# Pairwise Comparison of Beta Diversity

Jung-Jin Lee

Jan 15, 2020

## Contents

## Read in data

```r
# Mapping file
# SampleID: sample identifier
s <- "meta.tsv" %>%
  read_tsv()
dim(s)
```

```
## [1] 140  26
```

```r
# Bacteria counts data
bac_counts <- "bacteria_counts.tsv" %>%
  read_tsv() %>%
  column_to_rownames(var = "Taxa") %>%
  t()
dim(bac_counts) # Each row is a SampleID and each column is a species
```

```
## [1]  140 4373
```

```r
# Bacteria props data
bac_props <- sweep(bac_counts, 1, rowSums(bac_counts), "/")
dim(bac_props)
```

```
## [1]  140 4373
```

```r
# Make a distance object using usedist::dist_make()
dist_beta <- dist_make(bac_props, bray_curtis)

# Make sure that SampleIDs are aligned
dist_beta <- dist_subset(dist_beta, s$SampleID)
all(attributes(dist_beta)$Labels == s$SampleID)
```

```
## [1] TRUE
```

## Overall beta diversity

```r
pcoa_result <- ape::pcoa(dist_beta)
pcoa_df <- s %>%
  left_join(pcoa_result$vectors[, c("Axis.1", "Axis.2")] %>%
              as.data.frame() %>%
              rownames_to_column("SampleID"),
            by = "SampleID")
pcoa_pct <- round(pcoa_result$values$Relative_eig*100, 1)
```

```r
# Overall PCoA plot
pcoa_df %>%
  ggplot(aes(Axis.1, Axis.2)) +
  geom_point(aes(color = Hospital)) +
  theme_bw() +
  xlab(paste0("PCoA axis 1 (", pcoa_pct[1], "%)")) +
  ylab(paste0("PCoA axis 2 (", pcoa_pct[2], "%)")) +
  theme(aspect.ratio = 1) +
  theme(legend.position = "bottom")
```
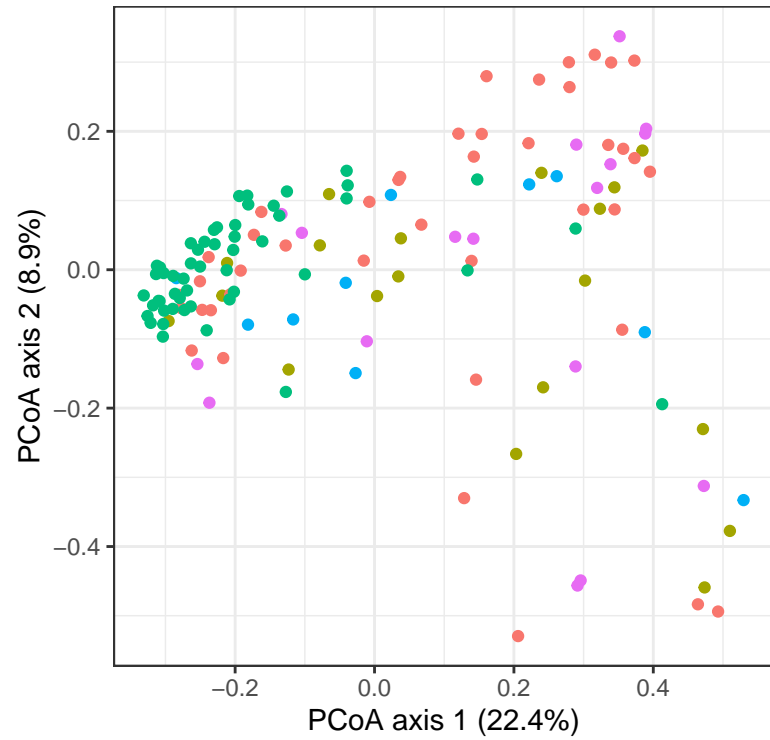
```
# PermANOVA
adonis(dist_beta ~ Hospital, data = s, permutation = 999)$aov.tab %>%
  as.data.frame() %>%
  rownames_to_column(var = "Term") %>%
  gt() %>%
  fmt_number(columns = c("SumsOfSqs", "MeanSqs", "F.Model")) %>%
  fmt_percent(columns = "R2")
```

| Term      | Df  | SumsOfSqs | MeanSqs | F.Model | R2      | Pr(>F) |
|-----------|-----|-----------|---------|---------|---------|--------|
| Hospital  | 4   | 4.81      | 1.20    | 4.28    | 11.25%  | 0.001  |
| Residuals | 135 | 37.96     | 0.28    | NA      | 88.75%  | NA     |
| Total     | 139 | 42.77     | NA      | NA      | 100.00% | NA     |

# Pairwise comparison of beta diversity

```r
pairwise_beta_plot <- function(s, dist_beta, color_by) {
  ###========
  # Step I:
  # Samples in dist_in may not appear in the metadata. Extract only common samples
  ###========
  common_samples <- intersect(s$SampleID, attributes(dist_beta)$Labels)
  s <- s %>%
    filter(SampleID %in% common_samples)
  dist_beta <- dist_subset(dist_beta, s$SampleID)

  ###========
  # Step II:
  # Treat color_by as a factor
  ###========
  color_by <- enquo(color_by)
  if (!is.factor(s %>% pull(!!color_by))) {
    s <- s %>%
      mutate(!!color_by := factor(!!color_by))
  }
  color_by_levels = s %>% pull(!!color_by) %>% levels()

  ###========
  # Step III:
  # Compute PCoA
  ###========
  pcoa_result <- ape::pcoa(dist_beta)
  pcoa_df <- s %>%
    left_join(pcoa_result$vectors[, c("Axis.1", "Axis.2")] %>%
                as.data.frame() %>%
                rownames_to_column("SampleID"),
              by = "SampleID")
  pcoa_pct <- round(pcoa_result$values$Relative_eig*100, 1)

  ###========
  # Step IV:
  # Create pairwise ordination
  ###========
  sub_df_list <- list()
  counter <- 0
  for (i in 1:(length(color_by_levels) - 1)) {
    Gr1 <- color_by_levels[i]
    for (j in (i + 1):length(color_by_levels)) {
      Gr2 <- color_by_levels[j]
      counter <- counter + 1
      curr_comparison <- paste0(Gr1, " vs ", Gr2)
      sub_df <- pcoa_df %>%
        filter(!!color_by %in% c(Gr1, Gr2)) %>%
        droplevels()
      sub_df_list[[counter]] <- sub_df %>%
        mutate(PairwiseBetaComparison = curr_comparison) %>%
        mutate_if(is.factor, as.character)
```

```r
    }
  }

  all_df <- bind_rows(sub_df_list) %>%
    mutate(PairwiseBetaComparison = factor(PairwiseBetaComparison,
                                           levels = unique(.$PairwiseBetaComparison)))

  g <- all_df %>%
    ggplot(aes(Axis.1, Axis.2)) +
    geom_point(aes(color = !!color_by)) +
    theme_bw() +
    xlab(paste0("PCoA axis 1 (", pcoa_pct[1], "%)")) +
    ylab(paste0("PCoA axis 2 (", pcoa_pct[2], "%)")) +
    theme(aspect.ratio = 1) +
    theme(legend.position = "bottom") +
    facet_wrap(~PairwiseBetaComparison)

  return(list(df = all_df, plot = g))
}
```
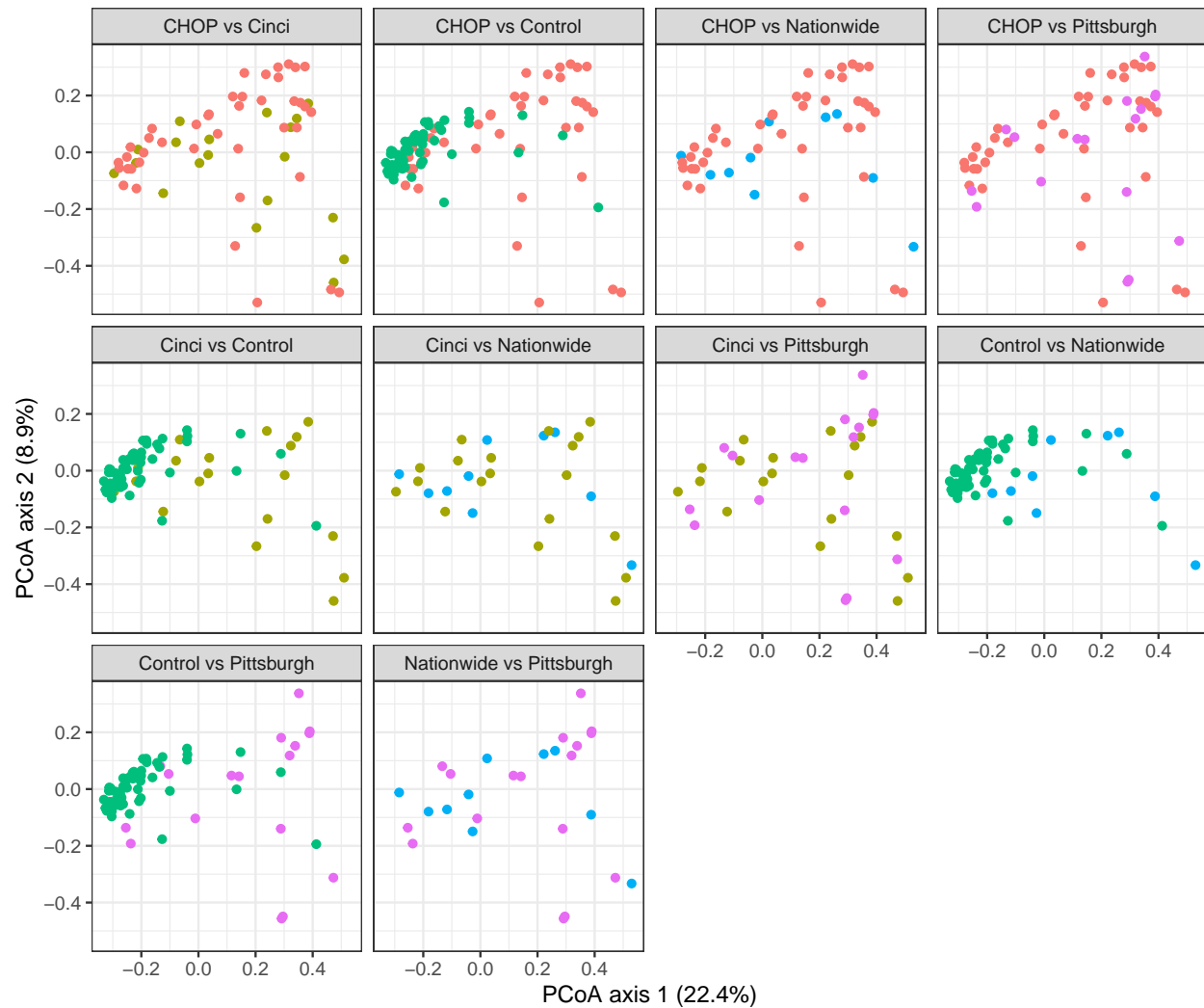
```r
plot <- pairwise_beta_plot(s, dist_beta, Hospital)$plot
df <- pairwise_beta_plot(s, dist_beta, Hospital)$df
```

```r
print(plot)
```

Hospital ● CHOP ● Cinci ● Control ● Nationwide ● Pittsburgh

```r
# Test result to be added
permanova_result_list <- list()
for (i in 1:length(unique(df$PairwiseBetaComparison))) {
  Comp <- unique(df$PairwiseBetaComparison)[[i]]
  sub_df <- df %>%
    filter(PairwiseBetaComparison %in% Comp) %>%
    droplevels()
  sub_dist <- dist_subset(dist_beta, sub_df$SampleID)
  if (all(attributes(sub_dist)$Label == sub_df$SampleID)) {
    set.seed(2020)
    permanova_out <- adonis(sub_dist ~ Hospital, data = sub_df, permutations = 999)$aov.tab
    permanova_result_list[[i]] <- data.frame(PairwiseBetaComparison = as.character(Comp),
                                             nominal_p_value = permanova_out[1, "Pr(>F)"],
                                             stringsAsFactors = FALSE)
  }
}
all_permanova_result <- bind_rows(permanova_result_list) %>%
  mutate(FDR = p.adjust(nominal_p_value, method = "BH")) %>%
```

```
  mutate(FDR = round(FDR, 3))

plot_annotated <- plot +
  geom_text(data = all_permanova_result,
            aes(x = -Inf, y = -Inf, label = paste0("p=", FDR)),
            hjust = -0.1, vjust = -0.6)
print(plot_annotated)
```