Replicating Ford et al.'s investigation on the immune consequences and SARS-CoV-2 antibody's binding affinity using an *in silico* approach

Hannah Zeru[12], Cameron Jones[3], Denis Jacob Machado[12]*
[1]Department of Bioinformatics and Genomics, University of North Carolina at Charlotte, NC, USA
[2]Computational Intelligence to Predict Health and Environmental Risks (CIPHER) Center, University of North Carolina at Charlotte, NC, USA
[3]Department of Computer Science, Eastern Michigan University, MI, USA
*Contact: dmachado@charlotte.edu

The 7th International Conference on Interdisciplinary Research on Computer Science, Psychology, and Education
December 27-29, 2023

UNIVERSITY OF NORTH CAROLINA CHARLOTTE

Table of Contents

The variants of SARS-CoV-2, the causative agent of COVID-19

- SARS-CoV-2 is a RNA virus
- Depending on the strain there are different COVID-19 symptoms
- This research mainly focused on Omicron variants

AI as a tool for genomic epidemiology

- Artificial Intelligence can be used to track the spread of new variants and predict if they will escape current treatments and vaccines
- New tools give us the ability to respond to outbreaks faster and inform policy makers almost in real time
- Ex. AlphaFold2 can predict 3D structures of novel variants and HADDOCK can estimate the binding affinity between viral proteins and neutralizing antibodies

AI as a tool for genomic epidemiology

- This study is replicating a study done early 2023 by Ford et al. (2023, doi: https://doi.org/10.3389/fviro.2023.1172027)
- Original study results:
  - SARS-CoV-2 Omicron variant XBB.1.5 could still be treated with current SARS-CoV-2 neutralizing antibodies
  - The genetic structure of XBB.1.5 is similar to other variants within the Omicron strain

AI as a tool for genomic epidemiology

- Our goals for this study:
  - Replicate previous study and get similar results (if not, find out what went wrong)
  - Document workflow and produce tutorial or flowchart
  - Test the tools used and compare speed, accuracy, and usability

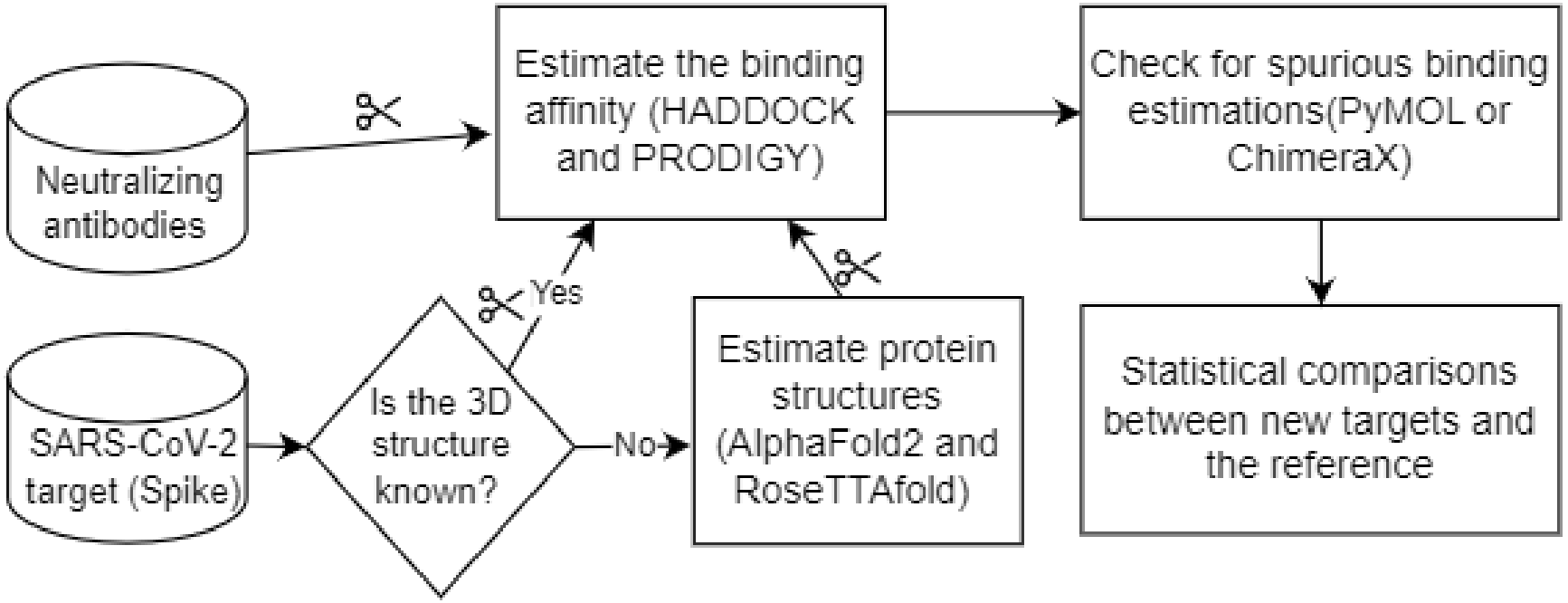
Selected antibodies and Spike data

- Selected nucleotide sequences of the variants' Spike genes were sourced from GenBank and RCSB Protein Data Bank
- Neutralizing antibodies were sourced from CoV-AbDab Database and the Protein Data Bank
- Sequences were extracted from the Spike sequences' Receptor Binding Domain (RBD)
- Antibody sequences were taken from the Fragment Antigen-Binding (FAB) region

Selected antibodies and Spike data

- SARS-CoV-2 variants selected
  - XBB.1.5
  - BM.1.1.1
  - B.1.1.529
  - BJ.1
  - SARS-CoV-2 Wuhan-Hu-1 (only for AlphaFold Colab vs ColabFold-mmseq2)
- Neutralizing antibodies selected:
  - LY-CoV555
  - LY-CoV1404
  - P5C3
  - COVOX-150
  - AZD1061
  - AZD8895
  - C110
  - EY6A
  - 58C6
  - CV38-142

Strategies for data preparation



Strategies for data preparation

1. Download nucleotide sequences of the SARS-CoV-2 Spike gene from GenBank
2. Download available crystallography results from RCSB Protein Data Bank
3. Predict 3D protein structures with AlphaFold2
    - Strategy 1: AlphaFold Colab
    - Strategy 2: ColabFold-mmseq2

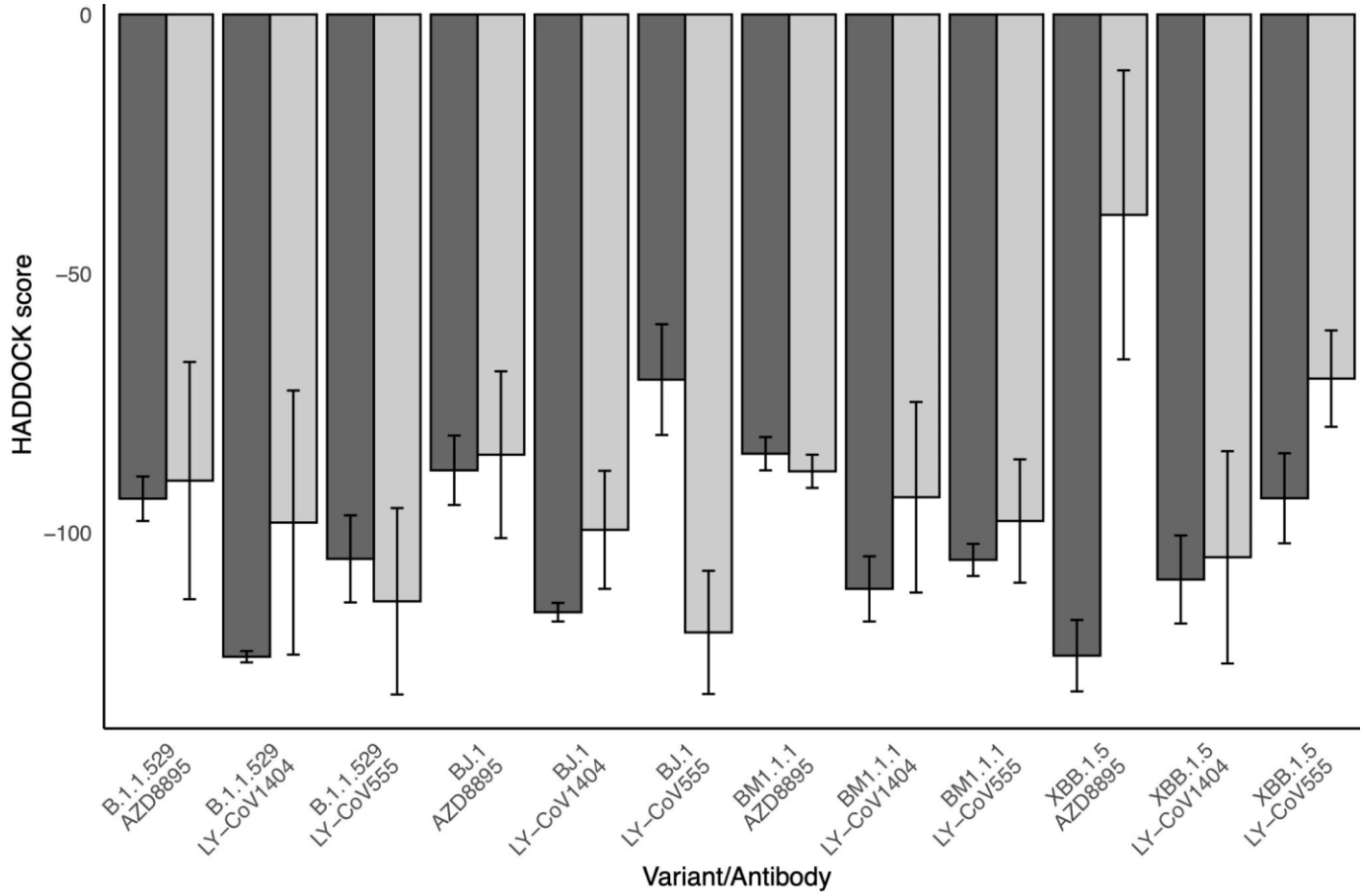
Protein-to-protein docking

1. Downloaded selected neutralizing antibodies from CoV-AbDab Database and the Protein Data Bank
2. Process data
    - Strategy 1: PyMOL (version 1.8)
    - Strategy 2: ChimeraX (version 1.6.1)

Protein-to-protein docking

1. Submit prepared Spike protein nucleotide sequences and neutralizing antibody FAB sequences to HADDOCK (version 2.4)
2. Analyze and compare results

Comparing results to Ford et al. (2023)

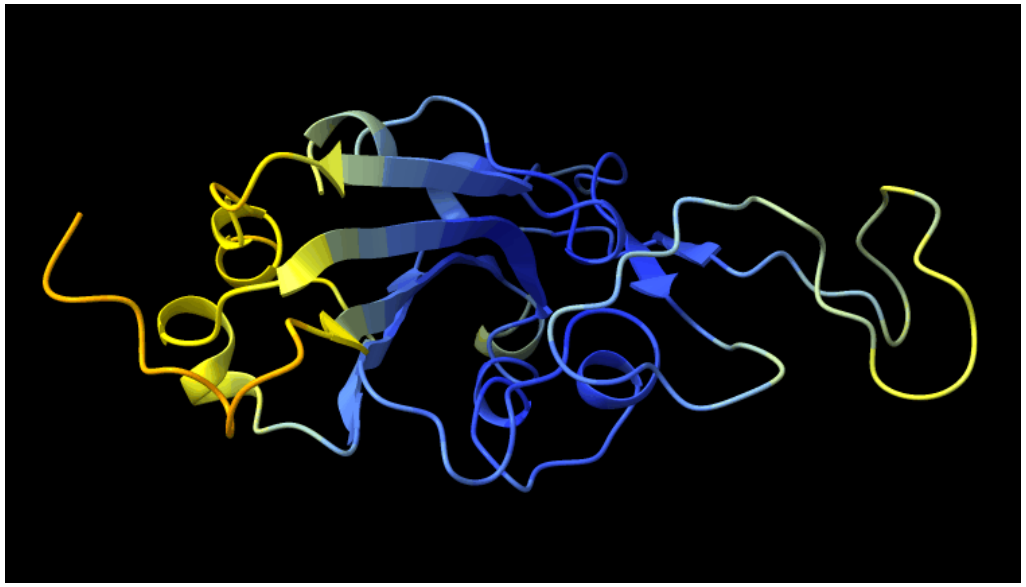


Light grey is Ford et al. and dark grey is our study. Results are not statistically significant.

Additional outcomes

- AlphaFold Colab vs. ColabFold-mmseq2: ColabFold-mmseq2 performed 20 times faster than AlphaFold Colab on average, with both having statistically insignificant results
- PyMOL (version 1.8) vs ChimeraX (version 1.6.1) : ChimeraX (version 1.6.1) was more useable compared to PyMOL (version 1.8)

Additional outcomes

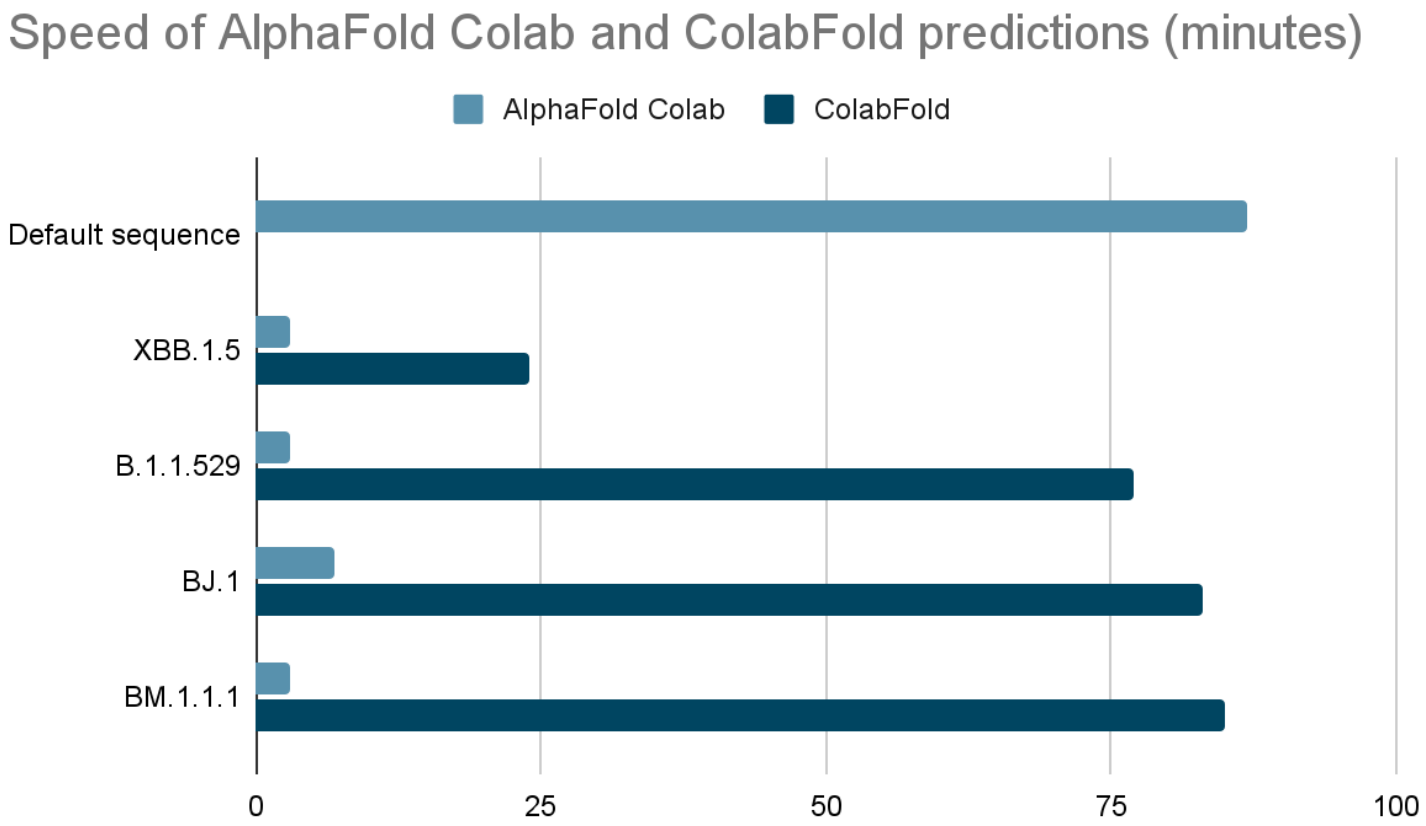


BJ.1 predicted 3D structure from AlphaFold Colab. Nucleotide sequence is isolated.

BJ.1 predicted 3D structure from ColabFold. Nucleotide sequence is not isolated but can be observed.

Blue means that the Predicted Local Distance Difference Test (PLDDT) is confident the predicted structure is correct, yellow/orange means the PLDDT has low confidence.

Additional outcomes



Additional outcomes

- PyMOL (version 1.8) vs ChimeraX (version 1.6.1):
  - Both can be used to isolate needed sequences from files
  - ChimeraX was used to visually analyze 3D predictions and crystallography structures
  - PyMOL was harder to set up compared to ChimeraX and failed to run on one laptop

Conclusions

- Summary of our results:
  - We documented the steps needed in a flowchart to help the Phyloinformatics lab when producing a future pipeline
  - We found that ChimeraX was a good alternative to PyMOL and reduced the needed programs
  - We analyzed the results and compared them to the original study, finding that the results are reproducible

UNIVERSITY OF NORTH CAROLINA CHARLOTTE

Thank you

Special thanks to Anastasiia Duchenko, Dr. Paula Prieto Oliveira, the Phyloinformatics lab (phyloinformatics.com) at UNC Charlotte, CIPHER lab at UNC Charlotte, the College and Computing Informatics department at UNC Charlotte, and the CCI's REU program funded by the NSF (NSF; Grant No. 2244424).

NSF

EASTERN MICHIGAN UNIVERSITY

UNIVERSITY OF NORTH CAROLINA CHARLOTTE

If you have any questions or comments, please contact us at:

Dr. Denis Jacob Machado: dmachado@charlotte.edu

Hannah Zeru: hzeru@charlotte.edu

Cameron Jones: cjone206@emich.edu

The full paper and resources used can be found here:

https://zenodo.org/records/10068319

UNIVERSITY OF NORTH CAROLINA CHARLOTTE