

Post-bioinformatic methods to identify and reduce the prevalence of artefacts in metabarcoding data

Lorna Drake¹, Jordan Cuff¹, Rebecca Young¹, Angela Marchbank¹, Elizabeth Chadwick¹, and William Symondson¹

¹Cardiff University College of Biomedical and Life Sciences

April 13, 2021

Abstract

Metabarcoding provides a powerful tool for investigating biodiversity and trophic interactions, but the high sensitivity of this methodology makes it vulnerable to errors, resulting in artefacts in the final data. Metabarcoding studies thus often utilise minimum sequence copy thresholds (MSCTs) to remove artefacts that remain in datasets; however, there is no consensus on best practice for the use of MSCTs. To mitigate erroneous reporting of results and inconsistencies, this study discusses and provides guidance for best-practice filtering of metabarcoding data for the ascertainment of conservative and accurate data. The most common MSCTs identified in the literature were applied to example datasets of Eurasian otter (*Lutra lutra*) and cereal crop spider (Araneae: Linyphiidae and Lycosidae) diets. Changes in both the method and threshold value considerably affected the resultant data. Of the MSCTs tested, it was concluded that the optimal method for the examples given combined a sample-based threshold with removal of maximum taxon contamination, providing stringent filtering of artefacts whilst retaining target data. Choice of threshold value differed between datasets due to variation in artefact abundance and sequencing depth, thus studies should employ controls (mock communities, negative controls with no DNA and unused MID-tag combinations) to select threshold values appropriate for each individual study.

Hosted file

Dietary metabarcoding artefacts MS.pdf available at <https://authorea.com/users/407506/articles/517802-post-bioinformatic-methods-to-identify-and-reduce-the-prevalence-of-artefacts-in-metabarcoding-data>