

GENOMICS PAPER

**Computational strategies to combat COVID-19
useful tools to accelerate SARS-CoV-2 and
coronavirus research**

Ahmed Amir Ahmed Ali

21-4-2021

Abstract :

SARS-CoV-2 (severe acute respiratory syndrome coronavirus 2) is a novel virus of the family Coronaviridae. The virus causes the infectious disease COVID-19. The biology of coronaviruses has been studied for many years. However, bioinformatics tools designed explicitly for SARS-CoV-2 have only recently been developed as a rapid reaction to the need for fast detection, understanding and treatment of COVID-19. To control the ongoing COVID-19 pandemic, it is of utmost importance to get insight into the evolution and pathogenesis of the virus. In this review, we cover bioinformatics workflows and tools for the routine detection of SARS-CoV-2 infection, the reliable analysis of sequencing data, the tracking of the COVID-19 pandemic and evaluation of containment measures, the study of coronavirus evolution, the discovery of potential drug targets and development of therapeutic strategies. For each tool, we briefly describe its use case and how it advances research specifically for SARS-CoV-2. All tools are free to use and available online, either through web applications or public code repositories

Introduction :

On 31 December 2019, the Wuhan Municipal Health Commission reported several cases of pneumonia in Wuhan (China) to the World Health Organization (<https://www.who.int/csr/don/05-january-2020-pneumonia-of-unknown-cause-china/en/>). The cause of these cases was a previously unknown coronavirus, now known as severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), which can manifest itself in the disease named COVID-19. At the time of writing (22 July 2020), nearly 15 million cases were reported worldwide, with over 600 000 deaths (<https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200722-covid-19-sitrep-184.pdf>). The group of Coronaviridae includes viruses with very long RNA genomes up to 33 000 nucleotides. SARS-CoV-2 belongs to the Sarbecovirus subgenus (genus: Betacoronavirus) and has a genome of approximately 30 000 nucleotides. In line with other members of Coronaviridae, SARS-CoV-2 has four main structural proteins: spike (S), envelope (E), membrane (M) and nucleocapsid (N). Further, several nonstructural proteins are encoded in the pp1a and pp1ab polyproteins, which are essential for viral replication. SARS-CoV-2 seems to use the human receptor ACE2 as its main entry, which has been observed for other Sarbecoviruses as well. The binding domains for ACE2 are located on the spike proteins, which further contain a novel furin cleavage site, associated with increased pathogenicity and transmission potential. In light of the COVID-19 pandemic, there has been a rapid increase in SARS-CoV-2-related research. It will be critical to get insight into the evolution and pathogenesis of the virus in order to control this pandemic.

Related works :

Researchers around the world are investigating SARS-CoV-2 sequence evolution on genome and protein level, tracking the pandemic using photodynamic and epidemiological models and examining potential drug targets. Laboratories are sharing SARS-CoV-2-related data with unprecedented speed. In light of this sheer amount of data many fundamental questions in SARS-CoV-2 research can only be tackled with the help of bioinformaticians. Adequate analysis of these data has the potential to boost discovery and inform both fundamental and applied science, in addition to public health initiatives. The natural first approach is to repurpose existing methods and resources. Simultaneously, the outbreak has had a huge impact on virus bioinformatics tools that have been developed recently and it is important to understand which tools are applicable to coronaviruses and which have been customized to address research questions related to SARS-CoV-2.