

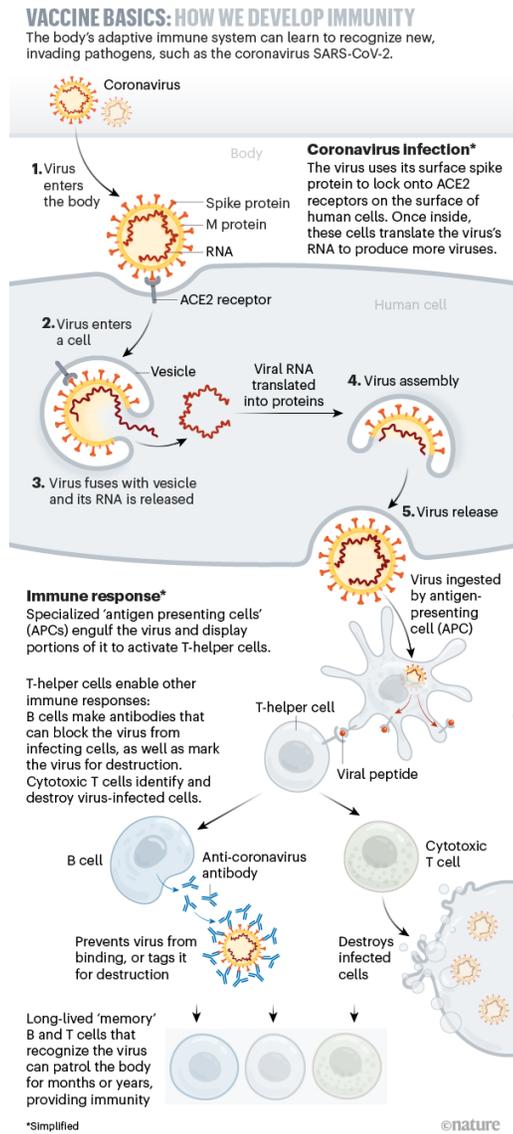
# Sequence Analysis of Pfizer and Moderna COVID-19 Vaccines



Benjamin King (benjamin.l.king@maine.edu)  
Department of Molecular and Biomedical Sciences

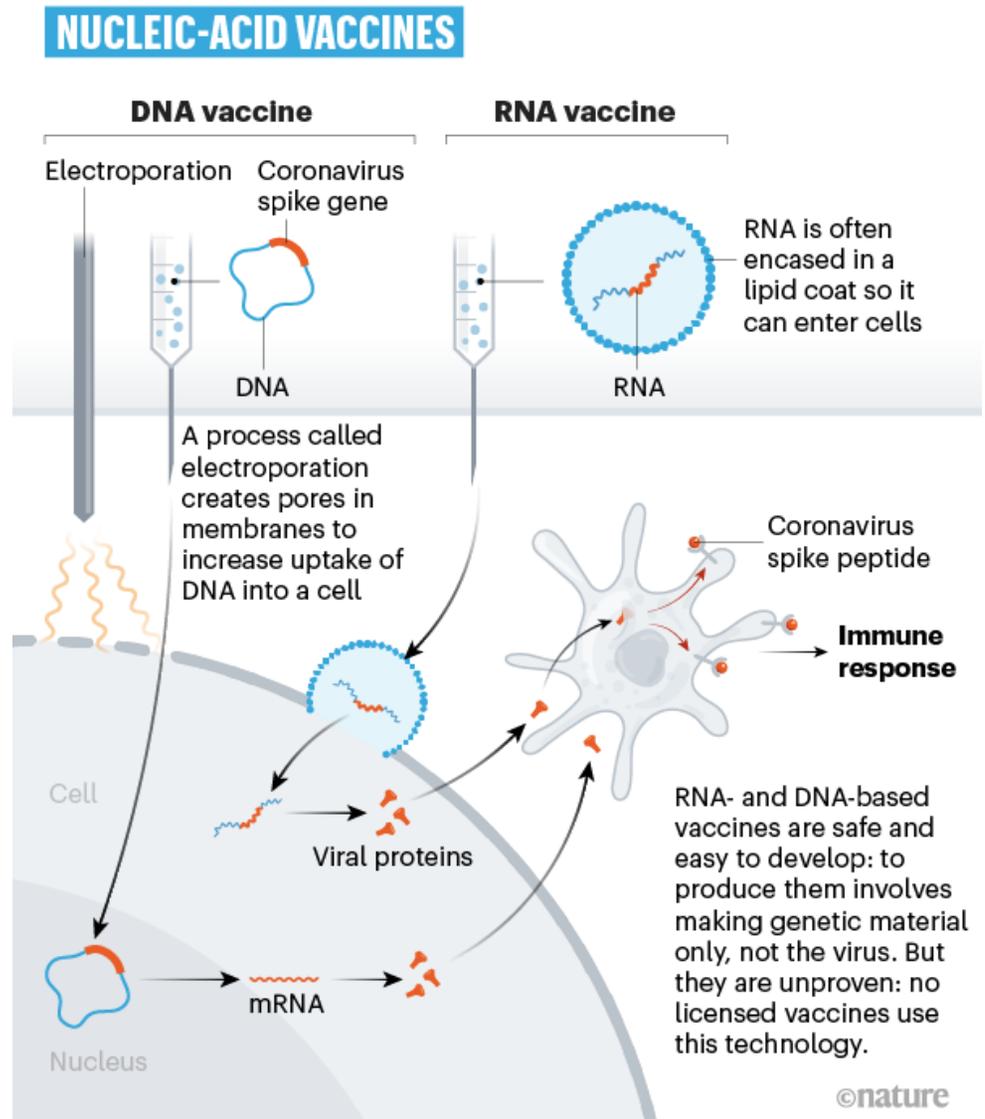


# SARS-CoV-2 Infection and Immune Response



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

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



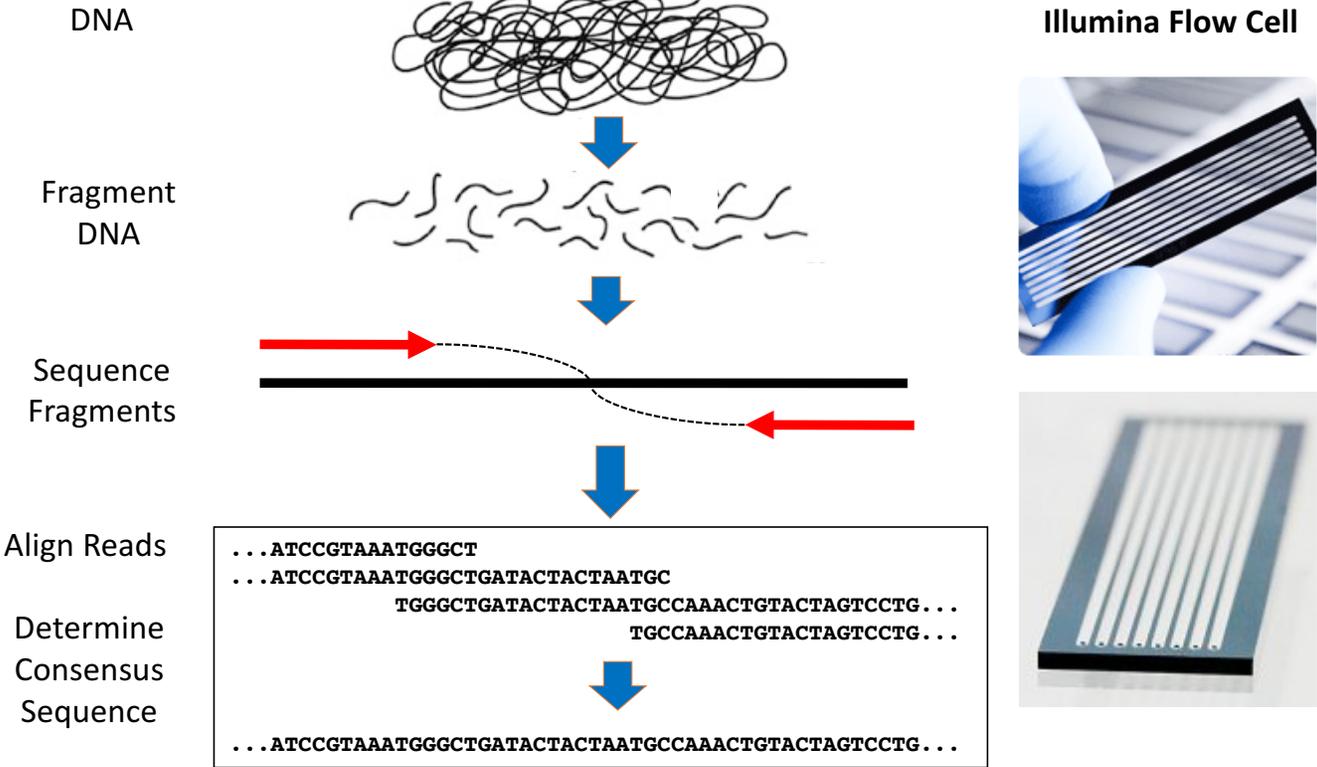
<https://www.nature.com/articles/d41586-020-01221-y>

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

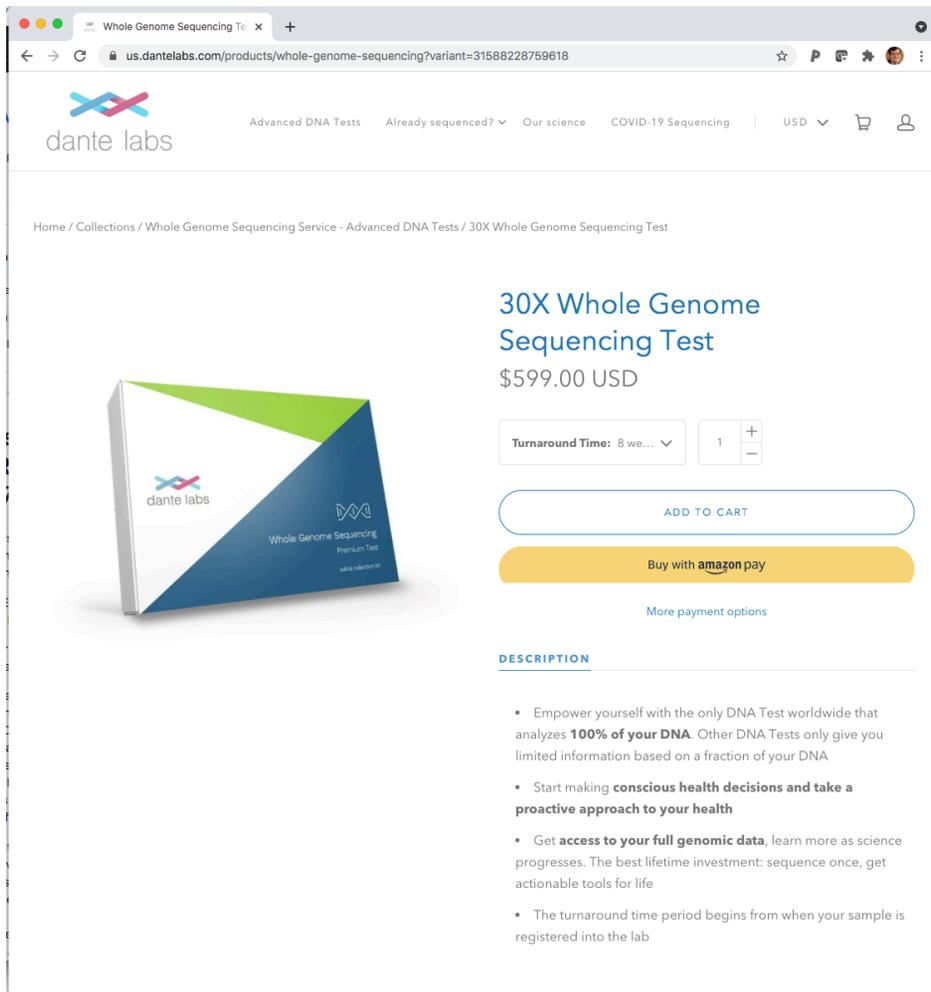
\$2.7B Over Decade → \$1000 in ~3 Days



# Whole Genome Shotgun Sequencing



# Your Own Genome Sequenced (30x Coverage)



The screenshot shows the Dante Labs website product page for the 30X Whole Genome Sequencing Test. The page includes the Dante Labs logo, navigation links, and a breadcrumb trail: Home / Collections / Whole Genome Sequencing Service - Advanced DNA Tests / 30X Whole Genome Sequencing Test. The product is displayed with a 3D rendering of the test kit box, which is white with green and blue accents and features the Dante Labs logo and the text "Whole Genome Sequencing Premium Test". The price is listed as \$599.00 USD. A dropdown menu shows a turnaround time of 8 weeks, and a quantity selector is set to 1. There are buttons for "ADD TO CART", "Buy with amazon pay", and "More payment options". A "DESCRIPTION" section contains the following text:

**DESCRIPTION**

- Empower yourself with the only DNA Test worldwide that analyzes **100% of your DNA**. Other DNA Tests only give you limited information based on a fraction of your DNA
- Start making **conscious health decisions and take a proactive approach to your health**
- Get **access to your full genomic data**, learn more as science progresses. The best lifetime investment: sequence once, get actionable tools for life
- The turnaround time period begins from when your sample is registered into the lab

## Black Friday Sale (Nov. 25, 2020)



The banner features a red background with a white torn-paper effect. On the left and right sides, there are images of laboratory equipment, including DNA sequencers. In the center, the Dante Labs logo is displayed above the text "dante labs". Below this, the text "Whole Genome" is written in white, followed by the price "\$149" in a large, bold white font.

## ILLUMINA MiSeq in Hitchner Hall at UMaine



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

The screenshot shows a GitHub repository page for the repository `NAalytics/Assemblies-of-putative-SARS-CoV2-spike-encoding-mRNA-sequences-for-vaccines-BNT-162b2-and-mRNA-1273`. The repository has 98 watchers, 3.4k stars, and 529 forks. The main branch is `main` with 1 branch and 0 tags. The repository contains several files, with `Figure1Figure2_032321.fasta` highlighted by a blue oval. The repository also has a README.md file, which is partially visible in the screenshot.

File Name	File Type	Last Commit
NAalytics Update README.md		21fac9b on Apr 14 23 commits
Assemblies of putative SARS-CoV2...	Add files via upload	4 months ago
Figure1Figure2_032321.fasta	FASTA	5 months ago
README.md	Update README.md	4 months ago

**README.md**

## Assemblies-of-putative-SARS-CoV2-spike-encoding-mRNA-sequences-for-vaccines-BNT-162b2-and-mRNA-1273

##version 0.21Beta 04/14/21: (updates intended to (i) clarify the clinical and research importance of sequence information and strand topology measurements, and (ii) clarify that the mRNA sequence is not a recipe to produce vaccine)##

Dae-Eun Jeong, Matthew McCoy, Karen Artiles, Orkan Ilbay, Andrew Fire\*, Kari Nadeau, Helen Park, Brooke Betts, Scott Boyd, Ramona Hoh, and Massa Shoura\*

**About**

RNA vaccines have become a key tool in moving forward through the challenges raised both in the current pandemic and in numerous other public health and medical challenges. With the rollout of vaccines for COVID-19, these synthetic mRNAs have become broadly distributed RNA species in numerous human populations. Despite their ubiquity, sequences ...

**Releases**

No releases published

**Packages**

No packages published

<https://github.com/NAalytics/Assemblies-of-putative-SARS-CoV2-spike-encoding-mRNA-sequences-for-vaccines-BNT-162b2-and-mRNA-1273>

# Pfizer Sequence

```
1 >Figure1_032321_Spike-encoding_contig_assembled_from_BioNTech/Pfizer_BNT-162b2_vaccine
2 GAGAAATAAGTATTCTTCTGGTCCCAACAGACTCAGAGAGAACCCGACCACTGTTCTGTTCTGGTGTCTGCTGCC
3 TCTGGTGTCCAGCCAGTGTGAACCTGACCAACAGCAACAGCTGCTCCAGCTACACCAACAGCTTACAGAGCGC
4 TGACTACCCCAAGGTTTCAGATCCAGCGTGTGACTCTACCCAGGACCTGTTCTGCTTCTTTCAGCAACGTG
5 ACCTGTTCCAGCCATCACGTGTCGGCCACCAATGGCAAGAGATTGACAAACCCGCTGCTCTCAACGACCG
6 GGTGACTTTGCCAGCAGCGAAGTCCAACATCATCAGAGGCTGGATCTTCCGCCACCACTGGACAGCAAGACCCAGA
7 GCCTGTGATCGTGAACACGCCCAACGTTGTCATCAAGTGTGGAGTTCAGTTCGCAACGACCCCTTCTGGGG
8 GTCTACTACCAAGAACCAACAGAGCTGGATGAAAGCGAGTTCGGGTGTACAGCAGCGCCAAACACTGCACCTCGA
9 GTACTGTCCCGACTTCTCTGATGGACTGGAAGGCAAGCAGGCAACTTCAAGAACCTGCGGAGTTCTGTTTTAAGA
10 ACATGACGGCTACTTCAAGATCTACAGCAAGCACCCTTCAACCTCGTGGGGATCTGCTCAGGGCTTCTCTGCT
11 CTGGAACCTCTGGTGGATCTGCCATCGGACTCAACATCACCGGTTTACAGACTGCTGGCTCTGACAGAACTACT
12 GACACTGGCGATAGCAGCGGATGGACAGTGGTGGCCGCTACTACTGGTGGCTACTGCGACTGAGAACTTCC
13 TGTGAAGTACAACGAAACCGCACTACCGACCCGTGGATTGTGCTCTGGATCTCTGAGCAGACAAGGTGACCC
14 CTGAAGTCTTACCGTGGAAAAGGACTCTACAGCAGCAAGCACTTCCGGGTGAGCCACCGAATCATCGTGGGTT
15 CCCAATATCACCAATCTGTGCCCTTGGGAGGTGTTCAATGACCACGATTGCTCTGTGTAGCCGTGAACCGGTA
16 AGCGGATCAGCAATTCGGTGGCGACTACTCGTGTGTACAACCTCCGCAAGCTTACGACCTTCAAGTCTACGGCGG
17 TCCCTCAAGCTGAACGACTGTGCTTCAAAAGCTGTACGGCAGCAGCTGTGATCTCGGGGAGATGAAGTGGCGGA
18 GATGGCCCTGGACAGCAGGCAAGATCGGCACTACAACCTACAAGCTGCGGCAAGCTTACCGGCTGTGATGCTCT
19 GGACAGCAACCACTGGACTCAAGTGGCGGCACTCAATTACTGTACCGGCTGTCTCGGAAGTCAACTGGAAG
20 CCTCTGAGCGGGACATCTCACAGAGATCTACGGCCGAGCACCCTTGAACGGCTGGAGGCTTCACTGCTGAA
21 CTTCCCACTGACTCTACGGCTTCAAGCCACAATGGCTGGCTATGACCTACAGAGTGTGGTCTGACTGCTCG
22 AACTGCTGATGCTCCCTGACAGAGTGGGGCTTGAAGAACCAACTCTGTAAGAACAAATGGTGAACCTTCAAC
23 TTTCAACGGCTGACCGGCAACCGGCTGTGACAGAGAGAACAAAGAAATCTTCCGCTTCAAGCAATTTGGCGGAACT
24 CGCGATACACAGAGCGCTTGAAGTCCCAACAGCACTGAAATCTGGACATCACCTTGCAGCTTCCGGGAGTGT
25 CTGATCACCCCTGGACCAACACAGCAATCAGGTGGAGTGTGACAGGACGTGAACGTGACCGAAGTGGCGG
26 GCAATCAGCGGATCAGCTGACACTACATGGGGGTGACTCAACCGGCAAGTGTGTTTACAGCAGAGCGGCTG
27 TCTGATCGGAGCGAGCAGTGAACAAATAGCTACGAGTGGACATCCCACTCGGCGTGGAAATCGCGCAGCTACCA
28 CAGACAAACAGCCCTCGGAGGCGAGAGCGTGGCCAGCAGAGCATATTGCTACAAATGTCTTGGCGCCGAG
29 AACAGCTGGCTACTCCAACACTATCGCTATCCCAACCACTTCACTACAGCTGACCAAGATCTCGCTGCTG
30 GCTCAAGCAAGACAGCGTGGACTGACCATGTACATCTGGCGGATTCACCGAGTGTCCAACCTGCTGGTAGT
31 ACGGAGCTTCTGCAACCGAGTGAATGAGGCTGACAGGATGGCTGGCAAGGACAGAAACCCCAAGAGTGTGCT
32 GCCCAAGTGAAGCAGATCTACAAGACCCCTCTAACAAGACTTCCGGCGCTTCAATTCAGCCAGATTTCTGCCGATC
33 TAGCAAGCCAGCAAGCGGAGCTTCACTGAGGACTGCTGTCTCAACAAGTGAACACTGGCCAGCGGCTTCAACAGC
34 AGTATGGCGATTGTCTGGCGGACTTGGCCAGGGATCTGATTTGGCCAGAAGTTTAAAGGACTGACAGTGTGCT
35 CCTCTGCTGACCGATGAGATGATCGCCAGTACACATCTGCTGCTGGCGGCAACACAAAGCGGCTGGCAATTTGG
36 ACGAGCGCCGCTCTGAGATCCCTTTGCTATGAGATGGCTACCGGTTCAACGGCATCGAGTGAACCAAGATGTC
37 TGTACAGAACAGAACTGATGCCAACAGTTCACAGCCATCGGCAAGTCCAGGACGCTGACAGCAGCAGCA
38 AGCGCCCTGGAAAGTGCAGGAGTGTGCAACAGAAAGTCCAGGACACTGAACACCTGGTCAAGCAGCTGCTCTCAA
39 CTTGGCGCATCAGCTCTGCTGAAAGATATCTGAGCAGACTGGACCTCTGAGCCGAGGTCAGATGACAGAC
40 TGATACAGCGAGACTGCAGAGCTCCAGACATACGTGACCCAGCAGCTGACAGAGCGCGGAGATTAGAGCTTCC
41 AATCTGGCCGCAACAGATGCTGAGTGTGCTGGGCGAGAGCAAGAGTGGACTTTGGCGCAAGGCTTACCACT
42 GATGAGCTTCCCTCAGTCTGCCCTCACGGGTGGTGTCTGACAGTGAATATGCTGGCTCAAGGAAGAAATTTCA
43 CACCGCTCCAGCATCTGCCACGACGGCAAGGCCACTTCTAGAGAAGCGTGTGCTGTTCAACGGCAACCTTTGG
44 TTGCTGACAGCGGAATTTCTAGAGCCAGATCATCAACAGCAGCAACACTTCTGCTGCGCAACTGCGAGCTGT
45 GATGGCAATTTGAACAACCTGCTGACAGCTCTGACCGCCAGCTGGACAGCTTCAAAGGAACTGGACAAGTACT
46 TTAAGAACACCAACAGCCCGGAGCTGGACTGGCGGATACAGGGAATCAATGCGAGCTGCTGAACATCCAGAAAG
47 ATCCAGCGGCTGACAGGCTGGCAAGAATCTGAACGAGAGCTGACTGACCTGCTCAAGAACTGGGAAGTACAGCAG
48 CATCAAGTGGCCCTGTACATCTGGCTGGCTTATCGCGGACTGATTGCCATGTGATGGTCAACATCATGCTGTT
49 GCATGACAGCTGCTGATGCTGCTGAAGGGCTGTGTAGCTGTGGAGCTGCTGCAAGTTCGACGAGGCAATTTCTG
50 CCGCTGCTGAAGGGCTGAACTGCACTACACATGATGACTGAGCTGGTACTGATGACGCAAGTCTAGTGGCCCTT
51 TCCGCTCTGGGCTACCAGGCTCTCCCGACTCGGGTCCAGGATGCTCCCACTCCACTGCGCCACTCACCACT
52 CTGACTGTTCCAGACCTCCCAAGCAGCAAGTCAAGCTCAAAAGCTTAGCTAGCACACCTCCCGGCAAGCAAC
53 CAGTGAATTAACCTTAGCAATAAAGCAAGATTTAATAGCTATAACCCAGGGTGGTCAATTTCTGCTGACCGAC
54 ACCCTGGAGTAGCA
55
```

# Moderna Sequence

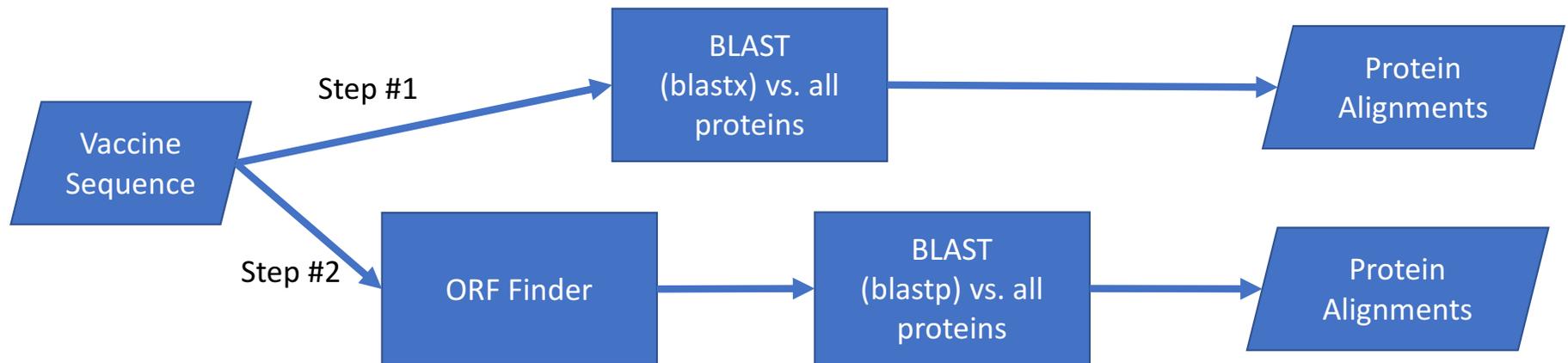
```
55 >Figure_2_32321_Spike-encoding_contig_assembled_from_Moderna_mRNA-1273_vaccine
56 GGGAAATAAGAGAGAAAGAGTAAGAAATAAAGACCCGGCGCCGACCACTGCTGTCTGGTGTCTGCTGCC
57 GCCCTGTGGAGCAGCGAGTGAACCTGACCCCGGACCCAGCTGCCACCGCACTCAACAGCTTACCCCGGG
58 GGGTCTACTACCCCAAGTGTTCGGAGCAGGCTGCTGACAGCAGCCAGGACCTGTTCTGCTTCTAGCAAG
59 GTGACTGTTCCAGCCATCACGTGAGCGGCCAACCGCACCAAGCGGTTGACAAACCCGCTGCTCTCAACGA
60 CGCGGTGACTTCCGACAGCAGAGAGCAACATCATCCGGGGCTGGATCTTCCGACCCACCTGGACAGCAAGACCC
61 AGAGCTGCTGATGGTAACGCCAACAGTGGTATCAAGTGTGGAGTCCAGTTCGCAACGACCCCTTCTG
62 GCGTGTACTACCAAGAACAAAGAGCTGGATGGAGGAGGTTCCGGGTACAGCAGCGCCAAACACTGCACCTT
63 GTCCCAAGGACTTCACTGGAGAGGGATCTACAGCAGCACTTCCGGGTGACAGCAGCGCCAAACACTGCACCT
64 CGATACGTGAGCAGCCCTTCTGATGGACTGGAGGCAAGCAGGCAACTTCAAGAACTGCGGGAGTCTGTGTTCA
65 AGAACACTGACGGCTACTCAAGATCTACAGCAAGCACCCAACTCAACCTGGTGGGATCTGCCAGGGCTTCTCA
66 GCCTGGAGCCCTGTGGACTGCCATCGGATCAACATCACCGGTTCAAGACTGCTGGCCCTGACCGGAGCTA
67 CTGACCCAGGCGACAGCAGCGGGTGGACAGCAGGCGGCTTACTACTGCTGGCTACTCGACGCCCGGACT
68 TCTGCTGAAGTACAAGAGAGCGCACTACCGACCGCTGGACTGCCCTGGACCTCTGAGCAGAGCAAGTGT
69 ACCCTGAAGAGTTCACCTGGAGAGGGATCTACAGCAGCAGCACTTCCGGGTGACGCCCAAGAGAGATCTGG
70 GTTCCCAACATCACCAAGCTGCTCCCTTGGGAGGTTTCAACGCCACCCGGTTCGCGAGGCTAGCTGGAAC
71 GGAAGGGATCAGCAACTGCTGGCCGACTACAGGCTGTGTAACAACGCGCCAGCTTCAACACTTCAAGTCTAC
72 GTGAGCCACCAAGCTGAACGCTGTGCTTACCAAGCTGTACCGCCAGCAGCTGTGATCGTGGCAGCAGAGTGG
73 GCGAGTGCACCCGGCGACAGCAAGTGGCAGCTACACTACAAGCTGCGCCAGGATTCACCGGCTGCTGATCG
74 CTGGAAGCAGCAACACTCGACAGCAAGTGGCGGCACTCAACTACTCTGACCGGCTTCCGAAAGACAACTG
75 AAGCCCTGAGCGGAGTACAGCAGCAGATCTCAAGCGGCTCCACCTTGAACGGCTGGAGGCTTCAACTG
76 CTACTCTCTGAGAGTACGGCTTCCAGCCCAACGCGGCTGGCTACAGCCCTACCGGTTGGTGTGAGCT
77 TGAGCTGCTGACCGCCAGCAGCGTGTGGCCCAAGAGAGCAACACTGCTGGAAGAACAAAGTGGGAACTTCA
78 AACTTCAAGGGCTTACCGGCAACCGGCTGTGACAGAGCAACAGAAATTTCTGCCCTTCAAGACTTCCGCGG
79 CATCGCCAGCACCAGCGCTGTGGGGATCCCAAGCCCTGGAGATCTGGACATCACCTTTCAGCTTCCGGCGG
80 TGAGGTGATCACCCAGGCAACCAACAGCAGCAACAGTGGCGGCTGTGATCAGGAGTGAAGTGCAGGAGTGGCC
81 TTGGCTATCCACCGGCAAGCTGACACCACTGCGGGTGTACAGCAGCCGAGCAAGTGAATTCAGAACCCGG
82 TTGCTGATCGGCGGAGCAGTGAACAGCAGTACAGTGGCAGTCCCATCGGCGGCACTGTGTCAGCTTACC
83 AGACAGCAGCAATTCACCCGGAGGCAAGGAGCTGGCCAGCAGAGCATCTCGCTACACTGAGCTGCGGCGCC
84 GAGAACGGTGGCTCAGCAACAGCAGTCCGCTCCCAACCACTTCACTACAGCTGACACCGGAGATCTGCTCC
85 CGTAGCATGACCAAGCAGCGGACTGACATGTACATCTGCGCGACAGCAGGTCAGGAGCAACCTGCTGCTGC
86 AGTAGCGGCTTCTGCAACCGACTGAACCGGCTTCCAGGCTGCGGCGGAGTGGAGTGGAGTGGAGGTTG
87 TCGCCAGTGAAGCAGTCTACAAGACCCCTCCATCAAGGACTTCCGGCGCTTCAACTCAGCAGATCTGCCCGA
88 CCCAGCAAGCCAGCAAGCGGAGCTTACAGGAGCTGCTGTTCAACAAGTGAACCTAGCCAGCAGCCGCTTAC
89 AGCAGTACCGGCTGCTGCGGAGATGAGCGCCGCGGCACTGATCTGCGCCAGAAGTTCAACGGCTGACCGT
90 CCTCCCTGCTGACAGCAGATGATCGCCAGTACACAGCGCCCTTTAGCGCCACTCAACCGGCGGAGACTT
91 CGGCGCTGGAGCGCTCTGAGATCCCTTCCGATGAGATGGCTACCGGTTCAACGGCTGCGGCTGACCGAGAAC
92 TGCTGACAGAACAGAAAGTGTGCGCAACAGTTCACAGCGCATCGGCAAGTCCAGGACAGCTGAGCAGCACCC
93 GCTAGCGCCCTGGCAAGCTGACAGAGCTGGTGAACAGAACGCCAGCCCTGAACACTGTTGAAGCAGCTGACAG
94 CAACTTGGCGGCTCAGCAGGCTGCTGAAGCAGCTCTGAGCGGCTGGACCTCCCGAGGCGGAGTGCAGATGAC
95 GCGTATCACTGGCGGCTGACAGCCTGACAGCTACGTGACCCAGCAGCTGATCGGGCCCGGAGATTCGGCCAGC
96 CCACACTGGCCGCAAGTGAAGTGGCTGCTGGCCAGAGCAAGCGGTTGACTTCCGGAGGAGGCTACCA
97 CCTGATGAGTTCCTCCAGAGCGCACCCAGGAGTGGTCTCTGCACTGACTGCTGCCCGCAGAGAAAGAACT
98 TCAACCGCCCGGCACTCTGCGCACGCGCAAGGCCACTTCCCGGGAGGGTGTGTTGAGCAAGCAGCCAC
99 TGGTGTGAGCCAGGGAATCTTCAAGCCCGAGTATCAACAGCCAGTATCAACAGCAGCAACACTTCTGAGCGG
100 GGTGATCGGATGTGAACAACCGTGTACGATCCCTGAGCCGAGCTGGAACAGTTCGAAGAGAGCTGGACAGT
101 ACTCAAGAACTCACAGCCGCGCTGACCTGGCGGACTGAGCGGACTCAAGCGGCTGAGTCACTCAAGAA
102 GAGATGATCGGCTGACAGGTTGGCAAGCAGTGAACAGGAGCTGATGACTGCTGAGGAGCTGGCAGATGAGCA
103 GTACATCAAGTGGCTGATCATCTGGCTGGCTTATCGCGGCTGATGCTGATGATGATGATGATGATGATGATG
104 GCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG
105 GAGCGCTGTTGAAGGGCTGAGCTGACTACACTGATAAATAGCTGGAGCTGCGGCTGATGCTTCCGCGCTT
106 GGGCTCCCCAGCCCTTCTCCCTTCTGCAACCGCTACCCCGTGGTCTTGAATAAAGTCTGAGTGGCGGCAAAA
107 AAAA
108
```

[https://github.com/NAalytics/Assemblies-of-putative-SARS-CoV2-spike-encoding-mRNA-sequences-for-vaccines-BNT-162b2-and-mRNA-1273/blob/main/Figure1Figure2\\_032321.fasta](https://github.com/NAalytics/Assemblies-of-putative-SARS-CoV2-spike-encoding-mRNA-sequences-for-vaccines-BNT-162b2-and-mRNA-1273/blob/main/Figure1Figure2_032321.fasta)



# Sequence Analysis Workflow

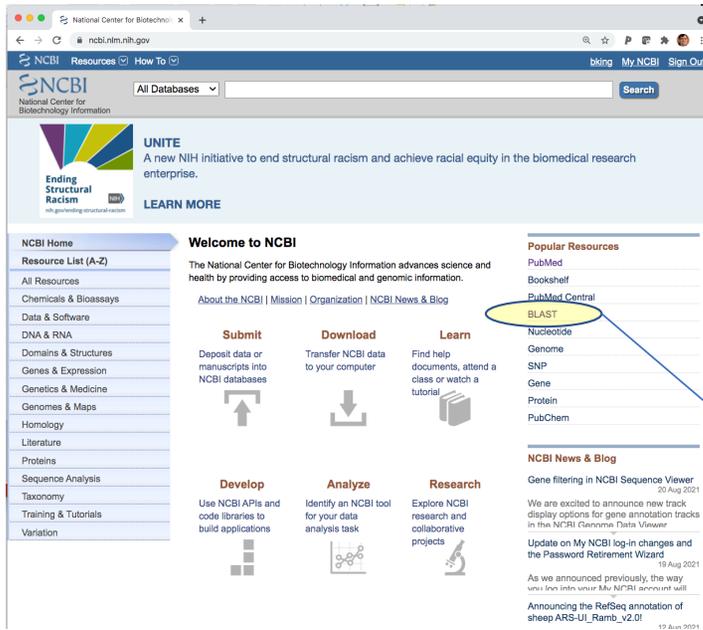
What viral protein(s) could be encoded by the sequences?



How similar are the sequences?

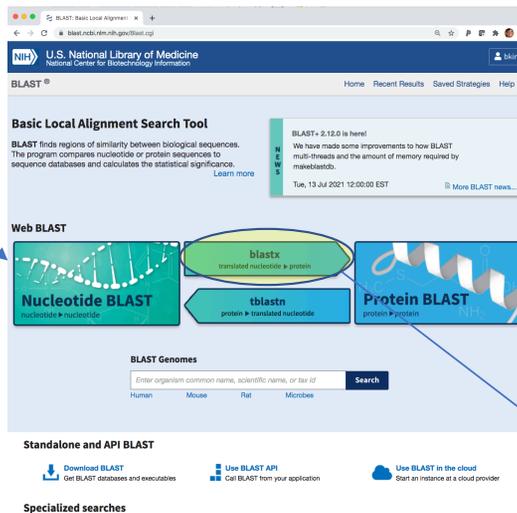


# BLASTX vs. All Proteins @ NCBI



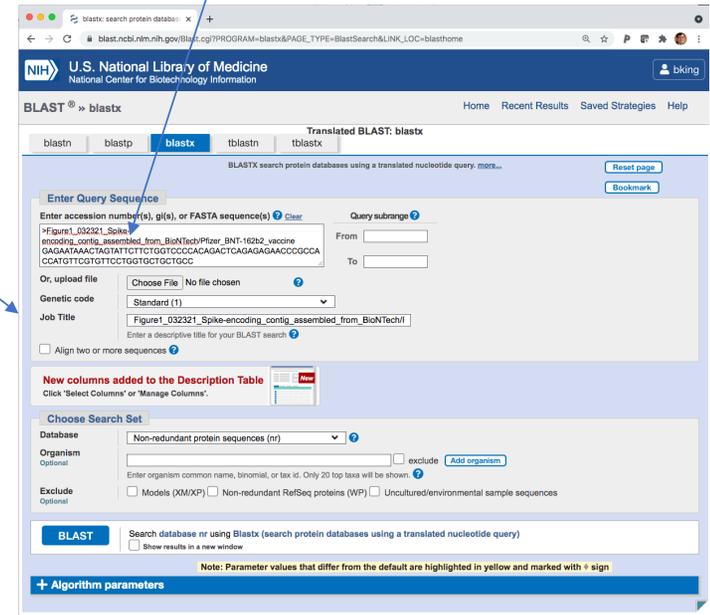
The screenshot shows the NCBI homepage with a navigation menu on the left. The 'Proteins' link is highlighted in red. The main content area features a 'Welcome to NCBI' message and several action buttons: 'Submit', 'Download', 'Learn', 'Develop', 'Analyze', and 'Research'. A yellow oval highlights the 'BLAST' link in the 'Popular Resources' section.

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



The screenshot shows the BLAST interface with a central diagram of the 'Web BLAST' tools. The diagram includes 'Nucleotide BLAST', 'blastx translated nucleotide > protein', 'tblastn protein > translated nucleotide', and 'Protein BLAST protein > protein'. A yellow oval highlights the 'blastx' tool, and a blue arrow points from it to the right-hand screenshot.

Paste the Pfizer sequence here



The screenshot shows the BLASTX search results page. The 'blastx' tool is selected. The 'Enter Query Sequence' section contains the Pfizer sequence: `>Figure1_032321_Spike-encoding_contig_assembled_from_BioNTech/Pfizer_BNT-162b2_vaccine (GAGATATAAAGTATATCTCTCTGGTCCGACAGAGCTGAGAGAGAAKCCCGCCA CCATGTCGTCTCTCGTGGCTGCTGCC`. The 'Database' is set to 'Non-redundant protein sequences (nr)'. The 'BLAST' button is highlighted in blue.

Paste the Pfizer sequence here

# ORF Finder @ NCBI

ORF Finder

Open Reading Frame Viewer

Sequence

ORFs found: 20 Genetic code: 1 Start codon: ATG only

Label	Strand	Frame	Start	Stop	Length (nt)
ORF1	+	1	55	3876	3822
ORF10	-	1	3764	3381	384
ORF11	-	1	437	165	273
ORF14	-	1	1814	1578	240
ORF11	-	1	2972	2783	180
ORF12	-	1	2585	2418	168
ORF7	+	2	3896	4054	159
ORF16	-	1	644	486	159
ORF19	-	2	3283	3107	147
ORF5	-	1	1000	1016	117

ORF1 (1273 aa) Display ORF as... Mark

BLAST

BLAST Database: Non-redundant protein sequences (nr)

Select "ORF1"

U.S. National Library of Medicine  
National Center for Biotechnology Information

BLAST® » blastp suite

Standard Protein BLAST

BLAST programs search protein databases using a protein query, most...

Enter Query Sequence

Enter accession number(s), gi(s), or FASTA sequence(s) Clear Query subrange

From: To:

Or, upload file Choose File No file chosen

Job Title Enter a descriptive title for your BLAST search

Align two or more sequences

Choose Search Set

Database Non-redundant protein sequences (nr)

Organism Optional Enter organism name or id-completions will be suggested Add organism

Exclude Optional Models (XM/XP) Non-redundant

Program Selection

Algorithm Quick BLAST (Accelerated protein-protein BLAST) PSI-BLAST (Position-Specific Iterative) PHI-BLAST (Pattern Hit Initiated) DELTA-BLAST (Domain Enhanced) Choose a BLAST algorithm

BLAST Search database nr using Blastp Show results in a new window

+ Algorithm parameters

BLAST® » blastp suite » results for RID-JGCGERE016

Job Title Protein Sequence

RID JGCGERE016 search expires on 08-26 21:11 pm Download All

Program BLASTP Citation

Database nr See details

Query ID I01Query\_81151

Description I01ORF1:55:3876 unnamed protein product

Molecule type amino acid

Query Length 1273

Other reports Distance tree of results Multiple alignment MSA viewer

Filter Results

Organism only top 20 will appear

Type common name, binomial, taxid or group name

+ Add organism

Percent Identity E value Query Coverage

Filter Reset

Descriptions Graphic Summary Alignments Taxonomy

Sequences producing significant alignments

Download Select columns Show 100

select all 100 sequences selected

Description	Scientific Name	Max Total Score	Query Cover	E	Phi	Acc.
SARS-CoV-2 spike in infection state (Severe acute respiratory syndrome coronavirus 2)	Severe acute rs...	2638	2638	100.0%	1283	GQ995621.1
S system of SARS-CoV-2 in complex bound with 4MB (Severe acute respiratory syndrome coronavirus 2)	Severe acute rs...	2637	2637	100.0%	1283	JGCG_A
Chain A, Spike glycoprotein (Severe acute respiratory syndrome coronavirus 2)	Severe acute rs...	2635	2635	100.0%	1283	JGCG_A
Spike glycoprotein (Severe acute respiratory syndrome coronavirus 2)	Severe acute rs...	2633	2633	100.0%	1282	NC045335.1
Distinct conformational states of SARS-CoV-2 spike protein (Severe acute respiratory syndrome coronavirus 2)	Severe acute rs...	2632	2632	100.0%	1310	GQ995621.1
Spike glycoprotein (Severe acute respiratory syndrome coronavirus 2)	Severe acute rs...	2631	2631	100.0%	1273	QJ25852.1
Structure of SARS-CoV-2 3D-2P full-length envelope spike trimer (3D structure) (Severe acute respiratory syndrome coronavirus 2)	Severe acute rs...	2631	2631	100.0%	1273	JJL_A
Spike glycoprotein (Severe acute respiratory syndrome coronavirus 2)	Severe acute rs...	2630	2630	100.0%	1273	YP_050724390.1
Spike glycoprotein (Severe acute respiratory syndrome coronavirus 2)	Severe acute rs...	2630	2630	100.0%	1273	QJ25852.1
Spike glycoprotein (Severe acute respiratory syndrome coronavirus 2)	Severe acute rs...	2630	2630	100.0%	1273	QJ25852.1
Spike glycoprotein (Severe acute respiratory syndrome coronavirus 2)	Severe acute rs...	2630	2630	100.0%	1273	QJ25852.1
Spike glycoprotein (Severe acute respiratory syndrome coronavirus 2)	Severe acute rs...	2630	2630	100.0%	1273	QJ25852.1
Spike glycoprotein (Severe acute respiratory syndrome coronavirus 2)	Severe acute rs...	2630	2630	100.0%	1273	QJ25852.1
Spike glycoprotein (Severe acute respiratory syndrome coronavirus 2)	Severe acute rs...	2630	2630	100.0%	1273	QJ25852.1
Spike glycoprotein (Severe acute respiratory syndrome coronavirus 2)	Severe acute rs...	2630	2630	100.0%	1273	QJ25852.1

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

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

Download GenPept Graphics Next Previous Descriptions

### 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
Sbjct 1		MFVFLVLLPLVSSQCVNLTTRTQLPPAYTNSFTRGVYYPDKVFRSSVLHSTQDLFLPFFS			60
Query 61		NVTWFHAIHVSNGTKRFDNPFVLPFNDGVYFASTEKSNIIIRGWIFGTTLDSTKQSLIV			120
Sbjct 61		NVTWFHAIHVSNGTKRFDNPFVLPFNDGVYFASTEKSNIIIRGWIFGTTLDSTKQSLIV			120
Query 121		NNATNVVIVKVEFQFCNDPFLGVYHKNKNSWMESEFRVYSSANNCTFEYVSQPFLMDLE			180
Sbjct 121		NNATNVVIVKVEFQFCNDPFLGVYHKNKNSWMESEFRVYSSANNCTFEYVSQPFLMDLE			180
Query 181		GKQGNFKNLREPVFKNIDGYFKIYSKHTPINLVRDLPQGFSALEPLVDLPIGINITRFQT			240
Sbjct 181		GKQGNFKNLREPVFKNIDGYFKIYSKHTPINLVRDLPQGFSALEPLVDLPIGINITRFQT			240
Query 241		LLALHRSYLTSGDSSSGWTAGAAAAYVGYLQPRTEFLKYNENGTITDAVDCALDPLSETK			300
Sbjct 241		LLALHRSYLTSGDSSSGWTAGAAAAYVGYLQPRTEFLKYNENGTITDAVDCALDPLSETK			300
Query 301		CTLKSFTVEKGIYQTSNFRVQPTESIVRFPNITNLCPPGEVFNATRFASVYAWNRKRISN			360
Sbjct 301		CTLKSFTVEKGIYQTSNFRVQPTESIVRFPNITNLCPPGEVFNATRFASVYAWNRKRISN			360
Query 361		CVADYSVLVNSASFSTFKCYGVSPTKLNDLCFTNVDYADSFVIRGDEVQRQIAPGQTGKIAD			420
Sbjct 361		CVADYSVLVNSASFSTFKCYGVSPTKLNDLCFTNVDYADSFVIRGDEVQRQIAPGQTGKIAD			420
Query 421		YNYKLPDDFTGCVIAWNSNLDKSVGGNYNYLRLFRKSNLKPFFERDISTEYIQAQSTPC			480
Sbjct 421		YNYKLPDDFTGCVIAWNSNLDKSVGGNYNYLRLFRKSNLKPFFERDISTEYIQAQSTPC			480
Query 481		NGVEGFNCYFPLQSYGFPQTNGVGYQPYRVVLSFELLHAPATVCGPKKSTNLVKNKCVN			540
Sbjct 481		NGVEGFNCYFPLQSYGFPQTNGVGYQPYRVVLSFELLHAPATVCGPKKSTNLVKNKCVN			540
Query 541		FNFNGLTGTGVLTSNKKFLPFQGFGRDIADTTDAVRDPQTLEILDITPCSFGGVSVITP			600
Sbjct 541		FNFNGLTGTGVLTSNKKFLPFQGFGRDIADTTDAVRDPQTLEILDITPCSFGGVSVITP			600
Query 601		GTNTSNQVAVLYQDVNCTEVPVAIHADQLTPTWRVYSTGNSVVFQTRAGCLIGAEHVNSY			660
Sbjct 601		GTNTSNQVAVLYQDVNCTEVPVAIHADQLTPTWRVYSTGNSVVFQTRAGCLIGAEHVNSY			660
Query 661		ECDIPIGAGICASYQTQTSNPRRARSVASQSI IAYTMSLGAENSVAYSNNISIAIPTNFTI			720
Sbjct 661		ECDIPIGAGICASYQTQTSNPRRARSVASQSI IAYTMSLGAENSVAYSNNISIAIPTNFTI			720
Query 721		SVTTEILFVSMTKTSVDCTMYICGDSSTECNSLLQYGSFCTQLNRALTGIAVEQDKNTQE			780
Sbjct 721		SVTTEILFVSMTKTSVDCTMYICGDSSTECNSLLQYGSFCTQLNRALTGIAVEQDKNTQE			780
Query 781		VFAQVKQIYKTPPIKDFGGFNFSQILPDPSPKSKRSFIEDLLFNKVTADAGFIKQYGDC			840
Sbjct 781		VFAQVKQIYKTPPIKDFGGFNFSQILPDPSPKSKRSFIEDLLFNKVTADAGFIKQYGDC			840

**Related Information**  
[Identical Proteins](#) - Identical proteins to 6ZOW\_A



Paste the Pfizer and Moderna sequences here

# Clustal Omega @ EBI

The screenshot shows the Clustal Omega web interface at <https://www.ebi.ac.uk/Tools/msa/clustalo/>. The page is titled "Clustal Omega" and "Multiple Sequence Alignment". It features a navigation bar with "EMBL-EBI", "Services", "Research", "Training", "Industry", and "About us". Below the navigation bar, there are tabs for "Input form", "Web services", "Help & Documentation", and "Bioinformatics Tools FAQ". The main content area is titled "Multiple Sequence Alignment" and includes an "Important note" about file size limits. The "STEP 1 - Enter your input sequences" section contains a text area where a set of DNA sequences has been pasted, including Pfizer BNT-162b2 and Moderna mRNA-1273 sequences. The "STEP 2 - Set your parameters" section shows the "OUTPUT FORMAT" set to "ClustalW with character counts". The "STEP 3 - Submit your job" section has a "Submit" button. At the bottom, there is a footer with "EMBL-EBI" logo and links to "Services", "Research", "Training", "Industry", and "About EMBL-EBI".

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

The screenshot shows the Clustal Omega web interface displaying the results of a sequence alignment job. The page title is "Results for job clustalo-I20210825-142414-0576-69617999-p1m". The navigation bar is the same as in the previous screenshot. Below the navigation bar, there are tabs for "Alignments", "Result Summary", "Phylogenetic Tree", "Results Viewers", and "Submission Details". The "Alignments" tab is selected, showing a "Download Alignment File" link and a "Clustal o(1.2.4) multiple sequence alignment" section. The alignment results are displayed as a text-based Clustal format, showing the alignment of the Pfizer BNT-162b2 and Moderna mRNA-1273 sequences. The alignment is shown in a grid format with sequence identifiers and positions on the left, and the aligned sequences on the right. The sequences are aligned with gaps (represented by dashes) to show the conserved regions. The alignment is shown for the first 1200 positions of the sequences.

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

GAGAAATAAACTAGTATTCTTCTGGTCCCCACAGACTCAGAGAGAACCOCGCCACTGTTTCGGTGTTCCTGGTGGTGTGCTGCCCTTGGTGTCCA  
 CCCAGTGTGTAACCTGACCAACAGAACACAGCTGCCCTCCAGCTACACCAACAGCTTACCAGAGGCGTGTACTACCCCGACAAGGTGTT  
 CAGATCCAGCGTGTGCATCTACCCAGGACCTGTTCCCTGCCCTTCTTCCAGCAACGTGACCTGGTCCACGCCATCCACGTGTCGGGACCC  
 AATGGCACCAGAGATTGCACAACCCCGTGTGCCCTTCAACGACGGGGTGTACTTTGCCAGCAGGAGAAGTCCAACATCATCAGAGGCT  
 SGTCTTCGGCACCACACTGGACAGCAAGACCAGAGCTGTGTGATGTAACAACGCCACCAACGTGGTTCATCAAAAGTGTGGAGTTCCA  
 GTTCTGCAACGACCCCTTCTGGGGGTCTACTACCACAAGAACAAAGAGCTGGATGGAAGCGAGTTCCGGGTGTACAGCAGCGCCAAC  
 AACTGCACCTTCCAGTACGTGTCCAGCCCTTCCCTGATGGACCTGGAAGGCAAGCAGGGCAACTCAAGAACCTCGCGAGTTTCGTGTTA  
 AGAACATCGACGGCTACTTCAAGATCTACAGCAAGCACACCCCTATCAACCTCGTCCGGGATCTGCCTCAGGGCTTCTCTGCTTGAACCC  
 CCTGGTGGATCTGCCATCGGCATCAACTCACCCGGTTTCCAGACACTGCTGGCCCTGCACAGAACTACTGACACCTGGCGATAGCAGC  
 AGCGGATGGCAGCTGTGTCCCGGCTTACTATGTGGGTACTCTGCACCTTAGAACCTTCTTGTGAAGTACAACGAGAAGCGGCACCATCA  
 CCGACCGGAGTGTGCTGTGATCTCTGAGCGAGACAAGTGCACCTGAAGTCTTCCAGCTTCCAGGAAAGGGCATCTACCCAGACCAG  
 CAACTTCCGGGTGCAGCCACCGAATCCATCGTGGGTTCGCCAATACCAATCTGTGCCCTTCCGGGAGGTGTCAATGCCACCAGA  
 TCCAGCTGTGTACCCCTGGAACCCGGAAGCGATCAGCAATTCGCTGGCCAGCTACTCGTGTCTTACAACCTCCGCCACCTTCAGCACC  
 TTAAGTGTCTACGGCTGTCCCAACCGACTGAAAGCGCTGTGCTTCAACAACGACTGACCCGACAGCTTCTGTATCCGGGGATGAAGT  
 CGCGCAGATTGCCCTTGGACAGACAGCAAGTTCGCCGACTCAACTTCAAGCTTCCCGCAGCAGCTTCCACCGCTGTGTGATTCCTTGGAA  
 AGCAACAACCTTGGACTTCAAAAGTCCGGCGCACTACAATTACCTGTACCCGGCTTCCGGAAGTCCAATCTGAAGCCCTTCGAGCGGGA  
 TCTCCACCGAGATCTATCAGCGCGGACACCCCTTGTACCGCGTGGAAAGCTTCAACTGCTACTTCCACCTGCAGTCTACGGCTTTC  
 GCCCACAATGGCTGGGTATCAGCCCTACAGATGGTGGTGTGAGCTTCAACTGTGTGATGCCCTGCCACAGTGTGGCGCCCTAAG  
 AALAGCAACAATCTCTGTGAAGAACAAATCGTGAACCTCAACTTCAACCGCTGACCGGCAACCGGCTCTGACAGAGCAACAAGT  
 TCTGTCCATTCGACGAGTTTGGCGGGATATCGCGATACCCAGACCGGCTTAGAGATCCCAAGCACTGGAATCTCGGATCACCCT  
 TCGAGCTTCGGCGGAGTGTCTGTGATCACCCTGGACCAACACCAGCAATCAGTGGCAGTGTGTACCAGGACTGAACCTGACCAGAA  
 GTGCCCTTGGCCATTCACCGGATCAGCTGACACCTACATGGCGGCTGTACTCCAGCGCAGCAATGTGTTCAGACCAAGCGGCTGT  
 TGATCGGAGCGGACGCTGAACAATAGCTACAGTGGCAGATCCCAATCGCGCTGGAATCTCGCGCAGCTACCAGACACAGACAACAG  
 CCCTCGGAGCCAGAGGCTGGCCAGCAGAGCATCATTGGCCACAAATGCTCTTGGCGCCGAGAACAGCGTGGCTACTTCCCAACA  
 TCTATCGCTATCCCAACACTTACCATCAGCGTGACACAGAGATCCTGCTGTGCTATGACCAAGACAGCGTGGACTGCACATGT  
 ACATCTGGCGGATCCCAAGTGTCTCAACCTGCTGCTGACAGTACGGCAGTGAATAGAGCCCTGACAGGGATCGC  
 CGTGGAAACAGGACAAGAACACCCAAAGAGTGTTCGCCCAAGTGAAGCAGATCTACAAGACCCCTTCTATCAAGGACTTCGGCGGCTTCAAT  
 TTCACCGAGTCTGCGCCATCTAGCAAGCCAGCAAGCGGAGCTTCACTGAGGACTGCTGTTCACAAAGTGCACCTGGCCGACCGCG  
 GCTTCATCAAGCAGTATGGCAGTGTCTGGCGGATGTCCCGCAGGGATGATTTGGCCGAGAAAGTTTAAAGCAGTGCAGTGTGCTG  
 TCTCTGCTGACCGATGAGATGATCGCCAGTACACATCTGCCCTGCTGGCCGCAACATCAAGCGGCTGGACATTTGGACAGGGCGC  
 GCTCTGCAGATCCCTTTGCTATCGCAGTGGCTACCGGTTCAACGGCATCGGAGTGAACCAAGATGTGCTGACGAGAACCAGAAAGTGA  
 TCGCAACAGGATTCACAGCGCCATTCGGCAAGATCCAGGACAGCTTGAAGCAGCAGCAAGCGCCCTGGGAAAGCTGCAGGAGCTGGTCAA  
 CCAGAAATGCCAGGCACTGAACACCTTGGTCAAGCAGCTTCTTCCAATTCGGCGCCATCAGCTTCTGTGCTGAACGATATCTGAGCAGA  
 CTGGACCTTCTGAGCGGAGGTGCAGATGCAGAGACTGATCACAGCCAGACTGCAGAGCTCCAGACATAGCTGACCCAGCAGCTGATCA  
 GAGCGCGGAGATTAGCGCTTGCACATCTGGCCGCCACCAAGATGTCTGAGTGTGTGCTGGGCCAGCAGAGAGTGGACTTTTGGCG  
 CAAAGGCTTACCACCTGATGAGCTTCCCTCAGCTGCCCTCAGGCGTGGTGTTCGACAGTGCACATATGTCGCCCTCAAGAGAAGAA  
 TTCACCCCGCTCCAGCCATCTGCCACGACGCAAGGCCACTTCTTAGAAGAGCGTGTTCGTGTCACAGCGCCACCAATGGTTCGTGA  
 CACAGCGGAATCTTACGAGCCCGAGATCATCACCACCGCAACACCTTCTGTCTTGGCAACTGCGACGCTGTGTATCGGCATTTGTGAACA  
 TACCCTGTACGACCTTCTGCACCCGAGCTGGACAGCTTCAAGAGCAACTGGACAAAGTACTTTAAGAACCACAAAGCCCGACGTGGAC  
 CTGGCGATATCAGCGGAATCAATGCCAGCGCTGTGAACATCCAGAAAGAGATCGCCGGCTGAACGAGTGGCCAGAAATCTGAACGAGA  
 GCCTGATCGACCTGCAAGAAGTGGGAAGTACGAGCAGTACAACAAGTGGCCCTGGTACATCTGGCTGGCTTATTCGCCGGACTGATTCG  
 CATCTGTGATGATCAACATCATGCTGTGTGATGACAGCTGTGTAGCTGCCCTGAAGGGCTGTGTAGCTGTGGCAGCTGTGCAAGTTC  
 GACGAGGAGATTTCTGAGCCGTGCTGAAGGGCTGAAACTGCACCTACACA **AGATGA**CTCGAGCTGGTACTGCATGCACGCAATGCTAGCT  
 GCCCTTTCCTGCTTGGTACCCGAGTCTCCCGACCTCGGGTCCAGGATAGCTCCCACTTCCACCTGCCACCTCCACCTCAGCAGCTCAGCAGCTCTG  
 TACTTCCAGACCTCCCAAGCAGCAGCAATGCAGCTCAAAAGCTTAAAGTGTAGCTGACCCACCCCGGAAACCGAGCTGAATTAACCTTT  
 AGCAATAAACGAAAGTTTAACTTAAGCTACTAATCAACCCAGGGTGTGTAAATTTCTGTCGACGCCACCCCTGAGGATAGCA

**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.**

GGGAAATAAGAGAGAAAAGAAGAGTAAAGAAGAAATAAAGACCCCGCCGCCACCATGTTTCGGTGTTCCTGGTGGTGTGCTGCCCTTGGTGA  
 CGAGCCAGTGCCTGAACTGACCAACCCGAGCCAGCTGCCACAGGCTACACCAACAGCTTACCCCGGGGCGTCTACTACCCCGACAAGGT  
 GTTCCGGAGAGCGTCTGCACAGCACCAGGACCTTCTCTGCCCTTCTTCCAGCAACGTGACCTGGTCCACGCCATCCACGTGAGCGCC  
 ACCAACCGGACCAAGCGGTTCGACAAACCCCGTGTGCCCTTCAACGACGGCGTGTACTTCCGACACCAGGAGAAGCAACATCATCCGGG  
 GCTGGATCTTCCGACCCACCTGGACAGCAAGACCCAGAGCTGTGTGATCGTGAATAACGCCACCAACGTTGGTGAATCAAGGTGTGGAGTT  
 CCAGTTCGCAACGACCCCTTCTGGCGGTGTACTACCAAGAAACAACAAGAGCTGGATGGAGAGCGAGTTCGGGGTGTACAGCAGCGCC  
 AACACTGCACCTTCCAGTACGTGAGCCAGCCCTTCTGTGATGGACCTGGAGGGCAAGCAGGGCAACTTCAAGAACCCTCGCGGAGTTCGTGT  
 TCAAGAATCATCGACGGCTACTTCAAGATCTACAGCAAGCACCCCAATCAACCTGGTGGGGATCTGCCCCAGGGCTTCTCAGCCCTGGA  
 GCCCTTGGTGGACTGCCATCGGCATCAACATCACCCGGTTCAGACCCCTGCTGGCCCTGCACCGGAGCTACTGACCCAGGGCAGACG  
 AGCAGCGGTGGACAGCAGCGCGGCTGTACTACTGTGGGTACTGTGACGCCCGGACCTTCTGTGAAAGTACAACGAGAAGCGGACCA  
 TCACCAGCCCGTGGACTGCCCTTGGACCCCTTGCAGCGAGCAAGTGCACCTTGAAGAGCTTACCCCTGGAGAAGGGCATCTACCAGAC  
 CAGCAACTTCCGGGTGCAGCCACCAGAGCATCTGTGGGTTCGCCAATCAACCACTGTGCCCTTCCGGCAGGTGTTCACAGCCACC  
 CGGTTCGACGCTGTACGCTGGAACCGGAAAGCGGATCAGCAACTCGTGGCCGACTACAGCTGTGTGTACACAGCGCCAGCTTACGCA  
 CCTTCAAGTGTACGGCTGAGCCCAACCAAGCTGAACACCTTGTGCTTCAACAACGTGTACGCCGACAGCTTCTGTATTCGTGGCGACGA  
 GGTGGCGCAGATCGCACCCGGCCAGACAGGCAAGATCGCGACTCAACACTCAAGCTGCCCGACAGCTTACCAGGCTGGTGTATCGCTGG  
 AACAGCAACACCTTGCAGAGCAAGTGGCGGCAACTACAACACTTACCTGTACCGGCTGTTCGGGAAGAGCAACCTGAAGCCCTTGCAGCGG  
 ACATCAGCAGGAGATCTACCAAGCGGCTCCACCCCTTGAACCGGCTGGAGGGCTTCAACTGCTACTTCCCTTGCAGAGCTACGGCT  
 CCAGCCCAACACCGGCTGGGTACCAAGCCCTACCGGCTGGTGGTGTGAGCTTCGAGCTGCAGCCCGGACCCAGCTGTGTGGCCCC  
 AAGAGAGCACAACCTGGTGAAGAACAGTGGTGAACCTTCAACGGCTTACCGGCAACCGGCTGTGACCGGACCGGCTGTGACCGAGCAACA  
 AATTCCTGCCCTTTCAGCAGTTCGGCCGGGACTTCGCCAGACCCAGCCAGCTTGGCGGATCCCAAGCCCTTGGAGATCTTGGACATCAC  
 CCTTGCAGCTTCCGGCGGTGAGCGTGTACCCAGCCACCAACACCAGCAACCGGTTGGCCGTGTGTACAGGAGCTGAACTGCACC  
 GAGTTCGCCCTTGGCCATTCACCGCCAGCAGCTGACACCCACTTGGCGGCTTACAGCAGCGGCAAGCTGTTCAGACCCCGGCGGCT  
 GCCTGATCGCGCGGAGCAGCTGAACAACAGCTACAGTGCAGCATCCCAATCGCGCGCCGATCTGTGGCAGCTACACAGCCAGACCA  
 TTCACCCCGGAGGGCAAGAGGCTGGCCAGCCAGAGCATCATCGCTACACCATGAGCTTGGCGCCGAGAACAGCGTGGCTTACAGCAAC  
 AACAGCATGCCATCCCAACACTTCAACATCAGCGTGACCCAGAGATTTCGCCGTGAGCATGACCAAGACCGGCTGGACTGCACCA  
 TGTACTTCTGGCGCAGCAGCCAGTGCAGCAACCTTGTGCTGTCAGTACGGCAGCTTTCGACCCAGCTGAACCGGCGCTTACCGGCA  
 CGCGTGGAGCAGGACAAGAACACCCAGGAGTGTTCGCCAGGTGAAGCAGATCAACAAGACCCCTCCCATCAAGGACTTCCGGCGCTT  
 AACTTTCGCGCAGATCTTGGCCAGCCACCCAGCAAGCCAGGAGCTTCACTGAGGACTGCTGTTCACAAAGTGAACCTTACCGGCGG  
 CGGCTTTCATCAAGCAGTACGGGCTGCTTCCCGCAGTACAGCCCGGGACTGTCTGCGCCAGAAAGTTCACCGCTGACCGTGTCT  
 GCCTCCCTGCTGACCGCAGGATGATCGCCAGTACACCGCCGCTGTAGCCGGAACCATCACCAGCGCTGGACTTTCGGCGCTGGA  
 CGCGCTTGCAGATCCCTTTCGCCATGCAGATGGCCCTACCGGTTCAACGGCATCGCGGTGACCCAGACAGCTGTGTACGAGAACCAGAAG  
 TGATCGCAACAGTTCACAGCGCCATCGGCAAGATCCAGGACAGCTTGAAGCAGCAGCTTGGCGCCCTGGGAAGCTGCAGGAGCTGGT  
 GAACGAGACCCGCGCCCTGAACACCTTGGTGAAGCAGCTGAGCAGCACTTCCGGCCATCAGCAGCTGTGAAAGCATCTCTGAGC  
 CGGTGGACCTCCCGAGGCGAGTGCAGATGCAGCGGCTGATCACTGGCCGGTGCAGAGCTGCAGACTACGTGACCCAGCAGCTGA  
 TCCCGGCGCCAGATTTCCGGCCAGCGCAACCTTGGCGCCACCAAGATGAGCGAGTGGCTGTGTTGGCCAGAGCAAGCGGTGGACTTCTG  
 CGCAAGGGCTTACCACCTGATGAGCTTCCCCAGAGCCACCCAGGAGTGGTGTTCCTGCAGCTGACCTACGTGCCCGCCAGGAGAAG  
 AACTTCAACCCCGCCAGCCATCTGCCAGCAGGCAAGGCCACTTTCGCCGGAGGGGCTGTGTGTGAGCAAGGCCACCCACTGGTTCG  
 TGACCCAGCGGACTTCTACGAGCCAGATCATCAACCCGCAACACCTTCTGAGCGGCAACTGCGAGCTGGTGTACGGCATCTGTGA  
 CAACACCGTGTACGATCCCTGCAGCCGAGCTGGACAGCTTCAAGGAGGAGTGGACAAGTACTTCAAGAATCACACCCAGCCCGACGTG  
 GACTTGGCGACATCAGCGGATCAACCCAGCGTGGTGAACATCCAGAAAGGAGATCGATCGCTGAAAGGAGTGGCCAAAGAACTGAAG  
 AGAGCTGATGACCTGCAGGAGCTGGGCAAGTACGAGCAGTACATCAAGTGGCCCTGGTACATCTGGCTGGGCTTCACTCCCGGCTGT  
 CGCCATCGTGTGATGACCATCATGCTGTGCTGATGACAGCAGCTGCTGCAGCTGCCGAAAGGGCTGTGACAGCTGGCGGACGCTGCTGCAAG  
 TTCAGCAGGAGCAGCAGCGGCGTGTGAAGGGCTGAAGCTGCACCTACAC **AGATAAAG**GCTGGAGCCTCGGTGGCCCTAGCTTCTGT  
 CCCCCTTGGGCTCCCCAGCCCTTCTCCCTTCTTGCACCCGTTACCCCGTGTGTTGAATAAAGTCTGAGTGGCGGCAAAAAAAAA

**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)

<https://github.com/NAlytics/Assemblies-of-putative-SARS-CoV2-spike-encoding-mRNA-sequences-for-vaccines-BNT-162b2-and-mRNA-1273/blob/main/Assemblies%20of%20putative%20SARS-CoV2-spike-encoding%20mRNA%20sequences%20for%20vaccines%20BNT-162b2%20and%20mRNA-1273.docx.pdf>

# 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



Registered Users EpiFlu™ EpiCoV™ My profile

EpiCoV™ Search Downloads Upload

Search

Accession ID:  Virus name:   complete  high coverage  
 Location:  Host:   low coverage excl  w/Patient status  
 Collection:  to  Submission:  to   collection date compl  
 Clade:  Lineage:  Substitutions:  Variants:

<input type="checkbox"/>	Virus name	Passage dt	Accession ID	Collection da	Submission C	<input type="checkbox"/>	Length	Host	Location	Originating
<input type="checkbox"/>	hCoV-19/USA/ME-HETL-J5855/2021	Original	EPI_ISL_3544587	2021-08-12	2021-08-21	<input type="checkbox"/>	29,727	Human	North America / U	Maine Heal
<input type="checkbox"/>	hCoV-19/USA/ME-HETL-J5854/2021	Original	EPI_ISL_3544586	2021-08-12	2021-08-21	<input type="checkbox"/>	29,727	Human	North America / U	Maine Heal
<input type="checkbox"/>	hCoV-19/USA/ME-HETL-J5826/2021	Original	EPI_ISL_3544585	2021-08-12	2021-08-21	<input type="checkbox"/>	29,564	Human	North America / U	Maine Heal
<input type="checkbox"/>	hCoV-19/USA/ME-HETL-J5825/2021	Original	EPI_ISL_3544584	2021-08-12	2021-08-21	<input type="checkbox"/>	29,588	Human	North America / U	Maine Heal
<input type="checkbox"/>	hCoV-19/USA/ME-HETL-J5824/2021	Original	EPI_ISL_3544583	2021-08-12	2021-08-21	<input type="checkbox"/>	29,518	Human	North America / U	Maine Heal
<input type="checkbox"/>	hCoV-19/USA/ME-HETL-J5823/2021	Original	EPI_ISL_3544582	2021-08-12	2021-08-21	<input type="checkbox"/>	29,518	Human	North America / U	Maine Heal
<input type="checkbox"/>	hCoV-19/USA/ME-HETL-J5822/2021	Original	EPI_ISL_3544581	2021-08-12	2021-08-21	<input type="checkbox"/>	29,518	Human	North America / U	Maine Heal
<input type="checkbox"/>	hCoV-19/USA/ME-HETL-J5864/2021	Original	EPI_ISL_3544580	2021-08-12	2021-08-21	<input type="checkbox"/>	29,676	Human	North America / U	Maine Heal
<input type="checkbox"/>	hCoV-19/USA/ME-HETL-J5663/2021	Original	EPI_ISL_3544579	2021-08-12	2021-08-21	<input type="checkbox"/>	29,676	Human	North America / U	Maine Heal
<input type="checkbox"/>	hCoV-19/USA/ME-HETL-J5657/2021	Original	EPI_ISL_3544578	2021-08-12	2021-08-21	<input type="checkbox"/>	29,728	Human	North America / U	Maine Heal
<input type="checkbox"/>	hCoV-19/USA/ME-HETL-J5656/2021	Original	EPI_ISL_3544577	2021-08-12	2021-08-21	<input type="checkbox"/>	29,676	Human	North America / U	Maine Heal
<input type="checkbox"/>	hCoV-19/USA/ME-HETL-J5655/2021	Original	EPI_ISL_3544576	2021-08-12	2021-08-21	<input type="checkbox"/>	29,518	Human	North America / U	Maine Heal
<input type="checkbox"/>	hCoV-19/USA/ME-HETL-J5653/2021	Original	EPI_ISL_3544575	2021-08-12	2021-08-21	<input type="checkbox"/>	29,728	Human	North America / U	Maine Heal
<input type="checkbox"/>	hCoV-19/USA/ME-HETL-J5853/2021	Original	EPI_ISL_3544574	2021-08-11	2021-08-21	<input type="checkbox"/>	29,676	Human	North America / U	Maine Heal
<input type="checkbox"/>	hCoV-19/USA/ME-HETL-J5848/2021	Original	EPI_ISL_3544573	2021-08-11	2021-08-21	<input type="checkbox"/>	29,728	Human	North America / U	Maine Heal
<input type="checkbox"/>	hCoV-19/USA/ME-HETL-J5847/2021	Original	EPI_ISL_3544572	2021-08-11	2021-08-21	<input type="checkbox"/>	29,726	Human	North America / U	Maine Heal
<input type="checkbox"/>	hCoV-19/USA/ME-HETL-J5845/2021	Original	EPI_ISL_3544571	2021-08-11	2021-08-21	<input type="checkbox"/>	29,676	Human	North America / U	Maine Heal
<input type="checkbox"/>	hCoV-19/USA/ME-HETL-J5851/2021	Original	EPI_ISL_3544570	2021-08-10	2021-08-21	<input type="checkbox"/>	29,714	Human	North America / U	Maine Heal
<input type="checkbox"/>	hCoV-19/USA/ME-HETL-J5844/2021	Original	EPI_ISL_3544569	2021-08-10	2021-08-21	<input type="checkbox"/>	29,725	Human	North America / U	Maine Heal
<input type="checkbox"/>	hCoV-19/USA/ME-HETL-J5843/2021	Original	EPI_ISL_3544568	2021-08-10	2021-08-21	<input type="checkbox"/>	29,727	Human	North America / U	Maine Heal
<input type="checkbox"/>	hCoV-19/USA/ME-HETL-J5841/2021	Original	EPI_ISL_3544566	2021-08-10	2021-08-21	<input type="checkbox"/>	29,676	Human	North America / U	Maine Heal
<input type="checkbox"/>	hCoV-19/USA/ME-HETL-J5840/2021	Original	EPI_ISL_3544565	2021-08-10	2021-08-21	<input type="checkbox"/>	29,728	Human	North America / U	Maine Heal
<input type="checkbox"/>	hCoV-19/USA/ME-HETL-J5768/2021	Original	EPI_ISL_3544564	2021-08-10	2021-08-21	<input type="checkbox"/>	29,676	Human	North America / U	Maine Heal
<input type="checkbox"/>	hCoV-19/USA/ME-HETL-J5767/2021	Original	EPI_ISL_3544563	2021-08-10	2021-08-21	<input type="checkbox"/>	29,676	Human	North America / U	Maine Heal
<input type="checkbox"/>	hCoV-19/USA/ME-HETL-J5766/2021	Original	EPI_ISL_3544562	2021-08-10	2021-08-21	<input type="checkbox"/>	29,728	Human	North America / U	Maine Heal
<input type="checkbox"/>	hCoV-19/USA/ME-HETL-J5765/2021	Original	EPI_ISL_3544561	2021-08-10	2021-08-21	<input type="checkbox"/>	29,714	Human	North America / U	Maine Heal
<input type="checkbox"/>	hCoV-19/USA/ME-HETL-J5662/2021	Original	EPI_ISL_3544560	2021-08-10	2021-08-21	<input type="checkbox"/>	29,676	Human	North America / U	Maine Heal
<input type="checkbox"/>	hCoV-19/USA/ME-HETL-J5661/2021	Original	EPI_ISL_3544559	2021-08-10	2021-08-21	<input type="checkbox"/>	29,727	Human	North America / U	Maine Heal
<input type="checkbox"/>	hCoV-19/USA/ME-HETL-J5660/2021	Original	EPI_ISL_3544558	2021-08-10	2021-08-21	<input type="checkbox"/>	29,714	Human	North America / U	Maine Heal
<input type="checkbox"/>	hCoV-19/USA/ME-HETL-J5551/2021	Original	EPI_ISL_3544557	2021-08-10	2021-08-21	<input type="checkbox"/>	29,728	Human	North America / U	Maine Heal
<input type="checkbox"/>	hCoV-19/USA/ME-HETL-J5550/2021	Original	EPI_ISL_3544556	2021-08-10	2021-08-21	<input type="checkbox"/>	29,728	Human	North America / U	Maine Heal
<input type="checkbox"/>	hCoV-19/USA/ME-HETL-J5839/2021	Original	EPI_ISL_3544555	2021-08-09	2021-08-21	<input type="checkbox"/>	29,200	Human	North America / U	Maine Heal

Total: 583 viruses

<< < 1 2 3 4 5 > >>

Important note: In the GISAID EpiFlu™ Database Access Agreement, you have accepted certain terms and conditions for viewing and using data regarding influenza viruses. To the extent the Database contains data relating to non-influenza viruses, the viewing and use of these data is subject to the same terms and conditions, and by viewing or using such data you agree to be bound by the terms of the GISAID EpiFlu™ Database Access Agreement in respect of such data in the same manner as if they were data relating to influenza viruses.