*PLOS Biology* Editorial Team

*PLOS*

1265 Battery Street, Suite 200

San Francisco, CA, USA 94111

Phone: +1 415 624 1200 | Fax: +1 415 546 4090

August 22th, 2022

Dear PLOS Biology Editorial Team,

We are pleased to submit the manuscript “Distinguishing Signal from Noise: Understanding the Patterns of Non-detections to Inform Accurate Quantitative Metabarcoding” for consideration as a research article at *PLOS Biology*. Key to the ability to derive quantitative estimates from metabarcoding studies is understanding the causes of non-detections, whereby a technical PCR replicate fails to detect a species observed in other replicates. Here, we demonstrate that the rate of non-detections among technical replicates is a function of both the template DNA concentration and species-specific amplification efficiency, allowing the ability to distinguish quantitative biological signals from noise in metabarcoding studies.

Recent work correcting for amplification biases in metabarcoding data have yielded quantitative estimates of template DNA concentrations. However, current models do not account for the high degree of variability observed among technical replicates, particularly non-detections. Thus, we develop a qualitative understanding of how non-detections arise in metabarcoding data using both simulated and empirical data. We demonstrate that the rate of non-detections among technical replicates is a function of both the template DNA concentration and species-specific amplification efficiency, explaining a substantial portion of observed noise in metabarcoding datasets. Consequently, we conclude that models for metabarcoding data need to incorporate both a deterministic component that describes the biases introduced by PCR amplification and a stochastic component representing the sampling of DNA sequences during PCR and sequencing steps.

Importantly, we highlight value of incorporating additional independent estimates (e.g. derived separately from metabarcoding) of amplification efficiencies and DNA concentration along with amplicon sequence data, providing for the application of routine statistical approaches and straightforward interpretation of observed metabarcoding sequence data. These results support the recent quantitative metabarcoding frameworks of Shelton et al. 2022, McLaren et al. 2019, and Silverman et al. 2021 which provide for the establishment of reliable estimates of abundance from amplicon sequence data. The ability to provide quantitative estimates from metabarcoding data will be critical for extending the application of this method to health and ecological questions.

The authors have no conflicts of interest to report. All data and code generated for this study will be made available in NCBI, Dryad, or GitHub upon acceptance. Thank you for considering this manuscript for publication in *PLOS Biology*.

Sincerely,

Zachary Gold, on behalf of all authors

Northwest Fisheries Science Center

National Marine Fisheries Service

2725 Montlake Blvd E,

Seattle, WA 98112

zachary.gold@noaa.gov

310-795-0020