Beta-diversity

Sterling Wright

2023-11-01

Load the necessary libraries

```r
library(ggpubr)
```

```
## Loading required package: ggplot2
```

```r
library(Tjazi)
library(scater)
```

```
## Loading required package: SingleCellExperiment

## Loading required package: SummarizedExperiment

## Loading required package: MatrixGenerics

## Loading required package: matrixStats

##
## Attaching package: 'MatrixGenerics'

## The following objects are masked from 'package:matrixStats':
##
##     colAlls, colAnyNAs, colAnys, colAvgsPerRowSet, colCollapse,
##     colCounts, colCummaxs, colCummins, colCumprods, colCumsums,
##     colDiffs, colIQRDiffs, colIQRs, colLogSumExps, colMadDiffs,
##     colMads, colMaxs, colMeans2, colMedians, colMins, colOrderStats,
##     colProds, colQuantiles, colRanges, colRanks, colSdDiffs, colSds,
##     colSums2, colTabulates, colVarDiffs, colVars, colWeightedMads,
##     colWeightedMeans, colWeightedMedians, colWeightedSds,
##     colWeightedVars, rowAlls, rowAnyNAs, rowAnys, rowAvgsPerColSet,
##     rowCollapse, rowCounts, rowCummaxs, rowCummins, rowCumprods,
##     rowCumsums, rowDiffs, rowIQRDiffs, rowIQRs, rowLogSumExps,
##     rowMadDiffs, rowMads, rowMaxs, rowMeans2, rowMedians, rowMins,
##     rowOrderStats, rowProds, rowQuantiles, rowRanges, rowRanks,
##     rowSdDiffs, rowSds, rowSums2, rowTabulates, rowVarDiffs, rowVars,
##     rowWeightedMads, rowWeightedMeans, rowWeightedMedians,
##     rowWeightedSds, rowWeightedVars

## Loading required package: GenomicRanges

## Loading required package: stats4

## Loading required package: BiocGenerics

##
## Attaching package: 'BiocGenerics'

## The following objects are masked from 'package:stats':
##
```

```
##      IQR, mad, sd, var, xtabs

## The following objects are masked from 'package:base':
##
##      anyDuplicated, aperm, append, as.data.frame, basename, cbind,
##      colnames, dirname, do.call, duplicated, eval, evalq, Filter, Find,
##      get, grep, grepl, intersect, is.unsorted, lapply, Map, mapply,
##      match, mget, order, paste, pmax, pmax.int, pmin, pmin.int,
##      Position, rank, rbind, Reduce, rownames, sapply, setdiff, sort,
##      table, tapply, union, unique, unsplit, which.max, which.min

## Loading required package: S4Vectors

##
## Attaching package: 'S4Vectors'

## The following object is masked from 'package:utils':
##
##      findMatches

## The following objects are masked from 'package:base':
##
##      expand.grid, I, unname

## Loading required package: IRanges

## Loading required package: GenomeInfoDb

## Loading required package: Biobase

## Welcome to Bioconductor
##
##      Vignettes contain introductory material; view with
##      'browseVignettes()'. To cite Bioconductor, see
##      'citation("Biobase")', and for packages 'citation("pkgname")'.

##
## Attaching package: 'Biobase'

## The following object is masked from 'package:MatrixGenerics':
##
##      rowMedians

## The following objects are masked from 'package:matrixStats':
##
##      anyMissing, rowMedians

## Loading required package: scuttle
```

```r
pseudocount <- 1

#### Niger PCoA analysis ####

Gobero_species <- read.csv("ExportedTables-Gobero-species-samples.qza/Gobero-species.csv",
    row.names = 1, header = TRUE, check.names = FALSE)
metadata2 <- read.table("Library_METADATA.txt", header = TRUE,
    check.names = TRUE)

Gobero_species_pseudo <- Gobero_species + pseudocount

Gobero_sp_clr <- clr_lite(counts = Gobero_species_pseudo, method = "const",
```

```
        samples_are = "cols")

df <- calculateMDS(Gobero_sp_clr, method = "euclidean")

euclidean_pcoa_df <- data.frame(pcoa1 = df[, 1], pcoa2 = df[,
    2])

eucl_Lib_pcoa_df <- cbind(euclidean_pcoa_df, metadata2$LibraryMethod,
    metadata2$ExtractionType)
```

Warning in data.frame(..., check.names = FALSE): row names were found from a
short variable and have been discarded

```
plot1 <- ggplot(data = eucl_Lib_pcoa_df, aes(x = pcoa1, y = pcoa2,
    color = eucl_Lib_pcoa_df$`metadata2$LibraryMethod`, shape = eucl_Lib_pcoa_df$`metadata2$ExtractionT
    geom_point() + labs(x = "Coordinate 1", y = "Coordinate 2",
    title = "PCoA for Niger Samples") + theme(title = element_text(size = 12),
    panel.grid = element_blank()) + scale_color_manual(name = "Library Method",
    values = c(DoubleStranded = "red", SingleStranded = "blue"),
    labels = c("Double Stranded", "Single Stranded")) + scale_shape_manual(name = "Extraction Method",
    values = c(PB_Dabney = 16, QG_Rohland = 17), labels = c("PB method",
        "QG method"))

#### Hungarian PCoA analysis ####

Hungarian_species <- read.csv("Hungary-species-samples.qza.csv",
    row.names = 1, header = TRUE, check.names = FALSE)

Hungarian_species_pseudo <- Hungarian_species + pseudocount

Hungarian_sp_clr <- clr_lite(counts = Hungarian_species_pseudo,
    method = "const", samples_are = "cols")

df2 <- calculateMDS(Hungarian_sp_clr, method = "euclidean")

euclidean_pcoa_df2 <- data.frame(pcoa1 = df2[, 1], pcoa2 = df2[,
    2])

eucl_Lib_pcoa_df2 <- cbind(euclidean_pcoa_df2, metadata2$LibraryMethod,
    metadata2$ExtractionType)
```

Warning in data.frame(..., check.names = FALSE): row names were found from a
short variable and have been discarded

```
plot2 <- ggplot(data = eucl_Lib_pcoa_df2, aes(x = pcoa1, y = pcoa2,
    color = `metadata2$LibraryMethod`, shape = `metadata2$ExtractionType`)) +
    geom_point() + labs(x = "Coordinate 1", y = "Coordinate 2",
    title = "PCoA for Hungarian Samples") + theme(title = element_text(size = 12),
    panel.grid = element_blank()) + scale_color_manual(name = "Library Method",
    values = c(DoubleStranded = "red", SingleStranded = "blue"),
    labels = c("Double Stranded", "Single Stranded")) + scale_shape_manual(name = "Extraction Method",
    values = c(PB_Dabney = 16, QG_Rohland = 17), labels = c("PB method",
        "QG method"))
```

Plot the results

```
ggarrange(plot1, plot2, labels = c("A", "B"), common.legend = TRUE,
    legend = "bottom", ncol = 2)
```

```
## Warning: Use of `` eucl_Lib_pcoa_df$`metadata2$LibraryMethod` `` is discouraged.
## i Use `metadata2$LibraryMethod` instead.

## Warning: Use of `` eucl_Lib_pcoa_df$`metadata2$ExtractionType` `` is discouraged.
## i Use `metadata2$ExtractionType` instead.

## Warning: Use of `` eucl_Lib_pcoa_df$`metadata2$LibraryMethod` `` is discouraged.
## i Use `metadata2$LibraryMethod` instead.

## Warning: Use of `` eucl_Lib_pcoa_df$`metadata2$ExtractionType` `` is discouraged.
## i Use `metadata2$ExtractionType` instead.
```
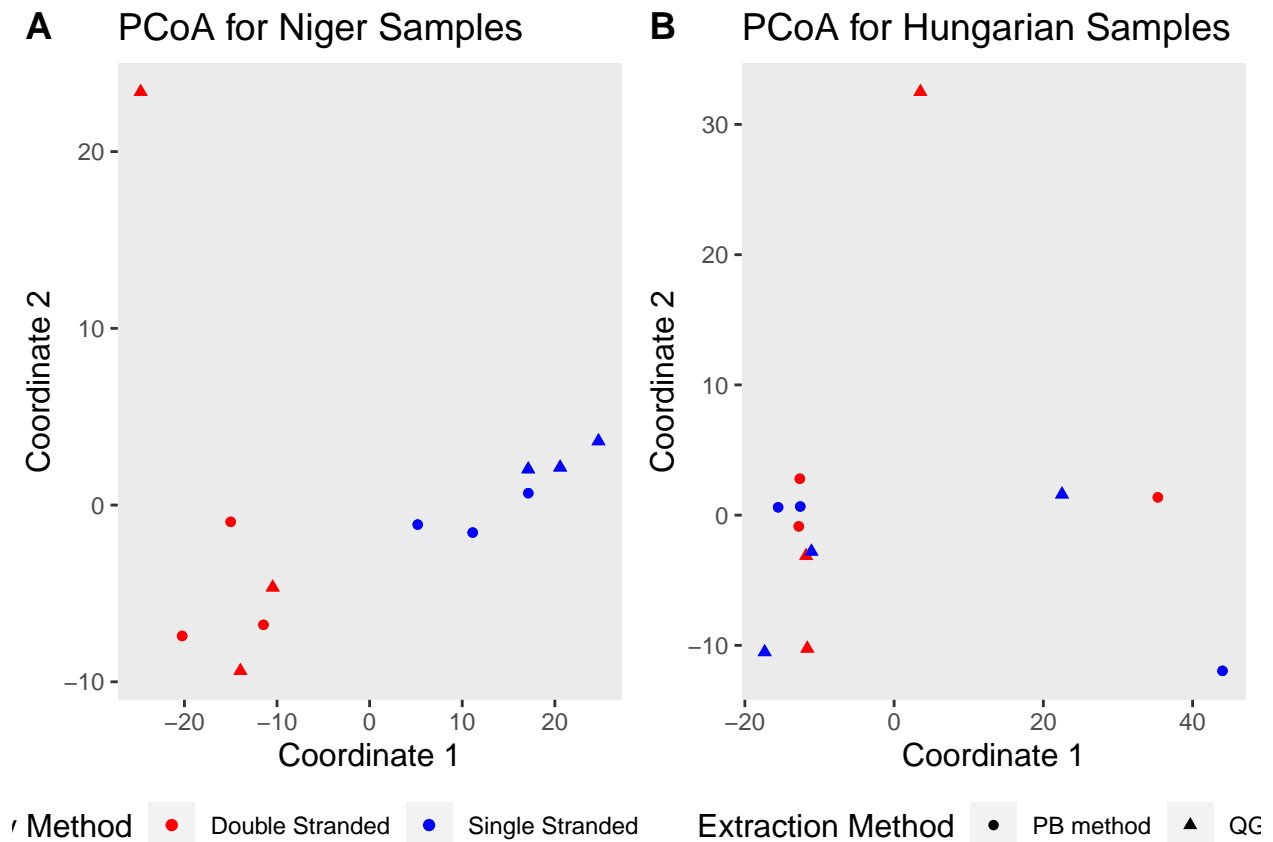

Figure 1: PCoA