New REPET improvements: REPET V4.0, making life easier

Johann Confais[1], Mariène Wan[1], Etienne Bardet*[1], and Hadi Quesneville[1]

[1]Unité de Recherche Génomique Info – Institut National de Recherche pour l'Agriculture, l'Alimentation et l'Environnement – INRAE, Centre de recherche de Versailles, bat.18 Route de Saint Cyr 78000 Versailles, France

Abstract

The detection and annotation of transposable elements (TEs) are now considered mandatory to any genome sequencing project.

To this aim, the REPET package integrates bioinformatics pipelines dedicated to detect, annotate and analyse TEs in genomic sequences. The two main pipelines are (i) TEdenovo, that search for interspersed repeats, build consensus sequences and classify them according to TE features and (ii) TEannot, which mines a genome with a library of TE sequences, for instance the one produced by the TEdenovo pipeline, to provide TE annotations.
The REPET package is in continuous improvement. Lately our effort have been focused on REPET's ease of use and installation. The new version 4.0 of REPET is now encapsulated in an snakemake framework and uses sSingularity/Apptainer images to handle it's dependency. This allows for seamless integration into HPC cluster environnements.
It enables also an easier code maintenance and evolution, which permit simpler collaborative developement.

Keywords: pipeline, genome annotation, transposable elements, eukaryote, bioinformatics

*Speaker