Chapter 2
Genomics and Evolution of Novel Corona Virus 2019

NCBI is one of the world's premier websites and a master database [1] that has direct access to the genetics and protein related data of most of the known organisms. NCBI is used to retrieve the complete genome of the organism SARS Corona Virus 2019. The following is the information regarding the same:

Genome Length: 29903 bp ss-RNA

Description of the sequence: Severe acute respiratory syndrome coronavirus 2 isolate Wuhan-Hu-1, complete genome

The accession no for whole genome sequence: MN908947.3

2.1 Important Regions Within the Novel Corona Virus Genome [2]

From the data of NCBI it can be inferred that the major proteins involved in the virus structure and function are ORFs and structural proteins M, S, E and N. The longest gene region in the viral genome codes for ORF1ab followed by the gene coding for S protein. Furthermore, certain important characteristics of the genomic region are identified as follows:

- The complete genome is from 1 to 29903 bp.
- The initial 1–265 region codes for UTR un translated or leader sequence.
- The region between 266 and 21555 codes for ORF1ab polyprotein of the bacterial genes.
- 21563–25384 region codes for one of the structural proteins S of the virus called as Surface glycoprotein.
- 25393–26220 region codes for the ORF3a gene.
- E protein or Envelope protein is coded by the region 26245–26472.

A. Kumar et al., *Novel Coronavirus 2019*,
SpringerBriefs in Forensic and Medical Bioinformatics,
https://doi.org/10.1007/978-981-15-7918-9_2