

Abstract

VANIR—NextFlow Pipeline for Viral Variant Calling and de Novo Assembly of Nanopore and Illumina Reads for High-Quality dsDNA Viral Genomes [†]

Joan Martí-Carreras and Piet Maes *

Laboratory of Clinical and Epidemiological Virology, Department of Microbiology, Rega Institute, Immunology and Transplantation, KU Leuven, Herestraat 49 Box 1040, BE3000 Leuven, Belgium; joan.marti@kuleuven.be

* Correspondence: piet.maes@kuleuven.be

[†] Presented at Viruses 2020—Novel Concepts in Virology, Barcelona, Spain, 5–7 February 2020.

Published: 3 July 2020

Abstract: Human cytomegalovirus (HCMV), like other herpes and dsDNA viruses, possesses unique properties derived from their genome architecture. The HCMV genome is composed of two unique domains: long (L) and short (S). Each domain contains a central unique region (U; thus, UL and US, respectively) and two repeated regions (thus, TRL/IRL and TRS/IRS). Recombination between repetitive regions is possible, yielding four possible genomic isomers, found in equimolar proportion in any viral infective population. Frequent recombination and an altered selective landscape can give rise to the persistence, if not fixation, of diverse variants in cultured HCMV isolates. This phenomenon has already been discovered in AD169 and Towne strains, characterizing a 10 kbp deletion (Δ UL/b') in commonly used viral strains. Other dsDNA viruses are known for their structural rearrangements and frequent recombination. VANIR (viral variant calling and de novo assembly using nanopore and illumina reads) is a novel analysis pipeline that benefits from both short-read (Illumina) and long-read sequencing technologies (Oxford Nanopore Technologies Ltd.) to assemble high-quality dsDNA viral genomes and detection of variants. Illumina and nanopore sequencing provide complementary information to the assembly and variant discovery. Assembly contiguity, structural variant, and repeat calling are greatly improved by nanopore read-length and base-calling and base confidence by Illumina reduced error rate and increased yield. This specialized bioinformatic analysis pipeline is encoded in the NextFlow pipeline manager and containerized in a Singularity image. This set-up allows for improved traceability, reproducibility, transportability, and speed. Through VANIR, novel point mutations and structural genome rearrangements are called from sequencing data, benefiting diversity research with attenuated lab-strains and wild-type viruses.

Keywords: DNA viruses; diversity; genomics; NGS; pipeline; containerization



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).