BITS :: Call for Abstracts 2022 - Oral communication

Туре	Oral communication
Session	Metagenomics
Title	benchdamic: benchmarking of differential abundance methods for microbiome data
All Authors	Calgaro M(1), Romualdi C(2), Risso D(3), Vitulo N(1)
Affiliation	
(1) Department of Biotechnology, University of Verona, Verona (2) Department of Biology, University of Dedevo	

Department of Biology, University of Padova, Padova

(3) Department of Statistical Sciences, University of Padova, Padova

Motivation

Recently, an increasing amount of methodological approaches have been proposed to tackle the complexity of metagenomics and microbiome data. In this scenario, reproducibility and replicability have become two critical issues, and the development of computational frameworks for the comparative evaluations of such methods is of utmost importance. Here, we present benchdamic, a Bioconductor package to benchmark methods for the identification of differentially abundant taxa.

Methods

The structure of benchdamic can be summarized into 4 main parts (Figure 1). Each section is developed to answer specific questions when comparing samples from two experimental groups, namely: i) the ability for a given statistical distribution to successfully fit the input data, with particular focus on sparsity and their count nature; ii) the ability of the DA methods to control the type I error; iii) the concordance among methods; and iv) the accuracy of the findings based on a priori biological knowledge. benchdamic builds on existing R/Bioconductor infrastructure packages: the primary input of benchdamic's main functions is a phyloseq object; however, for some functions a count table can also be used. Ready-to-use normalization and DA methods included in benchdamic are based on the edgeR, DESeq2, limma-voom, metagenomeSeq, ALDEx2, corncob, MAST, Seurat, and zinbwave packages.

Results

We envision two main uses of our package: (i) for practitioners interested in performing DA analysis on a new dataset, benchdamic can be used to identify the best DA methods among those already in the literature; (ii) for method developers interested in benchmarking their new approach, benchdamic can be used as an impartial tool to evaluate the relative merits of the new method compared to what is already available. The future of benchdamic is oriented to the addition of new aspects of analysis, e.g., new normalization methods and new DA approaches, to make it an always up-to-date tool including the most recent methods.

benchdamic is available as an open-source package through the Bioconductor project. The package includes a vignette with a detailed tutorial. The future of benchdamic is oriented to the addition of new aspects of analysis, e.g., new normalization methods and new DA approaches, to make it an always up-to-date tool including the most recent methods.

Info

filename

benchdamic abstract.png

Figure

benchdamic: benchmarking of differential abundance methods on microbiome data

Issue: Many approaches have been proposed for Differential Abundance (DA) analysis in metagenomics, it is widely recognised that the perfect method does not exist

Solution:

A careful exploratory data analysis is necessary to methodological address choices. We benchmarked methods from bulk RNAseq, metagenomics, and single-cell RNA-seq metagenomics data.

Application:

benchdamic is a new R/ Bioconductor package that provides a computational framework to guide researchers in the selection of the method that best fits their data.











IMPLEMENTATION

Goodness of fit

Question: Which are the parametric distributions that are able to fit both the proportion of zeros and the counts in your data? Input: phyloseq object or count matrix Working function: FitModels Output: plotRMSE, plotMD **Type I Error Control**

Question: Which are the DA methods that are able to control the number of false positives in your data? Input: phyloseq object Working function: createMocks, runMocks, createTIEC

Output: plotFPR, plotKS, plotQQ Concordance

Question: If multiple DA methods are run on the same data, would they be concordant?

Input: phyloseq object

Working function: createSplits, runSplits, createConcordance Output: plotConcordance

Enrichment

Question: If some prior knowledge about your experiment is available, would the findings be coherent with that knowledge?

Input: phyloseq object Working function: runDA, createEnrichment, createPositives Output: plotEnrichment, plotContingency, plotMutualFindings, plotPositives

Availability	https://bioconductor.org/packages/benchdamic/
Corresponding Author	
Name, Surname	Matteo, Calgaro
Email	matteo.calgaro_01@univr.it
Submitted on	11.05.2022

Società Italiana di Bioinformatica C.F. / P.IVA 97319460586 E-mail bits@bioinformatics.it Sede legale Viale G. Mazzini, 114/B - 00195 Roma Website bioinformatics.it

message generated by sciencedev.com for https://bioinformatics.it 17:33:19 11.05.2022