A pile of pipelines: an overview of the bioinformatics software for metabarcoding data analyses

Ali Hakimzadeh1, Alejandro Abdala Asbun2, Davide Albanese3, Maria Bernard4, Dominik Buchner5, Benjamin Callahan6, Gregory Caporaso7, Emily Curd8, Christophe DJEMIEL9, Mikael Brandström Durling10, Vasco Elbrecht5, Zachary Gold11, Hyun Gweon12, Mehrdad Hajibabaei13, Falk Hildebrand14, Vladimir Mikryukov1, Eric Normandeau15, Ezgi Ozkurt14, Jonathan M. Palmer16, Géraldine Pascal17, Teresita Porter13, Daniel Straub18, Martti Vasar1, Tomáš Větrovský19, Haris Zafeiropoulos20, and Sten Anslan1

1University of Tartu Institute of Ecology and Earth Sciences
2NIOZ Royal Netherlands Institute for Sea Research
3Fondazione Edmund Mach Istituto Agrario di San Michele all’Adige
4Université Paris-Saclay
5University of Duisburg-Essen
6North Carolina State University College of Veterinary Medicine
7Northern Arizona University
8University of Vermont
9INRA Dijon
10Swedish University of Agricultural Sciences
11University of California San Diego
12Center for Ecology and Hydrology
13University of Guelph
14Quadram Institute Bioscience
15Université Laval `Institut de Biologie Intégrative et des Systèmes
16Northern Research Station, US Forest Service
17GenPhySE
18University of Tübingen
19Institute of Microbiology Czech Academy of Sciences
20KU Leuven Rega Institute for Medical Research

February 11, 2023

Abstract

Environmental DNA (eDNA) metabarcoding has gained growing attention as a strategy for monitoring biodiversity in ecology. However, taxa identifications produced through metabarcoding require sophisticated processing of high-throughput sequencing data from taxonomically informative DNA barcodes. Various sets of universal and taxon-specific primers have been developed, extending the usability of metabarcoding across archaea, bacteria, and eukaryotes. Accordingly, a multitude of metabarcoding data analysis tools and pipelines have also been developed. Often, several developed workflows are designed to process the same amplicon sequencing data, making it somewhat puzzling to choose one amongst the plethora of existing pipelines. However, each pipeline has its own specific philosophy, strengths, and limitations, which should be considered depending on the aims of...
any specific study, as well as the bioinformatics expertise of the user. In this review, we outline the input data requirements, supported operating systems, and particular attributes of thirty-one amplicon processing pipelines with the goal of helping users to select a pipeline for their metabarcoding projects.

Hosted file

Hosted file

Table 1. Pipelines list .xlsx available at https://authorea.com/users/585035/articles/623934-a-pile-of-pipelines-an-overview-of-the-bioinformatics-software-for-metabarcoding-data-analyses