Salma Adel Fathy

Third year [Medical Informatics Department]

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

Abstract

COVID-19 developed into a global pandemic. Scientists must constantly refresh and update the data sets to keep up with these changes. To address these issues, we created CoV-Seq, an integrated web server that allows for the quick and easy analysis of SARS-CoV-2 genomes. Seq methods are written in Python and JavaScript. We have a web address. CoV-Seq determines gene boundaries and detects genetic variants from a new sequence, which are presented in an interactive genome visualizer and downloadable for further study. For high-throughput processing, a command-line interface is available. In addition, we compiled all SARS-CoV-2 sequences that were publicly accessible . The web server includes an interactive framework for analyzing SARS-CoV-2 unique sequences.

Introduction

SARS-CoV-2, a novel coronavirus, has triggered a viral pneumonia outbreak. SARS-CoV-2 had infected nearly 33 million people worldwide and killed nearly a million people. Scientists sequenced SARS-CoV-2 genomes from various patients to better understand its evolution and genetics. 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 genetic variants while filtering low-quality sequences, removing duplicates, performing sequence alignment, and identifying and filtering low-quality sequences. To fix these issues, we created the CoV-Seq framework. A data processing system that takes FASTA sequences and produces variant callers in variant call format (VCF) and open reading frame (ORF) predictions is part of CoV-Seq. Both of the findings are available for download for further review. We also have a present predominantly for increased processing in settings. We compiled SARS-CoV-2 molecules from the Global Initiative on Exchanging Bird Flu Sample, the Biotechnology Information, the European Nucleic acid Database, and China National GeneBank to make data sharing easier.

Related works

-Shean RC, Makhsous N, Stoddard GD, Lin MJ, Greninger AL. VAPiD: a lightweight cross-platform viral annotation pipeline and identification tool to facilitate virus genome submissions to NCBI GenBank. BMC Bioinformatics 2019 Jan 23;20(1):48 [https://bmcbioinformatics.biomedcentral.com/articles/10.1186/s12859-019-2606-y]

[https://dx.doi.org/10.1186/s12859-019-2606-y] [https://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&list_uids=30674273&dopt=Abstract]

https://publichealth.jmir.org/2020/4/e24661/?utm_source=TrendMD&utm_medium=cpc&utm_campaign=JMIR_TrendMD_0

Carla Mavian et al., JMIR Public Health Surveill, 2020 .

https://bioinform.jmir.org/2021/1/e25995/citations?utm_source=TrendMD&utm_medium=cpc&utm_campaign=JMIR_Bioinformatics_and_Biotechnology_TrendMD_0

Emilio Mastriani et al., JMIR Bioinformatics and Biotechnology, 2021.

https://publichealth.jmir.org/2020/4/e23542?utm_source=TrendMD&utm_medium=cpc&utm_campaign=JMIR_TrendMD_0

Peter Forster et al., J Med Internet Res, 2020.

https://www.pnas.org/content/117/38/23652?utm_source=TrendMD&utm_medium=cpc&utm_campaign=Proc_Natl_Acad_Sci_U_S_A_TrendMD_1

Bethany Dearlove et al., Proc Natl Acad Sci U S A, 2020.