

La reproductibilité au service de la Biologie computationnelle

jeudi 23 mai 2019 10:45 (45 minutes)

Les outils de reproductibilité choisis ont tous des équivalents fonctionnels (docker/singularity, snakemake/nextflow, jupyter/Rmarkdown, ...)

Une étude récente publiée dans Nature a montré que près de 70% des expériences en Biologie ne sont pas reproductibles. Il est donc indispensable de mettre en place des bonnes pratiques afin de garantir l'intégrité des données et la reproductibilité des résultats d'analyse. Concernant les données, les principes FAIR-data sont de plus en plus utilisés. Ces mêmes principes peuvent être détournés au service des analyses pour garantir des résultats identiques à partir d'un même jeu de données et au cours du temps. L'objectif de cette présentation est de proposer un panel de fonctionnalités permettant de rendre reproductible une analyse complète de bioinformatique. L'exemple présenté a pour but de sélectionner des gènes qui ne se comportent pas de la même façon entre deux conditions expérimentales. Les fonctionnalités présentées ne sont pas dépendantes de cet exemple. En effet, elles peuvent être appliquées à n'importe quelle autre question biologique. Brièvement, nous récupérons les données depuis les bases de données publiques (ENA/SRA), nous réalisons une analyse reproductible avec un système de workflow (snakemake) dans un environnement virtuel (docker) dont l'ensemble du code, versionné (git), est disponible en open source (Github et dockerhub). La visualisation des résultats est dynamique (shiny app) et un rapport (Rmarkdown) en pdf ou html est disponible. Il regroupe les résultats de l'analyse et détaille l'ensemble des paramètres choisis par l'utilisateur.

Auteur principal: DENECKER, T. (I2BC, Gif-Sur-Yvette)

Co-auteur: TOFFANO-NIOCHE, C. (I2BC, Gif-Sur-Yvette)

Orateur: DENECKER, T. (I2BC, Gif-Sur-Yvette)