Improving Metabarcoding Taxonomic Assignment: A Case Study of Fishes in a Large Marine Ecosystem

Zachary Gold[1], Emily Curd[1], Kelly Goodwin[2], Emma Choi[3], Benjamin Frable[3], Andrew Thompson[4], Harold J Walker Jr[3], Ronald Burton[3], Dovi Kacev[3], and Paul Barber[1]

[1]UCLA
[2]Atlantic Oceanographic and Meteorological Laboratory
[3]University of California San Diego Scripps Institution of Oceanography
[4]Southwest Fisheries Science Center

February 23, 2021

Abstract

DNA metabarcoding is an important tool for molecular ecology. However, its effectiveness hinges on the quality of reference sequence databases and classification parameters employed. Here we evaluate the performance of MiFish 12S taxonomic assignments using a case study of California Current Large Marine Ecosystem fishes to determine best practices for metabarcoding. Specifically, we use a taxonomy cross-validation by identity framework to compare classification performance between a global database comprised of all available sequences and a curated database that only includes sequences of fishes from the California Current Large Marine Ecosystem. We demonstrate that the curated, regional database provides higher assignment accuracy than the comprehensive global database. We also document a tradeoff between accuracy and misclassification across a range of taxonomic cutoff scores, highlighting the importance of parameter selection for taxonomic classification. Furthermore, we compared assignment accuracy with and without the inclusion of additionally generated reference sequences. To this end, we sequenced tissue from 605 species using the MiFish 12S primers, adding 253 species to GenBank's existing 550 California Current Large Marine Ecosystem fish sequences. We then compared species and reads identified from seawater environmental DNA samples using global databases with and without our generated references, and the regional database. The addition of new references allowed for the identification of 16 native taxa and 17.0% of total reads from eDNA samples, including species with vast ecological and economic value. Together these results demonstrate the importance of comprehensive and curated reference databases for effective metabarcoding and the need for locus-specific validation efforts.

Hosted file

Gold-fish_card_manuscript_draft_20210218.docx available at https://authorea.com/users/330161/articles/510400-improving-metabarcoding-taxonomic-assignment-a-case-study-of-fishes-in-a-large-marine-ecosystem