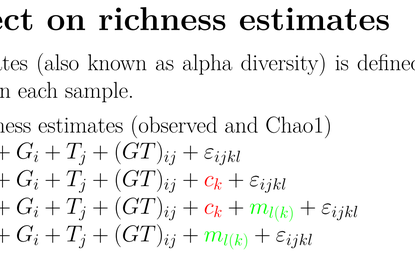


Figure 2: Modeling scheme to study the cage effect on OTU appearance

$$Z_{ijk} = \begin{cases} 1, & Y_{i(j-1)k} = 0 \quad \& \quad Y_{ijk} = 1 \\ 0, & \text{otherwise} \end{cases}$$

$$Z_{ijk} \sim \text{Bin}(\pi_i)$$

M_0: $g(\pi_i) = \beta_0 + \beta_1 Time_j + \beta_2 N_{bl_k} + \varepsilon_i$
M_1: $g(\pi_i) = \beta_0 + \beta_1 Time_j + \beta_2 N_{bl_k} + c_k + \varepsilon_i$

- These models were run on OTU, genus, and family level.
- P-value are adjusted for multiple tests corrections using Benjamini-Hochberg.

Observed richness estimates

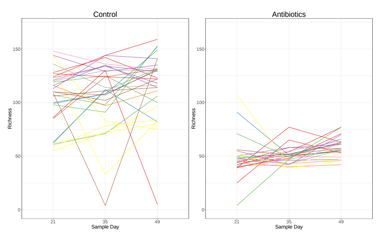


Figure 3: The antibiotic group has relatively lower and less varied absolute richness as compared to the control group. Each curve represents a mouse. Same color indicates common cage.

Modeling the cage effect

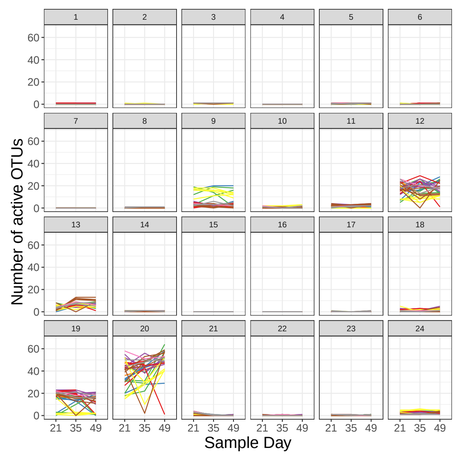
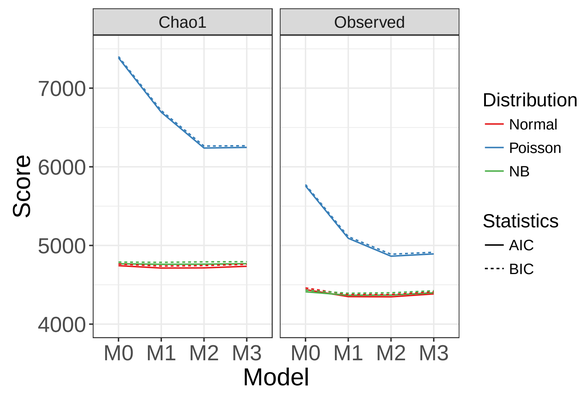


Figure 4: The y axis indicates the number of active OTUs for each family in each mouse. The 348 OTUs were grouped into 24 different families. Only some families were active during observation period.

Modeling the cage effect on richness estimates



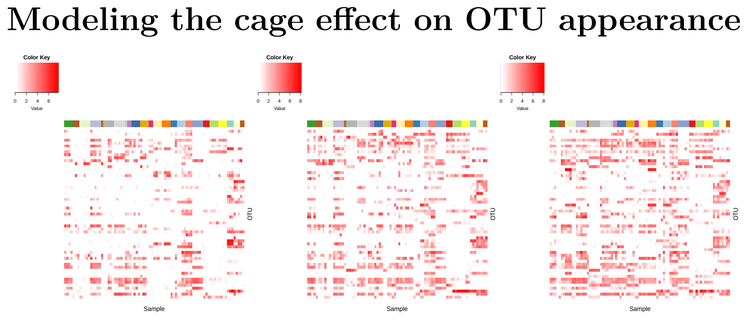
In both absolute richness as well as Chao1 estimates, the Poisson models have the highest AIC and BIC as compared to models with normal and negative binomial distribution assumption. This may be due to the overdispersion issue when the richness estimates are modeled under the Poisson distribution assumption. 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 the absolute richness estimates, the overdispersion measure for M1, M2, and M3 respectively are 3.38, 2.53, 2.52.

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

Estimate	Dist	Cage effect	Mouse effect on cage	Mouse effect
Chao1	Normal	< 0.0001	1.0000	0.0016
Chao1	NB	0.0023	1.0000	0.3323
Observed	Normal	< 0.0001	0.0282	< 0.0001
Observed	NB	< 0.0001	1.0000	0.0002

Under normal and negative binomial distribution assumption, the cage effect is strongly observed on the absolute as well as Chao1 richness estimates over time, while the subject specific (mouse) effect is no longer observable when the cage effect has been placed in the model.

Modeling the cage effect on OTU appearance



(a) Day 21 (b) Day 35 (c) Day 49

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 cage 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. Based on this finding, we were interested in modeling the cage effect on OTU appearance.

Other than modeling the cage effect using generalized linear mixed effect model (GLMM), the Fisher's exact test was also used to test the association between the appearance of the OTU between two consecutive time points (from Day 21 to Day 35, and from Day 35 to Day 49). The success event is defined the same as in GLMM. The presence of the same OTU in co-located mice was defined as the OTU also detected in at least one other mouse sharing the same cage at the earlier and/or current time point.

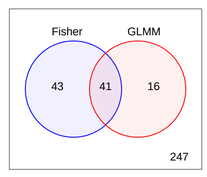


Figure 6: Venn diagram of OTUs that are strongly associated with the cage, detected by Fisher's exact test and GLMM

Using Fisher's exact test, 84 OTUs were detected of having association with the cage, while GLMM detected only 57. Fisher's and GLMM agreed on the strong association of cage and OTU appearance in 41 OTUs. The GLMM was also run on genus and family levels. On the genus level, 7 out of 50 genera have a strong association to cage. And on family level, the cage effect was strongly observed in 5 out of 30 families.

R Usage

The analysis presented in this presentation have been performed in R version 3.3.3, and the following codes were used to analyse the data.

Data set preparation

```
dat_new <- read.delim("nod_data.txt")

## Making an OTU matrix
otu <- dat_new[c(1:3)]
otu$Abundance <- as.numeric(otu$Abundance)
otu_wide <- dcast(otu, OTU ~ SampleID, value.name=Abundance)

otu_matrix <- as.matrix(otu_wide[, -c(1)])
dat <- otu_table(otu_matrix, taxa_are_rows = TRUE)
rownames(dat) <- otu_wide$OTU

## Making Sample Data Table
dat_new1 <- dat_new[SampleID, ]
sample.data <- dat_new1[, -c(3, 18:28)]
rownames(sample.data) <- sample.data$SampleID

## Making the Taxa Information Table
OTUID <- match(rownames(dat), dat_new[, 1])

subset.taxa <- dat_new[OTUID, ]
taxa.table <- subset.taxa[, c(1, 22:28)]
rownames(taxa.table) <- taxa.table$OTU

taxa.table <- taxa.table[, -c(1)]
taxa.table1 <- tax_table(taxa.table)
View(taxa.table1)
rownames(taxa.table1) <- subset.taxa$OTU
colnames(taxa.table1) <- c("Kingdom", "Phylum", "Class", "Order", "Family", "Genus", "Species", "taxa")

## Making phyloseq data format
data_ori <- phyloseq(dat, taxa.table1, sample_data(sample.data))
save (data_ori, file = "Data/data_ori.rda")
```

Richness estimates

```
richness <- estimate_richness(data_ori, measures = c("Observed", "Chao1", "Shannon", "InvSimpson"))
```

Model building

```
## Linear mixed effect model
m0 <- lm(alpha_chao1 ~ Antibiotics + SampleDay + Antibiotics*SampleDay,
         data = chao1)

## Generalized linear mixed effect model
m1_nb <- glmmadmb(alpha_chao1 ~ Antibiotics + SampleDay + Antibiotics*SampleDay + (1|cage),
                  data = chao1, family = "nbinom")
m3_pois <- glmer(alpha_chao1 ~ Antibiotics + SampleDay + Antibiotics*SampleDay + (1|Mouse),
                 family = "poisson", data = chao1)
```

Check the random effects

```
## m1 is the model with random effect and m0 is the model without the random effect
anova(m1, m0)
```

Cage effect on the OTU appearance

```
fit <- function(x) {
  d <- as.data.frame(x)
  d$Cage <- as.factor(d$Cage)
  fit <- try(glm(Appear ~ Day + nbl, data = d, family = binomial()), silent = TRUE)
  if(inherits(fit, "try-error")) {fit <- NULL}
  return(fit)
}

glm.fit <- NULL

for(i in seq(0, 340, 10)){                          # total number of OTU: 348
  if(i == 340) {
    otuToSelect <- otu[(i+1):(i+8)]
  } else {
    otuToSelect <- otu[(i+1):(i+10)]
  }

  d <- nbl.data[which(nbl.data$OTU %in% otuToSelect), ]
  glm.fit1  <- dlply(d, "OTU", fit, .parallel = TRUE)
  glm.fit <- c(glm.fit, glm.fit1)
  if(i%%100 == 0) save(glm.fit, file="Output/glm_nbl.rda")
  print(i)
}
```

Conclusion

- Only 6 out of 24 families are active with high abundances during the seven weeks of sample collections.
- The cage has a strong effect on absolute as well as Chao1 richness estimates over time in both the control 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 is added to the model, these variations are no longer meaningful in explaining the total variations.
- In the control group only, there were 41 out of 348 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.

Interuniversity Institute for Biostatistics and statistical Bioinformatics