Rarefy your data

Patrick D. Schloss[†]

† To whom corresponsdence should be addressed:

pschloss@umich.edu

Department of Microbiology & Immunology

University of Michigan

Ann Arbor, MI 48109

10 Abstract

11 Importance

Introduction

- What is rarefaction? History, reason for rarefaction

    - Repeated down sampling of datasets to a common number of observations to calculate the average value to ascertain the expected value of a metric for the metric under study; typically richness

    - Control for unneven sampling effort

    - Methods vary in their sensitivity to uneven sampling

- Reasons behind "rarefaction is inadmissable"

    - Weird simulation

- Alternative approaches and claims

    - sampling invariance

- Goal of this study

Results

Datasets

- Represented diverse communities

- High quality V4 sequence data that consisted of paired 250 nt reads

- Wide distribution in the number of reads within each dataset (Figure S1, Table 1)

Null models

- Null models were samplings, without replacement, of all samples from a study that have been pooled together and sampled to the same depth of sampling

- Created 100 random sets of samples per dataset

- Would expect a zero correlation between alpha diversity metric and number of sequences in the community

- Rarefaction was the only approach that resulted in a near-zero correlation (Figure 1)

36 • Would expect a zero correlation between beta diversity metrics and the difference in the number of

37 sequences in the samples going into the distance calculation

38 • Rarefaction was the only approach that resulted in a near-zero correlation (Figure 2)

39 **Result when sampling effort is confounded with treatment group**

40 • Used Null model

41 • Created two treatment groups for each study by dividing samples based on whether they were above

42 or below the median number of sequences per sample

43 • Used Wilcoxon test to detect differences between treatment groups for alpha diversity, would expect

44 ~5% of random sets to be significant

45 • Rarefaction the only approach that reliably resulted in the expected Type I error rate

46 • Used AMOVA test to detect differences between treatment groups for beta diversity, would expect

47 ~5% of random sets to be significant

48 • Rarefaction the only approach that reliably resulted in the expected Type I error rate

49 **Result when known effect size imposed between two groups of samples**

50 • For each dataset I divided samples into two groups. The probability distribution for the first group was

51 determined by the null model and the second was determined by xxxxxxxxxxx

52 • These parameters were selected because they produced results were some of the simulations for

53 the rarefied data did not result in a significant test

54 • Fraction of tests that were significant would be a measure of statistical power

55 • Used Wilcoxon test to detect differences between treatment groups for alpha diversity

56 • All tests resulted in reduced power to detect differences relative to the power of using rarefaction

57 • Used AMOVA test to detect differences between treatment groups for beta diversity

58 • All tests resulted in reduced power to detect differences relative to the power of using rarefaction

4

- This simulation would work well for measures based on the structure of the community, but lacked resolution for those based on the richness or membership of commmunities. For an additional simulation, the first treatment group followed a null distribution for the second, XX% of OTUs were removed from the distribution

- Tests for richness and membership (i.e. Jaccard) resutled in reduced power relative to rarefaction

Effect of sampling depth on metrics

- Simulated rarefying null model datasets to retain varying percentage of samples
- High sampling depth resulted in less variation between samples for alpha and beta diversity

Re-analysis of previously published datasets

- Human study: Did not alter effect size or significance of alpha or beta diversity, but did result in reduced effect sizes for measures of richness and non-parametric estimators of richness; breakaway detected a difference

Discussion

- Rarefy your data
- Problems with recommended methods...
    - Many recommended methods are borrowed from gene expression analysis
    - Meaning of zeroes in data - structural vs. below limit of detection
- Factors that determine what number of sequences to rarefy to
- Need better methods of pooling libraries that result in more even distribution of sequences across samples
- Rarefy your data

Materials and Methods

Data availability.

Reproducible data analysis.

83 **Acknowledgements.**

84 References

Figure 1.

86 **Figure S1.**