

UnCoVar: A Reproducible and Scalable Workflow for Transparent and Robust SARS-CoV-2 Variant Calling and Lineage Assignment

Inhalt

Genome sequencing became an indispensable tool to characterize the ongoing COVID-19 pandemic. We present a benchmark data set of 54 patient samples, sequenced with several different in-vitro approaches on two sequencing platforms, a comprehensive benchmark of existing workflows and introduce UnCoVar, an open-source bioinformatics workflow for analyzing SARS-CoV-2 sequencing data for patient samples and environments. The fully automated workflow assembles SARS-CoV-2 genomes, identifies lineages and offers high resolution variant calling. The workflow also includes multiple analysis approaches suitable for assembled genomes, variant-based consensus genomes and unassembled reads. UnCoVar includes extensive quality control and automated generation of a comprehensive report, that provides valuable insights for data consumers, for both researchers and clinicians. The software provides a configurable, user-friendly, scalable, and reproducible pipeline for SARS-CoV-2 genome sequence data analysis. It is implemented with Snakemake and Python and a containerized version is available. Redundant analysis paths of the workflow ensure robust results, producing submission-ready high-quality SARS-CoV-2 genome sequences, enabling molecular surveillance of the pandemic. The open-source code is available under a BSD 2-clause license at github.com/IKIM-Essen/uncover.

Keywords

SARS-CoV-2, Workflow, Variant Calling, Lineage Assignment, Next Generation Sequencing

Registration-ID code

ZOO23-627

Professional Status of the Speaker

PhD Student

Junior Scientist Status

Yes, I am a Junior Scientist.

Thema Einordnung: Epidemiology and Secondary Data Use

Typ des Beitrags: Poster presentation