Exploring the data that explores the oceans: working towards robust eDNA workflows for ocean wildlife monitoring

Jessica Pearce$^1$, Philipp Bayer$^1$, Adam Bennett$^1$, Eric Raes$^1$, Marcelle Ayad$^1$, Shannon Corrigan$^1$, Matthew Fraser$^1$, Denise Anderson$^2$, Priscila Goncalves$^1$, Benjamin Callahan$^3$, Michael Bunce$^4$, Stephen Burnell$^1$, and Sebastian Rauschert$^1$

$^1$Minderoo Foundation
$^2$INSiGENe Pty Ltd
$^3$North Carolina State University at Raleigh
$^4$New Zealand Department of Conservation

August 28, 2023

Abstract

In the face of a looming ecological crisis, effective and efficient biomonitoring methods will be critical to the conservation of marine environments. Environmental DNA (eDNA) is now established for detecting species presence within ecosystems and is increasingly being recognised as a tool for marine management. However, concerns surrounding reproducibility and standardisation has led to a lack of confidence and uptake by some end-users as the method transitions from the research domain into biomonitoring practice. Here we incorporate DADA2 into our automatable, containerised, and fully reproducible amplicon (metabarcoding) workflow and highlight how changing analytical steps and parameters may impact inferences from eDNA-derived biodiversity data. By pooling all samples for amplicon sequence variant (ASV) inference, compared to the DADA2 default of independent sample inference, we were able to recover significantly higher ASV counts per sample and directly increased detection of ASVs across replicates and sites. This led to a higher number of species detections and revealed known species as ecological drivers only when ASV inference was pooled across samples. Although, for the purpose of this study, we have focused on the manipulation of a single parameter for a single analytical tool, our findings reiterate the importance of ensuring both are selected in a manner appropriate to the research questions being addressed, and that suitability for comparisons to previously generated datasets is considered. Finally, we provide guidance for robust data processing with the aim to make eDNA more effective, transparent, and useful for management.

Hosted file

ASV inference
DADA2
Demultiplex
Cutadapt
Raw reads
Taxonomic assignment
BLASTN
Read quality control
Seqkit
ASV curation
LULU
Ecology
Taxonomic curation
LCA (eDNAFlow)
Phyloseq
object
Metadata
Raw data curation
Generate ASVs
Taxonomy

A

B

C

D

E

F

Mode
-independent
-pooled
-pseudo

Observed ASVs
Cocos Islands
Rowley Shoals
NW WA

Samples

Samples

Samples

Observed unique species

Mode independent pooled pseudo

Samples

Samples

Samples