

SARS-CoV-2 Practicals – Manipulating databases October 2024

NCBI GenBank is a large repository of nucleic sequences hosted in the USA. Every day the sequence data is synchronized between the European database EMBL-EBI/ENA, the Japanese database DDBJ/SRA and GenBank.

The reference sequence

- Go to GenBank (<https://www.ncbi.nlm.nih.gov/genbank/>) and search for the reference sequence of SARS-CoV-2.
- What is the length of this sequence?
- What information can you find in Genbank about the patient from which this sequence was obtained?
- In which publication was this sequence published?
- Each sequence in GenBank has an accession number. What is the accession number for the SARS-CoV-2 reference sequence?

Note that the current accession number ends with .2. This means that there has been an update of the sequence in GenBank.

Go to https://www.ncbi.nlm.nih.gov/nuccore/NC_045512.1?report=gireshist to compare the two versions.

- When was the second version of the sequence published on NCBI?
- Is there a difference in the reference authors list?
- Save the sequence as a FASTA file and examine the text file.

NCBI Taxonomy

NCBI Taxonomy is a curated classification and nomenclature for all of the organisms in the public sequence databases.

<https://www.ncbi.nlm.nih.gov/taxonomy>

- What is the taxon ID for SARS-CoV-2?
- Among the Orthocoronaviridae, how many groups are defined besides the alphacoronaviruses and betacoronaviruses?

Finding related sequences

BLAST (Basic Local Alignment Search Tool) finds regions of similarity between biological sequences. The program compares nucleotide or protein sequences to sequence databases and calculates the statistical significance. <https://blast.ncbi.nlm.nih.gov/Blast.cgi>

- Use nucleotide BLAST to find other coronaviruses related to SARS-CoV-2.

Hint: you have to exclude an organism.

The first sequences found by BLAST that are most closely related are synthetic constructs and sequences whose taxon field was incorrectly entered.

- Based on BLAST, what is the name of the virus that is most closely related to SARS-CoV-2 (outside of humans)?
- What is the percentage identity indicated between this virus and SARS-CoV-2?
- On which animal host species was this virus collected?
- By which research team?
- When and where was this virus sampled?