Phylogenetics

Dynamic assessment of microbial ecology (DAME): a web app for interactive analysis and visualization of microbial sequencing data

Brian D. Piccolo^{1,2,*}, Umesh D. Wankhade^{1,2}, Sree V. Chintapalli^{1,2},
Sudeepa Bhattacharyya^{1,2}, Luo Chunqiao² and Kartik Shankar^{1,2}

¹Arkansas Children's Nutrition Center, Little Rock, AR 72202, USA and ²Department of Pediatrics, University of Arkansas for Medical Sciences, Little Rock, AR 72202, USA

*To whom correspondence should be addressed.

Associate Editor: Janet Kelso

Received on August 2, 2017; revised on October 6, 2017; editorial decision on October 23, 2017; accepted on October 25, 2017

Abstract

Summary: Dynamic assessment of microbial ecology (DAME) is a Shiny-based web application for interactive analysis and visualization of microbial sequencing data. DAME provides researchers not familiar with R programming the ability to access the most current R functions utilized for ecology and gene sequencing data analyses. Currently, DAME supports group comparisons of several ecological estimates of α -diversity and β -diversity, along with differential abundance analysis of individual taxa. Using the Shiny framework, the user has complete control of all aspects of the data analysis, including sample/experimental group selection and filtering, estimate selection, statistical methods and visualization parameters. Furthermore, graphical and tabular outputs are supported by R packages using D3.js and are fully interactive.

Availability and implementation: DAME was implemented in R but can be modified by Hypertext Markup Language (HTML), Cascading Style Sheets (CSS), and JavaScript. It is freely available on the web at <https://acnc-shinyapps.shinyapps.io/DAME/>. Local installation and source code are available through Github (<https://github.com/bdpiccolo/ACNC-DAME>). Any system with R can launch DAME locally provided the shiny package is installed.

Contact: bdpiccolo@uams.edu

1 Introduction

There is enormous interest in understanding the role of commensal bacteria in health and disease (Fung *et al.*, 2017; Lynch and Pedersen, 2016; Zmora *et al.*, 2016). Thus, a variety of research scientists continue to acquire microbial sequencing data, with data analysis and interpretation fast becoming rate-limiting steps. Microbial ecology data also pose unique challenges not generally observed in more traditional sequencing outputs (e.g. taxonomic hierarchy and overdispersion). Furthermore, complex study paradigms from clinical or basic research studies necessitates a flexible analysis tool that can provide dynamic and real-time adjustment of analyses and visualizations.

We have developed a new web-based Shiny application, dynamic assessment of microbial ecology (DAME), to address the needs of

health research scientists and others incorporating microbial sequencing data into their research. Although there have been several graphical user interfaced (GUI) and web-based applications that analyse microbial sequencing data (Beck *et al.*, 2015; Caporaso *et al.*, 2010; McMurdie and Holmes, 2015; Zakrzewski *et al.*, 2016), DAME distinguishes itself in several ways: (i) it provides the ability to select/deselect experimental groups and individual samples in real time, (ii) it incorporates both statistical analysis and visualization for common ecological measures and (iii) provides fully interactive tables and graphical outputs. Altogether, DAME is designed to provide basic and clinical scientists who are not familiar with R programming, the ability to utilize R functionalities and conduct exploratory analyses of microbial sequencing data within a flexible and interactive GUI.

2 Methods

2.1 Getting started

DAME is configured primarily for 16S rRNA gene amplicon sequencing data and requires two files: a Biological Observation Matrix (BIOM) file with sequencing reads combined with taxonomy details (accepts both JavaScript Object Notation (JSON) and Hierarchical Data Format 5 (HDF5) formats) and a comma-separated values (CSV) file containing experimental metadata with at least one column with sample ID labels matching those used in the BIOM file. We recommend using the map file used during the QIIME workflow, but it is not necessary. As of the current release, only classification data can be utilized in the metadata CSV.

2.2 Importing and selecting data

Upon uploading both BIOM and CSV files, DAME will scan all columns in the CSV file, identify the column in the CSV file that matches sample IDs in the BIOM file, sync both files together and then render interactive widgets that control which data are to be analysed. Users have complete control over the provided metadata and can deselect classifiers and/or samples that may not be pertinent to their analysis (Fig. 1). This feature is not isolated to the initial data import and users can re-select or deselect other classifiers based on downstream analyses.

Descriptive statistics of sample prevalence, sequence reads and Operational Taxonomic Units (OTU) of imported data and finalized data used in the downstream analysis are provided in an interactive table using the DT package (Xie, 2016). These tables default to

phylum level statistics but can show other taxonomic levels. They can also be downloaded in Excel, PDF or CSV formats. In addition, the imported and finalized sequence data or percent relative abundance data for a specific taxonomic level can be downloaded in CSV format.

2.3 Alpha diversity

DAME allows users to select one or more taxonomic levels, alpha diversity parameters and experimental groups simultaneously. Currently, DAME supports several options for statistical comparisons of selected groups. Graphical plots and statistical results are managed with a control button and will render results based on the selected taxonomic levels. DAME utilizes the rbokeh package (Hafen and Continuum Analytics, 2016) to render interactive graphics detailing either the distribution of the data or to check the assumptions of parametric statistical tests. Box plots are the default plot option, but bar plots with standard error bars are available for parametric based tests. Additionally, Q-Q plots and Fitted versus Residual plots can be rendered when a one-way analysis of variance (ANOVA) or multi-factor ANOVA are selected. All plots are fully interactive with zooming, movable plotting canvas and hovering text options. Mean/median and dispersion values are provided with the results of the statistical test in an interactive table alongside of the barplot using the DT package (Xie, 2016). Selection of multiple taxonomic levels will render graphical and tabular outputs for all levels individually. Interactive features of graphical plots and tables within a taxonomic level are independent, i.e. interaction in a single plot will not affect adjacent plots or a similar plot in a different taxonomic level. Raw α -diversity values (per sample) are also available to download in CSV format.

2.4 Beta-diversity

Similar to the Alpha-Diversity tab, widgets for taxonomic levels and experimental group(s) selection are controlled by a beta-diversity control button. DAME performs permutational multivariate analysis of variance (PERMANOVA) on beta-diversity measures using the `adonis2()` function from the `vegan` package. Multiple dissimilarity, distance or phylogenetic-based indices are provided to measure beta-diversity along with several ordination options. Currently, DAME defaults the graphical output to a 2-dimensional scatterplot of the first two components of the selected ordination, using the `scatterD3` package (Barnier *et al.*, 2016). Users can manipulate the graphical window (zoom, scroll, reset, resize points, toggle/resize labels, etc.). Selection of beta diversity parameters and ordination parameters update the scatterplots in real time. A DT table of the PERMANOVA results is also provided adjacent to the scatterplot. Again, selecting multiple taxonomic levels will render plots and tables for all taxonomic levels and these outputs are independent of each other. Downloadable options are available for graphical and tabular outputs.

2.5 Differential-abundance

DAME utilizes negative binomial regression for the differential abundance of individual taxa (McMurdie and Holmes, 2014) using the R package DESeq2 (Love *et al.*, 2014). Users select taxonomic level(s) and the specific test (Wald test for pairwise group comparisons or likelihood ratio test for overall experimental group comparisons). Tabular output of DESeq2 results and boxplots are controlled in real time by widgets controlling the experimental group and pairwise comparisons (if Wald test was selected) (Fig. 2). The DESeq2 outputs and boxplots are fully controlled by the user and can be customized to show results for other experimental groups, pairwise

Fig. 1. Example of DAME Import tab. After importing data, DAME will provide descriptive statistics regarding sample prevalence, sequence reads and Operational Taxonomic Units (OTU) and also provide filtering options for imported metadata. Deselecting or selecting classifiers and/or samples can be updated at any time by pressing the *Finalize Import and Filters* control button. Similar descriptive statistics are provided for the finalized data

Fig. 2. Example of Differential-Abundance tab. Group selections within the Alpha-Diversity, Beta-Diversity and Differential-Abundance tabs are reflective of group level selections made in Import Data tab. For example, the Select Group(s): widget displayed in this figure will only display groups with >1 levels retained. Alpha-Diversity and Beta-Diversity tabs follow a similar linear workflow: tabular and graphical outputs are displayed for each selected taxonomic level selected

comparisons, and specific taxa. DAME uses an interactive DT table to print the output from the `DESeq2::results()` function, including taxa labels, log₂-fold changes and adjusted *P*-values (Fig. 2). All taxa can be selected for boxplot visualization by either total reads or % abundance (Fig. 1). The Differential-Abundance follows a similar architecture to the other tabs (e.g. rendering of multiple taxonomic levels) and all results and interactive plots are downloadable.

3 Conclusion

DAME was created to cater to basic and clinical scientists in order to provide an engrossing and dynamic experience for rapid microbiome data analysis. The current release provides the fundamental analyses required for publications and is rooted in the most up-to-date statistical methods. Future releases are under development and will incorporate continuous data for covariate adjustment in NB models in addition to correlation statistics.

Acknowledgements

We would like to thank the authors of the following R packages in no particular order of importance: shiny (Chang et al., 2017), shinyjs (Attali, 2016), DT (Xie, 2016), V8 (Ooms, 2017), biomformat (McMurdie and Paulson, 2016), ape (Paradis et al., 2004), pbapply (Solymos and Zawadzki, 2017), tibble (Müller and Wickham, 2017), reshape2 (Wickham, 2007), phyloseq (McMurdie and Holmes, 2013), dplyr (Wickham et al., 2017), vegan (Oksanen et al., 2016), DESeq2 (Love et al., 2014), scatterD3 (Barnier et al.,

2016), RColorBrewer (Neuwirth, 2014) and rbokeh (Hafen and Continuum Analytics, Inc., 2016). We are also indebted to the R user community.

Funding

This work has been supported by United States Department of Agriculture-Agricultural Research Service Project 6026-51000-010-05S and National Center for Research Resources and the National Center for Advancing Translational Sciences of the National Institute of Health Grant UL1TR001449.

Conflict of Interest: none declared.

References

- Attali,D. (2016) shinyjs: Easily Improve the User Experience of Your Shiny Apps in Seconds. R package version 0.9. <https://CRAN.R-project.org/package=shinyjs>.
- Barnier,J. et al. (2016) scatterD3: D3 JavaScript Scatterplot from R. R package version 0.8.1. <https://CRAN.R-project.org/package=scatterD3>.
- Beck,D. et al. (2015) Seed: a user-friendly tool for exploring and visualizing microbial community data. *Bioinformatics*, 31, 602–603.
- Caporaso,J.G. et al. (2010) QIIME allows analysis of high-throughput community sequencing data. *Nat. Methods*, 7, 335–336.
- Chang,W. et al. (2017) shiny: Web Application Framework for R. R package version 1.0.3. <https://CRAN.R-project.org/package=shiny>.
- Fung,T.C. et al. (2017) Interactions between the microbiota, immune and nervous systems in health and disease. *Nat Neurosci*, 20, 145–155.
- Hafen,R. and Continuum Analytics, Inc. (2016) rbokeh: R Interface for Bokeh. R packaged version 0.5.0. <https://CRAN.R-project.org/package=rbokeh>.
- Love,M.I. et al. (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.*, 15, 550.
- Lynch,S.V. and Pedersen,O. (2016) The human intestinal microbiome in health and disease. *N. Engl. J. Med.*, 375, 2369–2379.
- McMurdie,P.J. and Holmes,S. (2013) Phyloseq: An R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS One*, 8, e61217.
- McMurdie,P.J. and Holmes,S. (2015) Shiny-phyloseq: web application for interactive microbiome analysis with provenance tracking. *Bioinformatics*, 31, 282–283.
- McMurdie,P.J. and Holmes,S. (2014) Waste not, want not: why rarefying microbiome data is inadmissible. *PLoS Comput. Biol.*, 10, e1003531.
- McMurdie,P.J. and Paulson,J.N. (2016) biomformat: An Interface Package for the BIOM File Format. <https://github.com/joe711/biomformat/>, <http://biom-format.org/>.
- Müller,K. and Wickham,H. (2017) tibble: Simple Data Frames. R package version 1.3.3. <https://CRAN.R-project.org/package=tibble>.
- Neuwirth,E. (2014) RColorBrewer: ColorBrewer Palettes. R package version 1.1-2. <https://CRAN.R-project.org/package=RColorBrewer>.
- Oksanen,J. et al. (2016) vegan: Community Ecology Package. R package version 2.4-1. <https://CRAN.R-project.org/package=vegan>.
- Ooms,J. (2017) V8: Embedded JavaScript Engine for R. R package version 1.5. <https://CRAN.R-project.org/package=V8>.
- Paradis,E. et al. (2004) APE: analyses of phylogenetics and evolution in R language. *Bioinformatics*, 20, 289–290.
- Solymos,P. and Zawadzki,Z. (2017) pbapply: Adding Progress Bar to “*apply” Functions. R package version 1.3-2. <https://CRAN.R-project.org/package=pbapply>.
- Wickham,H. et al. (2017) dplyr: A Grammar of Data Manipulation. R package version 0.7.3. <https://CRAN.R-project.org/package=dplyr>.
- Wickham,H. (2007) Reshaping data with the reshape package. *J. Stat. Softw.*, 21, 1–20.
- Xie,Y. (2016) DT: A Wrapper of the JavaScript Library “DataTables”. R package version 0.2. <https://CRAN.R-project.org/package=DT>.
- Zakrzewski,M. et al. (2016) Calypso: a user-friendly web-server for mining and visualizing microbiome? Environment interactions. *Bioinformatics*, 8, btw725.
- Zmora,N. et al. (2016) Taking it personally: personalized utilization of the human microbiome in health and disease. *Cell Host Microbe*, 19, 12–20.