EukRef: Annotating the taxonomy of 18S rRNA genesequences since 2015

Javier del Campo, Marta Noguera and the PR2 team

INSTITUT de BIOLOGIA EVOLUTIVA

CSIC
upf.

UNIVERSITY OF MIAMI
ROSENSTIEL
SCHOOL of MARINE &
ATMOSPHERIC SCIENCE

EUK REF

November 9, 2022
PR2 Zoom meeting

2015   2019   2023
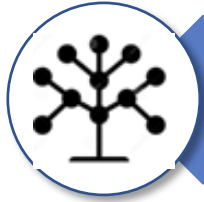
Introduction
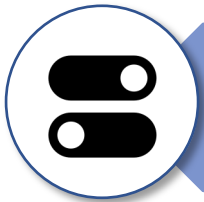
PR2 curation

Search in PR2 database and run the alignment pipeline.

Build phylogenetic tree.

Make several iterative modifications.

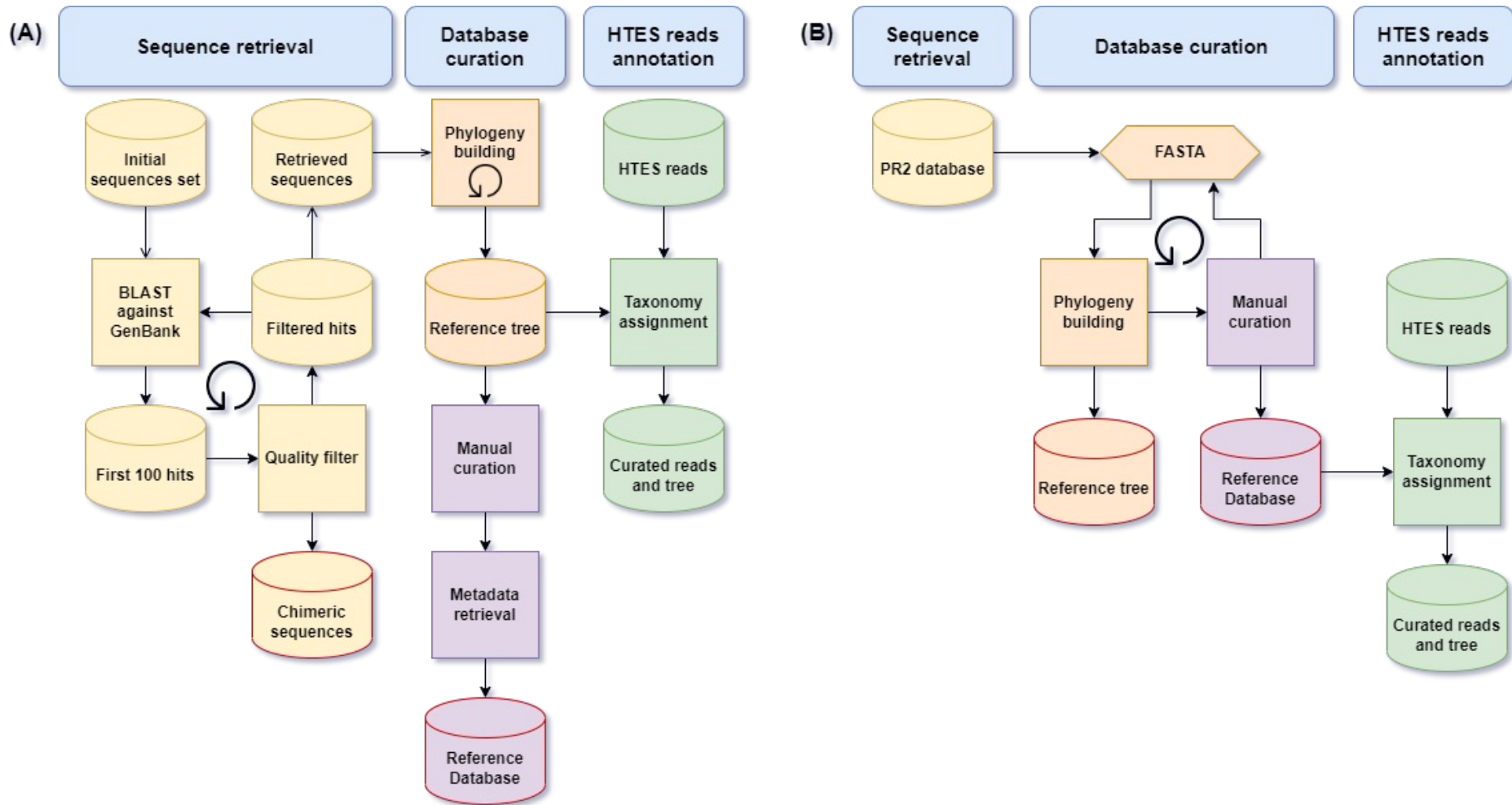Download the reference tree and the reference database.

Democratise the access to phylogenetic based annotation to taxonomists

Transferring the knowledge from taxonomists to curate reference databases

Comunity

Pipeline simplification

Pipeline improvements

Steps	EukRef	PR2 curation
Sequence retrieval process	Iterative sequences retrieval process from GenBank by using BLAST	Downloading the target lineage sequences from the PR2 database
Database curation	Executing lines of code in the terminal	In the application interface by clicking
Hands-on time	From two days to a week	From 30 minutes to 2 hours

Software needed

R packages

The PR2 curation process

Input: Download initial set of sequences from the PR2 database.

Clusterize sequences with VSEARCH.

Alignment of sequences with MAFFT (with outgroups).

Trimming with trimAl.

Phylogenetic inference with RAxML.

Inicial phylogenetic tree.

Manual curation in the PR2 curation interface.

Outputs: Curated dataset

1 2 3 4 5 6 7 8