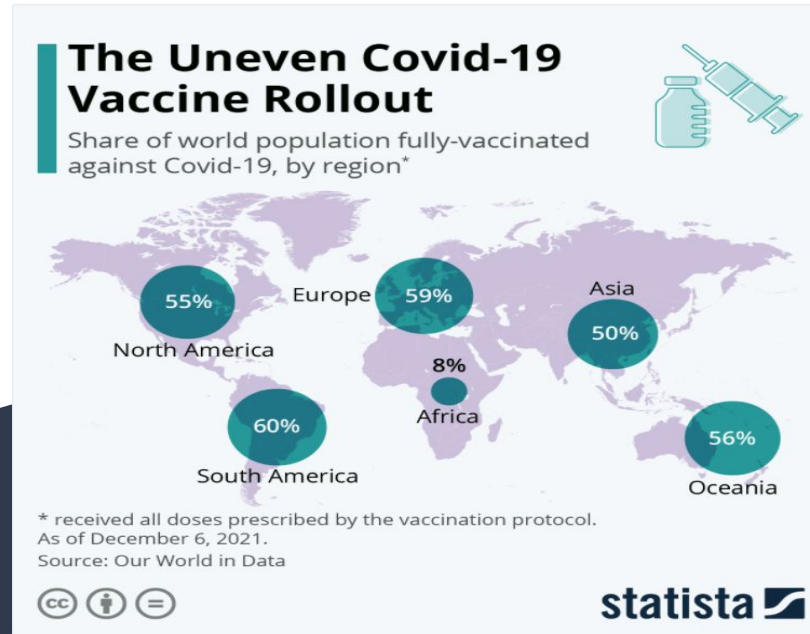


Designing an mRNA vaccine against SARS-CoV-2 variants BA.1 and BA.2

Ben, Joel,
Gil and Aggie



Comparing Pfizer vs Moderna Vaccines

- Given DNA sequence, transcribe to RNA (using BioPython)
- Identify spike region (between 5UTR and 3UTR) using start (AUG) and end codons (UGA, UAG, UAA)
- Translate spike region to amino acid sequence

For Pfizer:

- The position of the initiation codon is: 54
- The position of the termination codon is: 3876
- The sequence used for the stop codon is: UGA
- The length is 3822 nucleotides

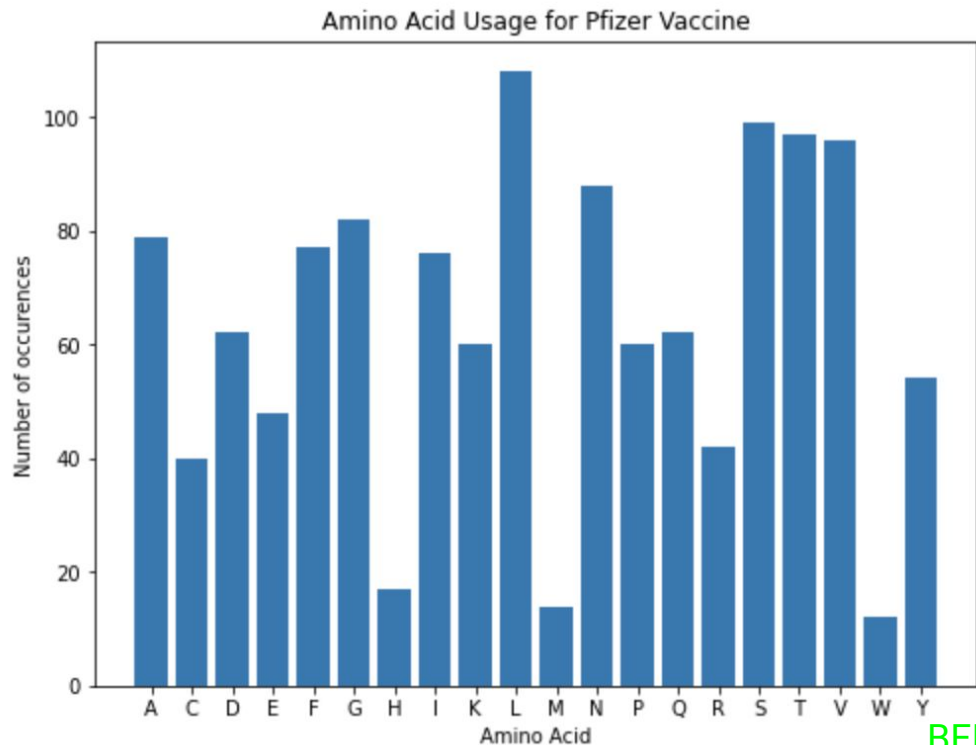
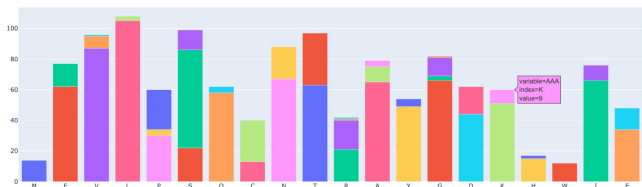
For Moderna:

- The position of the initiation codon is: 57
- The position of the termination codon is: 3879
- The sequence used for the stop codon is: UGA
- The length is 3822 nucleotides

- Identical in protein level but not in nucleotide level: different codons give same amino acid so its implementation decision on which codons to select due to trials and protein stability

Pfizer Vaccine breakdown

- Analyzed the amino acid makeup of the Pfizer vaccine and count each
- For each codon, what amino acid it translates to: Jupyter notebook for interaction
- Pfizer uses multiple codon mapping for each amino acid



Pfizer's approach to Wuhan strain

- Wuhan Strain length is 29903 nucleotides: spike region is from 21563 to 25384 (3821 total)
 - Spike region was written in NCBI website
- The GC content of the Wuhan spike region is 37.3%, while Pfizer is 57%
 - Higher GC content provides greater stability and resistance to denaturation (lose molecular structure) to allow it to withstand temperatures in transport
- When comparing amino acid spike regions for Wuhan and Pfizer, there were 2 amino acid differences at positions 985 and 986
 - Where Wuhan has KV, Pfizer replaced it with [PP](https://cen.acs.org/pharmaceuticals/vaccines/tiny-tweak-behind-COVID-19/98/138) <https://cen.acs.org/pharmaceuticals/vaccines/tiny-tweak-behind-COVID-19/98/138>
- Virus-Vaccine Lookup: Iterate through every codon in Wuhan spike, identify which codon it maps to in the vaccine
 - Most of the codons in the virus were replaced with GC heavy codons in the vaccine

```
{ 'AAA': { 'AAA': 6, 'AAG': 31, 'CCU': 1 },  
  'AAC': { 'AAC': 26, 'AAU': 8 },  
  'AAG': { 'AAA': 3, 'AAG': 20 },  
  'AAU': { 'AAC': 41, 'AAU': 13 },  
  'ACA': { 'ACA': 13, 'ACC': 27 },  
  'ACC': { 'ACA': 1, 'ACC': 9 },
```


BA.2 Vaccine Creation Process

- To create the vaccine:
 - Iterated through Wuhan RNA spike region (ignoring the artificially filled positions) to extract the codons
 - Compared the Wuhan codon to the BA.2 codon at the same position:
 - If it was the same codon, copy the vaccine codons from Pfizer
 - If it was not the same, use the virus vaccine lookup table to extract the best codon
 - For codon CGA (not found within Wuhan), convert it to CGC for GC content
 - Insert 5UTR and 3UTR regions from Pfizer
- Our vaccine has 60.5% GC content, compared to Pfizer with 57%, thus making it more resistant

Our BA.2 vaccine is:

GAGAAUAAACUAGUAUUCUUCUGGUCCCCACAGACUCAGAGAGAACCCGCCACCAUG

GIL

BA.1 Analysis & Comparison

- From the 88,000 BA.1 variant samples, we randomly selected 20 and used local alignment to identify the spike region
 - We found that, on average, the spike region length was 3813 nucleotides
 - We ensured that the samples used had no unknown amino acids (X) in the sequence
- Most entries were collected by the US, then followed by Bahrain (same as in BA.2)
- When comparing Wuhan with BA.1 spike region, there were over 1130 differences
 - By artificially filling in 5 amino acids to the BA.1 spike and 2 to the Wuhan spike, we were able to reduce it to 35 differences between the two
- The BA.1 vaccine was created in the same way as BA.2 (ignoring different codons)

Wuhan Spike

MFVFLVLLPLVSSQCVNLTTTRTQLPPAYTNSFTRGVYYPDKVFRSSVLHSTQDLFLPFFSNVTWFHAIHVSGTNGTKRFDNPVLPFNDGVYFASTE

BA.1 Spike

MFVFLVLLPLVSSQCVNLTTTRTQLPPAYTNSFTRGVYYPDKVFRSSVLHSTQDLFLPFFSNVTWFHVI--SGTNGTKRFDNPVLPFNDGVYFASIE

Scroll to the right to spot differences

There are 35 differences at positions: [66, 68, 69, 94, 141, 142, 143, 144, 210, 211, 212, 213, BEN

BA.1 vs Ba.2 Variants

- Upon initial observation, there was 188 differences between the variants
 - There was 5 amino acids in BA.1 that were missing or modified in BA.2 and 5 amino acids in BA.2 that were missing or modified in BA.1 at various positions
 - Should be filled in relation to the previous
- After, there were 27 differences between the variants

```
copy_bal_protein = str(bal_amino_seq)
copy_ba2_protein = str(ba2_amino_seq)

copy_ba2_protein = copy_ba2_protein[0:23]+'///'+copy_ba2_protein[23:]

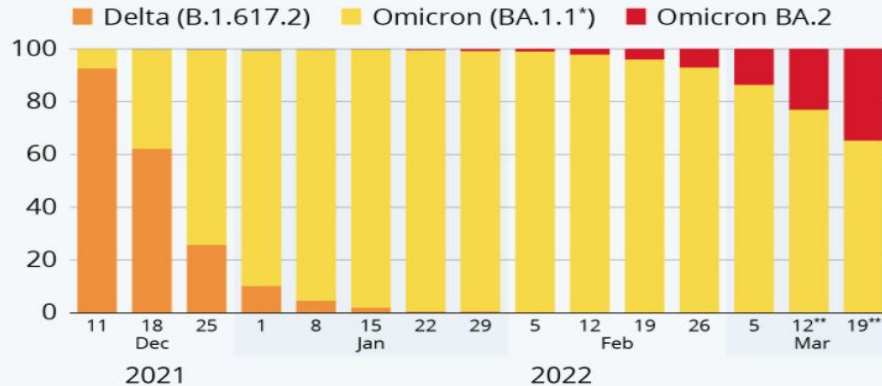
copy_bal_protein = copy_bal_protein[0:68]+'--'+copy_bal_protein[68:]
copy_bal_protein = copy_bal_protein[0:142]+'---'+copy_bal_protein[142:]

copy_ba2_protein = copy_ba2_protein[0:214]+'//'+copy_ba2_protein[214:]
```


<https://www.statista.com/chart/27076/coronavirus-variants-in-the-us/>

Omicron BA.2 Variant to Become Dominant in the U.S.

Viral lineage of newly detected Covid-19 infections in the U.S. per week (in percent)



The End!

Questions?