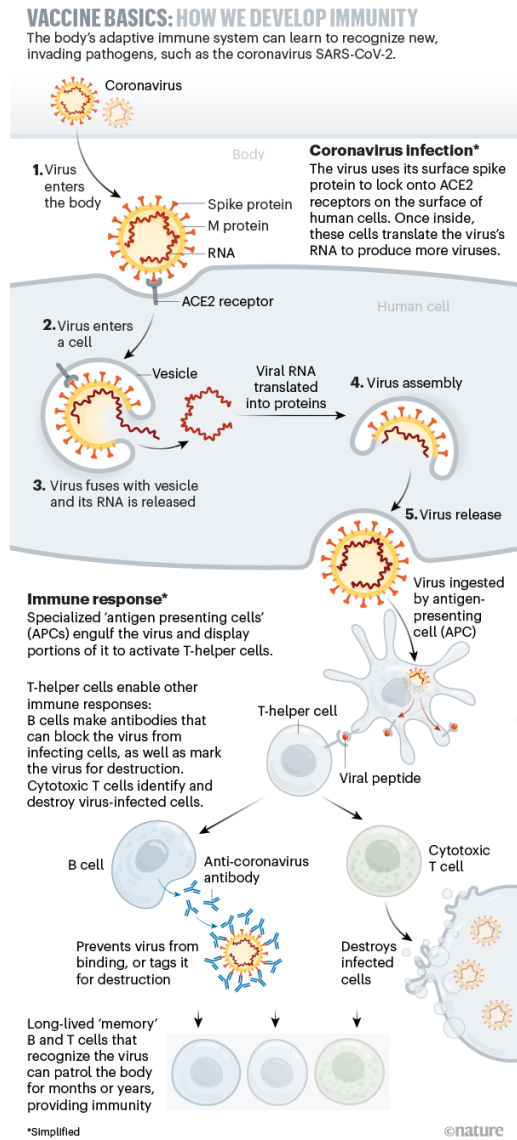# Sequence Analysis of Pfizer and Moderna COVID-19 Vaccines

Benjamin King (benjamin.l.king@maine.edu)

Department of Molecular and Biomedical Sciences
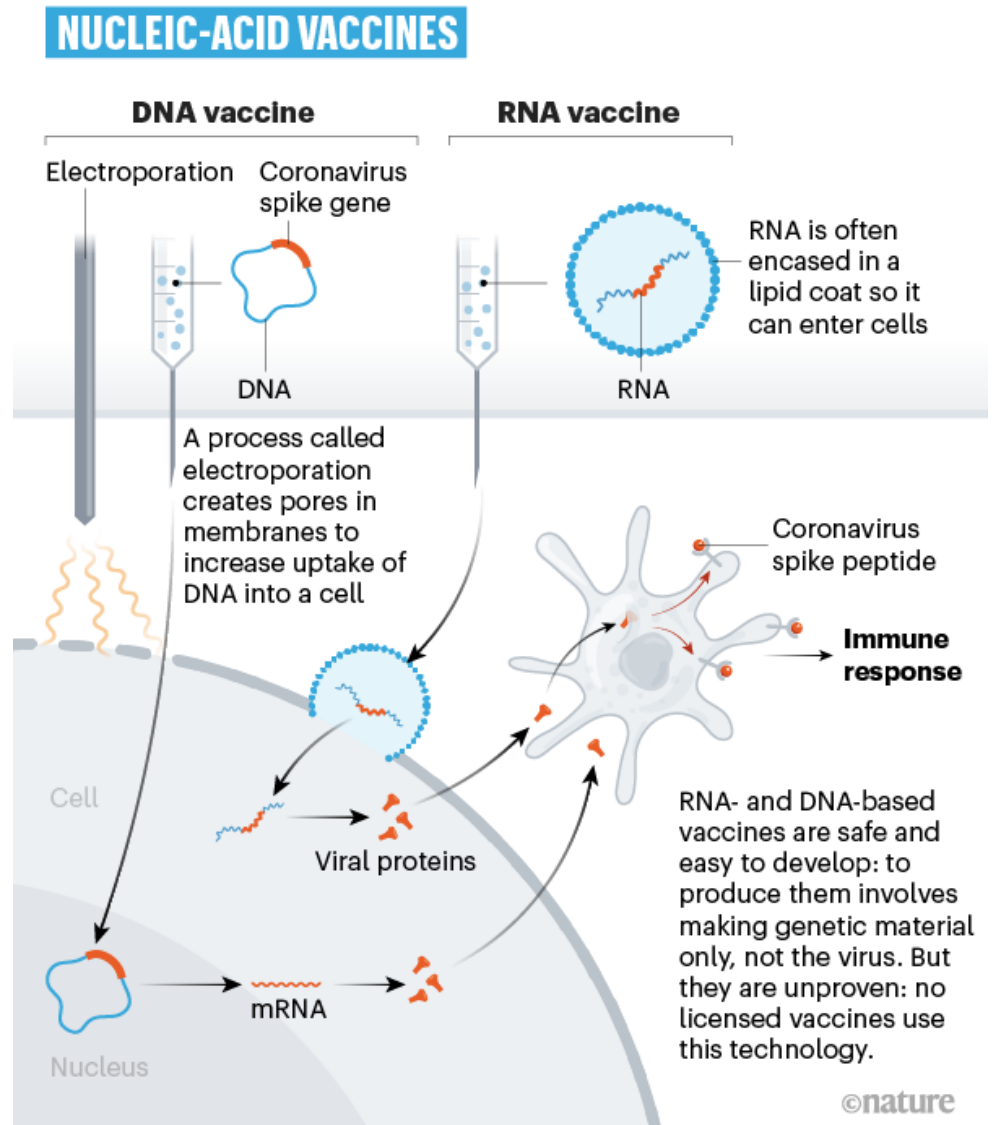
THE UNIVERSITY OF MAINE
1865

SARS-CoV-2 Infection and Immune Response

VACCINE BASICS: HOW WE DEVELOP IMMUNITY

The body's adaptive immune system can learn to recognize new, invading pathogens, such as the coronavirus SARS-CoV-2.

Coronavirus

Body

1. Virus enters the body
- Spike protein
- M protein
- RNA

ACE2 receptor

Coronavirus infection*
The virus uses its surface spike protein to lock onto ACE2 receptors on the surface of human cells. Once inside, these cells translate the virus's RNA to produce more viruses.

Human cell

2. Virus enters a cell

Vesicle

3. Virus fuses with vesicle and its RNA is released

Viral RNA translated into proteins

4. Virus assembly

5. Virus release

Virus ingested by antigen-presenting cell (APC)

Immune response*
Specialized 'antigen presenting cells' (APCs) engulf the virus and display portions of it to activate T-helper cells.

T-helper cells enable other immune responses:
B cells make antibodies that can block the virus from infecting cells, as well as mark the virus for destruction.
Cytotoxic T cells identify and destroy virus-infected cells.

T-helper cell

Viral peptide

B cell

Anti-coronavirus antibody

Prevents virus from binding, or tags it for destruction

Cytotoxic T cell

Destroys infected cells

Long-lived 'memory' B and T cells that recognize the virus can patrol the body for months or years, providing immunity

*Simplified

©nature

https://www.nature.com/articles/d41586-020-01221-y

RNA COVID-19 Vaccines Rely on Cell to Express Spike Protein

NUCLEIC-ACID VACCINES

DNA vaccine

Electroporation — Coronavirus spike gene

DNA

A process called electroporation creates pores in membranes to increase uptake of DNA into a cell

RNA vaccine

RNA is often encased in a lipid coat so it can enter cells

RNA

Cell

Viral proteins

Coronavirus spike peptide

Immune response

RNA- and DNA-based vaccines are safe and easy to develop: to produce them involves making genetic material only, not the virus. But they are unproven: no licensed vaccines use this technology.

Nucleus

mRNA

©nature

The $1000 Genome Is Revolutionizing How We Study Biology and Practice Medicine

$2.7B Over Decade   → $1000 in ~3 Days

Whole Genome Shotgun Sequencing

DNA

Fragment DNA

Sequence Fragments

Align Reads

```
...ATCCGTAAATGGGCT
...ATCCGTAAATGGGCTGATACTACTAATGC
          TGGGCTGATACTACTAATGCCAAACTGTACTAGTCCTG...
                            TGCCAAACTGTACTAGTCCTG...
```

Determine Consensus Sequence

```
...ATCCGTAAATGGGCTGATACTACTAATGCCAAACTGTACTAGTCCTG...
```

Illumina Flow Cell
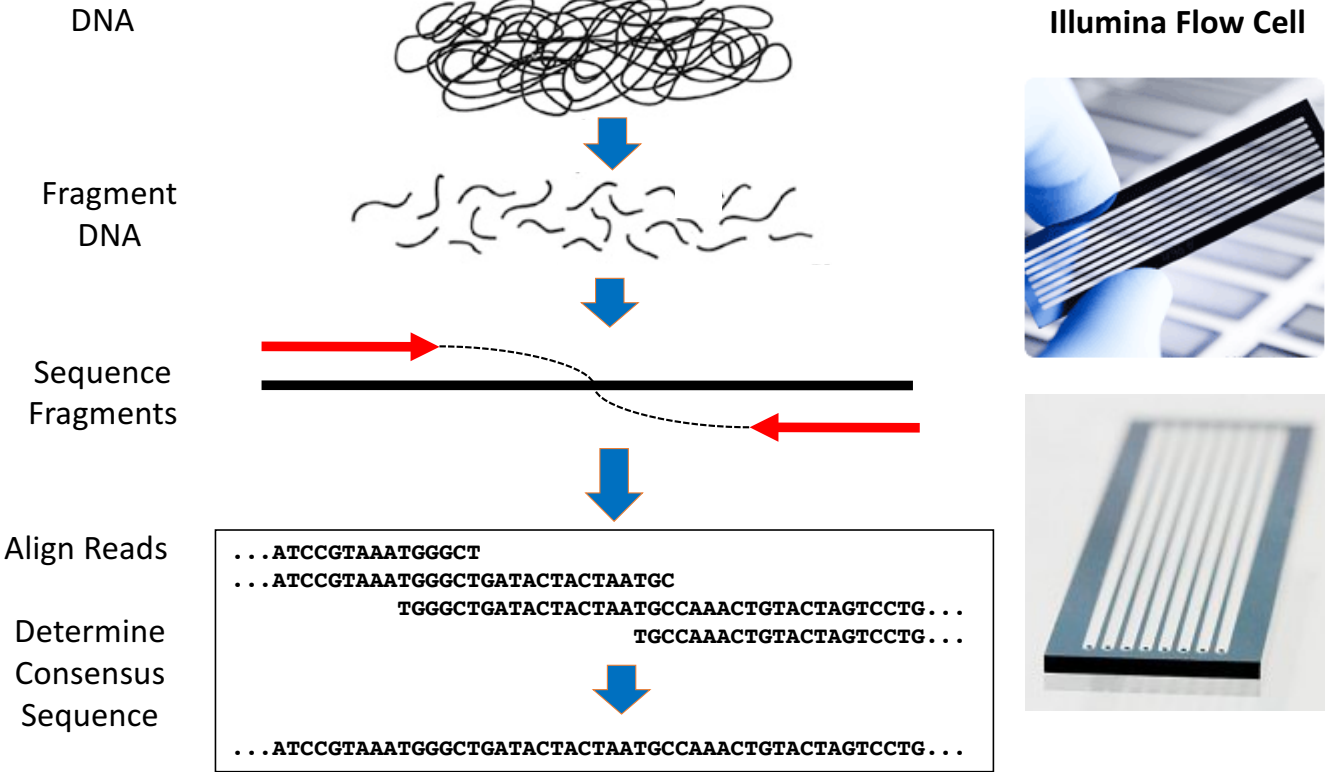
Your Own Genome Sequenced (30x Coverage)

Black Friday Sale (Nov. 25, 2020)

Illumina MiSeq in Hitchner Hall at UMaine

What are the sequences of the RNA in the Pfizer and Moderna COVID-19 Vaccines?

Pfizer Sequence

Moderna Sequence

>Figure1_032321_Spike-encoding_contig_assembled_from_BioNTech/Pfizer_BNT-162b2_vaccine

>Figure_2_32321_Spike-encoding_contig_assembled_from_Moderna_mRNA-1273_vaccine

What do we want to learn about the sequences?

-
-
-
-
-
-
-
-
-
-
-

Sequence Analysis Workflow
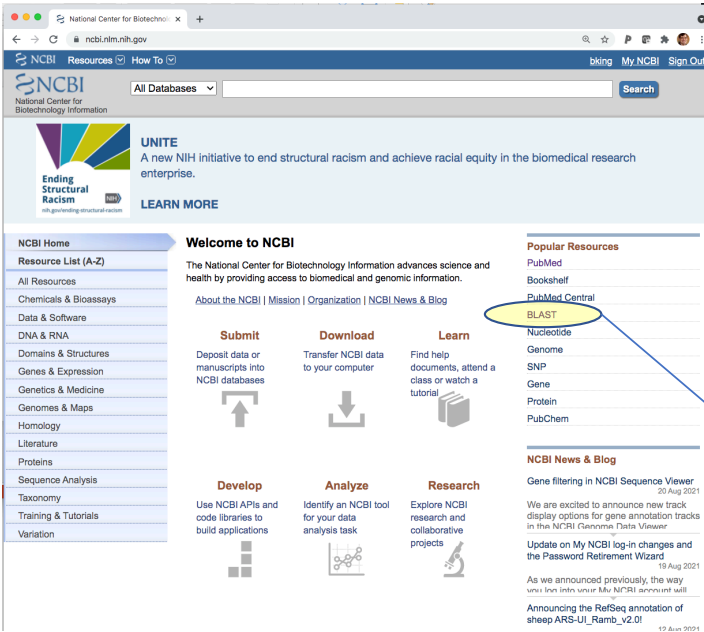
What viral protein(s) could be encoded by the sequences?

```
                          Step #1          BLAST                    Protein
                    ┌─────────────────►  (blastx) vs. all  ───────► Alignments
                    │                      proteins
  Vaccine           │
  Sequence ─────────┤
                    │   Step #2                          BLAST              Protein
                    └─────────────────►  ORF Finder ───► (blastp) vs. all ─► Alignments
                                                          proteins
```
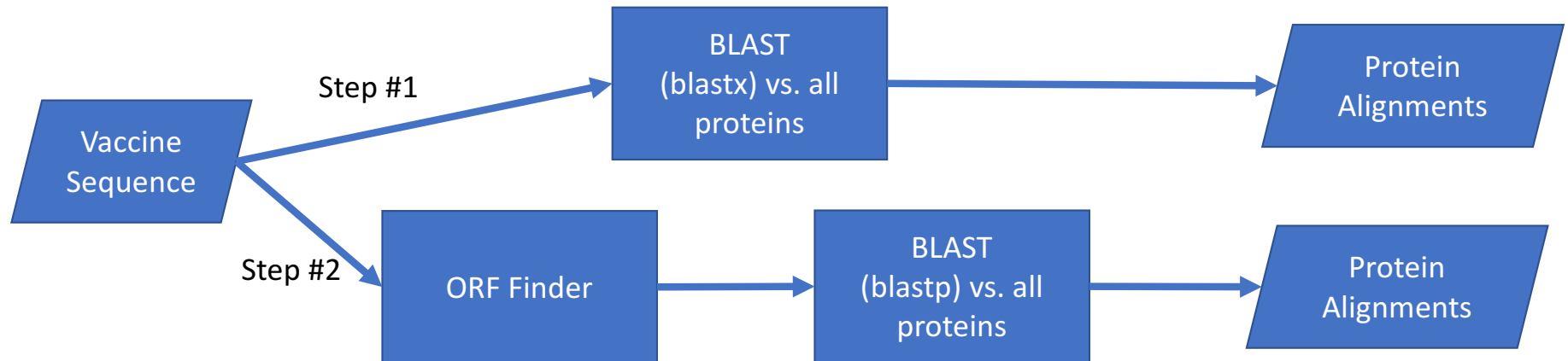
How similar are the sequences?

```
                          Clustal Omega
  Vaccine                 Multiple                Nucleotide
  Sequences ───────────►  Sequence    ─────────►  Alignments
                          Alignment
```
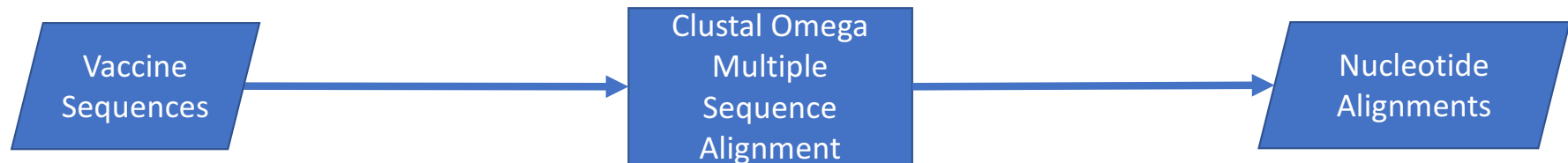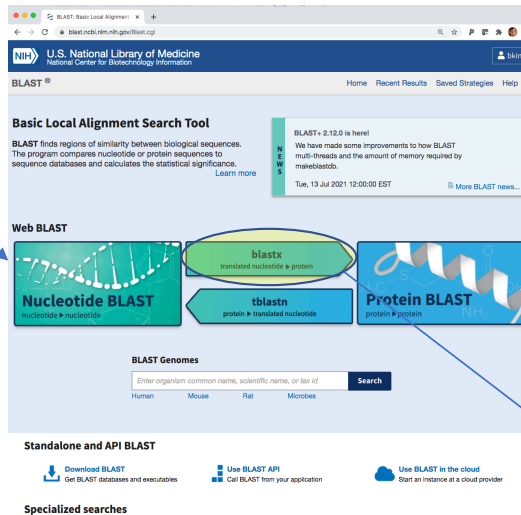
BLASTX vs. All Proteins @ NCBI

https://www.ncbi.nlm.nih.gov/

Paste the Pfizer sequence here

ORF Finder @ NCBI

Paste the Pfizer sequence here

Select "ORF1"

https://www.ncbi.nlm.nih.gov/orffinder/

Run BLAST with "Non-redundant protein sequences (nr)"

SARS-CoV-2 spike in prefusion state [Severe acute respiratory syndrome coronavirus 2]

Sequence ID: **6ZOW_A**   Length: **1273**   Number of Matches: **1**

See 9 more title(s) ⌄   See all Identical Proteins(IPG)

Range 1: 1 to 1273   GenPept   Graphics   ▼ Next Match   ▲ Previous Match

Score	Expect	Method	Identities	Positives	Gaps
2638 bits(6837)	0.0	Compositional matrix adjust.	1273/1273(100%)	1273/1273(100%)	0/1273(0%)

```
Query  1     MFVFLVLLPLVSSQCVNLTTRTQLPPAYTNSFTRGVYYPDKVFRSSVLHSTQDLFLPFFS  60
             MFVFLVLLPLVSSQCVNLTTRTQLPPAYTNSFTRGVYYPDKVFRSSVLHSTQDLFLPFFS
Sbjct  1     MFVFLVLLPLVSSQCVNLTTRTQLPPAYTNSFTRGVYYPDKVFRSSVLHSTQDLFLPFFS  60

Query  61    NVTWFHAIHVSGTNGTKRFDNPVLPFNDGVYFASTEKSNIIRGWIFGTTLDSKTQSLLIV  120
             NVTWFHAIHVSGTNGTKRFDNPVLPFNDGVYFASTEKSNIIRGWIFGTTLDSKTQSLLIV
Sbjct  61    NVTWFHAIHVSGTNGTKRFDNPVLPFNDGVYFASTEKSNIIRGWIFGTTLDSKTQSLLIV  120

Query  121   NNATNVVIKVCEFQFCNDPFLGVYYHKNNKSWMESEFRVYSSANNCTFEYVSQPFLMDLE  180
             NNATNVVIKVCEFQFCNDPFLGVYYHKNNKSWMESEFRVYSSANNCTFEYVSQPFLMDLE
Sbjct  121   NNATNVVIKVCEFQFCNDPFLGVYYHKNNKSWMESEFRVYSSANNCTFEYVSQPFLMDLE  180

Query  181   GKQGNFKNLREFVFKNIDGYFKIYSKHTPINLVRDLPQGFSALEPLVDLPIGINITRFQT  240
             GKQGNFKNLREFVFKNIDGYFKIYSKHTPINLVRDLPQGFSALEPLVDLPIGINITRFQT
Sbjct  181   GKQGNFKNLREFVFKNIDGYFKIYSKHTPINLVRDLPQGFSALEPLVDLPIGINITRFQT  240

Query  241   LLALHRSYLTPGDSSSGWTAGAAAYYVGYLQPRTFLLKYNENGTITDAVDCALDPLSETK  300
             LLALHRSYLTPGDSSSGWTAGAAAYYVGYLQPRTFLLKYNENGTITDAVDCALDPLSETK
Sbjct  241   LLALHRSYLTPGDSSSGWTAGAAAYYVGYLQPRTFLLKYNENGTITDAVDCALDPLSETK  300

Query  301   CTLKSFTVEKGIYQTSNFRVQPTESIVRFPNITNLCPFGEVFNATRFASVYAWNRKRISN  360
             CTLKSFTVEKGIYQTSNFRVQPTESIVRFPNITNLCPFGEVFNATRFASVYAWNRKRISN
Sbjct  301   CTLKSFTVEKGIYQTSNFRVQPTESIVRFPNITNLCPFGEVFNATRFASVYAWNRKRISN  360

Query  361   CVADYSVLYNSASFSTFKCYGVSPTKLNDLCFTNVYADSFVIRGDEVRQIAPGQTGKIAD  420
             CVADYSVLYNSASFSTFKCYGVSPTKLNDLCFTNVYADSFVIRGDEVRQIAPGQTGKIAD
Sbjct  361   CVADYSVLYNSASFSTFKCYGVSPTKLNDLCFTNVYADSFVIRGDEVRQIAPGQTGKIAD  420

Query  421   YNYKLPDDFTGCVIAWNSNNLDSKVGGNYNYLYRLFRKSNLKPFERDISTEIYQAGSTPC  480
             YNYKLPDDFTGCVIAWNSNNLDSKVGGNYNYLYRLFRKSNLKPFERDISTEIYQAGSTPC
Sbjct  421   YNYKLPDDFTGCVIAWNSNNLDSKVGGNYNYLYRLFRKSNLKPFERDISTEIYQAGSTPC  480

Query  481   NGVEGFNCYFPLQSYGFQPTNGVGYQPYRVVVLSFELLHAPATVCGPKKSTNLVKNKCVN  540
             NGVEGFNCYFPLQSYGFQPTNGVGYQPYRVVVLSFELLHAPATVCGPKKSTNLVKNKCVN
Sbjct  481   NGVEGFNCYFPLQSYGFQPTNGVGYQPYRVVVLSFELLHAPATVCGPKKSTNLVKNKCVN  540

Query  541   FNFNGLTGTGVLTESNKKFLPFQQFGRDIADTTDAVRDPQTLEILDITPCSFGGVSVITP  600
             FNFNGLTGTGVLTESNKKFLPFQQFGRDIADTTDAVRDPQTLEILDITPCSFGGVSVITP
Sbjct  541   FNFNGLTGTGVLTESNKKFLPFQQFGRDIADTTDAVRDPQTLEILDITPCSFGGVSVITP  600

Query  601   GTNTSNQVAVLYQDVNCTEVPVAIHADQLTPTWRVYSTGSNVFQTRAGCLIGAEHVNNSY  660
             GTNTSNQVAVLYQDVNCTEVPVAIHADQLTPTWRVYSTGSNVFQTRAGCLIGAEHVNNSY
Sbjct  601   GTNTSNQVAVLYQDVNCTEVPVAIHADQLTPTWRVYSTGSNVFQTRAGCLIGAEHVNNSY  660

Query  661   ECDIPIGAGICASYQTQTNSPRRARSVASQSIIAYTMSLGAENSVAYSNNSIAIPTNFTI  720
             ECDIPIGAGICASYQTQTNSPRRARSVASQSIIAYTMSLGAENSVAYSNNSIAIPTNFTI
Sbjct  661   ECDIPIGAGICASYQTQTNSPRRARSVASQSIIAYTMSLGAENSVAYSNNSIAIPTNFTI  720

Query  721   SVTTEILPVSMTKTSVDCTMYICGDSTECSNLLLQYGSFCTQLNRALTGIAVEQDKNTQE  780
             SVTTEILPVSMTKTSVDCTMYICGDSTECSNLLLQYGSFCTQLNRALTGIAVEQDKNTQE
Sbjct  721   SVTTEILPVSMTKTSVDCTMYICGDSTECSNLLLQYGSFCTQLNRALTGIAVEQDKNTQE  780

Query  781   VFAQVKQIYKTPPIKDFGGFNFSQILPDPSKPSKRSFIEDLLFNKVTLADAGFIKQYGDC  840
             VFAQVKQIYKTPPIKDFGGFNFSQILPDPSKPSKRSFIEDLLFNKVTLADAGFIKQYGDC
Sbjct  781   VFAQVKQIYKTPPIKDFGGFNFSQILPDPSKPSKRSFIEDLLFNKVTLADAGFIKQYGDC  840
```

Related Information

Identical Proteins - Identical proteins to 6ZOW_A

Paste the Pfizer and Moderna
sequences here

Clustal Omega @ EBI

https://www.ebi.ac.uk/Tools/msa/clustalo/

Figure 1: Spike-encoding contig assembled from BioNTech/Pfizer BNT-162b2 vaccine.

Cyan: Putative 5' UTR
Green: Start Codon
Yellow: Signal Peptide
Orange: Spike encoding region
Red: Stop codon(s)
Purple: 3' UTR
Blue: Start of polyA region (incomplete)

Figure 2: Spike-encoding contig assembled from Moderna mRNA-1273 vaccine.

Cyan: Putative 5' UTR
Green: Start Codon
Yellow: Signal Peptide
Orange: Spike encoding region
Red: Stop codon(s)
Purple: 3' UTR
Blue: Start of polyA region (incomplete)

SARS-CoV-2 Genome Sequencing Update

- 3,035,314 million genomes worldwide

- 812,520 genomes from US

- 5,537 genomes from Maine
  - **583 are Variant of Concern(VOC) Delta**
  - 939 VOC Alpha ("UK Variant")
  - 296 VOC Iota
  - 121 VOC Gamma
  - 108 VOC Zeta
  - 8 VOC Eta
  - 4 VOC Beta
  - 1 VOC Lambda
  - 0 VOC Kappa

VOC Alpha 202012/01 GRY (B.1.1.7) first detected in the UK
VOC Beta GH/501Y.V2 (B.1.351+B.1.351.2+B.1.351.3) first detected in South Africa
VOC Gamma GR/501Y.V3 (P.1+P.1.x) first detected in Brazil/Japan
VOC Delta G/478K.V1 (B.1.617.2+AY.x) first detected in India
VOI Zeta GR/484K.V2 (P.2) first detected in Brazil
VOI Eta G/484K.V3 (B.1.525) first detected in UK/Nigeria
VOI Iota GH/253G.V1 (B.1.526) first detected in USA/New York
VOI Kappa G/452R.V3 (B.1.617.1) first detected in India
VOI Lambda GR/452Q.V1 (C.37) first detected in Peru

Aug. 25, 2021

GISAID

| Registered Users | EpiFlu™ | EpiCoV™ | My profile |

EpiCoV™   Search   Downloads   Upload

Search

Accession ID		Virus name		☐ complete	☐ high coverage
Location	North America / USA / Maine...	Host		☐ low coverage excl	☐ w/Patient status
Collection		to		Submission		to		☐ collection date compl
Clade	all	Lineage		Substitutions		Variants	VOC Delta G/478K.V1 (l	Reset	Fulltext ▲

☐	Virus name	Passage de	Accession ID	Collection da	Submission D	ⓘ	Length	Host	Location	Originating
☐	hCoV-19/USA/ME-HETL-J5855/2021	Original	EPI_ISL_3544587	2021-08-12	2021-08-21	ⓘ	29,727	Human	North America / U	Maine Heal
☐	hCoV-19/USA/ME-HETL-J5854/2021	Original	EPI_ISL_3544586	2021-08-12	2021-08-21	ⓘ	29,727	Human	North America / U	Maine Heal
☐	hCoV-19/USA/ME-HETL-J5826/2021	Original	EPI_ISL_3544585	2021-08-12	2021-08-21	ⓘ	29,564	Human	North America / U	Maine Heal
☐	hCoV-19/USA/ME-HETL-J5825/2021	Original	EPI_ISL_3544584	2021-08-12	2021-08-21	ⓘ	29,568	Human	North America / U	Maine Heal
☐	hCoV-19/USA/ME-HETL-J5824/2021	Original	EPI_ISL_3544583	2021-08-12	2021-08-21	ⓘ	29,518	Human	North America / U	Maine Heal
☐	hCoV-19/USA/ME-HETL-J5823/2021	Original	EPI_ISL_3544582	2021-08-12	2021-08-21	ⓘ	29,518	Human	North America / U	Maine Heal
☐	hCoV-19/USA/ME-HETL-J5822/2021	Original	EPI_ISL_3544581	2021-08-12	2021-08-21	⚠	29,518	Human	North America / U	Maine Heal
☐	hCoV-19/USA/ME-HETL-J5664/2021	Original	EPI_ISL_3544580	2021-08-12	2021-08-21	⚠	29,676	Human	North America / U	Maine Heal
☐	hCoV-19/USA/ME-HETL-J5663/2021	Original	EPI_ISL_3544579	2021-08-12	2021-08-21	ⓘ	29,676	Human	North America / U	Maine Heal
☐	hCoV-19/USA/ME-HETL-J5657/2021	Original	EPI_ISL_3544578	2021-08-12	2021-08-21	ⓘ	29,728	Human	North America / U	Maine Heal
☐	hCoV-19/USA/ME-HETL-J5656/2021	Original	EPI_ISL_3544577	2021-08-12	2021-08-21	◇	29,676	Human	North America / U	Maine Heal
☐	hCoV-19/USA/ME-HETL-J5655/2021	Original	EPI_ISL_3544576	2021-08-12	2021-08-21	◇	29,518	Human	North America / U	Maine Heal
☐	hCoV-19/USA/ME-HETL-J5653/2021	Original	EPI_ISL_3544575	2021-08-12	2021-08-21	ⓘ	29,728	Human	North America / U	Maine Heal
☐	hCoV-19/USA/ME-HETL-J5853/2021	Original	EPI_ISL_3544574	2021-08-11	2021-08-21	◇	29,676	Human	North America / U	Maine Heal
☐	hCoV-19/USA/ME-HETL-J5848/2021	Original	EPI_ISL_3544573	2021-08-11	2021-08-21	ⓘ	29,728	Human	North America / U	Maine Heal
☐	hCoV-19/USA/ME-HETL-J5847/2021	Original	EPI_ISL_3544572	2021-08-11	2021-08-21	ⓘ	29,726	Human	North America / U	Maine Heal
☐	hCoV-19/USA/ME-HETL-J5845/2021	Original	EPI_ISL_3544571	2021-08-11	2021-08-21	ⓘ	29,676	Human	North America / U	Maine Heal
☐	hCoV-19/USA/ME-HETL-J5851/2021	Original	EPI_ISL_3544570	2021-08-10	2021-08-21	ⓘ	29,714	Human	North America / U	Maine Heal
☐	hCoV-19/USA/ME-HETL-J5844/2021	Original	EPI_ISL_3544569	2021-08-10	2021-08-21	ⓘ	29,725	Human	North America / U	Maine Heal
☐	hCoV-19/USA/ME-HETL-J5843/2021	Original	EPI_ISL_3544568	2021-08-10	2021-08-21	◇	29,727	Human	North America / U	Maine Heal
☐	hCoV-19/USA/ME-HETL-J5841/2021	Original	EPI_ISL_3544566	2021-08-10	2021-08-21	ⓘ	29,676	Human	North America / U	Maine Heal
☐	hCoV-19/USA/ME-HETL-J5840/2021	Original	EPI_ISL_3544565	2021-08-10	2021-08-21	ⓘ	29,728	Human	North America / U	Maine Heal
☐	hCoV-19/USA/ME-HETL-J5768/2021	Original	EPI_ISL_3544564	2021-08-10	2021-08-21	◇	29,676	Human	North America / U	Maine Heal
☐	hCoV-19/USA/ME-HETL-J5767/2021	Original	EPI_ISL_3544563	2021-08-10	2021-08-21	⚠	29,676	Human	North America / U	Maine Heal
☐	hCoV-19/USA/ME-HETL-J5766/2021	Original	EPI_ISL_3544562	2021-08-10	2021-08-21	ⓘ	29,728	Human	North America / U	Maine Heal
☐	hCoV-19/USA/ME-HETL-J5765/2021	Original	EPI_ISL_3544561	2021-08-10	2021-08-21	ⓘ	29,714	Human	North America / U	Maine Heal
☐	hCoV-19/USA/ME-HETL-J5662/2021	Original	EPI_ISL_3544560	2021-08-10	2021-08-21	◇	29,676	Human	North America / U	Maine Heal
☐	hCoV-19/USA/ME-HETL-J5661/2021	Original	EPI_ISL_3544559	2021-08-10	2021-08-21	ⓘ	29,727	Human	North America / U	Maine Heal
☐	hCoV-19/USA/ME-HETL-J5660/2021	Original	EPI_ISL_3544558	2021-08-10	2021-08-21	ⓘ	29,714	Human	North America / U	Maine Heal
☐	hCoV-19/USA/ME-HETL-J5551/2021	Original	EPI_ISL_3544557	2021-08-10	2021-08-21	ⓘ	29,728	Human	North America / U	Maine Heal
☐	hCoV-19/USA/ME-HETL-J5550/2021	Original	EPI_ISL_3544556	2021-08-10	2021-08-21	ⓘ	29,728	Human	North America / U	Maine Heal
☐	hCoV-19/USA/ME-HETL-J5839/2021	Original	EPI_ISL_3544555	2021-08-09	2021-08-21	⚠	29,200	Human	North America / U	Maine Heal

Total: 583 viruses   << < 1 2 3 4 5 > >>   ☑ Select   Analysis   Download