

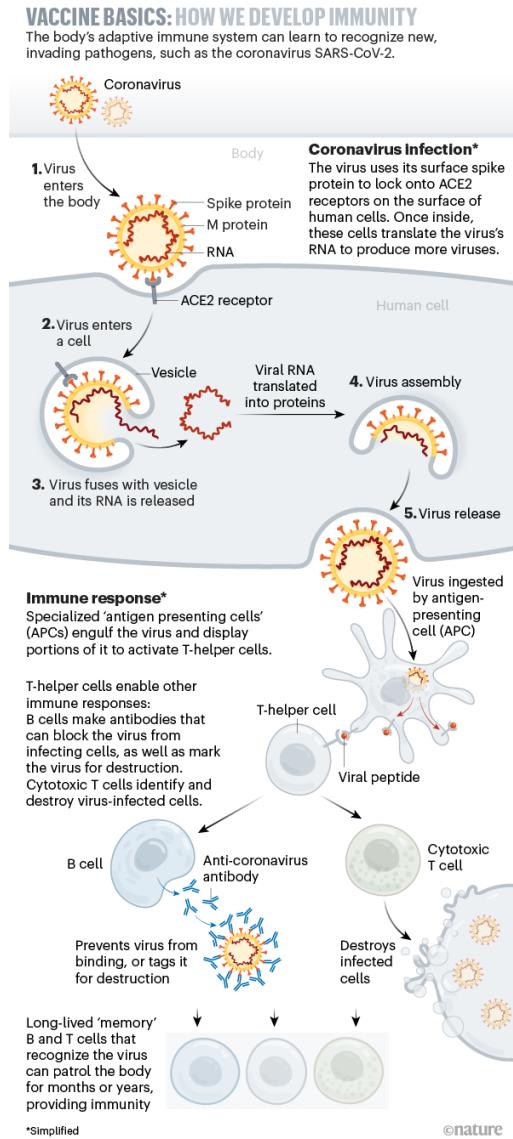
# Sequence Analysis of Pfizer and Moderna COVID-19 Vaccines



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



# SARS-CoV-2 Infection and Immune Response



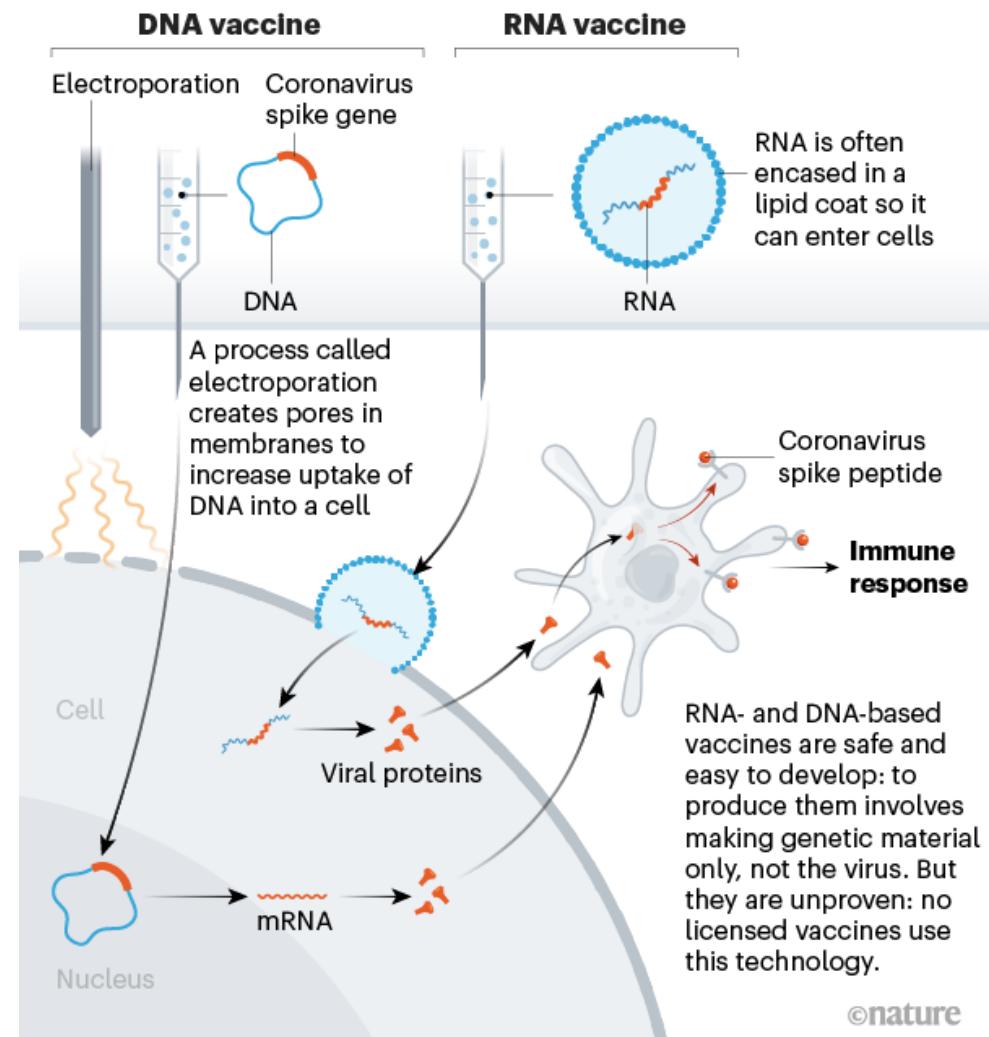
<https://www.nature.com/articles/d41586-020-01221-y>

©nature

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

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

## NUCLEIC-ACID VACCINES

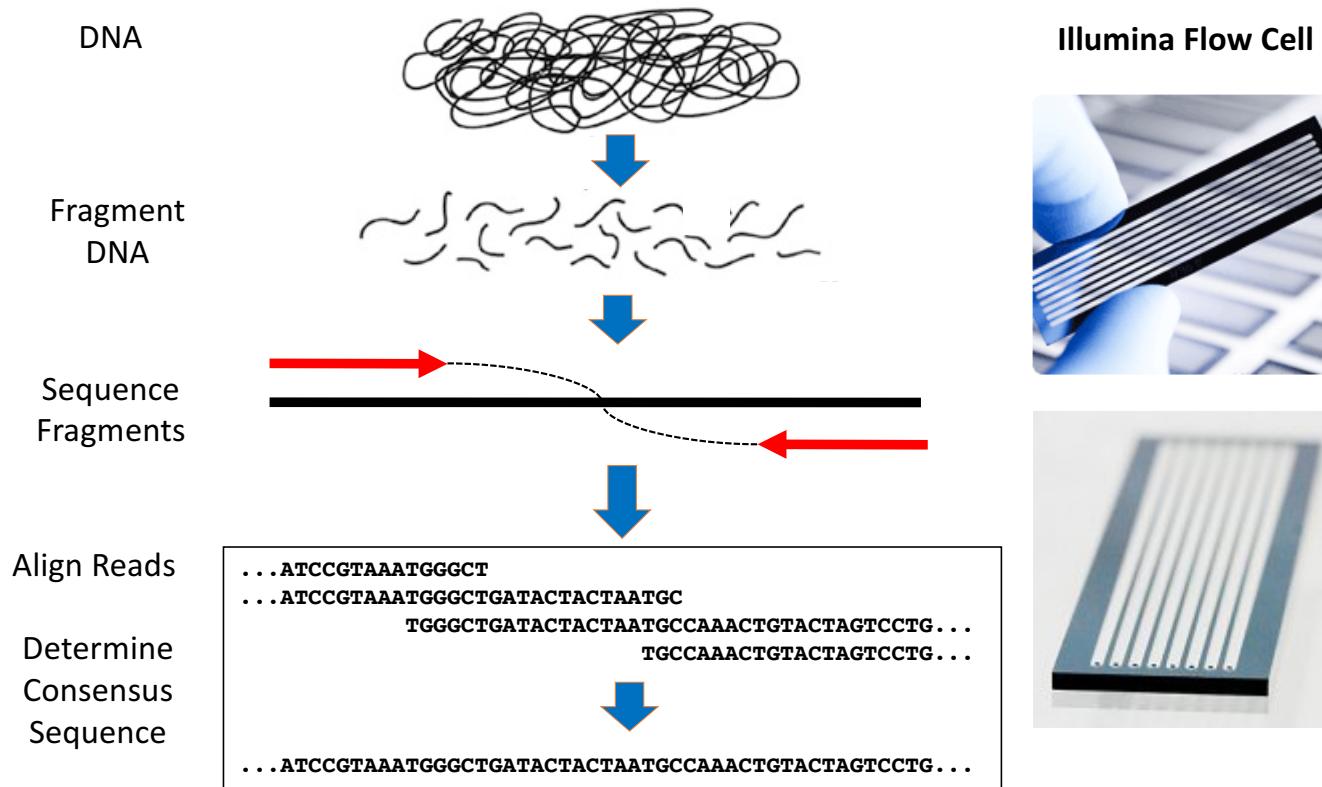


# The \$1000 Genome Is Revolutionizing How We Study Biology and Practice Medicine

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



# Whole Genome Shotgun Sequencing



# Your Own Genome Sequenced (30x Coverage)

The screenshot shows a web browser displaying the Dante Labs website at [us.dantelabs.com/products/whole-genome-sequencing?variant=31588228759618](https://us.dantelabs.com/products/whole-genome-sequencing?variant=31588228759618). The page features a large image of the test kit box, which is white with blue and green accents and labeled "dante labs" and "Whole Genome Sequencing Premium Test". To the right of the image, the product title "30X Whole Genome Sequencing Test" is displayed in blue, followed by the price "\$599.00 USD". Below the price are buttons for "Turnaround Time: 8 weeks", quantity selection (set to 1), "ADD TO CART", and "Buy with amazon pay". A "More payment options" link is also present. The "DESCRIPTION" section contains a bulleted list of benefits, including analyzing 100% of your DNA, making conscious health decisions, and getting full genomic data.

30X Whole Genome Sequencing Test  
\$599.00 USD

Turnaround Time: 8 weeks 1 + -

ADD TO CART

Buy with amazon pay

More payment options

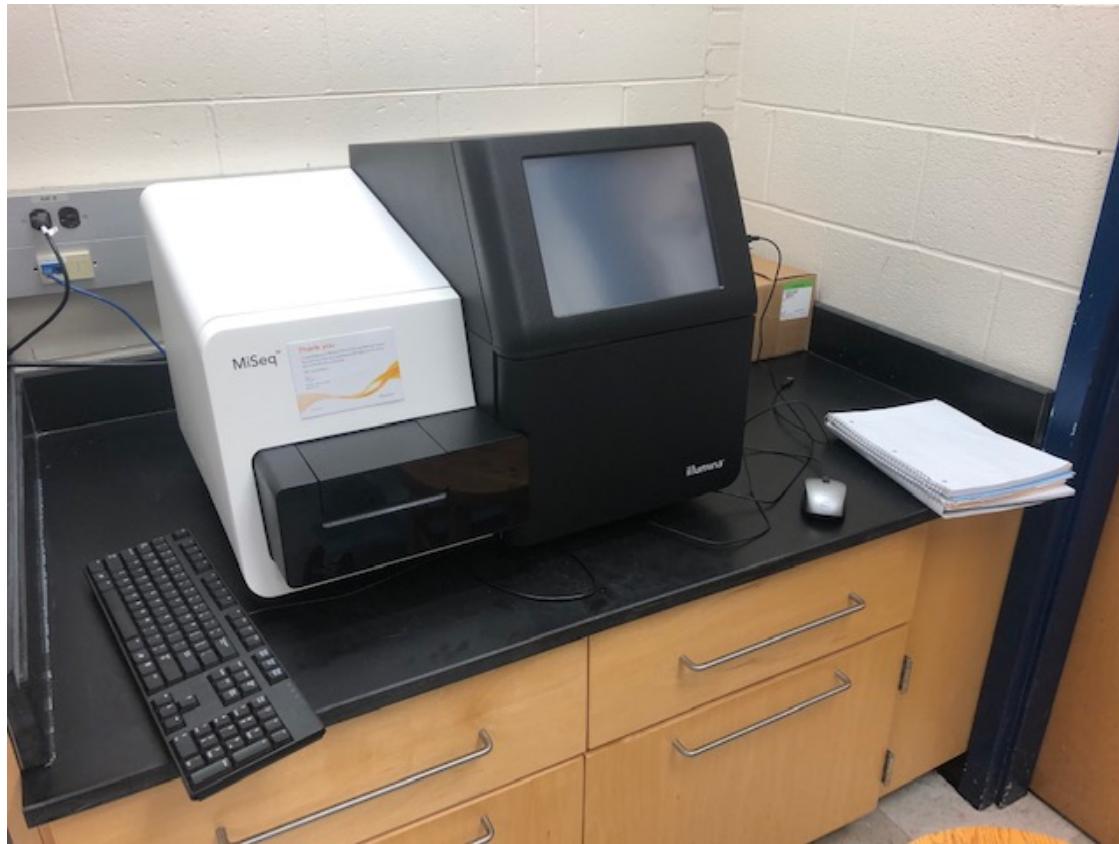
**DESCRIPTION**

- 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

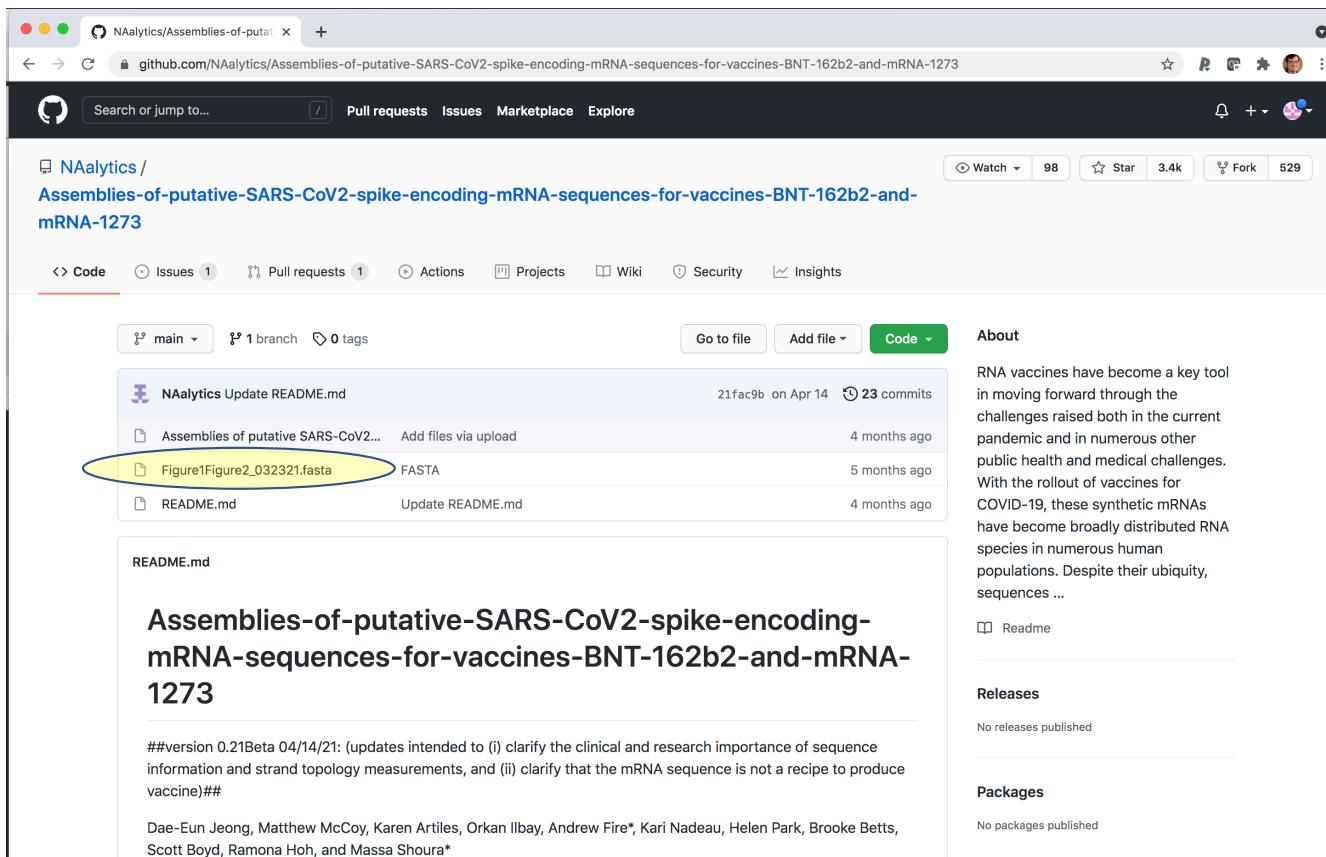
## Black Friday Sale (Nov. 25, 2020)



## 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 "NAalytics/Assemblies-of-putative-SARS-CoV2-spike-encoding-mRNA-sequences-for-vaccines-BNT-162b2-and-mRNA-1273". The repository has 98 stars, 3.4k forks, and 529 open issues. The "Code" tab is selected, showing a list of files. A yellow oval highlights the file "Figure1Figure2\_032321.fasta". The "About" section discusses the significance of RNA vaccines in the COVID-19 pandemic.

**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

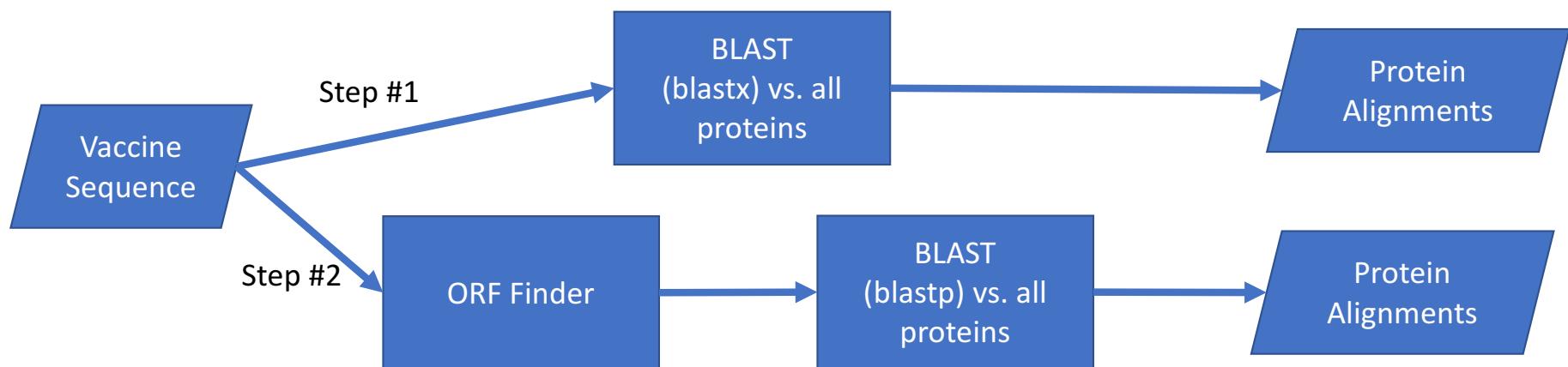
# Moderna Sequence

# What do we want to learn about the sequences?

- •  
•  
•  
•  
•  
•  
•  
•  
•  
•

# 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 main NCBI homepage. At the top right, there is a 'Search' bar and a 'My NCBI' sign-in link. Below the search bar, there's a 'Popular Resources' section with links to PubMed, Bookshelf, PubMed Central, and BLAST. The 'BLAST' link is circled in yellow. The 'Welcome to NCBI' section contains links for 'Submit', 'Download', 'Learn', 'Develop', 'Analyze', and 'Research'. The 'Learn' section specifically highlights the BLAST tool. The 'NCBI News & Blog' section has a recent post about the NCBI Sequence Viewer.

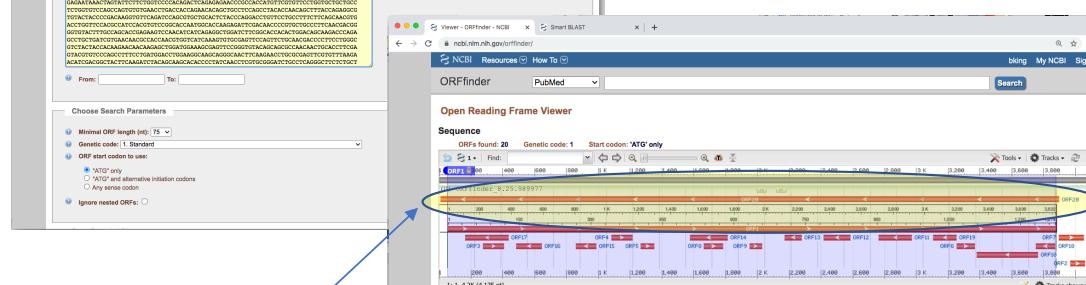
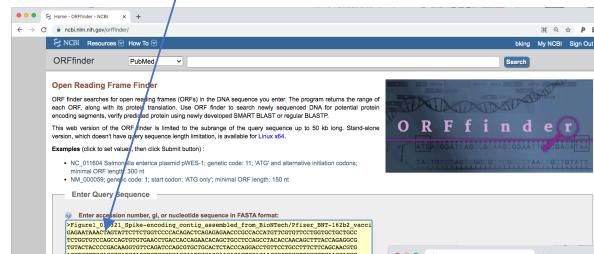
This screenshot shows the 'Basic Local Alignment Search Tool' (BLAST) search interface. The title bar says 'BLAST - Basic Local Alignment Search Tool'. The main area is titled 'Web BLAST' and features four search options: 'Nucleotide BLAST' (nucleotide > nucleotide), 'blastx' (translated nucleotide > protein), 'tblastn' (protein > translated nucleotide), and 'Protein BLAST' (protein > protein). Below these are sections for 'BLAST Genomes' (with a search bar for organism names) and 'Standalone and API BLAST' (with download and API links). The bottom navigation bar includes 'BLAST', 'blastx', 'tblastn', and 'blast'.

Paste the Pfizer sequence here

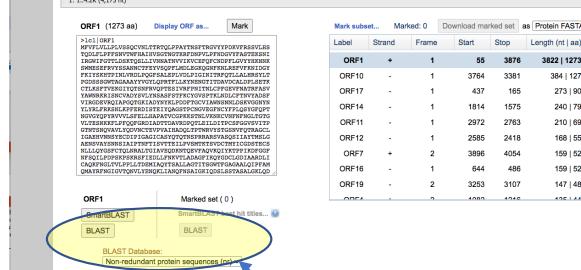
This screenshot shows the 'BLAST' search interface with a specific search configuration. The title bar says 'blastx' and the sub-title is 'Translated BLAST: blastx'. The 'Enter Query Sequence' field contains the sequence: '+Figure1\_032321\_Spike\_encoding\_config\_assembled\_from\_BioNTechPfizer\_BNT-162b2\_vaccine GAGATTAACCTAGTINTCTCTGTCGCCACAGACTCAGAAGAACCCGCGCA CGATGTCTGTTCTGCTGTCCTGCCTGCTGCGTCGG'. The 'From' and 'To' fields are empty. Below the sequence entry, there are fields for 'Or, upload file', 'Genetic code', and 'Job Title'. The 'Job Title' field contains 'Figure1\_032321\_Spike-encoding\_config\_assembled\_from\_BioNTech/i'. The 'Choose Search Set' section is expanded, showing options for 'Database' (set to 'Non-redundant protein sequences (nr)'), 'Organism' (optional), 'Exclude' (optional), and 'Algorithm parameters'. The 'BLAST' button is highlighted in blue.

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

Paste the Pfizer sequence here



Select "ORF1"



# ORF Finder @ NCBI

This screenshot shows the NCBI BLAST search interface. The 'Enter Query Sequence' field contains the sequence 'MVFIVLVLPLVSQCVNLTRITQLPPAYNTSPTRGVYPPKVKFRRSVLLHSTQQDFLPPFSNVLVWHPAHTSGTNGTKPDPVLPFDNDQVYFASTEKSNIWRGFTTLDKTSQSLVNNATN'. The 'Database' dropdown is set to 'Non-redundant protein sequences (nr)'. The 'Algorithm' dropdown is set to 'Quick BLASTP (Accelerated protein-protein BLAST)'. The search results table shows significant alignments with various SARS-CoV-2 proteins.

Scientific Name	Max Score	Total Cover	E	Ident	Len	Accesion	
SARS-CoV-2 spike in cell-free state [Severe acute respiratory syndrome coronavirus 2]	2638	100%	0.0	100.00%	1273	EU2000_1	
SARS-CoV-2 in complex bound with RAB [Severe acute respiratory syndrome coronavirus 2]	2637	2637	100%	0.0	100.00%	1282	TCDL_A
Chain A, Spike glycoprotein [Severe acute respiratory syndrome coronavirus 2]	2635	3635	100%	0.0	99.92%	1283	BCN88530.1
surface glycoprotein [Severe acute respiratory syndrome coronavirus 2]	2633	3633	100%	0.0	99.84%	1282	BCN88531
Distant conformational states of SARS-CoV-2 spike protein [Severe acute respiratory syndrome coronavirus 2]	2632	2632	100%	0.0	99.84%	1310	EU0836_1
surface glycoprotein [Severe acute respiratory syndrome coronavirus 2]	2631	2631	100%	0.0	99.76%	1273	QF73647.1
Structure of SARS-CoV-2 3D-2P full-length trifusion spike trimer (C3 symmetry) [Severe acute respiratory syndrome coronavirus 2]	2630	2630	100%	0.0	99.84%	1273	YP_00724390.1
surface glycoprotein [Severe acute respiratory syndrome coronavirus 2]	2630	2630	100%	0.0	99.76%	1273	QIIS3867.1
surface glycoprotein [Severe acute respiratory syndrome coronavirus 2]	2630	2630	100%	0.0	99.76%	1273	QIIS2721.1
surface glycoprotein [Severe acute respiratory syndrome coronavirus 2]	2630	2630	100%	0.0	99.76%	1273	QKU32833.1
surface glycoprotein [Severe acute respiratory syndrome coronavirus 2]	2630	2630	100%	0.0	99.76%	1273	QKU3280.1

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

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

Viewer - ORFfinder - NCBI    Protein BLAST: search protein    NCBI Blast:Protein Sequence

[Download](#) [GenPept](#) [Graphics](#)

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](#)

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 MVFVLVLLPLVSSQCVNLTTRTQLPPAYTNSFTRGVYYPDKVFRSSVLHSTQDLFPLPFFS 60  
Sbjct 1 MVFVLVLLPLVSSQCVNLTTRTQLPPAYTNSFTRGVYYPDKVFRSSVLHSTQDLFPLPFFS 60

Query 61 NVTWPHAIHVSGTNGTKRFDPNVLPPFDGVDYFASTEKSNIIRGWIFGTTLDSKTSQLLIV 120  
Sbjct 61 NVTWPHAIHVSGTNGTKRFDPNVLPPFDGVDYFASTEKSNIIRGWIFGTTLDSKTSQLLIV 120

Query 121 NNATNVVIKVCEFOFCNDPFLGVVYHKNNKSWMESEFRVYSSANNCTFEYVSQPFLMDLE 180  
Sbjct 121 NNATNVVIKVCEFOFCNDPFLGVVYHKNNKSWMESEFRVYSSANNCTFEYVSQPFLMDLE 180

Query 181 GKQGNFKNLREFVFKNIDGYFKIYSKHTPINLVRDLPPQFSALEPLVDPPIGINITRQQT 240  
Sbjct 181 GKQGNFKNLREFVFKNIDGYFKIYSKHTPINLVRDLPPQFSALEPLVDPPIGINITRQQT 240

Query 241 LLAHLRSYLTPGDSGGWTAGAAAAYVGQLQPRTFLLKYNEENGTTDAVDCALDPLSETK 300  
Sbjct 241 LLAHLRSYLTPGDSGGWTAGAAAAYVGQLQPRTFLLKYNEENGTTDAVDCALDPLSETK 300

Query 301 CTLKSFTVEKGIIYQTSNFRVQPTESIVRFPNITNLCPGEVFNATRFASVYAWNKRKISN 360  
Sbjct 301 CTLKSFTVEKGIIYQTSNFRVQPTESIVRFPNITNLCPGEVFNATRFASVYAWNKRKISN 360

Query 361 CVADYSVLYNSASFSTFKCYGVSPKTLNDLCFTNVYADSFVIRGDEVVRQIAPGQTGKIA 420  
Sbjct 361 CVADYSVLYNSASFSTFKCYGVSPKTLNDLCFTNVYADSFVIRGDEVVRQIAPGQTGKIA 420

Query 421 YNYKLPDDFTGCVIAWNSSNNLDSKVGNNYNYLRLFRKSNLKFEDISTEIQYQGSTPC 480  
Sbjct 421 YNYKLPDDFTGCVIAWNSSNNLDSKVGNNYNYLRLFRKSNLKFEDISTEIQYQGSTPC 480

Query 481 NGVEGFNCYFPQLQSYGFQPTNGVGYQPYRVVVLSEELLHAPATVCGPKKSTNLVKNKCVN 540  
Sbjct 481 NGVEGFNCYFPQLQSYGFQPTNGVGYQPYRVVVLSEELLHAPATVCGPKKSTNLVKNKCVN 540

Query 541 FNFNGLTGTVLTSNKKFLPQQFGRDIADTTDAVRDPQTLEILDITPCSFVGSVITP 600  
Sbjct 541 FNFNGLTGTVLTSNKKFLPQQFGRDIADTTDAVRDPQTLEILDITPCSFVGSVITP 600

Query 601 GTNTSNQAVLYQDVNCTEVPVAIHADQLPTWRVYTSGSNVFQTRAGCLIGAEHVNNSY 660  
Sbjct 601 GTNTSNQAVLYQDVNCTEVPVAIHADQLPTWRVYTSGSNVFQTRAGCLIGAEHVNNSY 660

Query 661 ECDIPIGAGICASYQTQNSPRRARSVASQSIIAYTMSLGAENSVAYSNNSSIAIPTNTI 720  
Sbjct 661 ECDIPIGAGICASYQTQNSPRRARSVASQSIIAYTMSLGAENSVAYSNNSSIAIPTNTI 720

Query 721 SVTTEILPVSMTKTSVDCMYICGDESTECNSNLLQYGSFCTQLNRALTGIAVEQDKNTQE 780  
Sbjct 721 SVTTEILPVSMTKTSVDCMYICGDESTECNSNLLQYGSFCTQLNRALTGIAVEQDKNTQE 780

Query 781 VFAQVKQIYKTPPIKDFGGNFSQILPDPSKPSKRSFIEDLLFNKVTLADAGFIKYQGDC 840  
Sbjct 781 VFAQVKQIYKTPPIKDFGGNFSQILPDPSKPSKRSFIEDLLFNKVTLADAGFIKYQGDC 840

[▼ Next](#) [▲ Previous](#) [◀ Descriptions](#)

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

Paste the Pfizer and Moderna sequences here

# Clustal Omega @ EBI

**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)

[https://github.com/NAalytics/Assemblies-of-putative-SARS-CoV2-spoke-encoding-mRNA-sequences-for-vaccines-BNT-162b2-and-mRNA-1273/blob/main/Assemblies%20of%20putative%20SARS-CoV2-spoke-encoding%20mRNA%20sequences%20for%20vaccines%20BNT-162b2%20and%20mRNA-1273.docx.pdf](https://github.com/NAalytics/Assemblies-of-putative-SARS-CoV2-spike-encoding-mRNA-sequences-for-vaccines-BNT-162b2-and-mRNA-1273/blob/main/Assemblies%20of%20putative%20SARS-CoV2-spoke-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

© 2008 - 2021 | Terms of Use | Privacy Notice | Contact  
You are logged in as Benjamin L King - logout

**EpiCoV™**

Registered Users EpiFlu™ **EpiCoV™** My profile

**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 G478K.V1.(I)  Reset  Fulltext ▲

<input type="checkbox"/>	Virus name	Passage or	Accession ID	Collection date	Submission date	Length	Host	Location	Originating
<input type="checkbox"/>	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
<input type="checkbox"/>	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
<input type="checkbox"/>	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
<input type="checkbox"/>	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
<input type="checkbox"/>	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
<input type="checkbox"/>	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
<input type="checkbox"/>	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
<input type="checkbox"/>	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
<input type="checkbox"/>	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
<input type="checkbox"/>	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
<input type="checkbox"/>	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
<input type="checkbox"/>	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
<input type="checkbox"/>	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
<input type="checkbox"/>	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
<input type="checkbox"/>	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
<input type="checkbox"/>	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
<input type="checkbox"/>	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
<input type="checkbox"/>	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
<input type="checkbox"/>	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
<input type="checkbox"/>	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
<input type="checkbox"/>	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
<input type="checkbox"/>	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
<input type="checkbox"/>	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
<input type="checkbox"/>	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
<input type="checkbox"/>	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
<input type="checkbox"/>	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
<input type="checkbox"/>	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
<input type="checkbox"/>	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
<input type="checkbox"/>	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
<input type="checkbox"/>	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
<input type="checkbox"/>	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
<input type="checkbox"/>	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

Important note: In the [GISAYD 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 [GISAYD EpiFlu™ Database Access Agreement](#) in respect of such data in the same manner as if they were data relating to influenza viruses.