Genomics paper in:

COV-Seq, a New Tool for SARS-CoV-2 Genome Analysis and Visualization: Development and Usability Study.

By:

Amira Abd Elkader Abd Allah

3rd grade, Bioinformatics department.

Mansours University

To be presented to:

Dr. Sara El Sayed El Metwally

22/4/2021

Abstract:

-COVID-19 quickly became a global pandemic after its discovery in late 2019. SARS-CoV-2 genomes are being sequenced and shared on public repositories at a rapid rate. To keep up with these changes, scientists must refresh and reclean data sets on a regular basis, which is a time-consuming and labor-intensive operation. Besides this, scientists may find it difficult to analyze SARS-CoV-2 genomes.

-To fix these issues, COV-Seq was created, an interactive web server that allows users to analyze SARS-CoV-2 genomes quickly and easily.

-Python and java script were used to implement COV-Seq. This article includes the URLs for the web server and source code.

Introduction:

To understand SARS-CoV-2 evolution and genomics and to address these challenges, we created the COV-Seq toolkit. A data analysis pipeline that takes FASTA sequences and produces variant callsets in variant call format (VCF) and open reading frame (ORF) predictions is part of COV-Seq. The pipeline detects and annotates low-quality sequences, eliminates redundant sequences, aligns sequences, and filters low-quality sequences We provide a web server [6] to enable the rapid analysis of custom sequences without any programming to quickly analyze custom sequences. An interactive genome visualizer and tabulated displays of genetic variants and ORF projections are included in the web interface. Which are available for download for further review. We also have a command-line interface for high-throughput processing in local environments. We compiled SARS-CoV-2 sequences from the Global Initiative on Sharing Avian Influenza Data (GISAID) [3], National Center for Biotechnology Information (NCBI) [4], European Nucleotide Archive (ENA) [5], and China National GeneBank to make data sharing easier (CNGB).

Related works:

1-VAPiD: a lightweight cross-platform viral annotation pipeline and identification tool to facilitate virus genome submissions to NCBI GenBank

-Viral Annotation Pipeline and iDentification (VAPiD), a lightweight and compact command-line method for viral genome annotation and GenBank deposition. Almost all unsegmented viral genomes can be annotated with VAPiD, Human immunodeficiency virus, human parainfluenza virus, human metapneumovirus, humacoronaviruses, human enteroviruses/rhinoviruses

measles virus, mumps virus, Hepatitis A-E virus, Chikungunya virus, dengue virus, and West Nile virus.

2-VIGOR, an annotation program for small viral genomes

-VIGOR (Viral Genome ORF Reader), a web application platform for influenza virus, rotavirus, rhinovirus, and coronavirus subtype gene prediction. VIGOR can reliably detect genome specific features such as frame changes, overlapping genes, embedded genes, and predict mature peptides within the range of a single polypeptide open reading frame based on sequence similarity searches. The software includes influenza and rotavirus genotyping capabilities.