Package 'ANCOMBC'

October 14, 2021

Type Package

Title Analysis of compositions of microbiomes with bias correction

Version 1.2.2

Description ANCOMBC is a package for normalizing the microbial observed abundance data due to unequal sampling fractions across samples, and identifying taxa (e.g. phyla, families, genera, species, etc.) that are differentially abundant with respect to the covariate of interest (e.g. study groups) between two or more groups of multiple samples.

Date 2021-08-13

License Artistic-2.0

Imports stats, MASS, nloptr, Rdpack, phyloseq, microbiome

Suggests knitr, tidyverse, testthat, DT, magrittr, qwraps2 (>= 0.5.0), rmarkdown

biocViews DifferentialExpression, Microbiome, Normalization, Sequencing, Software

BugReports https://github.com/FrederickHuangLin/ANCOMBC/issues

URL https://github.com/FrederickHuangLin/ANCOMBC

VignetteBuilder knitr

RdMacros Rdpack

Encoding UTF-8

RoxygenNote 7.1.1

git_url https://git.bioconductor.org/packages/ANCOMBC

git_branch RELEASE_3_13

git_last_commit fa2dd53

git_last_commit_date 2021-08-13

Date/Publication 2021-10-14

Author Huang Lin [cre, aut] (<https://orcid.org/0000-0002-4892-7871>), Shyamal Das Peddada [aut] (<https://orcid.org/0000-0002-5014-6513>)

Maintainer Huang Lin <HUL40@pitt.edu>

R topics documented:

---

ancombc *Differential abundance (DA) analysis for microbial absolute abundance data.*

Description

Determine taxa whose absolute abundances, per unit volume, of the ecosystem (e.g. gut) are significantly different with changes in the covariate of interest (e.g. the group effect). The current version of ancombc function implements Analysis of Compositions of Microbiomes with Bias Correction (ANCOM-BC) in cross-sectional data while allowing the adjustment of covariates.

Usage

```
ancombc(
  phyloseq,
  formula,
  p_adj_method = "holm",
  zero_cut = 0.9,
  lib_cut = 0,
  group = NULL,
  struc_zero = FALSE,
  neg_lb = FALSE,
  tol = 1e-05,
  max_iter = 100,
  conserve = FALSE,
  alpha = 0.05,
  global = FALSE
)
```

Arguments

phyloseq	a phyloseq-class object, which consists of a feature table (microbial observed abundance table), a sample metadata, a taxonomy table (optional), and a phylogenetic tree (optional). The row names of the metadata must match the sample names of the feature table, and the row names of the taxonomy table must match the taxon (feature) names of the feature table. See phyloseq for more details.
formula	the character string expresses how the microbial absolute abundances for each taxon depend on the variables in metadata.
p_adj_method	method to adjust p-values by. Default is "holm". Options include "holm", "hochberg", "hommel", "bonferroni", "BH", "BY", "fdr", "none". See p.adjust for more details.

zero_cut	a numerical fraction between 0 and 1. Taxa with proportion of zeroes greater than zero_cut will be excluded in the analysis. Default is 0.90.
lib_cut	a numerical threshold for filtering samples based on library sizes. Samples with library sizes less than lib_cut will be excluded in the analysis. Default is 0, i.e. do not filter any sample.
group	the name of the group variable in metadata. Specifying group is required for detecting structural zeros and performing global test.
struc_zero	whether to detect structural zeros. Default is FALSE.
neg_lb	whether to classify a taxon as a structural zero in the corresponding study group using its asymptotic lower bound. Default is FALSE.
tol	the iteration convergence tolerance for the E-M algorithm. Default is 1e-05.
max_iter	the maximum number of iterations for the E-M algorithm. Default is 100.
conserve	whether to use a conservative variance estimate of the test statistic. It is recommended if the sample size is small and/or the number of differentially abundant taxa is believed to be large. Default is FALSE.
alpha	level of significance. Default is 0.05.
global	whether to perform global test. Default is FALSE.

Details

The definition of structural zero can be found at ANCOM-II. Setting neg_lb = TRUE indicates that you are using both criteria stated in section 3.2 of ANCOM-II to detect structural zeros; otherwise, the algorithm will only use the equation 1 in section 3.2 for declaring structural zeros. Generally, it is recommended to set neg_lb = TRUE when the sample size per group is relatively large (e.g. > 30).

Value

a list with components:

- feature_table, a data.frame of pre-processed (based on zero_cut and lib_cut) microbial observed abundance table.
- zero_ind, a logical matrix with TRUE indicating the taxon is identified as a structural zero for the specified group variable.
- samp_frac, a numeric vector of estimated sampling fractions in log scale (natural log). Note that for each sample, if it contains missing values for any variable specified in the formula, the corresponding sampling fraction estimate for this sample will return NA since the sampling fraction is not estimable with the presence of missing values.
- resid, a matrix of residuals from the ANCOM-BC log-linear (natural log) model. Rows are taxa and columns are samples.
- delta_em, estimated bias terms through E-M algorithm.
- delta_wls, estimated bias terms through weighted least squares (WLS) algorithm.
- res, a list containing ANCOM-BC primary result, which consists of:
  - beta, a data.frame of coefficients obtained from the ANCOM-BC log-linear (natural log) model.

  - se, a data.frame of standard errors (SEs) of beta.
  - W, a data.frame of test statistics. W = beta/se.
  - p_val, a data.frame of p-values. P-values are obtained from two-sided Z-test using the test statistic W.
  - q_val, a data.frame of adjusted p-values. Adjusted p-values are obtained by applying p_adj_method to p_val.
  - diff_abn, a logical data.frame. TRUE if the taxon has q_val less than alpha.
- res_global, a data.frame containing ANCOM-BC global test result for the variable specified in group, each column is:
  - W, test statistics.
  - p_val, p-values, which are obtained from two-sided Chi-square test using W.
  - q_val, adjusted p-values. Adjusted p-values are obtained by applying p_adj_method to p_val.
  - diff_abn, A logical vector. TRUE if the taxon has q_val less than alpha.

Author(s)

Huang Lin

References

Kaul A, Mandal S, Davidov O, Peddada SD (2017). "Analysis of microbiome data in the presence of excess zeros." *Frontiers in microbiology*, **8**, 2114.

Lin H, Peddada SD (2020). "Analysis of compositions of microbiomes with bias correction." *Nature communications*, **11**(1), 1–11.

Examples

```
#================Build a Phyloseq-Class Object from Scratch==================
library(phyloseq)

otu_mat = matrix(sample(1:100, 100, replace = TRUE), nrow = 10, ncol = 10)
rownames(otu_mat) = paste0("taxon", 1:nrow(otu_mat))
colnames(otu_mat) = paste0("sample", 1:ncol(otu_mat))


meta = data.frame(group = sample(LETTERS[1:4], size = 10, replace = TRUE),
                  row.names = paste0("sample", 1:ncol(otu_mat)),
                  stringsAsFactors = FALSE)

tax_mat = matrix(sample(letters, 70, replace = TRUE),
                 nrow = nrow(otu_mat), ncol = 7)
rownames(tax_mat) = rownames(otu_mat)
colnames(tax_mat) = c("Kingdom", "Phylum", "Class", "Order",
                      "Family", "Genus", "Species")

OTU = otu_table(otu_mat, taxa_are_rows = TRUE)
META = sample_data(meta)
TAX = tax_table(tax_mat)
```

```
physeq = phyloseq(OTU, META, TAX)

#=======================Run ANCOMBC Using a Real Data=======================

library(phyloseq)
library(microbiome)
library(tidyverse)
data(GlobalPatterns)

# Aggregate to phylum level
phylum_data = aggregate_taxa(GlobalPatterns, "Phylum")
# The taxonomy table
tax_mat = as(tax_table(phylum_data), "matrix")

# Run ancombc function
out = ancombc(phyloseq = phylum_data, formula = "SampleType",
              p_adj_method = "holm", zero_cut = 0.90, lib_cut = 1000,
              group = "SampleType", struc_zero = TRUE, neg_lb = FALSE,
              tol = 1e-5, max_iter = 100, conserve = TRUE,
              alpha = 0.05, global = TRUE)

res = out$res
res_global = out$res_global
```

Index