Ravel_2013_BV

```
library(SummarizedExperiment)
library(dplyr)
library(tidyr)
library(tibble)
library(purrr)
library(ggplot2)
source("../R/functions_da.R")
```

The data comes from the article of Ravel et. al. 2013.

Reference:

Ravel, J., Brotman, R.M., Gajer, P. et al. Daily temporal dynamics of vaginal microbiota before, during and after episodes of bacterial vaginosis. Microbiome 1, 29 (2013). https://doi.org/10.1186/2049-2618-1-29

```
se <- readRDS("Ravel_2013_16S_BV_se.rds")
se
```

```
## class: SummarizedExperiment
## dim: 151 1657
## metadata(0):
## assays(2): counts relative_abundance
## rownames(151): Lactobacillus iners Gardnerella vaginalis ... Aerococcus
##   viridans Streptococcus mutans
## rowData names(6): superkingdom phylum ... family genus
## colnames(1657): s3.w1d1 s3.w1d2 ... s135.w10d7 s135.w11d1
## colData names(20): subjectID dayInStudy ... MENSTRU3 Total.Reads
```

Exploring sample metadata

```
sample_metadata <- se %>%
    colData() %>%
    as.data.frame() %>%
    rownames_to_column(var = "Sample") %>%
    as_tibble() %>%
    mutate(
        subjectID = as.factor(subjectID),
        pH = as.double(pH),
        Nugent_Score = as.integer(Nugent_Score),
        ## Is the sample positive to either SBV or ABV using Amsel?
        amsel_pos = case_when(SBV == "1" | ABV == "1" ~ 1, TRUE ~ 0),
        ## Classification according to the Amsel and the article's criteria
        amsel_res = case_when(
            SBV == "1" ~ "SBV", ABV == "1" ~ "ABV", TRUE ~ "Neg"
        ),
        nugent_res = case_when(
            ## Classification based on Nugent scores
```

```
              Nugent_Score <= 3 ~ "Negative",
              Nugent_Score >=4 & Nugent_Score <= 6 ~ "Intermediate",
              Nugent_Score >= 7 ~ "Positive"
          ),
          nugent_res = factor(
              nugent_res, levels = c("Negative", "Intermediate", "Positive")
          )
      )
```

```
## Warning in mask$eval_all_mutate(quo): NAs introduced by coercion
```

```
## Warning in mask$eval_all_mutate(quo): NAs introduced by coercion
```

```
glimpse(sample_metadata)
```

```
## Rows: 1,657
## Columns: 24
## $ Sample        <chr> "s3.w1d1", "s3.w1d2", "s3.w1d3", "s3.w1d5", "s3.w1d6", "~
## $ subjectID     <fct> 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3,~
## $ dayInStudy    <dbl> 1, 2, 3, 5, 6, 8, 9, 10, 11, 12, 14, 15, 16, 17, 18, 19,~
## $ DIA_WEEK      <dbl> 1, 1, 1, 1, 1, 2, 2, 2, 2, 2, 2, 3, 3, 3, 3, 3, 3, 3, 4,~
## $ DIA_DAY       <dbl> 1, 2, 3, 5, 6, 1, 2, 3, 4, 5, 7, 1, 2, 3, 4, 5, 6, 7, 1,~
## $ AGE           <dbl> 25, 25, 25, 25, 25, 25, 25, 25, 25, 25, 25, 25, 25, 25, ~
## $ ETHa          <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,~
## $ SBV           <dbl> 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,~
## $ ABV           <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,~
## $ BV_medication <dbl> 0, 0, 0, 0, 0, 0, 1, 1, 1, 1, 1, 1, 0, 0, 0, 0, 0, 0, 0,~
## $ pH            <dbl> NA, 4.0, 4.0, 4.0, 5.3, 4.7, 4.0, 4.0, 4.0, 4.0, 4.0, 4.~
## $ Nugent_Score  <int> 1, 4, 1, 6, 6, 7, 7, 7, 4, 4, 1, 3, 3, 3, 4, 3, 1, 1, 1,~
## $ VAG_ODOR      <chr> "NA", "0", "0", "0", "0", "1", "1", "0", "0", "0", "0", ~
## $ VAG_IRR       <chr> "NA", "0", "0", "0", "0", "0", "0", "0", "0", "0", "0", ~
## $ VAG_ITCH      <chr> "NA", "0", "0", "0", "0", "0", "0", "0", "0", "0", "0", ~
## $ VAG_BURN      <chr> "NA", "0", "0", "0", "0", "0", "0", "0", "NA", "0", "0",~
## $ VAG_DIS       <chr> "NA", "0", "0", "0", "0", "1", "0", "0", "NA", "1", "1",~
## $ MENSTRU1      <chr> "NA", "NA", "NA", "NA", "NA", "NA", "NA", "NA", "NA", "N~
## $ MENSTRU2      <chr> "NA", "NA", "NA", "NA", "NA", "NA", "NA", "NA", "NA", "N~
## $ MENSTRU3      <chr> "NA", "NA", "NA", "NA", "NA", "NA", "NA", "NA", "NA", "N~
## $ Total.Reads   <dbl> 206, 328, 253, 156, 131, 131, 142, 253, 293, 280, 320, 4~
## $ amsel_pos     <dbl> 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,~
## $ amsel_res     <chr> "Neg", "Neg", "Neg", "Neg", "Neg", "SBV", "Neg", "Neg", ~
## $ nugent_res    <fct> Negative, Intermediate, Negative, Intermediate, Intermed~
```

```
sample_metadata %>%
    count(subjectID, nugent_res, name = "count") %>%
    filter(!is.na(nugent_res)) %>%
    ggplot(aes(subjectID, count)) +
    geom_col(aes(fill = nugent_res), position = "stack") +
    geom_text(
        aes(label = count, group = nugent_res),
        position = position_stack(vjust = 0.5)
    ) +
    theme_bw() +
    theme(
        axis.text.x = element_text(angle = 45, hjust = 1)
    )
```
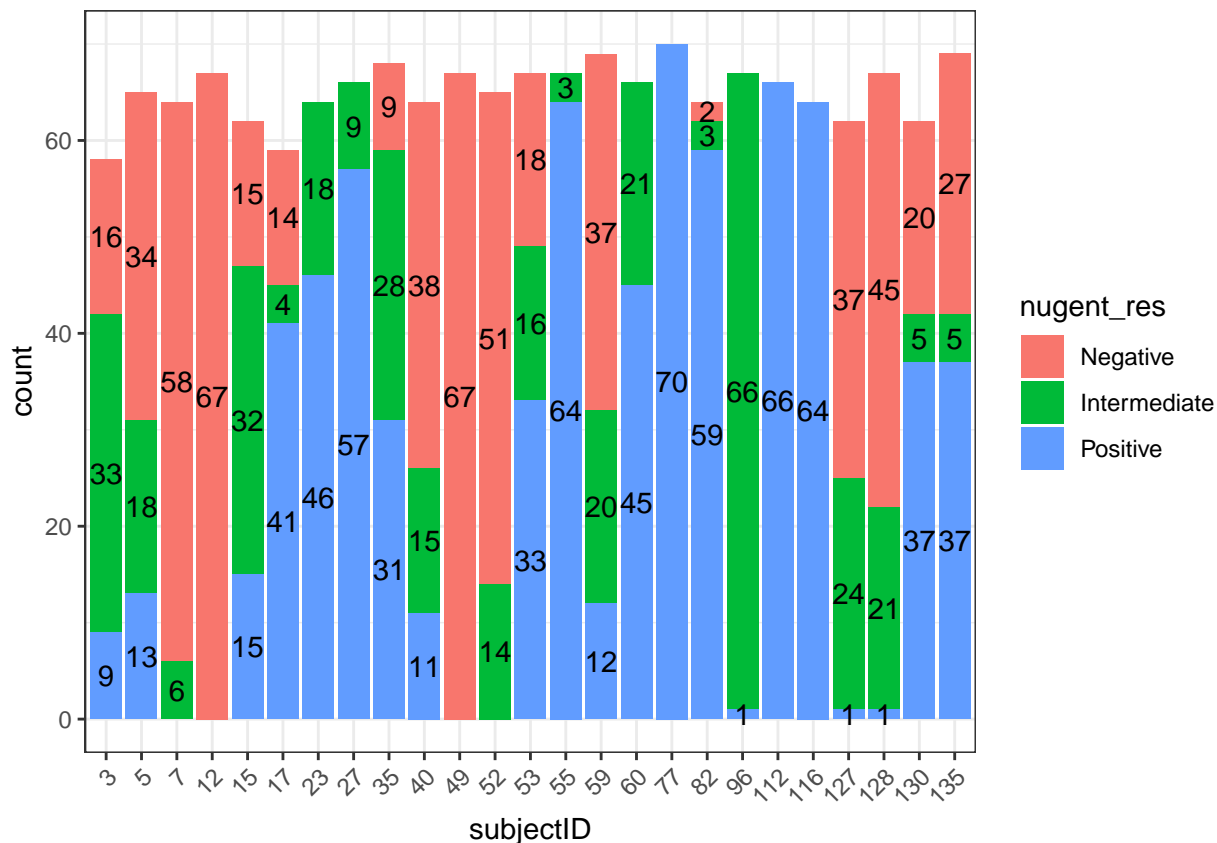
Select subjects

Subjects with nine or more samples classified in both positive and negative BV were selected.

```r
subject_neg_pos_counts <- sample_metadata %>%
  filter(
      !is.na(nugent_res),
      nugent_res != "Intermediate",
  ) %>%
  count(subjectID, nugent_res) %>%
  pivot_wider(names_from = nugent_res, values_from = n) %>%
  drop_na() %>%
  filter(Negative >= 9, Positive >= 9)

subject_neg_pos_counts
```

```
## # A tibble: 10 x 3
##    subjectID Negative Positive
##    <fct>        <int>    <int>
## 1 3              16        9
## 2 5              34       13
## 3 15             15       15
## 4 17             14       41
## 5 35              9       31
## 6 40             38       11
## 7 53             18       33
```

```
##  8  59               37        12
##  9 130               20        37
## 10 135               27        37
```

```
subject_neg_pos_counts %>%
  summarise(across(c(Negative, Positive), sum))
```

```
## # A tibble: 1 x 2
##   Negative Positive
##      <int>    <int>
## 1      228      239
```

```
selected_sample_metadata <- sample_metadata %>%
    filter(
        subjectID %in% subject_neg_pos_counts$subjectID,
        nugent_res != "Intermediate"
    )
selected_samples <- selected_sample_metadata$Sample
```

Nugent score vs pH of selected subjects

There should be a clear positive correlation pattern between pH and Nugent score.

```
selected_sample_metadata %>%
    mutate(pH = as.double(pH), Nugent_Score = as.integer(Nugent_Score)) %>%
    suppressWarnings() %>%
    filter(!is.na(pH), !is.na(Nugent_Score)) %>%
    ggplot(aes(Nugent_Score, pH )) +
    geom_boxplot(aes(group = Nugent_Score, color = nugent_res)) +
    # geom_point(size = .25, position = "jitter", shape = 1) +
    geom_smooth() +
    scale_x_continuous(breaks = seq(0, 10, 1)) +
    theme_bw()
```
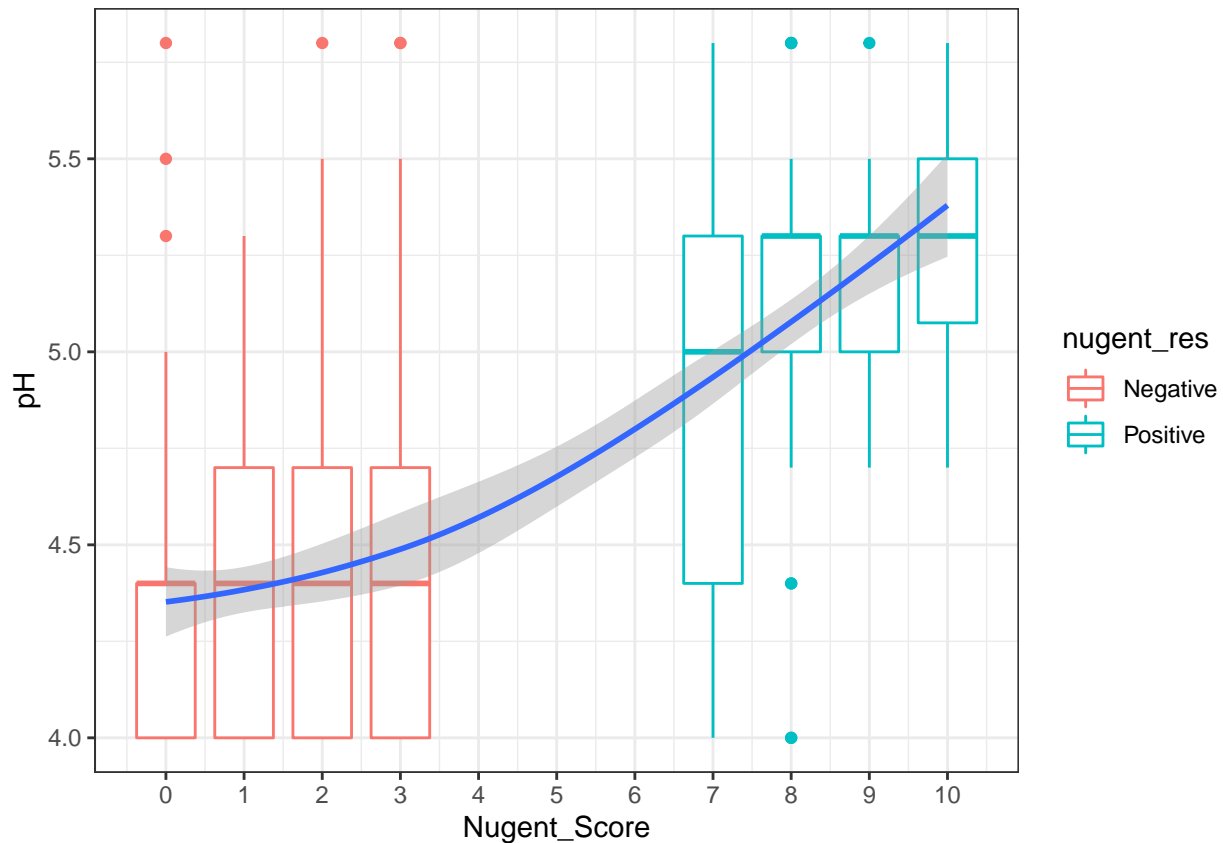
Summarize by genus

```
se_genus <- mia::agglomerateByRank(se, rank = "genus")
se_genus
```

```
## class: SummarizedExperiment
## dim: 63 1657
## metadata(0):
## assays(2): counts relative_abundance
## rownames(63):
##   Firmicutes_Bacilli_Lactobacillales_Lactobacillaceae_Lactobacillus
##   Actinobacteria_Actinomycetia_Bifidobacteriales_Bifidobacteriaceae_Gardnerella
##   ...
##   Proteobacteria_Gammaproteobacteria_Moraxellales_Moraxellaceae_Acinetobacter
##   Proteobacteria_Gammaproteobacteria_Pseudomonadales_NA_NA
## rowData names(6): superkingdom phylum ... family genus
## colnames(1657): s3.w1d1 s3.w1d2 ... s135.w10d7 s135.w11d1
## colData names(20): subjectID dayInStudy ... MENSTRU3 Total.Reads
```

Get taxonomy table:

```
taxa_genus <- se_genus %>%
    rowData() %>%
    as.data.frame %>%
    rownames_to_column(var = "tax_name") %>%
    as_tibble()
```

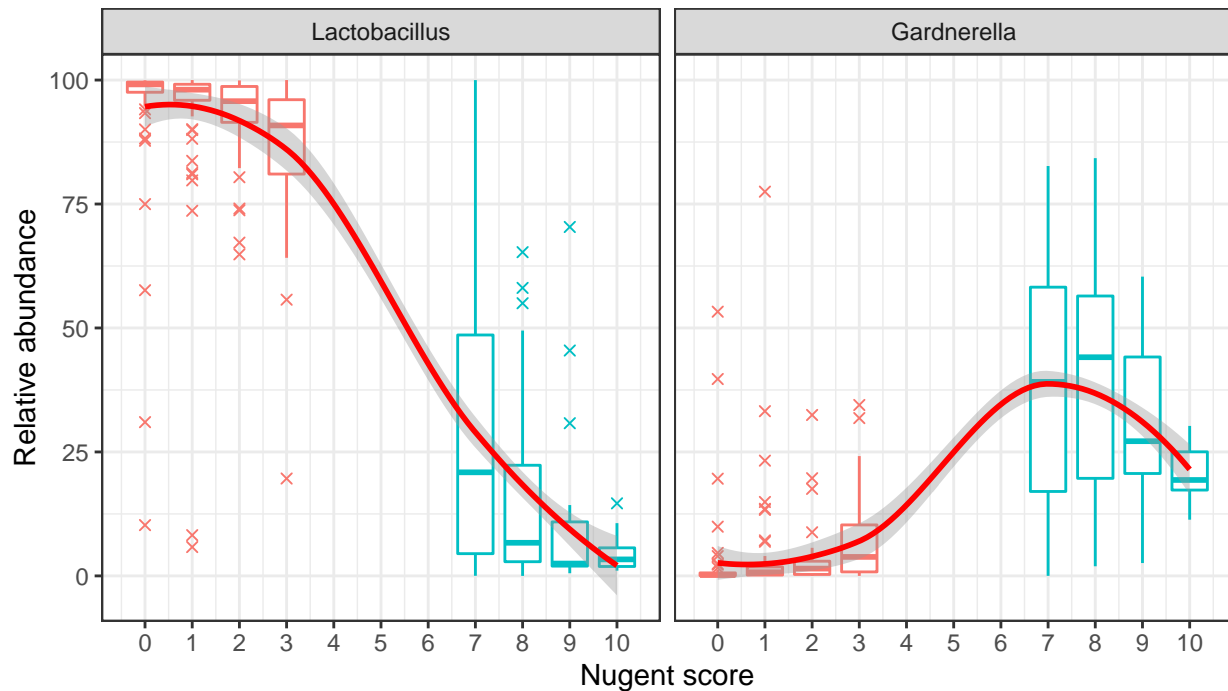Combine sample metadata, relative abundance, and taxonomy into a single tibble:

```r
tidy_table_genus <- assays(se_genus)$relative_abundance %>%
    as.data.frame() %>%
    rownames_to_column(var = "Taxa") %>%
    as_tibble() %>%
    pivot_longer(names_to = "Sample", values_to = "Abundance", 2:last_col()) %>%
    right_join(selected_sample_metadata, by = "Sample") %>%
    right_join(taxa_genus, by = c("Taxa" = "tax_name")) %>%
    relocate(Sample, Taxa, Abundance)
```

Nugent score vs relative abundance of Lactubacillus

```r
taxa_genera <- c(
    "Lactobacillus$",
    "Gardnerella$"
)
taxa_regex <- paste0("(", paste0(taxa_genera, collapse = "|"), ")")

tidy_table_genus %>%
    filter(grepl(taxa_regex, Taxa)) %>%
    select(Sample, Taxa, Abundance, pH, Nugent_Score, nugent_res) %>%
    mutate(
        Taxa = sub("^.+_", "", Taxa),
        Taxa = factor(Taxa, levels = sub("\\$$", "", taxa_genera)),
    ) %>%
    drop_na() %>%
    suppressWarnings() %>%
    ggplot(aes(Nugent_Score, Abundance)) +
    geom_boxplot(
        aes(group = Nugent_Score, color = nugent_res), outlier.shape = 4
    ) +
    # geom_point(size = .25, position = "jitter") +
    geom_smooth(color = "red") +
    facet_wrap(~Taxa, ncol = 2) +
    scale_x_continuous(breaks = seq(0,10,1)) +
    scale_color_discrete(name = "BV diagnostic (Nugent score)") +
    labs(
        y = "Relative abundance", x = "Nugent score",
        title = "Ravel et. al., 2013"
    ) +
    theme_bw() +
    theme(
        legend.position = "bottom"
    )
```

Ravel et. al., 2013

```r
selected_se <- se[,selected_samples]
taxa_lgl <- rowSums(assays(selected_se)$counts) > 0
selected_se <- selected_se[taxa_lgl, ]
nugent_scores <- colData(selected_se)$Nugent_Score
colData(selected_se)$BV <-
  ifelse(nugent_scores >= 7, "Positive BV", "Negative BV")
table(colData(selected_se)$BV)
```

```
##
## Negative BV Positive BV
##         259         208
```

Differential expression analysis

Some variables

```r
mode(assays(selected_se)$counts) <- "integer"
grp <- "BV"
ref <- "Negative BV"

deseq_results <- deseq2Poscounts(selected_se, grp = grp, ref = ref)

deseq_results %>%
  filter(ADJPVAL <=0.1)
```

```
## # A tibble: 57 x 4
##    TAXA                    FC    PVAL  ADJPVAL
##    <chr>                <dbl>   <dbl>    <dbl>
```

7

```
##  1 Lactobacillus iners       -3.83  2.73e-54 2.76e-52
##  2 Gardnerella vaginalis      2.14  1.46e-27 1.84e-26
##  3 Lactobacillus crispatus   -3.40  2.52e- 9 9.08e- 9
##  4 Atopobium vaginae          2.90  5.19e-34 1.31e-32
##  5 Lactobacillus jensenii    -5.12  1.20e-25 1.21e-24
##  6 Megasphaera sp. type 1     2.07  3.15e-12 1.51e-11
##  7 BVAB2                      3.89  8.20e-47 4.14e-45
##  8 Aerococcus christensenii  0.572 1.01e- 2 1.93e- 2
##  9 Bifidobacteriaceae         1.80  1.20e-15 7.45e-15
## 10 Lactobacillus gasseri     -5.51  6.36e-15 3.38e-14
## # ... with 47 more rows
```