LKS Faculty of Medicine
School of Biomedical Sciences
香港大學生物醫學學院

How sequencing is used to identify SARS-COV2 strains

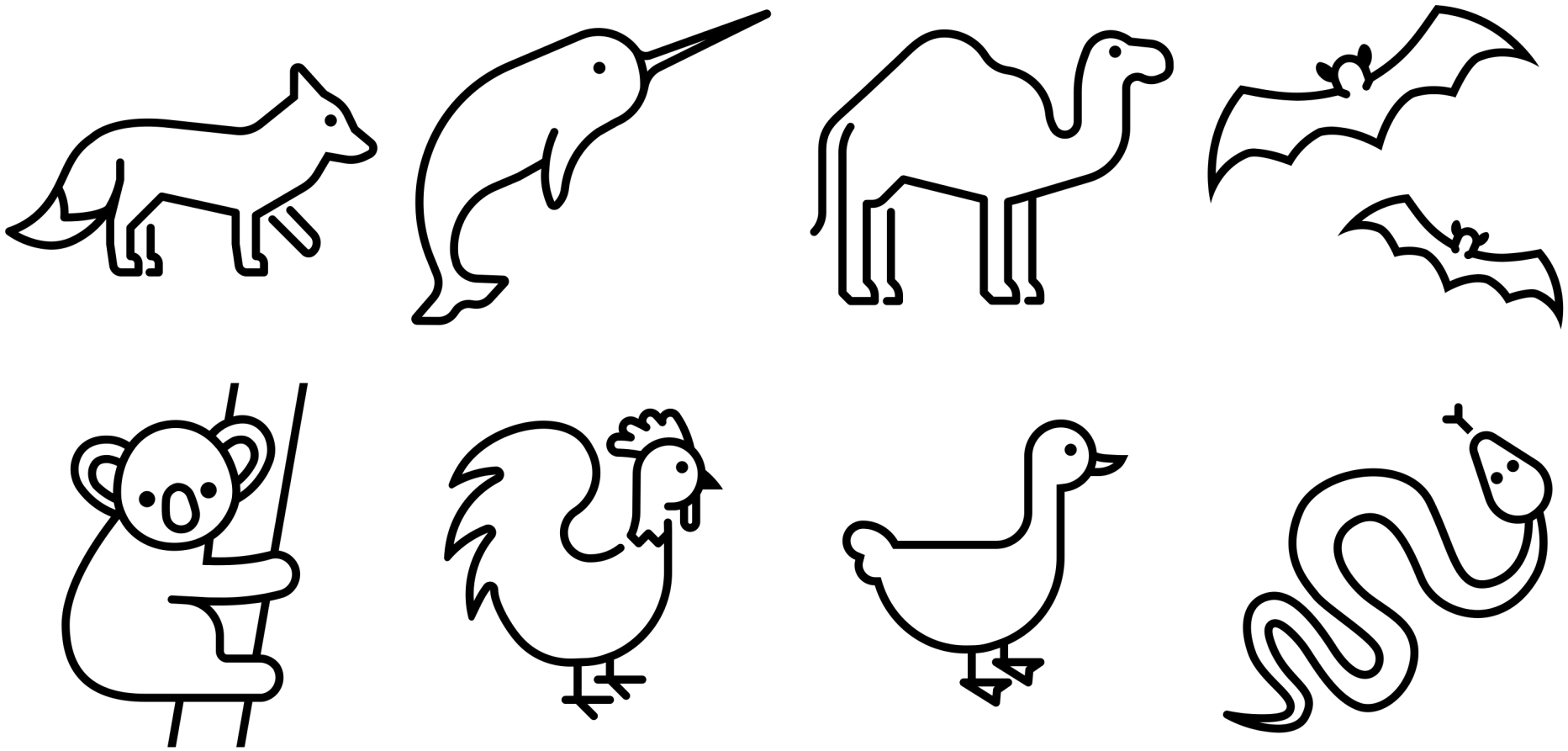
HKU STEM Data Science Lab

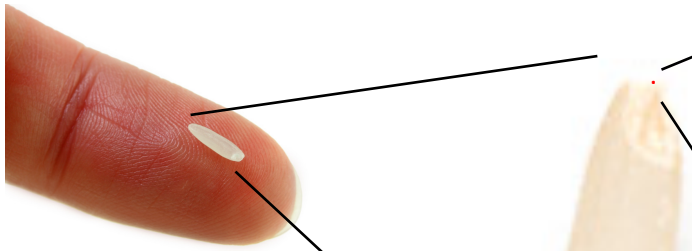
11th July 2022

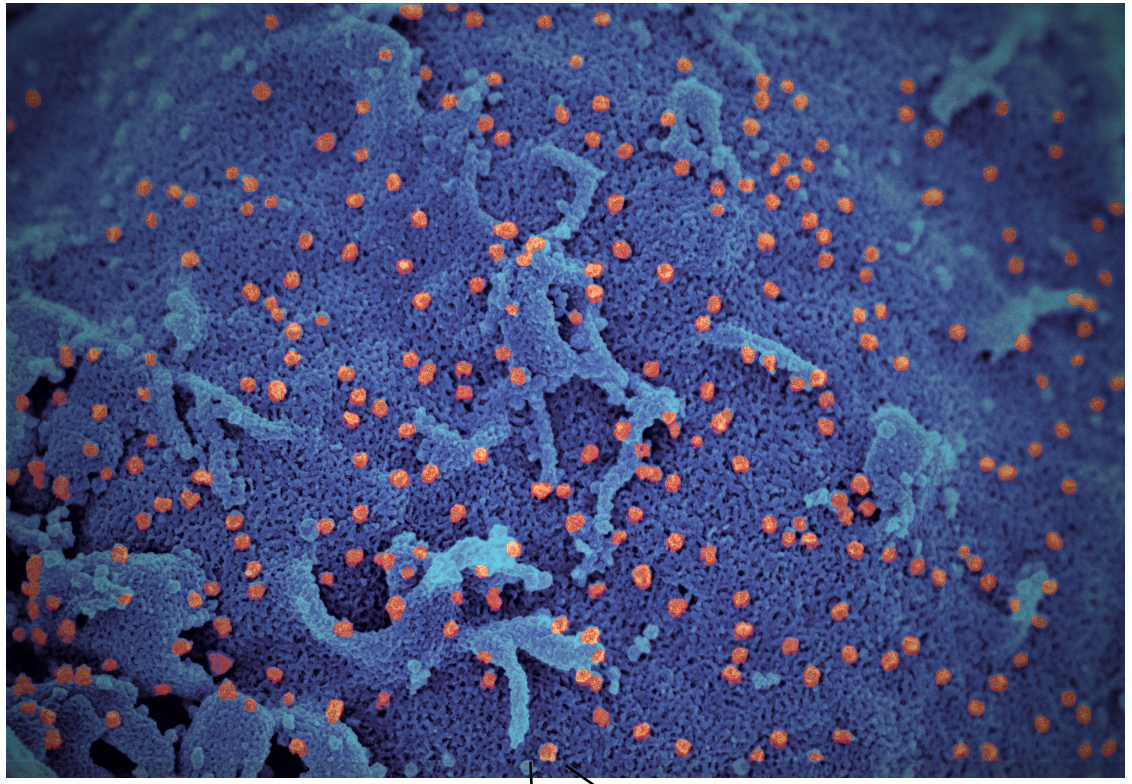
Dr Jason Wong & Dr Joshua Ho

The University of Hong Kong

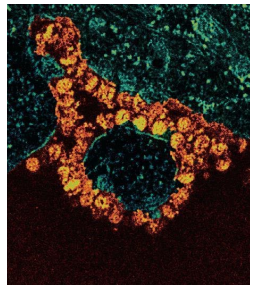
How do we identify animals?

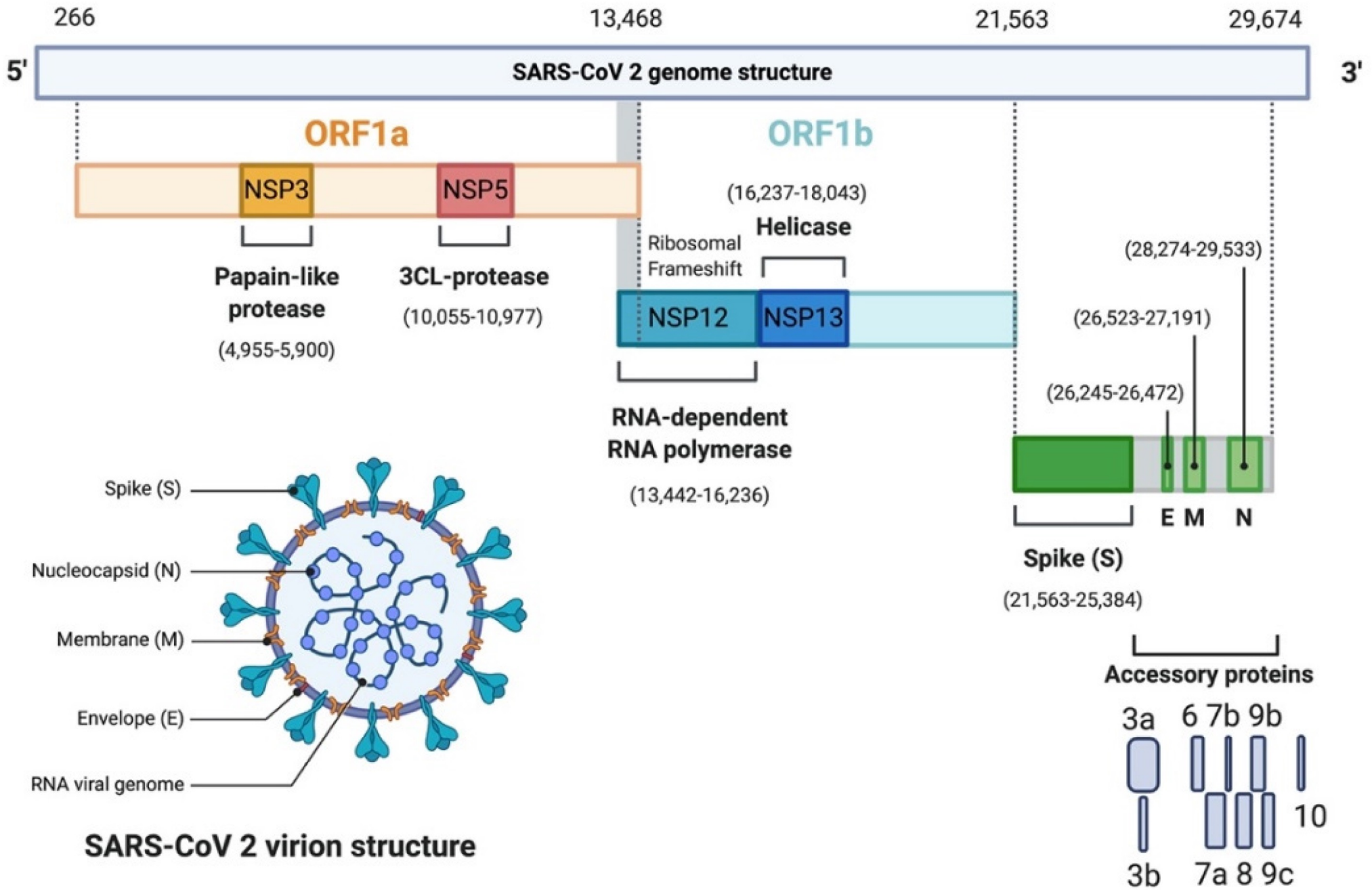






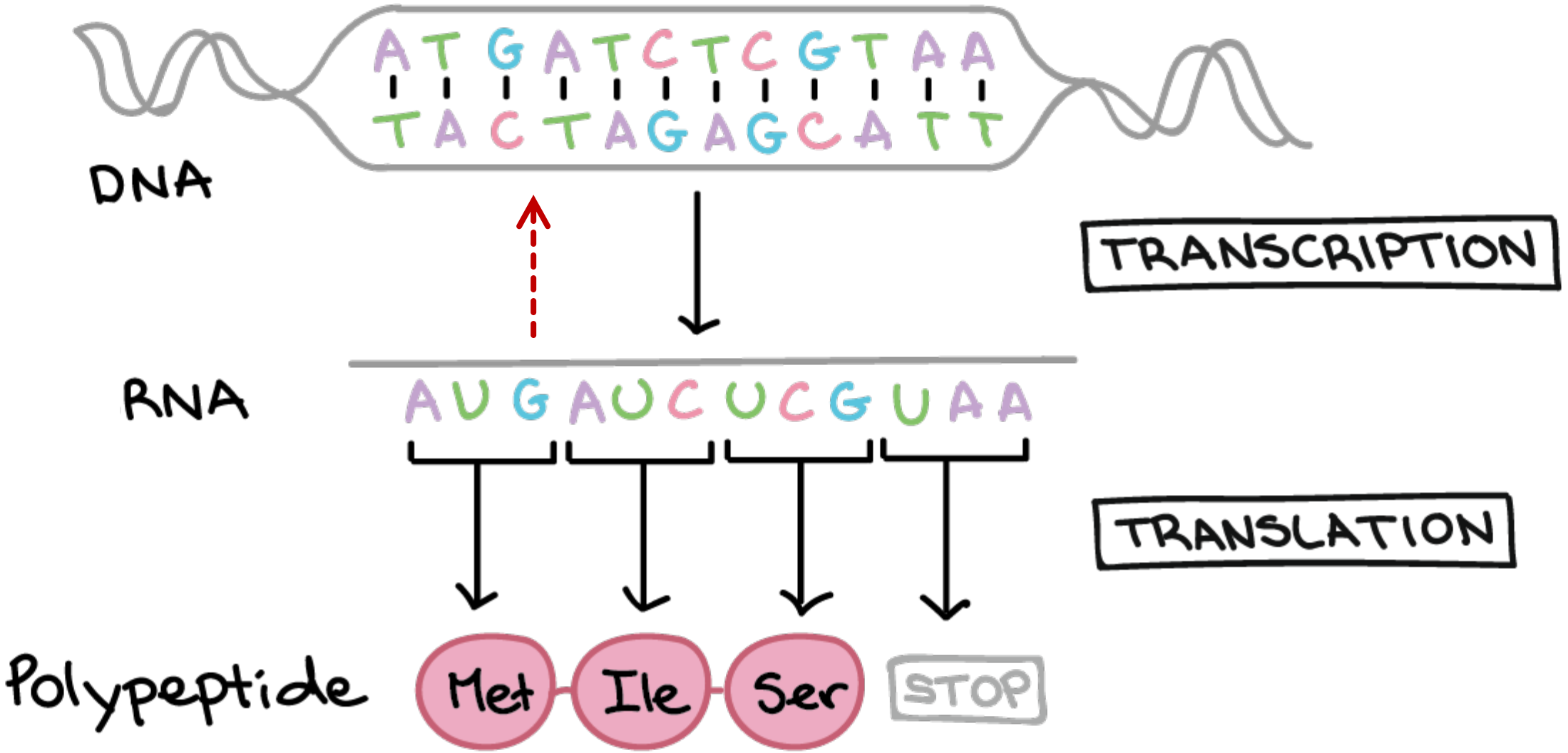
Virus are very very very small
(0.00001 cm or 100 nm)





Alanagreh et al. (2020) Pathogens

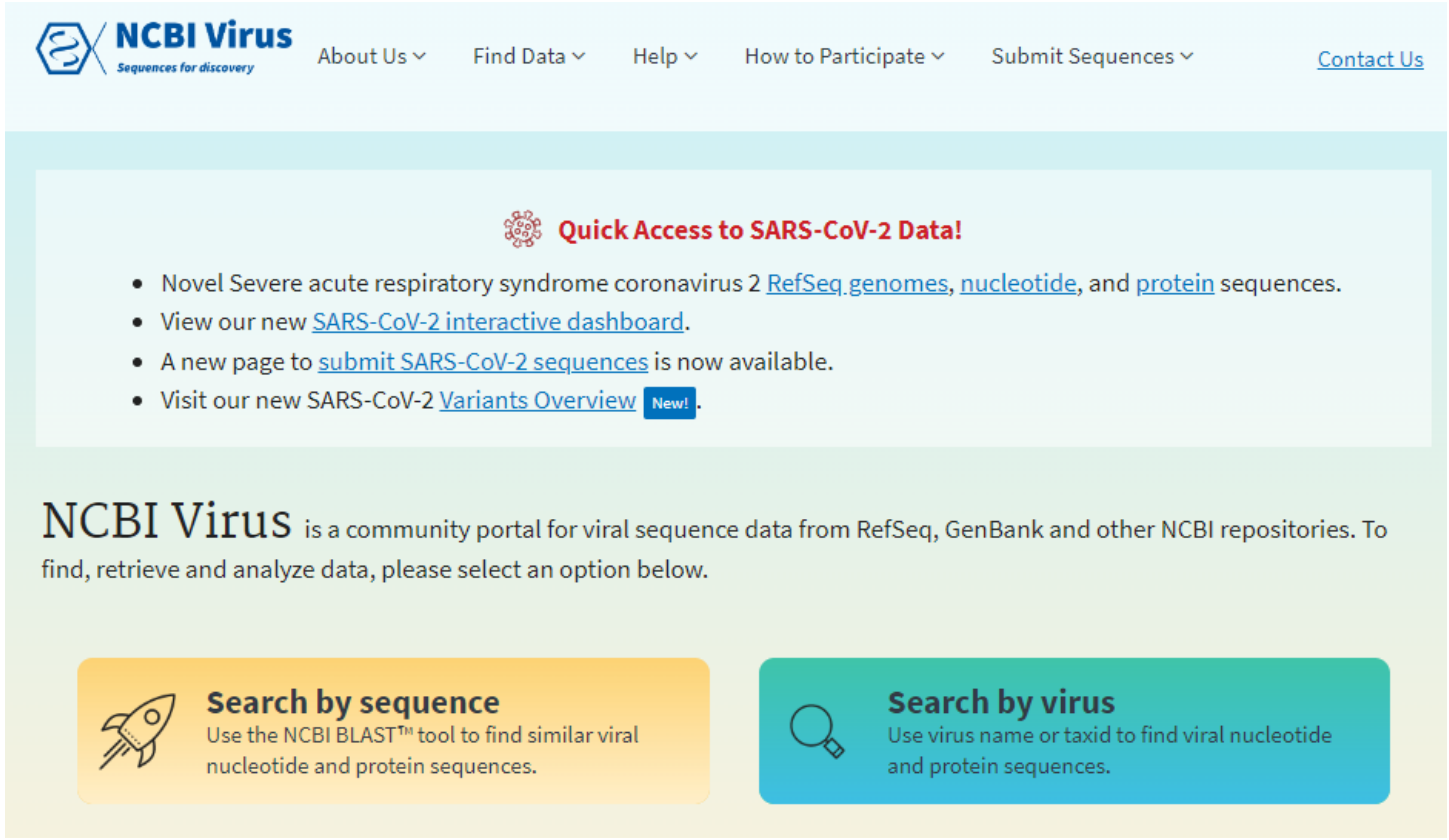
THE CENTRAL DOGMA



https://www.khanacademy.org/science/ap-biology/gene-expression-and-regulation/translation/a/intro-to-gene-expression-central-dogma

Reference SARS-CoV2 sequence (Wuhan strain)

- Open NCBI virus webpage in your browser (https://www.ncbi.nlm.nih.gov/labs/virus/)
- Click on Refseq genome (https://www.ncbi.nlm.nih.gov/nuccore/NC_045512.2)



Severe acute respiratory syndrome coronavirus 2 isolate Wuhan-Hu-1, complete genome

NCBI Reference Sequence: NC_045512.2

FASTA Graphics

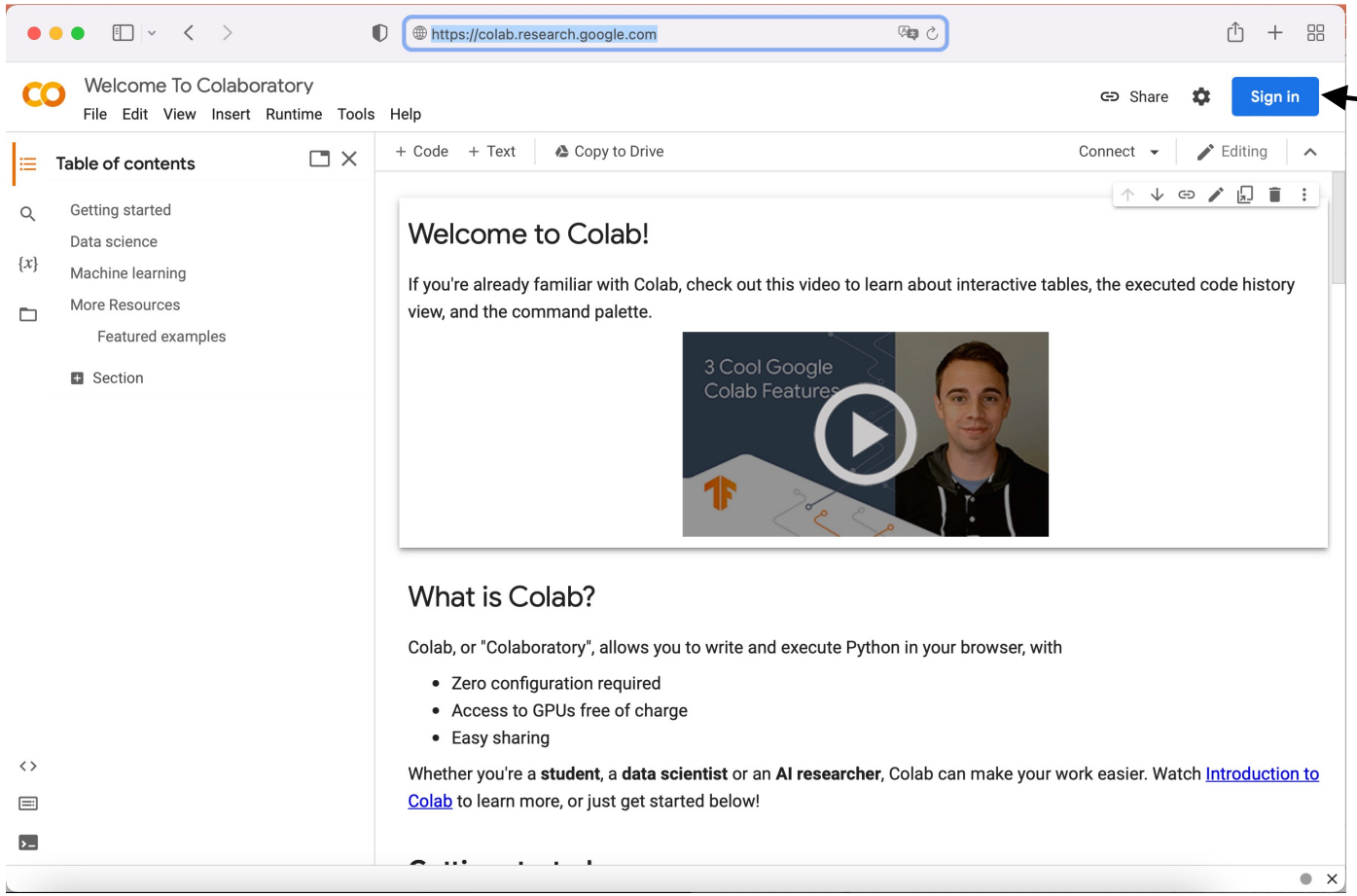
Go to:

```
LOCUS       NC_045512              29903 bp ss-RNA     linear   VRL 18-JUL-2020
DEFINITION  Severe acute respiratory syndrome coronavirus 2 isolate Wuhan-Hu-1,
            complete genome.
ACCESSION   NC_045512
VERSION     NC_045512.2
DBLINK      BioProject: PRJNA485481
KEYWORDS    RefSeq.
SOURCE      Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2)
  ORGANISM  Severe acute respiratory syndrome coronavirus 2
            Viruses; Riboviria; Orthornavirae; Pisuviricota; Pisoniviricetes;
            Nidovirales; Cornidovirineae; Coronaviridae; Orthocoronavirinae;
            Betacoronavirus; Sarbecovirus.
REFERENCE   1  (bases 1 to 29903)
  AUTHORS   Wu,F., Zhao,S., Yu,B., Chen,Y.M., Wang,W., Song,Z.G., Hu,Y.,
            Tao,Z.W., Tian,J.H., Pei,Y.Y., Yuan,M.L., Zhang,Y.L., Dai,F.H.,
            Liu,Y., Wang,Q.M., Zheng,J.J., Xu,L., Holmes,E.C. and Zhang,Y.Z.
  TITLE     A new coronavirus associated with human respiratory disease in
            China
  JOURNAL   Nature 579 (7798), 265-269 (2020)
   PUBMED   32015508
  REMARK    Erratum:[Nature. 2020 Apr;580(7803):E7. PMID: 32296181]
REFERENCE   2  (bases 13476 to 13503)
  AUTHORS   Baranov,P.V., Henderson,C.M., Anderson,C.B., Gesteland,R.F.,
            Atkins,J.F. and Howard,M.T.
  TITLE     Programmed ribosomal frameshifting in decoding the SARS-CoV genome
  JOURNAL   Virology 332 (2), 498-510 (2005)
   PUBMED   15680415
REFERENCE   3  (bases 29728 to 29768)
  AUTHORS   Robertson,M.P., Igel,H., Baertsch,R., Haussler,D., Ares,M. Jr. and
            Scott,W.G.
  TITLE     The structure of a rigorously conserved RNA element within the SARS
            virus genome
  JOURNAL   PLoS Biol. 3 (1), e5 (2005)
   PUBMED   15630477
REFERENCE   4  (bases 29609 to 29657)
  AUTHORS   Williams,G.D., Chang,R.Y. and Brian,D.A.
```

Genome data science workshop

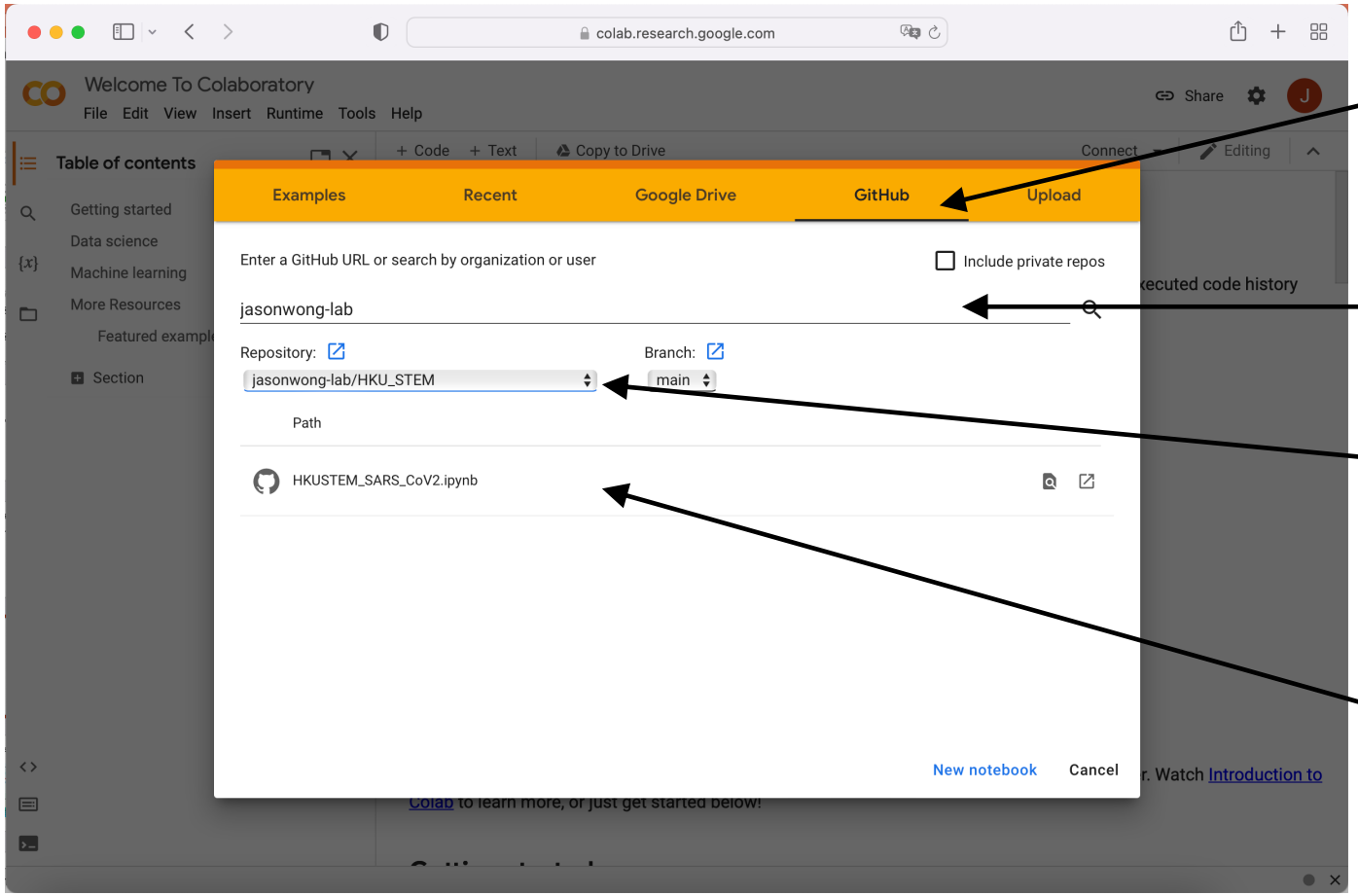
1) Set up Google Colab (5 minutes)
2) Multiple sequence alignment of Coronaviruses (15 minutes)
3) Alignment of Spike protein (10 minutes)
4) Visualise Spike protein (10 minutes)

Google Colab (https://colab.research.google.com)



Sign in with your Google account

Open jupyter notebook from GitHub

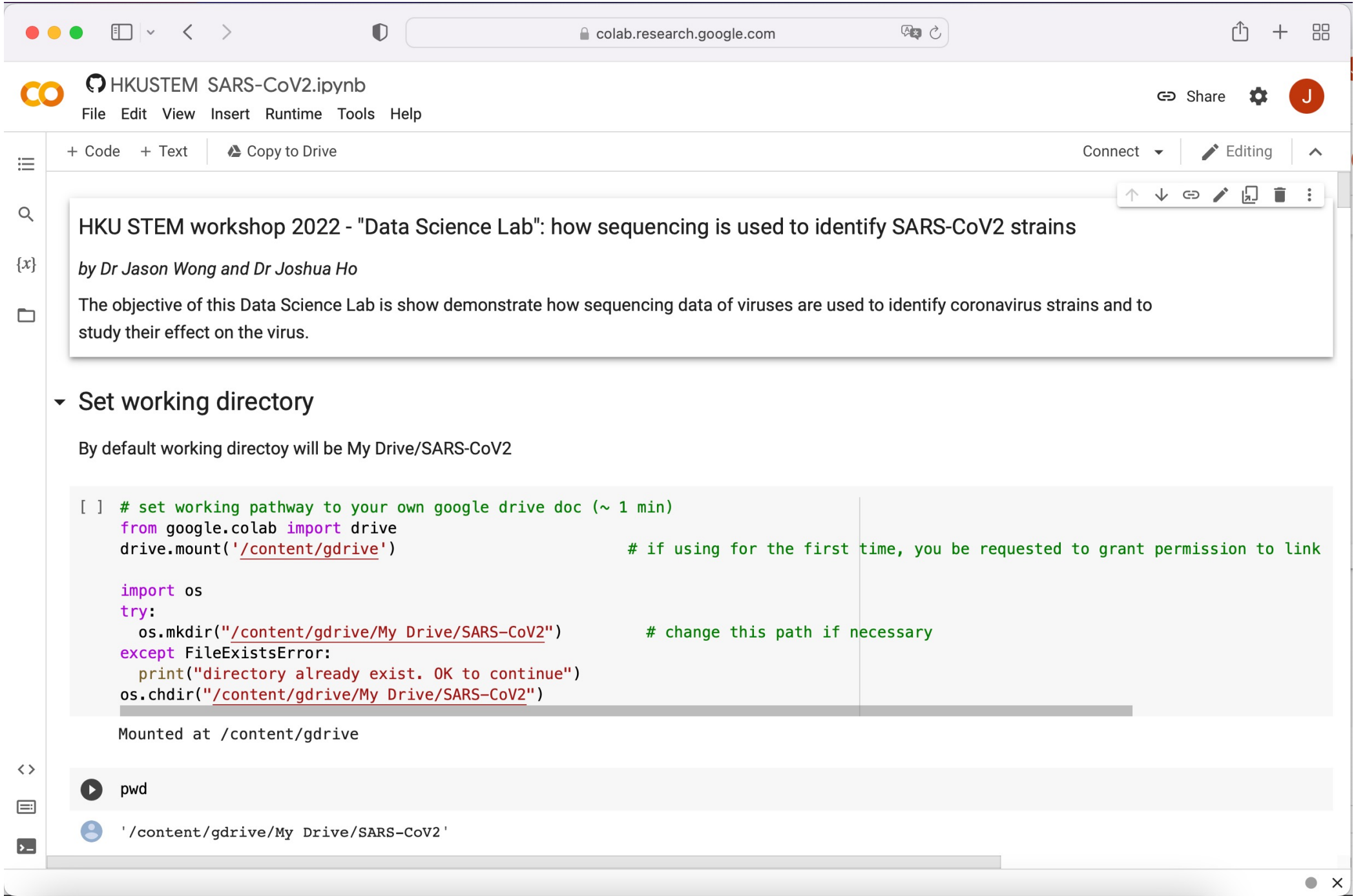


After logging it a window should open to open a notebook. If not go to File->Open notebook
Go to the GitHub tab

Search for “jasonwong-lab”

Select jasonwong-lab/HKU_STEM

Double click HKUSTEM_SARS_CoV2.ipynb


colab.research.google.com
HKUSTEM SARS-CoV2.ipynb
File Edit View Insert Runtime Tools Help
Share
J
+ Code + Text Copy to Drive
Connect
Editing
HKU STEM workshop 2022 - "Data Science Lab": how sequencing is used to identify SARS-CoV2 strains
by Dr Jason Wong and Dr Joshua Ho
The objective of this Data Science Lab is show demonstrate how sequencing data of viruses are used to identify coronavirus strains and to study their effect on the virus.
Set working directory
By default working directoy will be My Drive/SARS-CoV2
[ ] # set working pathway to your own google drive doc (~ 1 min)
from google.colab import drive
drive.mount('/content/gdrive') # if using for the first time, you be requested to grant permission to link
import os
try:
os.mkdir("/content/gdrive/My Drive/SARS-CoV2") # change this path if necessary
except FileExistsError:
print("directory already exist. OK to continue")
os.chdir("/content/gdrive/My Drive/SARS-CoV2")
Mounted at /content/gdrive
pwd
'/content/gdrive/My Drive/SARS-CoV2'

Resources

- NCBI SARS-CoV2 resource (https://www.ncbi.nlm.nih.gov/labs/virus/vssi/#/sars-cov-2)
- PDB spike protein (https://pdb101.rcsb.org/motm/246)
- Understanding evolution and phylogenetics (https://evolution.berkeley.edu/)