Bioinformatics
doi.10.1093/bioinformatics/xxxxxx
Advance Access Publication Date:
Manuscript Category

OXFORD

Original Paper

Template for preparing a submission to Bioinformatics using RMarkdown

Quang P. Nguyen [1,2], Anne G. Hoen [1,2], H. Robert Frost [2],

[1]Department of Epidemiology, Geisel School of Medicine at Dartmouth College
[2]Department of Biomedical Data Science, Geisel School of Medicine at Dartmouth College

To whom correspondence should be addressed. E-mail: bob@email.com

Associate Editor: XXX

Received on XXX; revised on XXX; accepted on XXX

Abstract

Motivation: This section should specifically state the scientific question within the context of the field of study.
Results: This section should summarize the scientific advance or novel results of the study, and its impact on computational biology.
Availability: This section should state software availability if the paper focuses mainly on software development or on the implementation of an algorithm. Examples are: 'Freely available on the web at XXX.' Website implemented in Perl, MySQL and Apache, with all major browsers supported'; or 'Source code and binaries freely available for download at URL, implemented in C++ and supported on linux and MS Windows'. The complete address (URL) should be given. If the manuscript describes new software tools or the implementation of novel algorithms the software must be freely available to non-commercial users. Authors must also ensure that the software is available for a full TWO YEARS following publication. The editors of Bioinformatics encourage authors to make their source code available and, if possible, to provide access through an open source license.
Contact:bob@email.com
Supplementary information: Supplementary data are available at Bioinformatics Online.

1 Introduction

2 Materials and Methods

3 Methods

3.1 Competitive Isometric Log-ratio (cILR)

The cILR method generates sample-specific enrichment scores for microbial sets using the isometric log-ratio transformation (Egozcue *et al.*, 2003). The cILR method takes two inputs:

- $\mathbf{X}$: n by p matrix of positive counts for p taxa and n samples measured through either targeted sequencing p matrix of positive counts for p taxa and n samples measured through either targeted sequencing (such as 16S rRNA) or whole genome shotgun sequencing. Usually $\mathbf{X}$ is generated from standard sequence processing pipelines such as DADA2 (Callahan *et al.*, 2016) and MetaPhlAn2 (Truong *et al.*, 2015).
- $\mathbf{A}$: p by m indicator matrix annotation the membership of each taxa p to m sets of interest. These sets can be Linnean taxonomic classifications annotated using databases such as SILVA (Quast *et al.*, 2013) or those based on more functionally driven categories such as the functional tropism of microbes ($a_{i,j} = 1$ indicates that microbe i belongs to set j

The cILR method generates one output:

- $\mathbf{E}$: n by m matrix indicating the enrichment score of m pre-defined sets identified in
- $\mathbf{E}$: n by m matrix indicating the enrichment score of m pre-defined sets identified in $\mathbf{A}$ across n samples.

The procedure is as follows:

1. **Compute the cILR statistic**: Let $\mathbf{M}$ be a n by m matrix of cILR scores. Let $\mathbf{M}_{i,k}$ be cILR scores for set k of sample i:

$$\mathbf{M}_{i,k} = \sqrt{\frac{\sum_k A_{ik}(p - \sum_k A_{ik})}{p}} \ln\left(\frac{g(\mathbf{X}_{i,j}|\mathbf{A}_{j,k} = 1)}{g(\mathbf{X}_{i,j}|\mathbf{A}_{j,k} \neq 1))}\right)$$

   where $g()$ is the geometric mean. This represents the ratio of the geometric mean of the relative abundance of taxa assigned to set k and remainder taxa.
2. **Compute the cILR statistic on permuted $\mathbf{X}$**: We seek to evaluate the empirical null distribution of the cILR statistic under H_o that relative abundances in $\mathbf{X}$ of members of set k are not enriched compared to those not in set k. Let $\mathbf{X}_p$ be the column permuted relative abundance matrix, and $\mathbf{M}_p$ be the corresponding cILR scores generated from $\mathbf{X}_p$.
3. **Fit Gaussian mixture distribution for each column of $\mathbf{M}_p$**
4. **Calculate finalized cILR scores as CDF values of the fitted mixture distribution**

Details for Method 1. Lorem ipsum ad nauseum. Introduce your topic. Lorem ipsum ad nauseum. Introduce your topic. Lorem ipsum ad nauseum. Introduce your topic. Lorem ipsum ad nauseum. Introduce your topic. Lorem ipsum ad nauseum.

Details for Method 1. Lorem ipsum ad nauseum. Introduce your topic. Lorem ipsum ad nauseum. Introduce your topic. Lorem ipsum ad nauseum. Introduce your topic. Lorem ipsum ad nauseum. Introduce your topic. Lorem ipsum ad nauseum.

Details for Method 1. Lorem ipsum ad nauseum. Introduce your topic. Lorem ipsum ad nauseum. Introduce your topic. Lorem ipsum ad nauseum. Introduce your topic. Lorem ipsum ad nauseum. Introduce your topic. Lorem ipsum ad nauseum.

Details for Method 1. Lorem ipsum ad nauseum. Introduce your topic. Lorem ipsum ad nauseum. Introduce your topic. Lorem ipsum ad nauseum. Introduce your topic. Lorem ipsum ad nauseum. Introduce your topic. Lorem ipsum ad nauseum.

3.2 Method 2

Details for Method 2. Lorem ipsum ad nauseum. Introduce your topic. Lorem ipsum ad nauseum. Introduce your topic. Lorem ipsum ad nauseum. Introduce your topic. Lorem ipsum ad nauseum. Introduce your topic. Lorem ipsum ad nauseum.

Details for Method 2. Lorem ipsum ad nauseum. Introduce your topic. Lorem ipsum ad nauseum. Introduce your topic. Lorem ipsum ad nauseum. Introduce your topic. Lorem ipsum ad nauseum. Introduce your topic. Lorem ipsum ad nauseum.

Details for Method 2. Lorem ipsum ad nauseum. Introduce your topic. Lorem ipsum ad nauseum. Introduce your topic. Lorem ipsum ad nauseum. Introduce your topic. Lorem ipsum ad nauseum. Introduce your topic. Lorem ipsum ad nauseum.

Details for Method 2. Lorem ipsum ad nauseum. Introduce your topic. Lorem ipsum ad nauseum. Introduce your topic. Lorem ipsum ad nauseum. Introduce your topic. Lorem ipsum ad nauseum. Introduce your topic. Lorem ipsum ad nauseum.

4 Discussion

Discussion of results. Lorem ipsum ad nauseum. Introduce your topic. Lorem ipsum ad nauseum. Introduce your topic. Lorem ipsum ad nauseum. Introduce your topic. Lorem ipsum ad nauseum. Introduce your topic. Lorem ipsum ad nauseum.

Discussion of results. Lorem ipsum ad nauseum. Introduce your topic. Lorem ipsum ad nauseum. Introduce your topic. Lorem ipsum ad nauseum. Introduce your topic. Lorem ipsum ad nauseum. Introduce your topic. Lorem ipsum ad nauseum.

Discussion of results. Lorem ipsum ad nauseum. Introduce your topic. Lorem ipsum ad nauseum. Introduce your topic. Lorem ipsum ad nauseum. Introduce your topic. Lorem ipsum ad nauseum. Introduce your topic. Lorem ipsum ad nauseum.

Discussion of results. Lorem ipsum ad nauseum. Introduce your topic. Lorem ipsum ad nauseum. Introduce your topic. Lorem ipsum ad nauseum. Introduce your topic. Lorem ipsum ad nauseum. Introduce your topic. Lorem ipsum ad nauseum.

5 Conclusion

Anything else? Lorem ipsum ad nauseum. Introduce your topic. Lorem ipsum ad nauseum. Introduce your topic. Lorem ipsum ad nauseum. Introduce your topic. Lorem ipsum ad nauseum. Introduce your topic. Lorem ipsum ad nauseum.

Anything else? Lorem ipsum ad nauseum. Introduce your topic. Lorem ipsum ad nauseum. Introduce your topic. Lorem ipsum ad nauseum. Introduce your topic. Lorem ipsum ad nauseum. Introduce your topic. Lorem ipsum ad nauseum.

Anything else? Lorem ipsum ad nauseum. Introduce your topic. Lorem ipsum ad nauseum. Introduce your topic. Lorem ipsum ad nauseum. Introduce your topic. Lorem ipsum ad nauseum. Introduce your topic. Lorem ipsum ad nauseum.

Anything else? Lorem ipsum ad nauseum. Introduce your topic. Lorem ipsum ad nauseum. Introduce your topic. Lorem ipsum ad nauseum. Introduce your topic. Lorem ipsum ad nauseum. Introduce your topic. Lorem ipsum ad nauseum.

Acknowledgements

These should be included at the end of the text and not in footnotes. Please ensure you acknowledge all sources of funding, see funding section below.

Details of all funding sources for the work in question should be given in a separate section entitled 'Funding'. This should appear before the 'Acknowledgements' section.

Funding

The following rules should be followed:

- The sentence should begin: 'This work was supported by …' -
- The full official funding agency name should be given, i.e. 'National Institutes of Health', not 'NIH' (full RIN-approved list of UK funding agencies)
- Grant numbers should be given in brackets as follows: '[grant number xxxx]'
- Multiple grant numbers should be separated by a comma as follows: '[grant numbers xxxx, yyyy]'
- Agencies should be separated by a semi-colon (plus 'and' before the last funding agency)
- Where individuals need to be specified for certain sources of funding the following text should be added after the relevant agency or grant number 'to [author initials]'.

An example is given here: 'This work was supported by the National Institutes of Health [AA123456 to C.S., BB765432 to M.H.]; and the Alcohol & Education Research Council [hfygr667789].'

Oxford Journals will deposit all NIH-funded articles in PubMed Central. See Depositing articles in repositories – information for authors

for details. Authors must ensure that manuscripts are clearly indicated as NIH-funded using the guidelines above.

References

Callahan, B. J., McMurdie, P. J., Rosen, M. J., Han, A. W., Johnson, A. J. A., and Holmes, S. P. (2016). DADA2: High-resolution sample inference from Illumina amplicon data. *Nature Methods*, **13**(7), 581–583.

Egozcue, J. J., Pawlowsky-Glahn, V., Mateu-Figueras, G., and Barcelo-Vidal, C. (2003). Isometric Logratio Transformations for Compositional Data Analysis. *Mathematical Geology*, page 22.

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J., and Glöckner, F. O. (2013). The SILVA ribosomal RNA gene database project: Improved data processing and web-based tools. *Nucleic Acids Research*, **41**(D1), D590–D596.

Truong, D. T., Franzosa, E. A., Tickle, T. L., Scholz, M., Weingart, G., Pasolli, E., Tett, A., Huttenhower, C., and Segata, N. (2015). MetaPhlAn2 for enhanced metagenomic taxonomic profiling. *Nature Methods*, **12**(10), 902–903.