metamicrobiomeR: an R package for analysis of microbiome relative abundance data using zero-inflated beta GAMLSS and meta-analysis across studies using random effect models

Nhan Thi Ho¹*, Fan Li²

1 Gertrude H. Sergievsky Center, Columbia University, New York City, NY, USA
2 Department of Pediatrics, University of California, Los Angeles, CA, USA

Authors’ emails: Nhan Thi Ho: nhanhocumc@gmail.com, Fan Li: fanli@mednet.ucla.edu

*Corresponding author: Nhan Thi Ho, Gertrude H. Sergievsky Center, Columbia University, 630 West 168th Street New York, New York 10032. Telephone: 212-305-2976; Fax: 212-305-2426; Email: nhanhocumc@gmail.com
ABSTRACT

Background: The rapid growth of high-throughput sequencing-based microbiome profiling has yielded tremendous insights into human health and physiology. Data generated from high-throughput sequencing of 16S rRNA gene amplicons are often preprocessed into composition or relative abundance. However, reproducibility has been lacking due to the myriad of different experimental and computational approaches taken in these studies. Microbiome studies may report varying results on the same topic, therefore, meta-analyses examining different microbiome studies to provide robust results are important. So far, there is still a lack of implemented methods to properly examine differential relative abundances of microbial taxonomies and to perform meta-analysis examining the heterogeneity and overall effects across microbiome studies.

Results: We developed an R package ‘metamicrobiomeR’ that applies Generalized Additive Models for Location, Scale and Shape (GAMLSS) with a zero-inflated beta (BEZI) family (GAMLSS-BEZI) for analysis of microbiome relative abundance datasets. Both simulation studies and application to real microbiome data demonstrate that GAMLSS-BEZI well performs in testing differential relative abundances of microbial taxonomies. Importantly, the estimates from GAMLSS-BEZI are log(odds ratio) of relative abundances between groups and thus are comparable between microbiome studies. As such, we also apply random effects meta-analysis models to pool estimates and their standard errors across microbiome studies. We demonstrate the meta-analysis workflow and highlight the utility of our package on four studies comparing gut microbiomes between male and female infants in the first six months of life.

Conclusions: GAMLSS-BEZI allows proper examination of microbiome relative abundance data. Random effects meta-analysis models can be directly applied to pool comparable estimates.
and their standard errors to evaluate the heterogeneity and overall effects across microbiome studies. The examples and workflow using our metamicrobiomeR package are reproducible and applicable for the analyses and meta-analyses of other microbiome studies.

**Keywords**: microbiome, relative abundance, GAMLSS, zero-inflated beta, meta-analysis, random effect, pooling estimates, infant, gender.
BACKGROUND

The rapid growth of high-throughput sequencing-based microbiome profiling has yielded tremendous insights into human health and physiology. However, interpretation of microbiome studies have been hampered by a lack of reproducibility in part due to the variety of different study designs, experimental approaches, and computational methods used [1, 2]. Microbiome studies may report varying results on the same topic. Therefore, meta-analyses examining different microbiome studies are critical to provide robust results. Meta-analysis studies pooling individual sample data across studies for pooled analysis of all samples or processing of all samples together followed by analysis of each study separately have revealed some consistent microbial signatures in certain conditions such as inflammatory bowel disease (IBD) and obesity [3–9]. Software has been developed for the analysis and meta-analysis of microbiome data [10]. However, these studies do not explicitly model microbiome relative abundance data using an appropriate statistical method and do not examined between-group-comparison overall pooled effects in the meta-analysis. For example, a recently published meta-analysis study examined differential relative abundances of bacterial taxonomies of each study separately using Wilcoxon’s tests but did not examine overall pooled effects across studies [8]. Moreover, meta-analysis pooling estimates across studies is not practical if studies utilize some methods such as Wilcoxon’s tests or Random Forest (RF) as there are no test statistics or estimates that are applicable for pooling.

Data generated from high-throughput sequencing of 16S rRNA gene amplicons are often preprocessed into relative abundance. Microbiome relative abundances are compositional data which range from zero to one and are generally zero-inflated. To test for differences in relative
abundance of microbial taxonomies between groups, methods such as bootstrapped non-parametric t-tests or Wilcoxon tests (not suitable for longitudinal data and covariate adjustment) [11–13] and linear or linear mixed effect models (LM) [14, 15] (suitable for longitudinal data and covariate adjustment) have been widely used. However, these methods do not address the actual distribution of the microbiome relative abundance data, which resemble a zero-inflated beta distribution. Transformations (e.g. arcsin square root) of relative abundance data to make it resemble continuous data to use in LM has been proposed by Morgan et al (implemented in MaAsLin software) [16] and has been widely used to test for differential relative abundances [17–20]. However, this adjustment does not address the inflation of zero values in microbiome relative abundance data. More recently, methods adapted from the RNA-seq field that account for zero inflation and utilize Poisson or negative binomial models have shown some promise in differential abundance testing of microbiome datasets [21, 22]. However, these methods are applicable to absolute abundance, which are count data and largely vary between studies due to the difference in experimental and computational approaches making the estimates from these methods incomparable across studies. Therefore, standard meta-analysis approaches may not be directly applicable to pool estimates and their standard errors from these methods across studies.

Here, we developed an R package “metamicrobiomeR” that applies Generalized Additive Models for Location, Scale and Shape (GAMLSS) [23] with a zero-inflated beta (BEZI) family (GAMLSS-BEZI) for the analysis of microbial taxonomy relative abundance data. With BEZI family, this model allows direct and proper examination of microbiome relative abundance data, which resemble a zero-inflated beta distribution. This model also allows adjustment for
covariates and can be used for both non-longitudinal and longitudinal studies. We evaluated the performance of GAMLSS-BEZI using simulation studies and real microbiome data. Importantly, the estimates from GAMLSS-BEZI are log(odds ratio) of relative abundances between groups and thus are comparable across microbiome studies and can be directly combined using standard meta-analysis approaches. As such, we apply random effect meta-analysis models to pool the estimates and standard errors as part of the “metamicrobiomeR” package. This approach allows examination of study-specific effects, heterogeneity between studies, and the overall pooled effects across studies. Finally, we provide examples and sample workflows for both components of the “metamicrobiomeR” package. Specifically, we use GAMLSS-BEZI to compare relative abundances of the gut microbial taxonomies of male versus female infants ≤ 6 months of age while adjusting for feeding status and infant age at time of sample collection and demonstrate the application of the random effect meta-analysis component on four studies of the infant gut microbiome.

IMPLEMENTATION

GAMLSS-BEZI for the analysis of bacterial taxa relative abundance and bacterial functional pathway relative abundance data

Relative abundances of bacterial taxa at various taxonomic levels (from phylum to genus or species) are obtained via the “summarize_taxa.py” script in QIIME1 [13]. Bacterial functional pathway abundances (e.g. Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway level 1 to 3) are obtained from metagenome prediction analysis using PICRUSt [24]. In the taxa.compare function, all bacterial taxa or pathway data are first filtered to retain features with mean relative abundance ≥ relative abundance threshold (e.g. ≥0.005%) and with prevalence ≥
prevalence threshold (e.g. present in ≥ 5% of the total number of samples). This pre-filtering step has been shown to improve performance of various differential abundance detection strategies [25]. A filtered data matrix is then modeled by GAMLSS-BEZI and (mu) logit link and other default options using the R package gamlss version 5.0-5 [23]. A subject random effect can be added to the model for longitudinal data. For performance evaluation, LM and LM with arcsin square-root transformation (LMAS) were also implemented in the function taxa.compare. Multiple testing adjustment can be done using different methods (False Discovery Rate (FDR) control by default). Below is an example call of the taxa.compare function:

```r
taxa.compare(taxtab=taxtab, propmed.rel="gamlss", transform="none",
comvar="gender",adjustvar=c("age.sample","feeding"),longitudinal="yes",
percent.filter=0.05,relabund.filter=0.00005, p.adjust.method="fdr")
```

For visualization purposes and subsequent meta-analysis, the output from taxa.compare comprises matrices containing coefficients, standard errors, p-values and multiple testing adjusted p-values of all covariates in the models for each bacterial taxon or pathway.

**Meta-analysis across studies using random effects models**

The adjusted estimates for bacterial taxa or pathway relative abundances from the GAMLSS models are log(odds ratio) of relative abundances between groups and thus are comparable across microbiome studies. Therefore, standard meta-analysis approaches can be directly applied. In the meta.taxa function, random effects meta-analysis models pooling adjusted estimates and standard errors with inverse variance weighting and the DerSimonian–Laird estimator for between-study variance are implemented to estimate the overall effects, corresponding 95% confidence intervals (95% CI) and heterogeneity across studies. A fixed
effect meta-analysis model is also implemented for comparison. Meta-analyses is performed only for taxa or pathways observed in ≥ a specified percentage threshold (e.g. 50%) of the total number of included studies. An example call to `meta.taxa` using the output data matrices combined from multiple calls to the `taxa.compare` function is shown below:

```
meta.taxa(taxcomdat=combined.taxa.compare.output,summary.measure="RR",pool.var="id",studylab="study",backtransform=FALSE,percent.meta=0.5,p.adjust.method="fdr")
```

The output from `meta.taxa` consists of pooled estimates, standard errors, 95% CI, pooled p-values and multiple testing adjusted pooled p-values of all covariates for each bacterial taxon or pathway.

**Additional microbiome measures**

Random Forest (RF) modeling of gut microbiota maturity has been widely used to characterize development of the microbiome over chronological time [12, 14, 26]. Adapting from the original approach of Subramanian et al [14], in the `microbiomeage` function, relative abundances of bacterial genera that were detected in the Bangladesh data [14] and in the data of other studies to be included were regressed against infant chronological age using a RF model on a predefined training dataset of the Bangladesh study. This predefined training set includes 249 samples collected monthly from birth to 2 years of age from 11 Bangladeshi healthy singleton infants. The RF training model fit based on relative abundances of these shared bacterial genera was then used to predict infant age on the test data of the Bangladesh study and the data of each other study to be included. The predicted infant age based on relative abundances of these shared bacterial genera in each study is referred to as gut microbiota age (Additional file 1).
All implemented functions in the *metamicrobiomeR* package are summarized in Table 1.

Table 1: Summary of all implemented functions in the *metamicrobiomeR* package.

<table>
<thead>
<tr>
<th>Functions</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>taxa.filter</td>
<td>Filter relative abundances of bacterial taxa or pathways using prevalence and abundance thresholds</td>
</tr>
<tr>
<td>taxa.meansdn</td>
<td>Summarize mean, standard deviation of abundances and number of subjects by groups for all bacterial taxa or pathways</td>
</tr>
<tr>
<td>taxa.mean.plot</td>
<td>Plot mean abundance by groups (from taxa.meansdn output)</td>
</tr>
<tr>
<td>taxa.compare</td>
<td>Compare relative abundances of bacterial taxa at all levels using Generalized Additive Models for Location, Scale and Shape (GAMLSS) with a zero-inflated beta (BEZI) family (GAMLSS-BEZI) or linear/linear mixed effect models (LM) or linear/linear mixed effect models with arcsin squareroot transformation (LMAS)</td>
</tr>
<tr>
<td>pathway.compare</td>
<td>Compare relative abundances of bacterial functional pathways at all levels using GAMLSS-BEZI or LM or LMAS. Compare of log(absolute abundances) of bacterial functional pathways at all levels using LM</td>
</tr>
<tr>
<td>taxonomy.show</td>
<td>Display the results of relative abundance comparison (from taxa.compare or pathway.compare outputs)</td>
</tr>
<tr>
<td>meta.taxa</td>
<td>Perform meta-analysis of relative abundance estimates of bacterial taxa or pathways (either from GAMLSS-BEZI or LM or LMAS) across studies (from combined taxa.compare/pathway.compare outputs of all included studies) using random effect and fixed effect meta-analysis models</td>
</tr>
<tr>
<td>meta.tab.show</td>
<td>Display meta-analysis results of bacterial taxa or pathway relative abundances (from meta.taxa output)</td>
</tr>
<tr>
<td>meta.niceplot</td>
<td>Produce nice combined heatmap and forest plot for meta-analysis results of bacterial taxa and pathway relative abundances (from metatab.show output)</td>
</tr>
<tr>
<td>alpha.compare</td>
<td>Calculate average alpha diversity indexes for a specific rarefaction depth, standardize and compare alpha diversity indexes between groups</td>
</tr>
<tr>
<td>microbiomeage</td>
<td>Predict microbiota age using Random Forest model based on relative abundances of bacterial genera shared with the Bangladesh study [14]</td>
</tr>
<tr>
<td>read.multi</td>
<td>Read multiple files in a path to R</td>
</tr>
</tbody>
</table>

**RESULTS and DISCUSSION**

**Performance of GAMLSS-BEZI: simulation studies**

Simulation studies were performed to evaluate type I error and power of GAMLSS-BEZI for testing differential relative abundances of microbial taxonomies as compared to LM with arcsin squareroot transformation (LMAS) (implemented in MaAsLin software [16]). LMAS was chosen for comparison with GAMLSS-BEZI because it is a commonly used approach for microbiome differential relative abundance testing and similarly to GAMLSS-BEZI, it allows covariate adjustment and can be used for longitudinal or non-longitudinal data. Simulations of
zero-inflated relative abundance data were based on the R package “gamlss.dist” version 5.0-3 [27].

**Type I error**

Three scenarios were simulated to mimic typical case-control microbiome studies of small (number of cases (n1)=number of controls (n2)=10), medium (n1=n2=50) and large (n1=n2=200) sample sizes. The same location parameter values (mu1=mu2=0.5), precision parameter values (sigma1=sigma2=5) and parameter values modelling the probability at zero (nu1=nu2=0.5) were used for case and control groups. These parameters were chosen to generate the data that mimics real microbiome relative abundance data used in our later examples. The simulations were repeated 10000 times for each sample size. Type I error was calculated for three different alpha levels of 0.01, 0.05 and 0.1. Under these conditions, type I errors were well controlled by both GAMLSS-BEZI and LMAS (Table 2).

Table 2. Type I error of GAMLSS-BEZI and LMAS: simulation studies.

<table>
<thead>
<tr>
<th>Sample size</th>
<th>GAMLSS-BEZI (Alpha level=0.01)</th>
<th>LMAS (Alpha level=0.01)</th>
<th>GAMLSS-BEZI (Alpha level=0.05)</th>
<th>LMAS (Alpha level=0.05)</th>
<th>GAMLSS-BEZI (Alpha level=0.1)</th>
<th>LMAS (Alpha level=0.1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>0.0137</td>
<td>0.0118</td>
<td>0.0698</td>
<td>0.0498</td>
<td>0.1138</td>
<td>0.0987</td>
</tr>
<tr>
<td>50</td>
<td>0.0097</td>
<td>0.0099</td>
<td>0.0508</td>
<td>0.0495</td>
<td>0.1030</td>
<td>0.0982</td>
</tr>
<tr>
<td>200</td>
<td>0.0104</td>
<td>0.0108</td>
<td>0.0517</td>
<td>0.0521</td>
<td>0.1037</td>
<td>0.1027</td>
</tr>
</tbody>
</table>

The same location parameter values (mu1=mu2=0.5), precision parameter values (sigma1=sigma2=5) and parameter values modelling the probability at zero (nu1=nu2=0.5) were used for case and control groups. Simulations were repeated 10000 times.

GAMLSS-BEZI: Generalized Additive Models for Location, Scale and Shape (GAMLSS) with a zero inflated beta (BEZI) family; LMAS: linear model with arcsin square root transformation (implemented in the software MaAsLin).

**Power**

We simulated 5000 bacterial species including 1000 species with no difference between case and control groups (mu1=mu2=0.5) and 4000 species with true difference between case and control groups. Four scenarios of true difference were used (1000 species each): mu1=0.5 vs. mu2=0.4,
mu1=0.5 vs. mu2=0.3, mu1=0.5 vs. mu2=0.2, mu1=0.2 vs. mu2=0.1. Other parameters were the same for case and control groups as above. A sample size of n=100 for both case and control groups was used in all scenarios. Performance of GAMLSS-BEZI and LMAS was evaluated based on the receiver operating characteristic (ROC) curve for identifying species with differential abundance between case and control groups. The analysis for the ROC and area under the curve (AUC) was done using the R package pROC version 1.10.0 [27]. GAMLSS-BEZI (AUC=91.1%, 95% CI=(90.2%, 91.9%)) significantly outperformed LMAS (AUC=86.2%, 95% CI=(85.2%, 87.3%)) (DeLong’s test p-value < 2.2e-16) (Figure 1).

Performance of GAMLSS-BEZI: application to real microbiome data

Using data from a study of Bangladeshi infants [14], which included 996 stool samples collected monthly from birth to 2 years of life in 50 subjects, we demonstrate that GAMLSS-BEZI outperforms LMAS in detecting differential relative abundances between various grouping variables.

Example 1: Comparison between non-exclusively breastfed (non-EBF) vs. exclusively breastfed (EBF) infants ≤ 6 months of age

Figure 2 showed average of relative abundance of bacterial phyla in non-EBF and EBF infants from birth to 6 months of age. A higher abundance of Proteobacteria, Firmicutes, and Bacteroidetes as well as a lower abundance of Actinobacteria was observed in non-EBF versus EBF infants. GAMLSS-BEZI was able to detect a significant difference in all four of these phyla whereas LMAS could only detect a significant difference in three phyla (Table 3).

Table 3. Results of GAMLSS-BEZI and LMAS: real microbiome data example 1 (non-EBF vs. EBF).
### Data from Bangladesh study.
Comparison of bacterial relative abundances at phylum level between non-exclusively breastfed (non-EBF) vs. exclusively breastfed (EBF) infants ≤ 6 months of age using GAMLSS-BEZI vs. LMAS.

Significant p-values (<0.05) are in bold. GAMLSS-BEZI: Generalized Additive Models for Location, Scale and Shape (GAMLSS) with a zero inflated beta (BEZI) family; LMAS: linear model with arcsin square root transformation (implemented in the software MaAsLin); FDR: false discovery rate.

**Example 2: Comparison between infants from 6 months to 2 years of age introduced to solid food after 5 months vs. before 5 months**

Figure 3 showed average of relative abundance of bacterial phyla in two groups of infants from 6 months to 2 years of age who was introduced to solid food after 5 months vs. before 5 months of life. Lower relative abundances of Firmicutes, Bacteroidetes and higher relative abundance of Actinobacteria was observed in infants with solid food introduction after 5 months.

GAMLSS-BEZI detected all three of these differences whereas LMEM could only detect a significant difference in one phylum (Table 4).

Example 1 and 2 demonstrate the increased sensitivity of GAMLSS-BEZI in detecting bacterial taxa with differential relative abundances as compared to LMAS.

**Table 4. Results of GAMLSS-BEZI and LMAS: real microbiome data example 2 (solid food introduction after 5 months vs. before 5 months of age).**

<table>
<thead>
<tr>
<th>Bacterial phyla</th>
<th>Estimate</th>
<th>95% Lower limit</th>
<th>95% Upper limit</th>
<th>p-value</th>
<th>FDR adjusted p-value</th>
<th>Estimate</th>
<th>95% Lower limit</th>
<th>95% Upper limit</th>
<th>p-value</th>
<th>FDR adjusted p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Actinobacteria</td>
<td>0.19</td>
<td>0.04</td>
<td>0.34</td>
<td>0.0119</td>
<td>0.0208</td>
<td>0.05</td>
<td>0.06</td>
<td>0.16</td>
<td>0.3451</td>
<td>0.3451</td>
</tr>
<tr>
<td>Bacteroidetes</td>
<td>-0.26</td>
<td>-0.42</td>
<td>-0.10</td>
<td>0.0018</td>
<td>0.0070</td>
<td>-0.05</td>
<td>-0.09</td>
<td>-0.01</td>
<td>0.027</td>
<td>0.1079</td>
</tr>
<tr>
<td>Firmicutes</td>
<td>-0.16</td>
<td>-0.30</td>
<td>-0.03</td>
<td>0.0156</td>
<td>0.0208</td>
<td>-0.04</td>
<td>-0.12</td>
<td>0.04</td>
<td>0.3168</td>
<td>0.3451</td>
</tr>
<tr>
<td>Proteobacteria</td>
<td>0.14</td>
<td>-0.02</td>
<td>0.30</td>
<td>0.0861</td>
<td>0.0861</td>
<td>0.02</td>
<td>-0.02</td>
<td>0.07</td>
<td>0.2916</td>
<td>0.3451</td>
</tr>
</tbody>
</table>

Data from Bangladesh study. Comparison of bacterial relative abundances at phylum level between infants from 6 months to 2 years of age with solid food introduction after 5 months vs. before 5 months using GAMLSS-BEZI vs. LMAS. Significant p-values (<0.05) are in bold. GAMLSS-BEZI: Generalized Additive Models for Location, Scale and Shape (GAMLSS) with a zero inflated beta (BEZI) family; LMAS: linear model with arcsin square root transformation (implemented in the software MaAsLin); FDR: false discovery rate.
Example 3: Effect of diarrheal episodes on bacterial phyla stratified by duration of exclusive breastfeeding (EBF)

Figure 4 showed average of relative abundance of bacterial phyla in groups of infants from 6 months to 2 years of age with vs. without diarrhea around the time of stool sample collection stratified by duration of EBF. In infants who received less than two months of EBF, a higher abundance of Firmicutes and a lower abundance of Actinobacteria was observed in the groups of infants with diarrhea vs. those without diarrhea (Figure 4, upper panel). GAMLSS-BEZI detected a significant difference in both Firmicutes and Actinobacteria. In contrast, in infants who received more than two months of EBF, no difference in relative abundance of any bacterial phylum was observed between those with diarrhea vs. those without diarrhea (Figure 4, lower panel) and GAMLSS-BEZI did not report any significant difference (Table 5). This example demonstrates that GAMLSS-BEZI detects differential abundances when there is observed difference and does not report difference when there is no observed difference.

Table 5. Results of GAMLSS-BEZI and LMAS: real microbiome data example 3 (diarrhea vs. no diarrhea stratified by duration of EBF).

<table>
<thead>
<tr>
<th>Bacterial phyla</th>
<th>GAMLSS-BEZI</th>
<th>LMAS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Estimate</td>
<td>95% Lower limit</td>
</tr>
<tr>
<td>Actinobacteria</td>
<td>-0.73</td>
<td>-1.12</td>
</tr>
<tr>
<td>Bacteroidetes</td>
<td>-0.29</td>
<td>-0.68</td>
</tr>
<tr>
<td>Firmicutes</td>
<td>0.49</td>
<td>0.15</td>
</tr>
<tr>
<td>Proteobacteria</td>
<td>-0.17</td>
<td>-0.54</td>
</tr>
</tbody>
</table>

In infants with duration of exbf <=2 months (diarrhea vs. no diarrhea comparison)

In infants with duration of exbf >2 months (diarrhea vs. no diarrhea comparison)

Data from Bangladesh study. Comparison of bacterial relative abundances at phylum level in infants from 6 months to 2 years of age with diarrhea vs. no diarrhea at the time of stool sample collection stratified by duration of exclusive breastfeeding. Significant p-values (<0.05) are in bold. GAMLSS-BEZI: Generalized Additive Models for Location, Scale and Shape (GAMLSS) with a zero inflated beta (BEZI) family; LMAS: linear model with arcsin square root transformation (implemented in the software MaAsLin); FDR: false discovery rate.
Illustration of meta-analysis workflow with real microbiome data from four studies

We show examples for the comparison of gut microbiomes between male vs. female infants ≤ 6 months of age adjusting for feeding status and infant age at time of sample collection. We illustrate the meta-analyses across microbiome studies with data from four studies (total number of stool samples=610 [female=339, male=271]). These four studies include one in Bangladesh [14], one in Haiti [11], and two in the USA (California and Florida [CA_FL] and North Carolina [NC]) [12, 28] (Table 6, Additional file 1).

Table 6. Summary of four studies included in meta-analysis.

<table>
<thead>
<tr>
<th>Data origin (study population), reference</th>
<th>Study design</th>
<th>Sample size (≤ 6 months of age)</th>
<th>Target region of 16S rRNA genes /sequence platform</th>
<th>Starting files used and data processing done in this project</th>
</tr>
</thead>
<tbody>
<tr>
<td>Haiti (Bender et al. 2016)</td>
<td>One time stool sample collection of 48 HIV negative infants with age varied from 0 to 6 months whose mothers were HIV negative (n=23) or HIV positive (n=25).</td>
<td>Total number of samples: 48 (female =25, male =21)</td>
<td>V4 /Illumina MiSeq</td>
<td>- Open OTU picking with UCLUST with 97% similarity using the Greengenes database (version 13.8)</td>
</tr>
<tr>
<td>USA (California and Florida) (Pannaraj et al. 2017)</td>
<td>Longitudinal stool sample collection of 113 healthy full-term infants at 0 to 7 days, 8 to 30 days, 31 to 90 days, 91 to 180 days.</td>
<td>Total number of samples: 221 (female=120, male=101)</td>
<td>V4 /Illumina MiSeq</td>
<td></td>
</tr>
<tr>
<td>USA (North Carolina) (Thompson et al. 2015)</td>
<td>Longitudinal stool sample collection of 6 healthy full term infants with varied age.</td>
<td>Total number of samples: 21 (female=14, male=7)</td>
<td>V1-2 /Roche GS FLX Titanium</td>
<td></td>
</tr>
</tbody>
</table>

* This study also contain 674 stool samples > 6 months of age which were used in the analyses comparing the performance of GAMLSS-BEZI (Generalized Additive Models for Location, Scale and Shape (GAMLSS) with a zero inflated beta (BEZI) family) vs. LMAS (linear model with arcsin square root transformation). Data from this study was downloaded from the authors’ website: https://gordonlab.wustl.edu/Subramanian_6_14/Nature_2014_Processed_16S_rRNA_datasets.html. Data from three other studies were obtained directly from the investigators.

Alpha diversity

For each study, the alpha.compare function imports the outputs from “alpha_rarefaction.py” QIIME1 script and calculates mean alpha diversity for different indices for each sample based on a user defined rarefaction depth. Mean alpha diversity indexes are standardized to have a mean of 0 and standard deviation of 1 to make these measures comparable across studies. Standardized alpha diversity indexes are compared between groups adjusting for covariates using LM. Meta-analysis across studies is then done and the results are displayed as a standard meta-analysis forest plot (Additional file 1). Our results showed that alpha diversity (four commonly used
indexes Shannon, Phylogenetic diversity whole tree, Observed species, Chao1) was not different between male and female infants ≤6 months of age in the meta-analysis of the four included studies (pooled standardized alpha diversity (Shannon index) difference (DD)= -0.01 standard deviation (sd), 95% CI=[-0.15; 0.12], p-value=0.8288) (Figure 5, Additional file 1).

**Microbiota age**

A RF model based on the relative abundance of the shared bacterial genera of the four included studies explained 96% of the variance related to chronological age in the training set and 67% of the variance related to chronological age in the test set of Bangladesh data. This performance is better than the original RF model proposed by Subramanian et al [14]. The predicted infant age in each included study based on relative abundance of the shared gut bacterial genera using this RF model is referred to as gut microbiota age. Standardized gut microbiota age are compared between groups adjusting for covariates using LM and meta-analysis across studies is then done. Our results showed that standardized microbiota age was significantly different between genders but in opposite directions in two studies with small sample sizes (Haiti and North Carolina). However, meta-analysis of all four studies revealed no significant difference in gut microbiota age between genders after adjusting for feeding status and infant age at time of sample collection (pooled standardized microbiota age difference (MD)= -0.06 sd, 95% CI=[-0.32; 0.20], p-value=0.6348) (Figure 6, Additional file 1).

**Bacterial taxa relative abundance**

The adjusted estimates (log(odds ratio) of relative abundance between genders) from GAMLSS-BEZI for each bacterial taxon and the pooled adjusted estimates across studies (meta-analysis)
are displayed as a heatmap (Figure 7, left panel). The adjacent forest plot (right panel) displays the pooled adjusted estimates and their 95%CI with different colors and shapes to reflect the magnitude of pooled p-values. Our meta-analyses of four included studies showed that, after adjusting for feeding status and infant age at time of sample collection, there was no significant difference in bacterial taxa relative abundance from phylum to genus levels between male vs. female infants ≤6 months of age after adjusting for multiple testing except for genus *Coprococcus* whose relative abundance was significantly higher in male vs. female infants (FDR adjusted pooled p-value<0.0001) (Figure 7, Additional file 1).

*Bacterial functional (KEGG) pathway relative abundance*

Meta-analysis of four included studies showed no difference in relative abundances of bacterial KEGG pathways at level 2 and level 3 between genders after adjusting for feeding status and infant age at time of sample collection and adjusting for multiple testing (Figure 8, Additional file 1).

Difference in gut microbial composition between genders has been reported in adults [29–31] and in some neonatal studies albeit with small sample sizes [32, 33]. However, the reported findings have varied between these studies. Our meta-analyses of four studies showed virtually no difference in gut microbiomes between male vs. female infants ≤6 months of age after adjusting for feeding status and infant age at time of sample collection as well as adjusting for multiple testing. There was one exception: relative abundance of *Coprococcus* was significantly higher in male vs. female infants. *Coprococcus* has been implicated in many conditions.
including hypertension and autism [34, 35], and the detected difference in our study may provide some insights into the known sex biases in health outcomes.

**CONCLUSION**

Our *metamicrobiomeR* package implemented GAMLSS-BEZI for analysis of microbiome relative abundance data and random effect meta-analysis models for meta-analysis across microbiome studies. The advantages of GAMLSS-BEZI are: 1) it directly and properly address the distribution of microbiome relative abundance data which resemble a zero-inflated beta distribution; 2) it has better power to detect differential relative abundances between groups than the commonly used approach LMAS; 3) the estimates from GAMLSS-BEZI are log(odds ratio) of relative abundances between groups and thus are directly comparable across studies. Moreover, standardization of alpha diversity indexes and predicted microbiota age also yields comparable metrics for cross-study comparisons. Random effects meta-analysis models can be directly applied to pool these comparable adjusted estimates and their standard errors across studies. This approach allows examination of study-specific effects, heterogeneity between studies, and the overall pooled effects across microbiome studies. The examples and workflow using our “*metamicrobiomeR*” package are reproducible and applicable for the analysis and meta-analysis of other microbiome studies.

**AVAILABILITY and REQUIREMENTS**

Project name: metamicrobiomeR

Project home page: [https://github.com/nhanhocu/metamicrobiomeR](https://github.com/nhanhocu/metamicrobiomeR)

Operating system(s): Platform independent
Programming language: R

Other requirements: R 3.4.2 or higher

License: GNU GPL v. 2.

Any restrictions to use by non-academics: none.

**ABBREVIATIONS**

BEZI: zero inflated beta

CA: California

CI: Confidence interval

DD: Diversity difference

EBF: exclusive breastfeeding (or exclusively breastfed)

FDR: False Discovery Rate

RF: Random Forest

FL: Florida

GAMLSS: Generalized Additive Models for Location, Scale and Shape

GAMLSS-BEZI: Generalized Additive Models for Location, Scale and Shape with a zero inflated beta family

KEGG: Kyoto Encyclopedia of Genes and Genomes

LM: linear/linear mixed effect models

LMAS: linear/linear mixed effect models with arcsin squareroot transformation

MD: microbiota age difference

NC: North Carolina

Non- EBF: non exclusive breastfeeding (or non exclusively breastfed)
DECLARATIONS

Ethics approval and consent to participate
Not applicable

Consent for publication
Not applicable

Availability of data and material
All data used in this study are included in these published articles and their supplementary information files (references: [11, 12, 14, 28]). The data from the Bangladesh study [14] were downloaded from the authors’ website:


The data from three other studies were obtained directly from the investigators. All example datasets, documentations and source code of the “metamicrobiomeR” package are available at https://github.com/nhanhocu/metamicrobiomeR.

Competing interests
The authors declare that they have no competing interests.

Funding
This work was supported by Mervyn W. Susser fellowship in the Gertrude H. Sergievsky Center, Columbia University Medical Center (to Nhan Thi Ho).

Authors’ contribution
NTH conceived the ideas, wrote the R package, documentations, performed the simulations and analyses with inputs from FL.

Acknowledgements
We would like to thank Dr. Grace M. Aldrovandi (University of California at Los Angeles) and Dr. M. Andrea Azcarate-Peril (University of North Carolina at Chapel Hill) for providing the data used in the examples.

REFERENCES


**Figure 1. ROC curves for identifying simulated bacterial species with differential relative abundances by GAMLSS-BEZI and LMAS**

5000 bacterial species were simulated including 1000 species with no difference between case and control groups (location parameter values for cases (mu1) = location parameter values for controls (mu2)=0.5) and 4000 species with true difference between case and control groups. Four scenarios for true difference were simulated (1000 species each): mu1=0.5 vs. mu=0.4, mu1=0.5 vs. mu2=0.3, mu1=0.5 vs. mu2=0.2, mu1=0.2 vs. mu2=0.1. Precision parameter values...
(sigma1=sigma2=5) and parameter values modeling the probability at zero (nu1=nu2=0.5) were the same for case and control groups. A sample size of n=100 for both case and control groups was used in all scenarios.

The figure shows ROC curves and AUC with 95% confidence interval.

ROC curve: receiver operating characteristic (ROC) curve; AUC: area under the curve;

GAMLSS-BEZI: Generalized Additive Models for Location, Scale and Shape (GAMLSS) with a zero inflated beta (BEZI) family; LMAS: linear model with arcsin squareroot transformation (implemented in the software MaAsLin).
Figure 2. Relative abundances of bacterial phyla in non-exclusively breastfed vs. exclusively breastfed infants ≤ 6 months of age.

Data from Bangladesh study.
Figure 3. Relative abundances of bacterial phyla in infants from 6 months to 2 years of age with solid food introduction after 5 months vs. before 5 months.

Data from Bangladesh study.
Figure 4. Relative abundance of bacterial phyla in infants from 6 months to 2 years of age with diarrhea vs. without diarrhea at the time of stool sample collection stratified by duration of exclusive breastfeeding (EBF).

Data from Bangladesh study.
Figure 5. Meta-analysis of four included studies: the difference in gut microbial diversity between male vs. female infants ≤ 6 months.

The difference in gut microbial alpha diversity (standardized Shannon index) between male vs. female infants ≤ 6 months old from each study and the pooled effect across four included studies (meta-analysis). Estimates for diversity difference and corresponding standard error from each study were from linear mixed effect models (for longitudinal data) or linear models (for non-longitudinal data) and were adjusted for feeding status and age of infants at sample collection. Pooled estimates of standardized diversity difference and their 95% CI were from random effect meta-analysis models with inverse variance weighting and DerSimonian–Laird estimator for between-study variance based on the adjusted estimates and corresponding standard errors of all included studies (results of fixed effect meta-analysis model are also showed).

USA: United States of America; CA: California; FL: Florida; NC: North Carolina; DD: standardized diversity difference; SE: Standard error.

<table>
<thead>
<tr>
<th>Study</th>
<th>DD</th>
<th>SE</th>
<th>Standardized diversity difference</th>
<th>DD</th>
<th>95%–CI</th>
<th>Weight (fixed)</th>
<th>Weight (random)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bangladesh</td>
<td>−0.00</td>
<td>0.0821</td>
<td></td>
<td>−0.00</td>
<td>[−0.16; 0.16]</td>
<td>70.0%</td>
<td>70.0%</td>
</tr>
<tr>
<td>Haiti</td>
<td>−0.01</td>
<td>0.2936</td>
<td></td>
<td>−0.01</td>
<td>[−0.58; 0.57]</td>
<td>5.5%</td>
<td>5.5%</td>
</tr>
<tr>
<td>USA(CA, FL)</td>
<td>−0.07</td>
<td>0.1434</td>
<td></td>
<td>−0.07</td>
<td>[−0.35; 0.21]</td>
<td>23.0%</td>
<td>23.0%</td>
</tr>
<tr>
<td>USA(NC)</td>
<td>0.18</td>
<td>0.5553</td>
<td></td>
<td>0.18</td>
<td>[−0.91; 1.27]</td>
<td>1.5%</td>
<td>1.5%</td>
</tr>
</tbody>
</table>

Fixed effect model
Random effects model

Heterogeneity: $I^2 = 0\%$, $r^2 = 0$, $\rho = 0.96$
Figure 6. Meta-analysis of four included studies: the difference in gut microbiota age between male vs. female infants ≤ 6 months.

The difference in gut (standardized) microbiota age between male vs. female infants ≤ 6 months old from each study and the pooled effect across four included studies (meta-analysis). Estimates for standardized microbiota age difference and corresponding standard error from each study were from linear mixed effect models (for longitudinal data) or linear models (for non-longitudinal data) and were adjusted for feeding status and age of infants at sample collection. Pooled estimates of standardized microbiota age difference and their 95% CI were from random effect meta-analysis models with inverse variance weighting and DerSimonian–Laird estimator for between-study variance based on the adjusted estimates and corresponding standard errors of all included studies (results of fixed effect meta-analysis model are also showed).

USA: United States of America; CA: California; FL: Florida; NC: North Carolina; MD:

<table>
<thead>
<tr>
<th>Study</th>
<th>MD</th>
<th>SE</th>
<th>Standardized microbiota age difference</th>
<th>MD</th>
<th>95%-CI</th>
<th>Weight (fixed)</th>
<th>Weight (random)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bangladesh</td>
<td>-0.10</td>
<td>0.0636</td>
<td></td>
<td>-0.10</td>
<td>[-0.22; 0.03]</td>
<td>71.4%</td>
<td>40.7%</td>
</tr>
<tr>
<td>Haiti</td>
<td>0.56</td>
<td>0.2723</td>
<td></td>
<td>0.56</td>
<td>[0.03; 1.09]</td>
<td>3.9%</td>
<td>15.4%</td>
</tr>
<tr>
<td>USA(CA, FL)</td>
<td>-0.10</td>
<td>0.1133</td>
<td></td>
<td>-0.10</td>
<td>[-0.32; 0.12]</td>
<td>22.5%</td>
<td>33.7%</td>
</tr>
<tr>
<td>USA(NC)</td>
<td>-0.73</td>
<td>0.3617</td>
<td></td>
<td>-0.73</td>
<td>[-1.44; -0.02]</td>
<td>2.2%</td>
<td>10.2%</td>
</tr>
</tbody>
</table>

Fixed effect model
Random effects model

Heterogeneity: $I^2 = 66\%$, $\tau^2 = 0.0385$, $p = 0.03$
Figure 7. Meta-analysis of four included studies: the difference in relative abundances of gut bacterial taxa between male vs. female infants ≤ 6 months of age.

A: Phylum level: heatmap of log(odds ratio) (log(OR)) of relative abundances of all gut bacterial phyla between male vs. female infants for each study and pooled estimates across all studies with 95% confidence intervals (95% CI) (forest plot). B: Genus level: heatmap of log(OR) of relative abundances of all gut bacterial genera between male vs. female infants for each study and pooled estimates across all studies with 95% CI (forest plot). All log(OR) estimates of each bacterial taxa from each study were from Generalized Additive Models for Location Scale and Shape (GAMLSS) with beta zero inflated family (BEZI) and were adjusted for feeding status and age of infants at sample collection. Pooled log(OR) estimates and 95% CI (forest plot) were from random effect meta-analysis models with inverse variance weighting and DerSimonian–Laird estimator for between-study variance based on the adjusted log(OR) estimates and corresponding standard errors of all included studies.

Pooled log(OR) estimates with pooled p-values<0.05 are in red and those with false discovery rate (FDR) adjusted pooled p-values <0.1 are in triangle shape.

Missing (unavailable) values are in white.

USA: United States of America; CA: California; FL: Florida; NC: North Carolina.
Figure 8. Meta-analysis of four included studies: the difference in relative abundances of gut microbial KEGG (Kyoto Encyclopedia of Genes and Genomes) pathways between male vs. female infants ≤ 6 months of age.

Heatmap of log(odds ratio) (log(OR)) of relative abundances of gut microbial KEGG pathways at level 3 between male vs. female infants for each study and pooled estimates of all studies with 95% confidence intervals (forest plot). Only pathways with pooled p-value <0.3 are shown. All log(OR) estimates of each pathway from each study were from Generalized Additive Models for Location Scale and Shape (GAMLSS) with beta zero inflated family (BEZI) and were adjusted for feeding status and age of infants at sample collection. Pooled log(OR) estimates and 95% confidence intervals (95%CI) (forest plot) were from random effect meta-analysis models with inverse variance weighting and DerSimonian–Laird estimator for between-study variance based on the adjusted log(OR) estimates and corresponding standard errors of all included studies. Pooled log(OR) estimates with pooled p-values<0.05 are in red and those with false discovery rate (FDR) adjusted pooled p-values <0.1 were in triangle shape.

USA: United States of America; CA: California; FL: Florida; NC: North Carolina.