

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

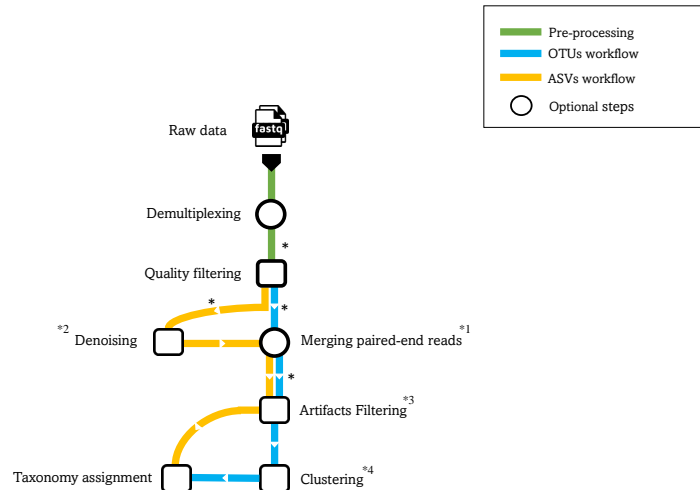
Abstract

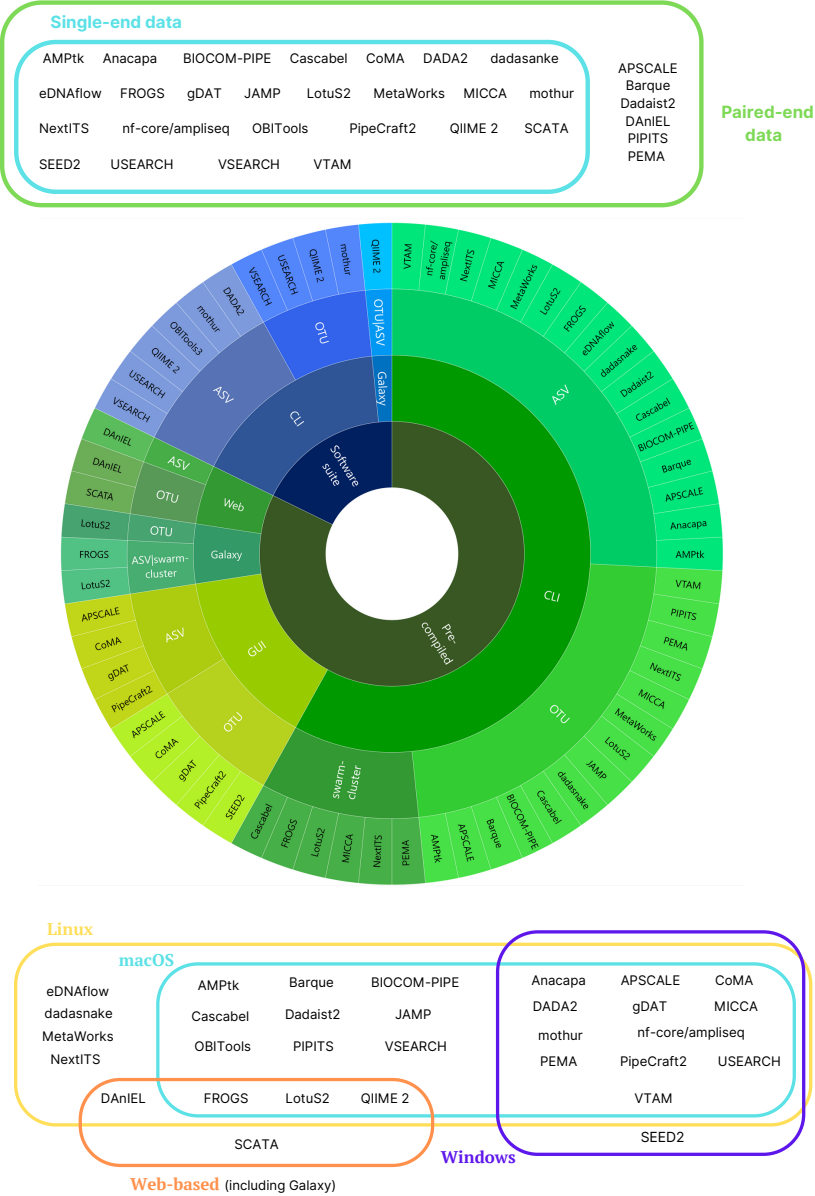
Environmental DNA (eDNA) metabarcoding has gained growing attention as a strategy for monitoring biodiversity in ecology. However, taxa identifications produced through metabarcoding require sophisticated processing of high-throughput sequencing data from taxonomically informative DNA barcodes. Various 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

Metabarcoding pipelines review paper.rtf available at <https://authorea.com/users/585035/articles/623934-a-pile-of-pipelines-an-overview-of-the-bioinformatics-software-for-metabarcoding-data-analyses>





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>