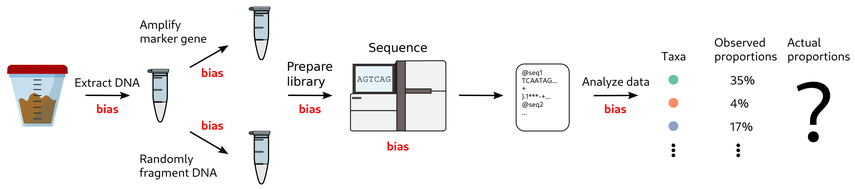
Countering experimental bias in microbiome measurements

Michael R. McLaren[1], David S. Clausen[2], Angie Mordant[1], Manuel Kleiner[1], Amy D. Willis[2], Benjamin J. Callahan[1]
[1]North Carolina State University [2]University of Washington ✉ m.mclaren42@gmail.com, @mikemc423

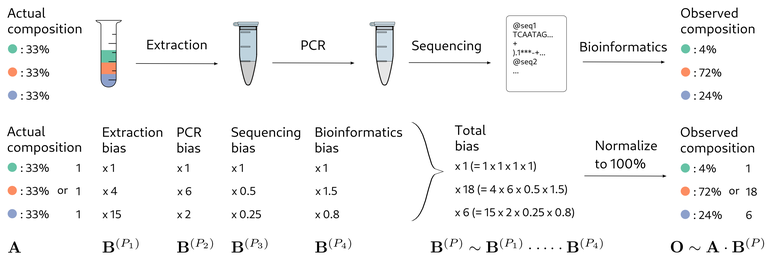
Summary

Each step in a marker-gene or shotgun metagenomics experiment has a **protocol-specific bias** that favors some taxa over others. Bias makes sequencing-based microbiome measurements **inaccurate** and **incomparable** across different experiments.



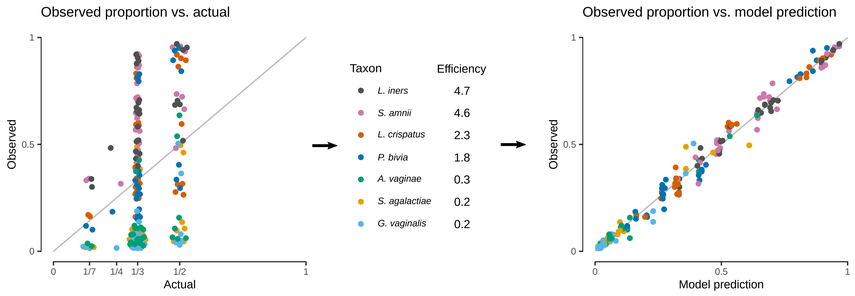
We derived a simple mathematical model of bias and showed that it accurately explained systematic error in measurements of defined bacterial communities. This model suggests new ways to measure the bias associated with specific taxa and protocols, correct bias using control measurements (calibration), and determine the effect of bias in common microbiome analyses.

Model: Bias multiplies across steps to form the total protocol's bias



Each bias vector denotes the **relative efficiency** with which different taxa are measured by one or more steps. Bias is independent of sample composition, but creates composition-dependent error in taxon proportions.

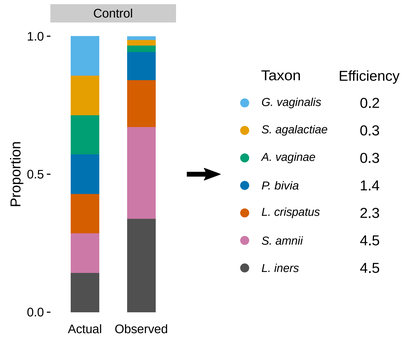
This model can explain large and ubiquitous systematic errors in measurements of defined ("mock") microbial communities



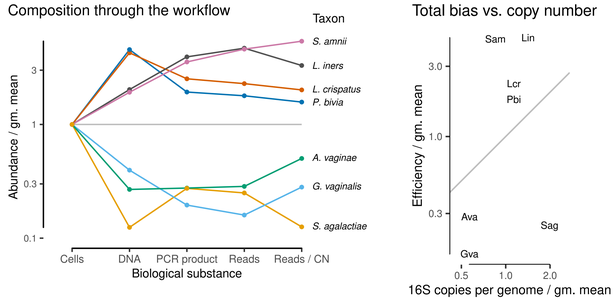
Brooks et al. (2015) measured 58 unique mixtures of 7 bacterial species with 16S amplicon sequencing. Each sample is an even mixture of cells (e.g., a 1:1:1 mixture of *L. iners*, *S. amnii*, and *P. bivia*). Our model explains 99% of the mean-squared error in the observed proportions with just 6 free parameters: the relative efficiencies of the 7 species. Analysis and figures are from McLaren et al. (2019); see McLaren et al. (2019) for additional validation in a shotgun dataset and Leopold and Busby (2020) for validation in mixtures of fungal DNA.

Estimating bias under our model provides a better way to evaluate and improve protocols

The bias (relative efficiencies) associated with K taxa can be estimated from a single control sample containing all K taxa



Example: In the Brooks et al. (2015) experiment, bias is dominated by DNA extraction and is not significantly reduced by 16S copy-number correction



Unlike other measures, the relative efficiencies are biophysically meaningful and are predicted by our model to still apply in samples with compositions that differ from the control samples. Suitable experimental designs allow estimating the bias associated with particular protocol steps. Data from Brooks et al. (2015); rightmost two figures from McLaren et al. (2019).

Challenges and open questions

Real microbiome data contains contamination from the environment, reagents, and other samples, which can make estimating bias difficult.

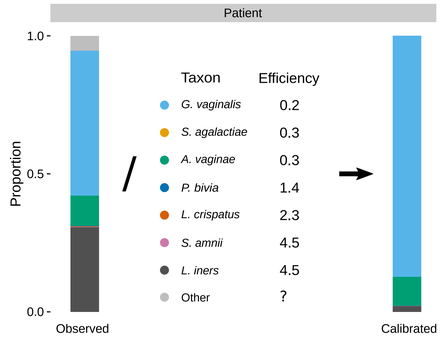
It is currently unknown how well bias estimates extrapolate to other labs, sample types, and taxa.

- How consistent is bias when different labs use the same nominal protocol?
- How consistent is bias across different sample types (e.g., different matrix chemistries)?
- Do taxa that are closely related or have similar phenotypes have similar efficiencies?

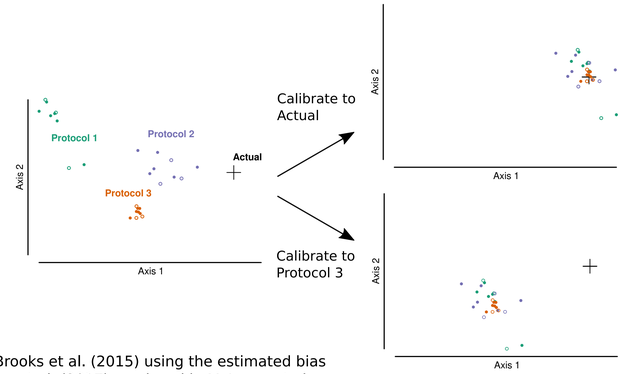
Our model must be tested over a wider range of taxa, sample types, and protocols to determine the experimental conditions for which bias is consistent across samples.

Bias estimated from controls can be used to calibrate unknown samples

Estimated relative efficiencies can be used to correct the bias associated with those taxa in the samples of interest



If only the bias between different protocols can be estimated, then calibration to a reference protocol can make their measurements quantitatively comparable



Left: Calibration of a patient vaginal microbiome data reported in Brooks et al. (2015) using the estimated bias from above. Right: Calibration of mock community data from Costea et al. (2017), analyzed in McLaren et al. (2019). For an example of calibration in a microbial ecology experiment, see Leopold and Busby (2020).

Challenges and open questions

Controls should be carefully constructed and tested to ensure that their bias reflects that of the target samples (see above).

Current methods only allow calibrating the relative abundances of taxa that are in the controls, effectively limiting calibration to synthetic-community experiments or environments dominated by a small number of culturable taxa.

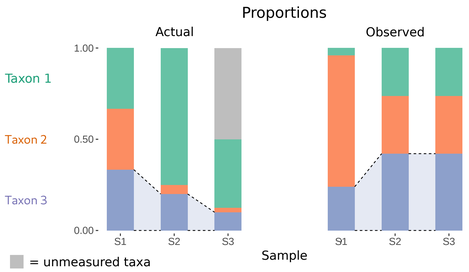
Using natural samples as controls can alleviate these concerns, but only provides comparable (not accurate) measurements.

New statistical methods may greatly improve the reliability and utility of calibration, especially by

- using phylogeny or phenotype to extrapolate bias estimates to taxa not in the controls
- accounting for other sources of measurement error in bias and calibration estimates.

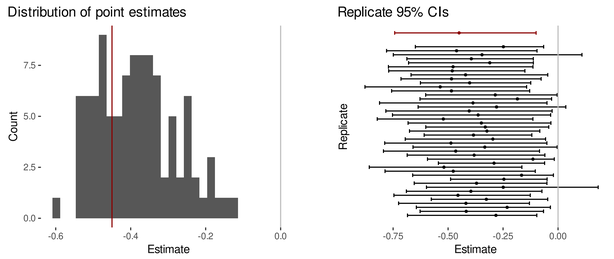
Our model can be used to rigorously evaluate the effect of bias on microbiome analyses

Bias does not necessarily "cancel out" when analyzing differences between samples measured by the same protocol



Hypothetical measurements of three samples with consistent bias. Taxa 1, 2, and 3 have relative efficiencies of 1:18:6, and other taxa (in gray) are not measured at all (a relative efficiency of 0). Bias leads to incorrect inferences about how taxon proportions change across samples and which samples have greater alpha diversity.

The sensitivity of a given result to bias can be analyzed by seeing how the result varies across hypothesized or randomly generated bias vectors



Bias sensitivity analysis for the estimated difference in alpha diversity (Shannon index) between the vaginal microbiomes of women with term versus preterm pregnancies (data from Callahan et al., 2017). Simulated "Actual" datasets were created by using randomly generated bias vectors to calibrate the observed data. The diversity analysis was then performed on these "Actual" datasets. The estimate remains negative across datasets, suggesting that the observed decrease (shown in red) is unlikely to be due to bias with these parameters.

Challenges and open questions

How often spurious results occur in practice is unknown and depends on the type of analysis, the distribution of efficiencies across taxa, and how true compositions vary across samples.

Our model predicts that analyses based on ratios of taxa, as used in Compositional Data Analysis, are invariant to bias, but commonly used zero-replacement and taxonomic aggregation strategies can violate this invariance.

References and attributions

Brooks JP, et al. 2015. BMC Microbiol 15:66. doi:10.1186/s12866-015-0351-6

Callahan BJ, et al. 2017. PNAS 114:9966–9971. doi:10.1073/pnas.1705899114

Costea PI, et al. 2017. Nat Biotechnol 35:1069–1076. doi:10.1038/nbt.3960

Leopold DR, Busby PE. 2020. Curr Biol 30:3260-3266.e5. doi:10.1016/j.cub.2020.06.011

McLaren MR, Willis AD, Callahan BJ. 2019. Elife 8:46923. doi:10.7554/eLife.46923

The icon of a fecal sample in Summary is from j4p4n on openclipart.org