

(Abstract)

Several NGS-based techniques have recently been used to investigate disease transmission and spread chains, improve the accuracy and speed of molecular diagnostic tests, and aid drug and vaccine development. COVID-19 is one of the most severe global threats in recent memory, posing major social and economic threats. The development of efficient and rapid sequencing methods to reconstruct the genomic sequence of SARS-CoV-2, the etiological agent of COVID-19, was critical for the development of diagnostic molecular tests as well as active steps and strategies to slow the spread of the pandemic. SARS-CoV-2 genomes can be sequenced using a variety of approaches and methods, as shown by the large number of sequences available. On the other hand, each technology and sequencing method has its own set of benefits and drawbacks. We will provide a fast, but hopefully thorough, overview of currently available platforms and methodological approaches for sequencing SARS-CoV-2 genomes in this review. There's also a list of existing repositories and databases that have SARS-CoV-2 genomic data and metadata. Finally, we provide general guidance and suggestions for sharing and archiving SARS-CoV-2 data and metadata, implying that a more efficient and uniform integration of current and future SARS-CoV-2-related data would greatly help the battle against this new pathogen. This goal should be aided by our 'vademecum' for the production and management of SARS-CoV-2-related sequencing data

(Introduction)

SARS-genome CoV-2's structure. (A) The SARS-CoV-2 genome structure. Labels signify the names of genes. The red circle represents the TRS-L. In the lower row, nsps derived from the processing of pp1a and pp1ab polyproteins are shown (B). Dotted lines connect the TRS-L to the body of each individual sgRNA. The specific gene product obtained from each individual sgRNA is represented by the coloured boxes and labels. By discontinuously transcribing these genes, sub-genomic mRNAs are formed, resulting in a nested sequence of sub-genomic transcripts (sgmRNAs). If their synthesis is prematurely terminated at unique transcription regulatory sequences (TRSs) upstream of each accessory gene, antisense RNAs are directed to continue synthesis of the complement of the 67–72 nt 'leader' at the extreme 5' end of the positive sense genomic RNA. Positive sense sgmRNAs are 5' and 3' coterminal with the genome sequence and are transcribed from negative sense sgmRNAs. Discontinuous transcription, which is mediated by sequence identity between a donor RNA (TRS-B) and hairpin structures present in the acceptor RNA (TRS-L), is thought to modulate long-distance RNA–RNA interactions (see also Figure 1B). summarises the coronavirus replication and transcription processes. NGS technologies have been shown to be useful for tracing outbreak origins, spreading, and transmission chains, as well as controlling the evolution of etiological agents, in recent experience with emerging infectious diseases such as SARS, MERS, Zika, and Ebola. After the COVID-19 pandemic, unprecedented efforts were made to develop effective real-time surveillance strategies based on SARS-CoV-2 genome sequencing, with over 100 000 SARS-CoV-2 genomes deposited in dedicated repositories like EpiCov and others. These results have already sparked many studies on the virus's evolutionary dynamics and the identification of variants with potential clinical significance. Given our lack of understanding of the complexities of virus replication and gene expression, the likelihood of RNA modifications of either RNA strand during replication or

transcription, and, not least, the need for consistent handling, labelling, and deposition of sequence data has arisen. These problems can only be solved by combining and organising efforts. The need for fast, unrestricted, and rapid access to large volumes of processed and, in many cases, raw molecular data is unparalleled in the case of SARS-CoV-2. In a succinct but hopefully informative manner, this review summarises the existing state of the art for NGS applications in SARS-CoV-2 genomics. We include an overview of the repositories and databases that provide access to SARS-CoV-2 genomic data and metadata, as well as general advice for their proper sharing and deposition, in addition to comprehensive reviews of currently accessible sequencing methods. By providing an easy and detailed vademecum for the processing and handling of COVID-19-related sequencing data, as well as a clear picture of the state of the art, we hope to contribute to more efficient and thorough curation, integration, and utilisation of SARS-CoV-2 sequencing data and metadata

