THE PENNSYLVANIA STATE UNIVERSITY
SCHREYER HONORS COLLEGE

DEPARTMENT OF BIOCHEMISTRY AND MOLECULAR BIOLOGY

Comparative Analysis Between SARS-CoV, MERS-CoV, and SARS-CoV-2: Genomic Similarity, Spike Protein Homology, and Government Funding as it Relates to Vaccine Development

BRIDGET FORESMAN
SPRING 2021

A thesis
submitted in partial fulfillment
of the requirements
for a baccalaureate degree
in Biochemistry and Molecular Biology
with honors in Biochemistry and Molecular Biology

Reviewed and approved* by the following:

Joel Segel
Assistant Professor of Health Policy and Administration
Thesis Supervisor

Teh-hui Kao
Distinguished Professor of Biochemistry and Molecular Biology
Honors Adviser

* Electronic approvals are on file.

ABSTRACT

As Severe Acute Respiratory Syndrome Coronavirus-2 (SARS-CoV-2) has undoubtedly changed day-to-day operations globally, with its associated disease Coronavirus Diseases-2019 (COVID-19) infecting over 116 million people and responsible for over 2.5 million deaths, analyzing the biochemical similarity of SARS-CoV-2 to its related pandemic-causing coronavirus strands Severe Acute Respiratory Syndrome Coronavirus (SARS-CoV) and Middle East Respiratory Syndrome Coronavirus (MERS-CoV) is essential for prevention and treatment of said disease. Understanding genomic details of SARS-CoV-2, in addition to previous, yet unsuccessful attempts at developing vaccines against SARS-CoV and MERS-CoV has allowed the incredibly rapid development of several effective vaccines against SARS-CoV-2. This paper will review the history of each coronavirus-responsible pandemic, analyze biochemical similarities, such as genome similarity and spike (S) protein structure of SARS-CoV, MERS-CoV, and SARS-CoV-2, the role the former two coronaviruses played in accelerating successful vaccine development against the latter, and finally, investigating the relationship between the mortality rate/transmissivity associated with the respective coronaviruses and the government funding of vaccine development.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

This thesis is dedicated to those who have supported me in any capacity along my academic journey. I extend my utmost gratitude to my thesis supervisor, Dr. Joel Segel, for his patience, kindness, and assistance through each step of my research analysis and writing process. I thank my honors advisor, Dr. Teh-hui Kao, for not only being one of the best teachers I have ever had, helping me to renew my passion for learning, but for his consistent support and guidance throughout my time at Penn State. To my family and friends—thank you for your unwavering love and encouragement, for without your support, I would be lost!

Finally, I especially thank my parents, for fostering my love of science from a young age, making countless sacrifices so to create a better life for me, and for empowering me to always strive for greatness. You are my heroes!

Chapter 1

Introduction

History of Coronavirus Related Pandemics

Severe Acute Respiratory Syndrome Coronavirus-2 (SARS-CoV-2), a novel coronavirus responsible for the coronavirus disease (COVID-19) pandemic, was first identified after the Wuhan Municipal Health Commission, China reported forty-four patients in Wuhan City, Hubei Province of China with pneumonia of unknown etiology to the World Health Organization (WHO) in December 2019.[1, 2] On January 12, 2020, China shared the genetic sequence of the virus responsible for these pneumonia-like symptoms, first identified as novel coronavirus nCoV and later renamed SARS-CoV-2. These symptoms included fever, difficulty breathing, as well as invasive pneumonic infiltrates in both lungs.[3] Since identification, patients with SARS-CoV-2 have experienced additional symptoms such as gastrointestinal issues—nausea, diarrhea, and vomiting, loss of smell and taste, sepsis, and multiorgan failure.[4] As of March 31, 2021, SARS-CoV-2 has infected over 128,540,982 people in 221 countries, resulting in 2,808,308 deaths.[5] This equates to a 2.2% over fatality rate. A mathematical model was used to calculate the average number of people an infected individual spreads the virus to, known as the R_0 value. For SARS-CoV-2, the R_0 value was 2.39-3.44, suggesting a high rate of transmission.[6] The identification of this highly transmissible, novel pathogenic coronavirus called attention to the Severe Acute Respiratory Syndrome Coronavirus (SARS-CoV) epidemic and the Middle East

Respiratory Syndrome Coronavirus (MERS-CoV) epidemic due to the similarities in genome and related symptoms of viral infection.[7]

SARS-CoV originated in Guangdong Province, China in November, 2002 and was later identified as a novel coronavirus in April, 2003 as a result of international collaboration.[8, 9] Similar to SARS-CoV-2, the disease was associated with an atypical pneumonia outbreak of unknown origin.[10] Within eight months, SARS-CoV spread to 26 countries via human-to-human transmission, and ultimately resulted in 8,096 cases and 774 deaths (9.56% fatality rate) (Table 1). For SARS-CoV, the R_0 value was calculated as less than 1, suggesting a lower rate of transmission in comparison to SARS-CoV-2.[5] This disease, however, led to the first pandemic of the twenty-first century.[7] While typical symptoms of SARS-CoV included fever, coughing, difficulty breathing, shortness of breath, and radiographic evidence of invasive pneumonic infiltrates, symptoms ranged from mild respiratory symptoms to severe acute respiratory disease and death. Additionally, some patients experienced gastrointestinal issues, such as nausea, diarrhea, and vomiting.[11] Note the similarities in symptoms of SARS-CoV and SARS-CoV-2 infection. While there is no commercially available vaccine for SARS-CoV, efforts from the scientific community to produce such a vaccine proved to be beneficial in the development of a SARS-CoV-2 vaccine.

MERS-CoV was identified as a novel coronavirus in June 2012 after a patient was hospitalized in Saudi Arabia for pneumonia and acute kidney injury. Since identification, the WHO has reported 2,494 total cases and 858 deaths (34.40% fatality rate) as a result of human-to-human transmission of MERS-CoV (Table 1. Case Analysis of Coronavirus Pandemics).[5] The R_0 value of MERS-CoV was 1.4-2.5, which is higher than that of SARS-CoV and lower than that of SARS-CoV-2, implying an intermediate rate of transmission relative to the other

coronaviruses.[5] Clinical symptoms of MERS-CoV included, but were not limited to, fever, cough, shortness of breath, as well as gastrointestinal symptoms, such as nausea, diarrhea, and vomiting. In serious MERS-CoV cases, patients experienced acute respiratory distress and multiorgan system failure.[12] It is important to comment on the 34.40% fatality rate, as this is significantly higher than the 9.56% fatality rate recorded from the SARS-CoV epidemic. This is likely due to the possibility of asymptomatic illness resulting from MERS-CoV infection and low diagnostic testing numbers, for there may have been people positive for MERS-CoV, presenting the disease asymptomatically, who were not tested.[13] Again, note the similarities between MERS-CoV and SARS-CoV-2 infection, both in symptom presentation and possibility of asymptomatic illness. As with SARS-CoV, there is no commercially available vaccine for MERS-CoV.

Table 1. Case Analysis of Coronavirus Pandemics

	SARS-CoV	MERS-CoV	SARS-CoV-2
Total Countries with Recorded Cases	26	27	221
Total Recorded Cases	8,096	2494	116,668,383
Total Recorded Deaths	774	858	2,592,047
Fatality Rate	9.56%	34.40%	2.22%
R_0 value	< 1	1.4-2.5	2.39-3.44

On January 12, 2020, China released the genetic sequence of the SARS-CoV-2 genome, allowing for the rapid development of diagnostic kits and propelling research for disease treatment and prevention.[3] This sequence analysis provided greater insight into the similarities and differences of SARS-CoV, MERS-CoV, and SARS-CoV-2, specifically by determining that SARS-CoV-2 has 82.45% entire genome sequence identity to SARS-CoV and 69.58% entire

genome sequence identity to MERS-CoV.[14] By understanding the similarities in biochemical features of these coronaviruses, such as host-pathogen interaction and molecular mechanism of pathogenicity, prior information collected about SARS-CoV and MERS-CoV can serve as a guide for treatment of SARS-CoV-2.

Biochemical Analysis of Coronavirus Structure and Pathogenicity

Coronaviruses are a group of pathogenic viruses having enveloped, positive-sense, single-stranded RNA (ssRNA+) genomes of size 26-32 kilobases (kb). Coronaviruses get their name from the crown-like appearance of club-shaped protein projections covering their spherical nucleocapsid, as 'corona' means crown in Latin. Belonging to the Coronaviridae family of order Nidovirales, coronaviruses are further classified into four subfamilies—α-, β-, γ-, and δ-coronaviruses, the former two primarily infecting mammals, latter two primarily infecting avian species.[14, 15] Of the subfamilies infecting mammals, there are seven Human coronaviruses (HCoVs) known to infect people. Four of these coronaviruses display mild pathogenicity resulting in mild, cold-like symptoms—HCoV-229E (α-coronavirus), HCoV-NL63 (α-coronavirus), HCoV-OC43(β-coronavirus), and HCoV-HKU1 (β-coronavirus). The other three, lethal coronaviruses infecting humans—SARS-CoV, MERS-CoV, SARS-CoV-2—result from the rapid genetic evolution of RNA viruses, specifically β-coronaviruses, that are known to infect other mammals.[15] Bats were identified as not only an important organism in inter-species transmission, but also as the likely ancestral host of SARS-CoV, MERS-CoV, and SARS-CoV-2, due to their species variation, long life-span, migration/flying patterns, and unique behavior of containing multiple species in a single colony.[16]

Specifically looking at genetic evolution of SARS-CoV, MERS-CoV, and SARS-CoV-2, it is necessary to understand their corresponding viral genome at a molecular level. In general, a viral genome encodes for proteins playing an important role in replication, transcription, structure, and infectivity. The overall organization of a coronavirus genome follows the format: 5'-[Open Reading Frame (ORF) 1a/b Replicase]-[Spike (S) Protein Region]-[Envelope (E) Protein Region]-[Membrane (M) Protein Region]-[Nucleocapsid (N) Protein Region]-3'.[14, 16]

The ORF 1a/b Replicase encodes for non-structural proteins important in the replication and transcription of viral RNA and is generally conserved across coronaviruses. The four regions following ORF 1a/b Replicase encode for the structural proteins S, E, M, and N, all of which are extremely important in virulence to host cells. It is within these regions that mutations contributing to virulence are observed, specifically seen when comparing the genomes of SARS-CoV, MERS-CoV, and SARS-CoV-2. [14, 16]

In terms of host specificity, the most important region in a coronavirus' genome is the S Protein Region, as it codes for proteins assisting in a virion's recognition of and entry into host cell. Specifically, the S Protein Region codes for a receptor binding domain—allows S protein to recognize and bind to receptors, a hydrophobic fusion protein—helps virion to fuse with membrane of host cell, heptad repeats—assist with structural modification of S protein to promote viral fusion, a transmembrane domain—allows virion to cross host cell membrane, and a cytoplasmic domain—important in transport of virion into host cell. The S protein is trimeric in nature, composed of three monomers, each with an S1 and an S2 domain. The receptor binding domain is found within the N-terminal, globular S1 domain, thus responsible for cell recognition and receptor attachment. The S2 domain contains the fusion protein, heptad repeats, transmembrane domain, and cytoplasmic domain, mediating fusion with cell membrane and

entry into host cell. [14, 16, 17] Research suggests a conformational change is required to expose the receptor binding domain of the trimer S protein, discussed further in Chapter 2: Coronavirus Vaccine Development.[18]

Genomic analysis has revealed the sequence coding for S proteins is highly variable among coronavirus strains, meaning it is generally more prone to mutation. While MERS-CoV recognizes and binds to Dipeptidyl peptidase 4 (DPP4), both SARS-CoV and SARS-CoV-2 S proteins recognize and bind to human Angiotensin Converting Enzyme-2 (ACE-2).[16] To analyze the homology of these S proteins, the National Center for Biotechnology Information (NCBI)'s GenBank was employed to determine the nucleotide sequence of SARS-CoV, MERS-CoV, and SARS-CoV-2's S protein coding region (CDS). The 'Reference Genome Sequence' was located for each strain using GenBank, which further provided the S proteins' CDS. [19-21] The corresponding nucleotide sequences were exported in FASTA format from GenBank to Clustal Omega (Table 3). In using Clustal Omega, the Multiple Sequence Alignment tool was employed to analyze sequence homology.[22]

Table 2.Analysis of SARS-CoV, MERS-CoV, SARS-CoV-2 Spike (S) Protein CDS Homology

GenBank Entry	NCBI Reference Sequence	S Protein CDS FASTA
SARS coronavirus Tor2, complete genome	NC_004718.3	>NC_004718.3:21492-25259 SARS coronavirus Tor2, complete genome ATGTTTATTTTCTTATTATTTCTTACTCTCACTAGTGGTAGTGACCTTGACCGGTGCACCACTTT TGATGATGTTCAAGCTCCTAATTACACTCAACATACTTCATCTATGAGGGGGGTTTACTATCCTG ATGAAATTTTTAGATCAGACACTCTTTATTTAACTCAGGATTTATTTCTTCCATTTTATTCTAAT GTTACAGGGTTTCATACTATTAATCATACGTTTGGCAACCCTGTCATACCTTTTAAGGATGGTAT TTATTTTGCTGCCACAGAGAAATCAAATGTTGTCCGTGGTTGGGTTTTTGGTTCTACCATGAACA ACAAGTCACAGTCGGTGATTATTATTAACAATTCTACTAATGTTGTTATACGAGCATGTAACTTT GAATTGTGTGACAACCCTTTCTTTGCTGTTTCTAAACCCATGGGTACACAGACACATACTATGAT ATTCGATAATGCATTTAATTGCACTTTCGAGTACATATCTGATGCCTTTTCGCTTGATGTTTCAG AAAAGTCAGGTAATTTTAAACACTTACGAGAGTTTGTGTTTAAAAATAAAGATGGGTTTCTCTAT GTTTATAAGGGCTATCAACCTATAGATGTAGTTCGTGATCTACCTTCTGGTTTTAACACTTTGAA ACCTATTTTTAAGTTGCCTCTTGGTATTAACATTACAAATTTTAGAGCCATTCTTACAGCCTTTT CACCTGCTCAAGACATTTGGGGCACGTCAGCTGCAGCCTATTTTGTTGGCTATTTAAAGCCAACT ACATTTATGCTCAAGTATGATGAAAATGGTACAATCACAGATGCTGTTGATTGTTCTCAAAATCC ACTTGCTGAACTCAAATGCTCTGTTAAGAGCTTTGAGATTGACAAAGGAATTTACCAGACCTCTA ATTTCAGGGTTGTTCCCTCAGGAGATGTTGTGAGATTCCCTAATATTACAAACTTGTGTCCTTTT GGAGAGGTTTTTAATGCTACTAAATTCCCTTCTGTCTATGCATGGGAGAGAAAAAAAATTTCTAA TTGTGTTGCTGATTACTCTGTGCTCTACAACTCAACATTTTTTTCAACCTTTAAGTGCTATGGCG TTTCTGCCACTAAGTTGAATGATCTTTGCTTCTCCAATGTCTATGCAGATTCTTTTGTAGTCAAG GGAGATGATGTAAGACAAATAGCGCCAGGACAAACTGGTGTTATTGCTGATTATAATTATAAATT GCCAGATGATTTCATGGGTTGTGTCCTTGCTTGGAATACTAGGAACATTGATGCTACTTCAACTG GTAATTATAATTATAAATATAGGTATCTTAGACATGGCAAGCTTAGGCCCTTTGAGAGAGACATA TCTAATGTGCCTTTCTCCCCTGATGGCAAACCTTGCACCCCACCTGCTCTTAATTGTTATTGGCC ATTAAATGATTATGGTTTTTACACCACTACTGGCATTGGCTACCAACCTTACAGAGTTGTAGTAC TTTCTTTTGAACTTTTAAATGCACCGGCCACGGTTTGTGGACCAAAATTATCCACTGACCTTATT AAGAACCAGTGTGTCAATTTTAATTTTAATGGACTCACTGGTACTGGTGTGTTAACTCCTTCTTC AAAGAGATTTCAACCATTTCAACAATTTGGCCGTGATGTTTCTGATTTCACTGATTCCGTTCGAG ATCCTAAAACATCTGAAATATTAGACATTTCACCTTGCGCTTTTGGGGGTGTAAGTGTAATTACA CCTGGAACAAATGCTTCATCTGAAGTTGCTGTTCTATATCAAGATGTTAACTGCACTGATGTTTC TACAGCAATTCATGCAGATCAACTCACACCAGCTTGGCGCATATATTCTACTGGAAACAATGTAT TCCAGACTCAAGCAGGCTGTCTTATAGGAGCTGAGCATGTCGACACTTCTTATGAGTGCGACATT CCTATTGGAGCTGGCATTTGTGCTAGTTACCATACAGTTTCTTTATTACGTAGTACTAGCCAAAA ATCTATTGTGGCTTATACTATGTCTTTAGGTGCTGATAGTTCAATTGCTTACTCTAATAACACCA TTGCTATACCTACTAACTTTTCAATTAGCATTACTACAGAAGTAATGCCTGTTTCTATGGCTAAA ACCTCCGTAGATTGTAATATGTACATCTGCGGAGATTCTACTGAATGTGCTAATTTGCTTCTCCA ATATGGTAGCTTTTGCACACAACTAAATCGTGCACTCTCAGGTATTGCTGCTGAACAGGATCGCA ACACACGTGAAGTGTTCGCTCAAGTCAAACAAATGTACAAAACCCCAACTTTGAAATATTTTGGT GGTTTTAATTTTTCACAAATATTACCTGACCCTCTAAAGCCAACTAAGAGGTCTTTTATTGAGGA CTTGCTCTTTAATAAGGTGACACTCGCTGATGCTGGCTTCATGAAGCAATATGGCGAATGCCTAG GTGATATTAATGCTAGAGATCTCATTTGTGCGCAGAAGTTCAATGGACTTACAGTGTTGCCACCT CTGCTCACTGATGATATGATTGCTGCCTACACTGCTGCTCTAGTTAGTGGTACTGCCACTGCTGG ATGGACATTTGGTGCTGGCGCTGCTCTTCAAATACCTTTTGCTATGCAAATGGCATATAGGTTCA ATGGCATTGGAGTTACCCAAAATGTTCTCTATGAGAACCAAAAACAAATCGCCAACCAATTTAAC AAGGCGATTAGTCAAATTCAAGAATCACTTACAACAACATCAACTGCATTGGGCAAGCTGCAAGA CGTTGTTAACCAGAATGCTCAAGCATTAAACACACTTGTTAAACAACTTAGCTCTAATTTTGGTG CAATTTCAAGTGTGCTAAATGATATCCTTTCGCGACTTGATAAAGTCGAGGCGGAGGTACAAATT GACAGGTTAATTACAGGCAGACTTCAAAGCCTTCAAACCTATGTAACACAACAACTAATCAGGGC TGCTGAAATCAGGGCTTCTGCTAATCTTGCTGCTACTAAAATGTCTGAGTGTGTTCTTGGACAAT CAAAAAGAGTTGACTTTTGTGGAAAGGGCTACCACCTTATGTCCTTCCCACAAGCAGCCCCGCAT GGTGTTGTCTTCCTACATGTCACGTATGTGCCATCCCAGGAGAGGAACTTCACCACAGCGCCAGC AATTTGTCATGAAGGCAAAGCATACTTCCCTCGTGAAGGTGTTTTTGTGTTTAATGGCACTTCTT GGTTTATTACACAGAGGAACTTCTTTTCTCCACAAATAATTACTACAGACAATACATTTGTCTCA GGAAATTGTGATGTCGTTATTGGCATCATTAACAACACAGTTTATGATCCTCTGCAACCTGAGCT TGACTCATTCAAAGAAGAGCTGGACAAGTACTTCAAAAATCATACATCACCAGATGTTGATCTTG GCGACATTTCAGGCATTAACGCTTCTGTCGTCAACATTCAAAAAGAAATTGACCGCCTCAATGAG GTCGCTAAAAATTTAAATGAATCACTCATTGACCTTCAAGAATTGGGAAAATATGAGCAATATAT TAAATGGCCTTGGTATGTTTGGCTCGGCTTCATTGCTGGACTAATTGCCATCGTCATGGTTACAA TCTTGCTTTGTTGCATGACTAGTTGTTGCAGTTGCCTCAAGGGTGCATGCTCTTGTGGTTCTTGC TGCAAGTTTGATGAGGATGACTCTGAGCCAGTTCTCAAGGGTGTCAAATTACATTACACATAA

Middle East respiratory syndrome-related coronavirus isolate HCoV-EMC/2012, complete genome	NC_019843.3	>NC_019843.3:21456-25517 Middle East respiratory syndrome-related coronavirus isolate HCoV-EMC/2012, complete genome ATGATACACTCAGTGTTTCTACTGATGTTCTTGTTAACACCTACAGAAAGTTACGTTGATGTAGG GCCAGATTCTGTTAAGTCTGCTTGTATTGAGGTTGATATACAACAGACTTTCTTTGATAAAACTT GGCCTAGGCCAATTGATGTTTCTAAGGCTGACGGTATTATATACCCTCAAGGCCGTACATATTCT AACATAACTATCACTTATCAAGGTCTTTTTCCCTATCAGGGAGACCATGGTGATATGTATGTTTA CTCTGCAGGACATGCTACAGGCACAACTCCACAAAAGTTGTTTGTAGCTAACTATTCTCAGGACG TCAAACAGTTTGCTAATGGGTTTGTCGTCCGTATAGGAGCAGCTGCCAATTCCACTGGCACTGTT ATTATTAGCCCATCTACCAGCGCTACTATACGAAAAATTTACCCTGCTTTTATGCTGGGTTCTTC AGTTGGTAATTTCTCAGATGGTAAAATGGGCCGCTTCTTCAATCATACTCTAGTTCTTTTGCCCG ATGGATGTGGCACTTTACTTAGAGCTTTTTATTGTATTCTAGAGCCTCGCTCTGGAAATCATTGT CCTGCTGGCAATTCCTATACTTCTTTTGCCACTTATCACACTCCTGCAACAGATTGTTCTGATGG CAATTACAATCGTAATGCCAGTCTGAACTCTTTTAAGGAGTATTTTAATTTACGTAACTGCACCT TTATGTACACTTATAACATTACCGAAGATGAGATTTTAGAGTGGTTTGGCATTACACAAACTGCT CAAGGTGTTCACCTCTTCTCATCTCGGTATGTTGATTTGTACGGCGGCAATATGTTTCAATTTGC CACCTTGCCTGTTTATGATACTATTAAGTATTATTCTATCATTCCTCACAGTATTCGTTCTATCC AAAGTGATAGAAAAGCTTGGGCTGCCTTCTACGTATATAAACTTCAACCGTTAACTTTCCTGTTG GATTTTTCTGTTGATGGTTATATACGCAGAGCTATAGACTGTGGTTTTAATGATTTGTCACAACT CCACTGCTCATATGAATCCTTCGATGTTGAATCTGGAGTTTATTCAGTTTCGTCTTTCGAAGCAA AACCTTCTGGCTCAGTTGTGGAACAGGCTGAAGGTGTTGAATGTGATTTTTCACCTCTTCTGTCT GGCACACCTCCTCAGGTTTATAATTTCAAGCGTTTGGTTTTTACCAATTGCAATTATAATCTTAC CAAATTGCTTTCACTTTTTTCTGTGAATGATTTTACTTGTAGTCAAATATCTCCAGCAGCAATTG CTAGCAACTGTTATTCTTCACTGATTTTGGATTACTTTTCATACCCACTTAGTATGAAATCCGAT CTCAGTGTTAGTTCTGCTGGTCCAATATCCCAGTTTAATTATAAACAGTCCTTTTCTAATCCCAC ATGTTTGATTTTAGCGACTGTTCCTCATAACCTTACTACTATTACTAAGCCTCTTAAGTACAGCT ATATTAACAAGTGCTCTCGTCTTCTTTCTGATGATCGTACTGAAGTACCTCAGTTAGTGAACGCT AATCAATACTCACCCTGTGTATCCATTGTCCCATCCACTGTGTGGGAAGACGGTGATTATTATAG GAAACAACTATCTCCACTTGAAGGTGGTGGCTGGCTTGTTGCTAGTGGCTCAACTGTTGCCATGA CTGAGCAATTACAGATGGGCTTTGGTATTACAGTTCAATATGGTACAGACACCAATAGTGTTTGC CCCAAGCTTGAATTTGCTAATGACACAAAAATTGCCTCTCAATTAGGCAATTGCGTGGAATATTC CCTCTATGGTGTTTCGGGCCGTGGTGTTTTTCAGAATTGCACAGCTGTAGGTGTTCGACAGCAGC GCTTTGTTTATGATGCGTACCAGAATTTAGTTGGCTATTATTCTGATGATGGCAACTACTACTGT TTGCGTGCTTGTGTTAGTGTTCCTGTTTCTGTCATCTATGATAAAGAAACTAAAACCCACGCTAC TCTATTTGGTAGTGTTGCATGTGAACACATTTCTTCTACCATGTCTCAATACTCCCGTTCTACGC GATCAATGCTTAAACGGCGAGATTCTACATATGGCCCCCTTCAGACACCTGTTGGTTGTGTCCTA GGACTTGTTAATTCCTCTTTGTTCGTAGAGGACTGCAAGTTGCCTCTTGGTCAATCTCTCTGTGC TCTTCCTGACACACCTAGTACTCTCACACCTCGCAGTGTGCGCTCTGTTCCAGGTGAAATGCGCT TGGCATCCATTGCTTTTAATCATCCTATTCAGGTTGATCAACTTAATAGTAGTTATTTTAAATTA AGTATACCCACTAATTTTTCCTTTGGTGTGACTCAGGAGTACATTCAGACAACCATTCAGAAAGT TACTGTTGATTGTAAACAGTACGTTTGCAATGGTTTCCAGAAGTGTGAGCAATTACTGCGCGAGT ATGGCCAGTTTTGTTCCAAAATAAACCAGGCTCTCCATGGTGCCAATTTACGCCAGGATGATTCT GTACGTAATTTGTTTGCGAGCGTGAAAAGCTCTCAATCATCTCCTATCATACCAGGTTTTGGAGG TGACTTTAATTTGACACTTCTAGAACCTGTTTCTATATCTACTGGCAGTCGTAGTGCACGTAGTG CTATTGAGGATTTGCTATTTGACAAAGTCACTATAGCTGATCCTGGTTATATGCAAGGTTACGAT GATTGCATGCAGCAAGGTCCAGCATCAGCTCGTGATCTTATTTGTGCTCAATATGTGGCTGGTTA CAAAGTATTACCTCCTCTTATGGATGTTAATATGGAAGCCGCGTATACTTCATCTTTGCTTGGCA GCATAGCAGGTGTTGGCTGGACTGCTGGCTTATCCTCCTTTGCTGCTATTCCATTTGCACAGAGT ATCTTTTATAGGTTAAACGGTGTTGGCATTACTCAACAGGTTCTTTCAGAGAACCAAAAGCTTAT TGCCAATAAGTTTAATCAGGCTCTGGGAGCTATGCAAACAGGCTTCACTACAACTAATGAAGCTT TTCAGAAGGTTCAGGATGCTGTGAACAACAATGCACAGGCTCTATCCAAATTAGCTAGCGAGCTA TCTAATACTTTTGGTGCTATTTCCGCCTCTATTGGAGACATCATACAACGTCTTGATGTTCTCGA ACAGGACGCCCAAATAGACAGACTTATTAATGGCCGTTTGACAACACTAAATGCTTTTGTTGCAC AGCAGCTTGTTCGTTCCGAATCAGCTGCTCTTTCCGCTCAATTGGCTAAAGATAAAGTCAATGAG TGTGTCAAGGCACAATCCAAGCGTTCTGGATTTTGCGGTCAAGGCACACATATAGTGTCCTTTGT TGTAAATGCCCCTAATGGCCTTTACTTCATGCATGTTGGTTATTACCCTAGCAACCACATTGAGG TTGTTTCTGCTTATGGTCTTTGCGATGCAGCTAACCCTACTAATTGTATAGCCCCTGTTAATGGC TACTTTATTAAAACTAATAACACTAGGATTGTTGATGAGTGGTCATATACTGGCTCGTCCTTCTA TGCACCTGAGCCCATTACCTCCCTTAATACTAAGTATGTTGCACCACAGGTGACATACCAAAACA TTTCTACTAACCTCCCTCCTCCTCTTCTCGGCAATTCCACCGGGATTGACTTCCAAGATGAGTTG GATGAGTTTTTCAAAAATGTTAGCACCAGTATACCTAATTTTGGTTCCCTAACACAGATTAATAC TACATTACTCGATCTTACCTACGAGATGTTGTCTCTTCAACAAGTTGTTAAAGCCCTTAATGAGT CTTACATAGACCTTAAAGAGCTTGGCAATTATACTTATTACAACAAATGGCCGTGGTACATTTGG CTTGGTTTCATTGCTGGGCTTGTTGCCTTAGCTCTATGCGTCTTCTTCATACTGTGCTGCACTGG TTGTGGCACAAACTGTATGGGAAAACTTAAGTGTAATCGTTGTTGTGATAGATACGAGGAATACG ACCTCGAGCCGCATAAGGTTCATGTTCACTAA

Severe acute respiratory syndrome coronavirus 2 isolate Wuhan-Hu-1, complete genome	NC_045512.2	>NC_045512.2:21563-25384 Severe acute respiratory syndrome coronavirus 2 isolate Wuhan-Hu-1, complete genomeATGTTTGTTTTTCTTGTTTTATTGCCACTAGTCTCTAGTCAGTGTGTTAATCTTACAAC CAGAACTCAATTACCCCCTGCATACACTAATTCTTTCACACGTGGTGTTTATTACCCTGACAAAG TTTTCAGATCCTCAGTTTTACATTCAACTCAGGACTTGTTCTTACCTTTCTTTTCCAATGTTACT TGGTTCCATGCTATACATGTCTCTGGGACCAATGGTACTAAGAGGTTTGATAACCCTGTCCTACC ATTTAATGATGGTGTTTATTTTGCTTCCACTGAGAAGTCTAACATAATAAGAGGCTGGATTTTTG GTACTACTTTAGATTCGAAGACCCAGTCCCTACTTATTGTTAATAACGCTACTAATGTTGTTATT AAAGTCTGTGAATTTCAATTTTGTAATGATCCATTTTTGGGTGTTTATTACCACAAAAACAACAA AAGTTGGATGGAAAGTGAGTTCAGAGTTTATTCTAGTGCGAATAATTGCACTTTTGAATATGTCT CTCAGCCTTTTCTTATGGACCTTGAAGGAAAACAGGGTAATTTCAAAAATCTTAGGGAATTTGTG TTTAAGAATATTGATGGTTATTTTAAAATATATTCTAAGCACACGCCTATTAATTTAGTGCGTGA TCTCCCTCAGGGTTTTTCGGCTTTAGAACCATTGGTAGATTTGCCAATAGGTATTAACATCACTA GGTTTCAAACTTTACTTGCTTTACATAGAAGTTATTTGACTCCTGGTGATTCTTCTTCAGGTTGG ACAGCTGGTGCTGCAGCTTATTATGTGGGTTATCTTCAACCTAGGACTTTTCTATTAAAATATAA TGAAAATGGAACCATTACAGATGCTGTAGACTGTGCACTTGACCCTCTCTCAGAAACAAAGTGTA CGTTGAAATCCTTCACTGTAGAAAAAGGAATCTATCAAACTTCTAACTTTAGAGTCCAACCAACA GAATCTATTGTTAGATTTCCTAATATTACAAACTTGTGCCCTTTTGGTGAAGTTTTTAACGCCAC CAGATTTGCATCTGTTTATGCTTGGAACAGGAAGAGAATCAGCAACTGTGTTGCTGATTATTCTG TCCTATATAATTCCGCATCATTTTCCACTTTTAAGTGTTATGGAGTGTCTCCTACTAAATTAAAT GATCTCTGCTTTACTAATGTCTATGCAGATTCATTTGTAATTAGAGGTGATGAAGTCAGACAAAT CGCTCCAGGGCAAACTGGAAAGATTGCTGATTATAATTATAAATTACCAGATGATTTTACAGGCT GCGTTATAGCTTGGAATTCTAACAATCTTGATTCTAAGGTTGGTGGTAATTATAATTACCTGTAT AGATTGTTTAGGAAGTCTAATCTCAAACCTTTTGAGAGAGATATTTCAACTGAAATCTATCAGGC CGGTAGCACACCTTGTAATGGTGTTGAAGGTTTTAATTGTTACTTTCCTTTACAATCATATGGTT TCCAACCCACTAATGGTGTTGGTTACCAACCATACAGAGTAGTAGTACTTTCTTTTGAACTTCTA CATGCACCAGCAACTGTTTGTGGACCTAAAAAGTCTACTAATTTGGTTAAAAACAAATGTGTCAA TTTCAACTTCAATGGTTTAACAGGCACAGGTGTTCTTACTGAGTCTAACAAAAAGTTTCTGCCTT TCCAACAATTTGGCAGAGACATTGCTGACACTACTGATGCTGTCCGTGATCCACAGACACTTGAG ATTCTTGACATTACACCATGTTCTTTTGGTGGTGTCAGTGTTATAACACCAGGAACAAATACTTC TAACCAGGTTGCTGTTCTTTATCAGGATGTTAACTGCACAGAAGTCCCTGTTGCTATTCATGCAG ATCAACTTACTCCTACTTGGCGTGTTTATTCTACAGGTTCTAATGTTTTTCAAACACGTGCAGGC TGTTTAATAGGGGCTGAACATGTCAACAACTCATATGAGTGTGACATACCCATTGGTGCAGGTAT ATGCGCTAGTTATCAGACTCAGACTAATTCTCCTCGGCGGGCACGTAGTGTAGCTAGTCAATCCA TCATTGCCTACACTATGTCACTTGGTGCAGAAAATTCAGTTGCTTACTCTAATAACTCTATTGCC ATACCCACAAATTTTACTATTAGTGTTACCACAGAAATTCTACCAGTGTCTATGACCAAGACATC AGTAGATTGTACAATGTACATTTGTGGTGATTCAACTGAATGCAGCAATCTTTTGTTGCAATATG GCAGTTTTTGTACACAATTAAACCGTGCTTTAACTGGAATAGCTGTTGAACAAGACAAAAACACC CAAGAAGTTTTTGCACAAGTCAAACAAATTTACAAAACACCACCAATTAAAGATTTTGGTGGTTT TAATTTTTCACAAATATTACCAGATCCATCAAAACCAAGCAAGAGGTCATTTATTGAAGATCTAC TTTTCAACAAAGTGACACTTGCAGATGCTGGCTTCATCAAACAATATGGTGATTGCCTTGGTGAT ATTGCTGCTAGAGACCTCATTTGTGCACAAAAGTTTAACGGCCTTACTGTTTTGCCACCTTTGCT CACAGATGAAATGATTGCTCAATACACTTCTGCACTGTTAGCGGGTACAATCACTTCTGGTTGGA CCTTTGGTGCAGGTGCTGCATTACAAATACCATTTGCTATGCAAATGGCTTATAGGTTTAATGGT ATTGGAGTTACACAGAATGTTCTCTATGAGAACCAAAAATTGATTGCCAACCAATTTAATAGTGC TATTGGCAAAATTCAAGACTCACTTTCTTCCACAGCAAGTGCACTTGGAAAACTTCAAGATGTGG TCAACCAAAATGCACAAGCTTTAAACACGCTTGTTAAACAACTTAGCTCCAATTTTGGTGCAATT TCAAGTGTTTTAAATGATATCCTTTCACGTCTTGACAAAGTTGAGGCTGAAGTGCAAATTGATAG GTTGATCACAGGCAGACTTCAAAGTTTGCAGACATATGTGACTCAACAATTAATTAGAGCTGCAG AAATCAGAGCTTCTGCTAATCTTGCTGCTACTAAAATGTCAGAGTGTGTACTTGGACAATCAAAA AGAGTTGATTTTTGTGGAAAGGGCTATCATCTTATGTCCTTCCCTCAGTCAGCACCTCATGGTGT AGTCTTCTTGCATGTGACTTATGTCCCTGCACAAGAAAAGAACTTCACAACTGCTCCTGCCATTT GTCATGATGGAAAAGCACACTTTCCTCGTGAAGGTGTCTTTGTTTCAAATGGCACACACTGGTTT GTAACACAAAGGAATTTTTATGAACCACAAATCATTACTACAGACAACACATTTGTGTCTGGTAA CTGTGATGTTGTAATAGGAATTGTCAACAACACAGTTTATGATCCTTTGCAACCTGAATTAGACT CATTCAAGGAGGAGTTAGATAAATATTTTAAGAATCATACATCACCAGATGTTGATTTAGGTGAC ATCTCTGGCATTAATGCTTCAGTTGTAAACATTCAAAAAGAAATTGACCGCCTCAATGAGGTTGC CAAGAATTTAAATGAATCTCTCATCGATCTCCAAGAACTTGGAAAGTATGAGCAGTATATAAAAT GGCCATGGTACATTTGGCTAGGTTTTATAGCTGGCTTGATTGCCATAGTAATGGTGACAATTATG CTTTGCTGTATGACCAGTTGCTGTAGTTGTCTCAAGGGCTGTTGTTCTTGTGGATCCTGCTGCAA ATTTGATGAAGACGACTCTGAGCCAGTGCTCAAAGGAGTCAAATTACATTACACATAA

In conducting a multiple sequence alignment of SARS-CoV, MERS-CoV, and SARS-CoV-2's S protein CDS, Clustal Omega provides a 'Percent Identity Matrix' as insight into sequence homology (Table 4). SARS-CoV and SARS-CoV-2 share 74.52% nucleotide sequence similarity in their S proteins, while MERS-CoV and SARS-CoV-2 only share 52.91% sequence similarity. As the S proteins of SARS-CoV and SARS-CoV-2 mediate viral genome entry into host cell through the same receptor, ACE-2, it is expected the sequences coding for the proteins are relatively conserved.

Table 3. Percent Identity Matrix of Spike (S) Protein Nucleotide Homology between SARS-CoV, MERS-CoV, and SARS-CoV-2

	SARS-CoV	**MERS-CoV**	**SARS-CoV-2**
SARS-CoV	100.00	52.96	74.52
MERS-CoV	52.96	100.00	52.91
SARS-CoV-2	74.52	52.91	100.00

The homology of the SARS-CoV, MERS-CoV, and SARS-CoV-2 S protein amino acid sequences were additionally analyzed using GenBank and Clustal Omega. GenBank allowed for identification of each strain's S protein amino acid sequence. This sequence was then exported to Clustal Omega in FASTA format, where the Multiple Sequence Alignment tool was used to analyze for sequence homology and conservation (Table 4).[19-22] The portion of the amino acid sequence corresponding to the receptor binding domain is highlighted in blue, as these specific sequences were also analyzed for homology and conservation.

Table 4. Analyzing Amino Acid Homology of SARS-CoV, MERS-CoV, and SARS-CoV-2 Spike (S) Protein CDS

GenBank Entry	NCBI Reference Sequence	S Protein CDS FASTA
spike glycoprotein [SARS coronavirus Tor2]	Entire S Protein and Receptor Binding Domain: YP_009825051.1	>YP_009825051.1 spike glycoprotein [SARS coronavirus Tor2] MFIFLLFLTLTSGSDLDRCTTFDDVQAPNYTQHTSSMRGVYYPDEIFRSDTLYLTQDLFLPFYSNVTG FHTINHTFGNPVIPFKDGIYFAATEKSNVVRGWVFGSTMNNKSQSVIIINNSTNVVIRACNFELCDNP FFAVSKPMGTQTHTMIFDNAFNCTFEYISDAFSLDVSEKSGNFKHLREFVFKNKDGFLYVYKGYQPID VVRDLPSGFNTLKPIFKLPLGINITNFRAILTAFSPAQDIWGTSAAAYFVGYLKPTTFMLKYDENGTI TDAVDCSQNPLAELKCSVKSFEIDKGIYQTSNFRVVPSGDVVRFPNITNLCPFGEVFNATKFPSVYAW ERKKISNCVADYSVLYNSTFFSTFKCYGVSATKLNDLCFSNVYADSFVVKGDDVRQIAPGQTGVIADY NYKLPDDFMGCVLAWNTRNIDATSTGNYNYKYRYLRHGKLRPFERDISNVPFSPDGKPCTPPALNCYW PLNDYGFYTTTGIGYQPYRVVVLSFELLNAPATVCGPKLSTDLIKNQCVNFNFNGLTGTGVLTPSSKR FQPFQQFGRDVSDFTDSVRDPKTSEILDISPCAFGGVSVITPGTNASSEVAVLYQDVNCTDVSTAIHA DQLTPAWRIYSTGNNVFQTQAGCLIGAEHVDTSYECDIPIGAGICASYHTVSLLRSTSQKSIVAYTMS LGADSSIAYSNNTIAIPTNFSISITTEVMPVSMAKTSVDCNMYICGDSTECANLLLQYGSFCTQLNRA LSGIAAEQDRNTREVFAQVKQMYKTPTLKYFGGFNFSQILPDPLKPTKRSFIEDLLFNKVTLADAGFM KQYGECLGDINARDLICAQKFNGLTVLPPLLTDDMIAAYTAALVSGTATAGWTFGAGAALQIPFAMQM AYRFNGIGVTQNVLYENQKQIANQFNKAISQIQESLTTTSTALGKLQDVVNQNAQALNTLVKQLSSNF GAISSVLNDILSRLDKVEAEVQIDRLITGRLQSLQTYVTQQLIRAAEIRASANLAATKMSECVLGQSK RVDFCGKGYHLMSFPQAAPHGVVFLHVTYVPSQERNFTTAPAICHEGKAYFPREGVFVFNGTSWFITQ RNFFSPQIITTDNTFVSGNCDVVIGIINNTVYDPLQPELDSFKEELDKYFKNHTSPDVDLGDISGINA SVVNIQKEIDRLNEVAKNLNESLIDLQELGKYEQYIKWPWYVWLGFIAGLIAIVMVTILLCCMTSCCS CLKGACSCGSCCKFDEDDSEPVLKGVKLHYT
spike protein [Middle East respiratory syndrome-related coronavirus]	Entire S Protein: YP_009047204.1 Receptor Binding Domain: AAP41037.1	>YP_009047204.1 spike protein [Middle East respiratory syndrome-related coronavirus] MIHSVFLLMFLLTPTESYVDVGPDSVKSACIEVDIQQTFFDKTWPRPIDVSKADGIIYPQGRTYSNIT ITYQGLFPYQGDHGDMYVYSAGHATGTTPQKLFVANYSQDVKQFANGFVVRIGAAANSTGTVIISPST SATIRKIYPAFMLGSSVGNFSDGKMGRFFNHTLVLLPDGCGTLLRAFYCILEPRSGNHCPAGNSYTSF ATYHTPATDCSDGNYNRNASLNSFKEYFNLRNCTFMYTYNITEDEILEWFGITQTAQGVHLFSSRYVD LYGGNMFQFATLPVYDTIKYYSIIPHSIRSIQSDRKAWAAFYVYKLQPLTFLLDFSVDGYIRRAIDCG FNDLSQLHCSYESFDVESGVYSVSSFEAKPSGSVVEQAEGVECDFSPLLSGTPPQVYNFKRLVFTNCN YNLTKLLSLFSVNDFTCSQISPAAIASNCYSSLILDYFSYPLSMKSDLSVSSAGPISQFNYKQSFSNP TCLILATVPHNLTTITKPLKYSYINKCSRLLSDDRTEVPQLVNANQYSPCVSIVPSTVWEDGDYYRKQ LSPLEGGGWLVASGSTVAMTEQLQMGFGITVQYGTDTNSVCPKLEFANDTKIASQLGNCVEYSLYGVS GRGVFQNCTAVGVRQQRFVYDAYQNLVGYYSDDGNYYCLRACVSVPVSVIYDKETKTHATLFGSVACE HISSTMSQYSRSTRSMLKRRDSTYGPLQTPVGCVLGLVNSSLFVEDCKLPLGQSLCALPDTPSTLTPR SVRSVPGEMRLASIAFNHPIQVDQLNSSYFKLSIPTNFSFGVTQEYIQTTIQKVTVDCKQYVCNGFQK CEQLLREYGQFCSKINQALHGANLRQDDSVRNLFASVKSSQSSPIIPGFGGDFNLTLLEPVSISTGSR SARSAIEDLLFDKVTIADPGYMQGYDDCMQQGPASARDLICAQYVAGYKVLPPLMDVNMEAAYTSSLL GSIAGVGWTAGLSSFAAIPFAQSIFYRLNGVGITQQVLSENQKLIANKFNQALGAMQTGFTTTNEAFQ KVQDAVNNNAQALSKLASELSNTFGAISASIGDIIQRLDVLEQDAQIDRLINGRLTTLNAFVAQQLVR SESAALSAQLAKDKVNECVKAQSKRSGFCGQGTHIVSFVVNAPNGLYFMHVGYYPSNHIEVVSAYGLC DAANPTNCIAPVNGYFIKTNNTRIVDEWSYTGSSFYAPEPITSLNTKYVAPQVTYQNISTNLPPPLLG NSTGIDFQDELDEFFKNVSTSIPNFGSLTQINTTLLDLTYEMLSLQQVVKALNESYIDLKELGNYTYY NKWPWYIWLGFIAGLVALALCVFFILCCTGCGTNCMGKLKCNRCCDRYEEYDLEPHKVHVH
surface glycoprotein [Severe acute respiratory syndrome coronavirus 2]	Entire S Protein and Receptor Binding Domain: YP_009724390.1	>YP_009724390.1 surface glycoprotein [Severe acute respiratory syndrome coronavirus 2] MFVFLVLLPLVSSQCVNLTTRTQLPPAYTNSFTRGVYYPDKVFRSSVLHSTQDLFLPFFSNVTWFHAI HVSGTNGTKRFDNPVLPFNDGVYFASTEKSNIIRGWIFGTTLDSKTQSLLIVNNATNVVIKVCEFQFC NDPFLGVYYHKNNKSWMESEFRVYSSANNCTFEYVSQPFLMDLEGKQGNFKNLREFVFKNIDGYFKIY SKHTPINLVRDLPQGFSALEPLVDLPIGINITRFQTLLALHRSYLTPGDSSSGWTAGAAAYYVGYLQP RTFLLKYNENGTITDAVDCALDPLSETKCTLKSFTVEKGIYQTSNFRVQPTESIVRFPNITNLCPFGE VFNATRFASVYAWNRKRISNCVADYSVLYNSASFSTFKCYGVSPTKLNDLCFTNVYADSFVIRGDEVR QIAPGQTGKIADYNYKLPDDFTGCVIAWNSNNLDSKVGGNYNYLYRLFRKSNLKPFERDISTEIYQAG STPCNGVEGFNCYFPLQSYGFQPTNGVGYQPYRVVVLSFELLHAPATVCGPKKSTNLVKNKCVNFNFN GLTGTGVLTESNKKFLPFQQFGRDIADTTDAVRDPQTLEILDITPCSFGGVSVITPGTNTSNQVAVLY QDVNCTEVPVAIHADQLTPTWRVYSTGSNVFQTRAGCLIGAEHVNNSYECDIPIGAGICASYQTQTNS PRRARSVASQSIIAYTMSLGAENSVAYSNNSIAIPTNFTISVTTEILPVSMTKTSVDCTMYICGDSTE CSNLLLQYGSFCTQLNRALTGIAVEQDKNTQEVFAQVKQIYKTPPIKDFGGFNFSQILPDPSKPSKRS FIEDLLFNKVTLADAGFIKQYGDCLGDIAARDLICAQKFNGLTVLPPLLTDEMIAQYTSALLAGTITS GWTFGAGAALQIPFAMQMAYRFNGIGVTQNVLYENQKLIANQFNSAIGKIQDSLSSTASALGKLQDVV NQNAQALNTLVKQLSSNFGAISSVLNDILSRLDKVEAEVQIDRLITGRLQSLQTYVTQQLIRAAEIRA SANLAATKMSECVLGQSKRVDFCGKGYHLMSFPQSAPHGVVFLHVTYVPAQEKNFTTAPAICHDGKAH FPREGVFVSNGTHWFVTQRNFYEPQIITTDNTFVSGNCDVVIGIVNNTVYDPLQPELDSFKEELDKYF KNHTSPDVDLGDISGINASVVNIQKEIDRLNEVAKNLNESLIDLQELGKYEQYIKWPWYIWLGFIAGL IAIVMVTIMLCCMTSCCSCLKGCCSCGSCCKFDEDDSEPVLKGVKLHYT

Again, the 'Percent Identity Matrix' created by Clustal Omega was used for analysis of the S protein and corresponding receptor binding domain amino acid homology between SARS-CoV, MERS-CoV, and SARS-CoV-2 (Table 5). As expected, SARS-CoV shares 77.38% of S protein amino acid sequence identity and 74.41% of receptor binding domain amino acid sequence identity with SARS-CoV-2.

Table 5. Percent Identity Matrix of SARS-CoV, MERS-CoV, and SARS-CoV-2 Entire Spike (S) Protein and Receptor Binding Domain Amino Acid Sequences

Format: (Percent Identity of Entire S Protein, Percent Identity of Receptor Binding Domain)

	SARS-CoV	**MERS-CoV**	**SARS-CoV-2**
SARS-CoV	(100.00, 100.00)	(32.27, 21.67)	(77.38, 74.41)
MERS-CoV	(32.27, 21.67)	(100.00, 100.00)	(31.93, 20.83)
SARS-CoV-2	(77.38, 74.41)	(31.93, 20.83)	(100.00, 100.00)

While exhibiting homology in the receptor binding domain, SARS-CoV and SARS-CoV-2 differ in five of the six amino acids identified as essential binding sites for ACE-2. These six residues are Y442, L472, N479, D480, T487, and Y491 in SARS-CoV and L455, F486, Q493, S494, N501, and Y505 in SARS-CoV-2. The property conservation of these five differing amino acid were investigated through Clustal Omega (Figure 1).[22]

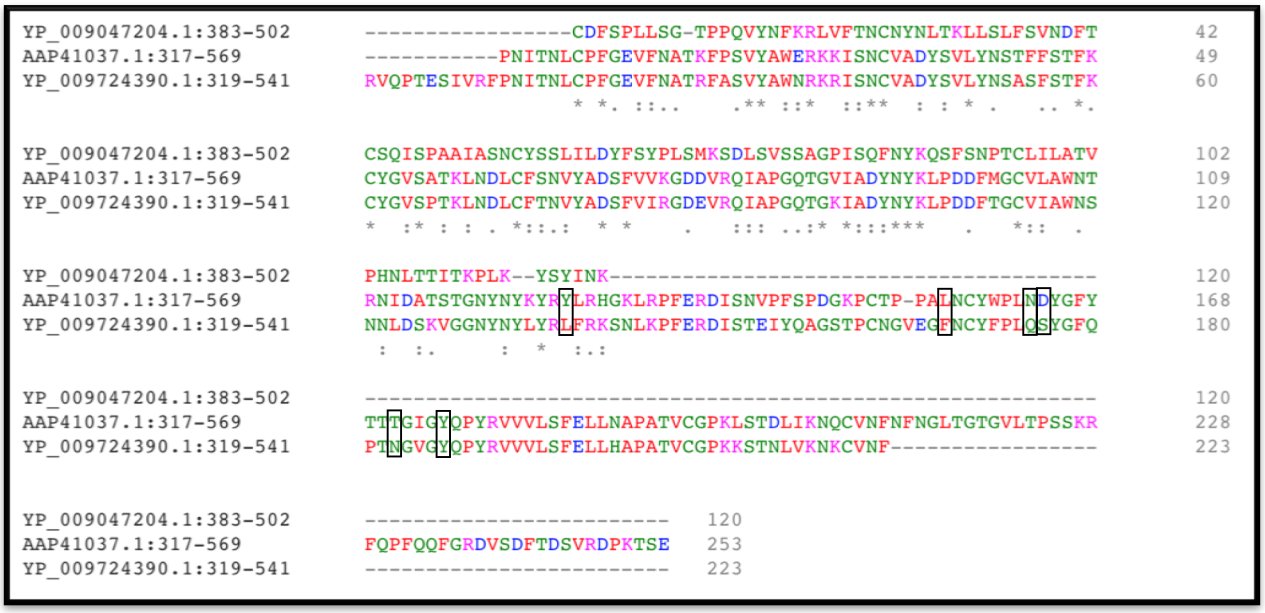

Figure 1. Multiple Sequence Alignment of SARS-CoV, MERS-CoV, and SARS-CoV-2 Spike (S) Protein Receptor Binding Domain

The alignment of SARS-CoV (AAP41037.1:317-569) and SARS-CoV-2 (YP_009724390.1) revealed conservation of property in three of the five amino acids essential for binding to human ACE-2. Specifically, the aligned amino acids L472 (SARS-CoV) and F486 (SARS-CoV-2) are both small, hydrophobic residues. N479 (SARS-CoV) and F493 (SARS-CoV-2) are both polar residues, as are T487 (SARS-CoV) and N501 (SARS-CoV-2). In terms of non-conserved residues, Y442 (SARS-CoV) and F455 (SARS-CoV-2), as well as D480(SARS-CoV) and S494 (SARS-CoV-2).[22] The discrepancy in amino acid property suggests either an evolutionary adaptation of SARS-CoV-2 to increase binding to ACE-2 or decreased importance of these two amino acid residues in binding to ACE-2.

Due to their role in mediating binding of virion to host cell, S proteins have been at the forefront of research regarding coronavirus vaccine development. Analyzing the homology of SARS-CoV, MERS-CoV, and SARS-CoV-2 S proteins provides greater insight into respective vaccine development efforts, specifically regarding the value provided by the two former.

Chapter 2

Coronavirus Vaccine Development

While there are no commercially available, United States Food and Drug Administration (FDA) approved vaccines against SARS-CoV and MERS-CoV, vaccines of various forms have been developed and tested in preclinical models, a few entering into the clinical trial stage (Phase 1). Vaccines based on live-attenuated virus, whole-inactivated virus, protein subunits, virus-like particles, DNA, and viral vectors were studied through SARS-CoV and MERS-CoV vaccine development efforts, primarily focused on using S proteins to elicit immune response. The most promising of these approaches being DNA vaccines and viral vector vaccines.[23, 24]

As for SARS-CoV-2, three vaccines, Pfizer-BioNTech, Moderna, and Johnson & Johnson/Janssen, are commercially available and FDA approved, the former two being mRNA vaccines, the latter being viral vector. These vaccines also use S proteins to elicit immune response.[25]

Information Provided by SARS-CoV and MERS Vaccine Development

The SARS-CoV and MERS-CoV DNA vaccines reaching the clinical trial phase contain a DNA plasmid encoding either full-length S proteins or the S1 subunit specifically. Using electroporation, administration of a short, local electrical pulse, the DNA plasmid enters cells and produces target protein/antigen to induce the production of neutralizing antibodies and activation of killer T cells. With success of these vaccines observed in both animal models (SARS-CoV and MERS-CoV) and humans (SARS-CoV), several companies investigated this

method to produce a DNA vaccine against SARS-CoV-2, in which four of these DNA vaccines entered the clinical trial stage. [23, 24]

The viral vector vaccines against SARS-CoV and MERS-CoV employ similar methodology to DNA vaccines, producing either the full-length S proteins or just the S1 subunit to induce immune response. However, in viral vector vaccines, the DNA encoding these proteins is incorporated into viral vectors as to exploit the infectivity of viruses, inducing a stronger immune response. The most commonly used, well studied viruses used in viral vector vaccines are adenovirus, modified vaccinia virus Ankara, and Venezuelan equine encephalitis virus. Vaccines based on these viral vectors demonstrated the ability to induce immune response against SARS-CoV and MERS-CoV, in addition to vaccines based on parainfluenza virus, vesicular stomatitis virus, Measle virus, and rabies virus. [23, 24] Looking at the development of viral vector SARS-CoV-2 vaccines, vaccines based on adenovirus, modified vaccinia virus Ankara, parainfluenza virus, vesicular stomatitis virus, Measle virus, and rabies virus were explored due to the extensive studies performed for SARS-CoV and MERS-CoV viral vector vaccine development. It is interesting to note that of the twelve SARS-CoV-2 viral vector vaccines reaching clinical trial phase, eight of these vaccines are based on adenovirus, with the Johnson & Johnson/Janssen viral vector vaccine receiving FDA approval for commercial distribution. [25]

It is evident that SARS-CoV and MERS-CoV vaccine development efforts provided valuable information regarding the importance of the S protein/S1 subunit in eliciting immune response against coronaviruses, expediting the process of SARS-CoV-2 vaccine development.

Implications of SARS-CoV and MERS-CoV Vaccine Development Efforts in SARS-CoV-2 Vaccine Development

As mentioned above, three vaccines against SARS-CoV-2 have been approved by the FDA and is being distributed in record numbers, today. The first two of these three vaccines produced, Pfizer-BioNTech and Moderna, are mRNA based, encoding mutant S protein to produce neutralizing antibodies and elicit immune response. This approach to vaccine development is relatively new and highly adaptable to new pathogens due to their ability to be developed using solely genomic sequence information. Both Pfizer-BioNTech and Moderna exploit this property of mRNA vaccines to encode for a mutant form of SARS-CoV-2 S protein, S-2P, that was studied in MERS-CoV vaccine development. [25] It was discovered that the mutation of two amino acid resides, V1060 and L1061, to proline resulted in a stabilized, pre-conformational change structure which has been shown to increase the immune response elicited by vaccination.[18] This mutant S protein was also encoded in the adenovirus used in the Johnson & Johnson/Janssen viral vector vaccine, a technique seen in both SARS-CoV and MERS-CoV vaccine development. [25]

It is important to note that while there are currently only three commercial vaccines against SARS-CoV-2 authorized by the FDA, over 150 SARS-CoV-2 vaccines have been developed, 77 of these vaccines still in the preclinical stage and 86 in the clinical trial stage.[26] Vaccine platforms of all types were investigated as potential treatment for SARS-CoV-2, with several whole-inactivated virus, protein subunit, DNA, RNA, and viral vector vaccines entering the final stages of clinical testing. The majority of these vaccines exploit the entire S protein, an S protein subunit, or the S protein receptor binding domain as an antigen to produce neutralizing antibodies and induce immune response. [25] Again, the importance of the S protein to

coronavirus infection is emphasized, as is the value provided by SARS-CoV and MERS-CoV vaccine development.

In addition to providing insight into the importance of the S protein in SARS-CoV-2 vaccine development, SARS-CoV and MERS-CoV vaccine development efforts drew attention to areas in which coronavirus vaccine development could be improved. For example, there were limited animal models suitable for SARS-CoV and MERS-CoV vaccine testing, as the animals showed limited disease manifestation and viral replication. [25] The implication of this finding in the development of SARS-CoV-2 vaccines is illustrated in the direct entrance of many vaccine candidates into clinical trials, entirely skipping preclinical testing. Additionally, many phases of clinical testing were combined so to further expedite the commercial release of a SARS-CoV-2 vaccine. The knowledge gained from vaccine development efforts for SARS-CoV and MERS-CoV allowed for the rapid development of successful SARS-CoV-2 vaccines. [25]

While the vaccine development efforts of SARS-CoV and MERS-CoV proved to be of extremely value in the successful, rapid development of a commercial SARS-CoV-2 vaccine, the expedited development process is largely due to the global impact of the COVID-19 pandemic. Further research was conducted to investigate the role government funding played in this rapid vaccine development.

Chapter 3

Investigating Government Involvement in Vaccine Development

The National Institutes of Health (NIH), founded in 1887, is an agency of the Public Health Service of the United States Department of Health and Human Services. [27] As the largest biomedical research agency in the world, the NIH is comprised of twenty-seven Institutes and Centers that conduct or support research in various fields of medicine with the goal of applying obtained knowledge to enhance health, reduce illness and disability, and prevent disease.[28, 29] The United States Congress provides twenty-four of these Institutes and Centers with direct funding, specifically providing the NIH with a 3% funding increase for the 2021 fiscal year that resulted in a total yearly budget of $42.93B.[30] This relationship was exploited to investigate government funding of SARS-CoV, MERS-CoV, and SARS-CoV-2 vaccine development, as the successful development of SARS-CoV-2 vaccines can be linked to the research done for SARS-CoV and MERS-CoV vaccine development.

On top of its dedication to promoting the health of the nation, the NIH highly regards public accountability and has integrated this into its mission as a U.S. federal agency.[31] In supporting this portion of its mission, the NIH allows public access to reports, data, and analyses of its research activities through the Research Portfolio Online Reporting Tools (RePORT). One of the most comprehensive tools included under the umbrella of RePORT is the RePORT Expenditures and Results (RePORTER) module, an electronic database available for users to investigate intramural and extramural NIH-funded research projects from 1985 to present.[32] Users are able to navigate RePORTER via an Advanced Projects Search, inputting specific data or key terms to narrow results. One can narrow the search based on 'Researcher and Organization' by selecting specific Fiscal Year(s), Principal Investigator(s), Organization, City,

State, Country, Congressional District, Department Type, and/or Organization Type. Users are also able to perform a 'Text Search,' in which RePORTER results are limited to projects including the entered key word(s) in their title, terms, and/or abstract. Finally, results can be further narrowed by specifying Agency/Institute/Center, NIH Spending Category, Funding Mechanism, Award Type, Award Size, Project Number/Application ID, Activity Code, Study Section, Program Officer, Project Start/End Date, Award Notice Date, and/or Funding Opportunity Announcement (FOA).[33] RePORTER allows users to customize the scope of their search, ranging from general to extremely specific.

NIH RePORTER provides information on projects not only funded by the NIH, but also on research projects funded by the Centers for Disease Control and Prevention (CDC), U.S. Department of Veterans Affairs (VA), Agency for Healthcare Research and Quality (AHRQ), Administration for Children and Families (ACF), Health Resources and Services Administration (HRSA), and U.S. Food and Drug Administration (FDA).[32] The inclusion of these United States federal agencies allows for a more comprehensive search of government funded research projects, subsequently allowing for a more accurate conclusion to be drawn regarding the relationship between government funding and specific subsets of biomedical research.

To further analyze the value provided by SARS-CoV and MERS-CoV vaccine development, NIH RePORTER was utilized to investigate how government funding of these developmental efforts compared to that of SARS-CoV-2 vaccine development.

Research Methods

To begin, the 'categories' of vaccine development to compare against the development of a SARS-CoV-2 vaccine was determined, ultimately looking to analyzing the cost of federal funding for each category. As this is a comparative analysis of SARS-CoV, MERS-CoV, and SARS-CoV-2, it was necessary to include SARS-CoV vaccine development and MERS-CoV vaccine development as categories. The categories were further divided after analyzing the developmental efforts for a SARS-CoV, MERS-CoV, and SARS-CoV-2 vaccine, subsequently deciding to narrow the scope of the research by adding the subsets of "____ mRNA Vaccine Development" and "____Viral Vector Vaccine Development," where the blanks were filled with the overarching category title (either SARS-CoV, MERS-CoV, or SARS-CoV-2). To serve as a baseline, or 'control', value to which data regarding federal funding of SARS-CoV-2 mRNA and viral vector vaccine development was compared, the federal funding costs of all mRNA and viral vector vaccine development projects prior to 2019 were also included in the analysis.

Once the intended areas of research were determined, category-specific methods of precisely narrowing RePORTER results to relevant projects were developed, tested, and revised if necessary. The Fiscal Year and Text Search features available through the RePORTER Advanced Reports Search were manipulated for desired results (Table 2).

Table 6. Search Criteria for Each Area of Research

Area of Research	Fiscal Year(s)	Text Search	Text Search Limited To:
mRNA Vaccine Development (pre-SARS-CoV-2)	1985-2018	mRNA vaccine development (and)	Project Title, Project Terms, Project Abstracts
Viral Vector Vaccine Development (pre-SARS-CoV-2)	1985-2018	Viral vector vaccine development (and)	Project Title, Project Terms, Project Abstracts
SARS-CoV Vaccine Development	1985-2018	SARS vaccine development (and)	Project Title, Project Terms, Project Abstracts
-----SARS-CoV mRNA Vaccine Development	1985-2018	SARS mRNA vaccine development (and)	Project Title, Project Terms, Project Abstracts
-----SARS-CoV Viral Vector Vaccine Development	1985-2018	SARS viral vector vaccine development (and)	Project Title, Project Terms, Project Abstracts
MERS-CoV Vaccine Development	1985-2018	MERS vaccine development (and)	Project Title, Project Terms, Project Abstracts
-----MERS-CoV mRNA Vaccine Development	1985-2018	MERS mRNA vaccine development (and)	Project Title, Project Terms, Project Abstracts
-----MERS-CoV Viral Vector Vaccine Development	1985-2018	MERS viral vector vaccine development (and)	Project Title, Project Terms, Project Abstracts
SARS-CoV-2 Vaccine Development	2019-2021	SARS vaccine development (and)	Project Title, Project Terms, Project Abstracts
-----SARS-CoV-2 mRNA Vaccine Development	2019-2021	SARS mRNA vaccine development (and)	Project Title, Project Terms, Project Abstracts
-----SARS-CoV-2 Viral Vector Vaccine Development	2019-2021	SARS viral vector vaccine development (and)	Project Title, Project Terms, Project Abstracts

The Fiscal Year feature of RePORTER Advanced Project Search allows users to tighten results by permitting the selection of specific years in which projects were funded. However, RePORTER records only begin with projects funded in fiscal year 1985 and searches are thereby limited to the years 1985-Present. This feature was utilized to distinguish the pre-COVID-19 areas of research: mRNA Vaccine Development (pre-SARS-CoV-2), Viral Vector Vaccine Development (pre-SARS-CoV-2), SARS-CoV (mRNA/Viral Vector) Vaccine Development, SARS-CoV mRNA Vaccine Development, SARS-CoV Viral Vector Vaccine Development, MERS-CoV Vaccine Development, MERS-CoV mRNA Vaccine Development, and MERS-CoV Viral Vector Vaccine Development, from the COVID-19 areas of research: SARS-CoV-2 Vaccine Development, SARS-CoV-2 mRNA Vaccine Development, and SARS-CoV-2 Viral Vector Vaccine Development. Each search method was developed and tested with the assumption that all relevant projects included in the pre-COVID-19 areas of research were funded prior to 2019. This assumption primarily prevented the overlap of SARS-CoV and SARS-CoV-2 related projects, as using Text Search alone resulted in the inclusion of SARS-CoV related projects with that of SARS-CoV-2.

To obtain all relevant projects pertaining to mRNA and viral vector vaccine development (pre-SARS-CoV-2), each respective search was conducted for the fiscal years 1985-2018. The same fiscal years were selected when investigating projects pertaining to SARS-CoV and MERS-CoV vaccine development. The decision to end each search of pre-COVID-19 areas of research at fiscal year 2018 was consistent with the assumption that all relevant projects included in the pre-COVID-19 areas of research were funded prior to 2019. While SARS-CoV and MERS-CoV were not of any relevance prior to 2002 and 2012, respectively, no difference was detected in results when the fiscal years were changed. Search methods for these areas were

held consistent with the other pre-COVID-19 areas of research, searching projects funded in fiscal years 1985-2018.

When searching for COVID-19 areas of research, the fiscal years 2019-2021 were selected, seeing that SARS-CoV-2 emerged in 2019 and is still extremely prevalent today. These selections were made as to again stay consistent with the assumption that all relevant projects of the pre-COVID-19 areas of research were funded prior to 2019, meaning all relevant projects of COVID-19 areas of research were awarded funding in fiscal years 2019, 2020, or 2021.

The Text Search feature of NIH RePORTER was utilized to further narrow results to include only projects relevant for each respective area of research. This feature allows users to directly search project titles, abstracts, and/or scientific terms for key words, using either “And,” “Or,” or “Advanced” logic. Through trial and error, it became clear that selecting “Or” logic resulted in the inclusion of non-relevant projects, selecting “Advanced” logic resulted in the exclusion of relevant projects, and selecting “And” logic, the default, resulted in primarily only relevant projects. Therefore, each search was conducted using the default "And” logic, which compiles projects in which all of the search terms are found. The project search was limited to project title, project terms, and project abstracts, as excluding even one of these categories withheld relevant projects from results, producing an incomplete dataset to analyze.

In crafting and testing the specific search terms used for each area of research, it was discovered that including “-CoV(-2)” when searching for SARS-CoV, MERS-CoV, and SARS-CoV-2 related projects inhibited the specificity of RePORTER and resulted in a result list linking all three types of coronaviruses. The “-CoV(-2)” portion of SARS-CoV, MERS-CoV, and SARS-CoV-2 was removed from the respective searches and subsequently increased the

relevancy and specificity of the resulting dataset. This was essentially the only modification made when altering the "Area of Research" title to create search terms.

It is interesting to note that the exact same terms when searching for projects related to SARS-CoV and SARS-CoV-2 vaccine development—"SARS (mRNA/viral vector) vaccine development." This exemplifies the importance of the Fiscal Year feature, as it was the only difference in search criteria when comparing the SARS-CoV (mRNA/viral vector) Vaccine Development area of research with its SARS-CoV-2 counterpart.

As these searches often resulted in hundreds to thousands of entries to sort through, it was difficult and infeasible to ensure relevancy of every individual project. When the search results totaled over 500 projects, projects were randomly selected from the resulting dataset to create a sample group with a sampling fraction of 1%. The abstract of each selected project was read to ensuring relevancy to the specific area of research. If at least 80% of the sample was determined to be relevant, the corresponding study population/ corresponding search results were categorized as 'valid.' If a search resulted in a total project number less than 500, the sampling fraction was altered from 1% to 10% and the threshold to determine validity was also changed from 80% to 90%. Once search results were verified as 'valid,' the cost analysis was conducted.

Using the crafted search criteria, NIH RePORTER was utilized to extract the total cost and total number of projects from the eleven areas of interest. This was easily accomplished by exporting the search results in an Excel format for further analysis. The exported spreadsheet contained data about each individual project's principal investigator(s)/project leader(s), organization, fiscal year, funding institute/center, project number, budget start and end dates, total cost of project/sub project, etc. Although originally confused about the difference in classification as project or sub project, it was determined that entries classified as projects are

NIH funded projects, while sub projects are funded by one of the other federal agencies—the CDC, VA, AHRQ, ACF, HRSA, or FDA—whose funding data is included within RePORTER. The data reported under 'Total Cost' and 'Total Cost (Sub Projects)' was combined/summed to determine the total amount granted to each area of research. Once each area of research's 'Total Amount Funded' was calculated through summation of the individual projects' 'Total Cost,' an average project cost was calculated by dividing the 'Total Amount Funded' by 'Total Project Number' (Table 3, Figure 1).

Results of NIH RePORTER Search

In employing NIH RePORTER to investigate government funding in eleven areas of interest, it was determined that SARS-CoV, MERS-CoV, and SARS-CoV-2 vaccine development had total government funding amounts of $231,105,912, $41,981,175, and $722,282,359, respectively (Table 7). The total amount of federal funding given to each research area was depicted graphically to emphasize the discrepancy in total funding when comparing SARS-CoV, MERS-CoV, and SARS-CoV-2 (Figure 2). Additionally, analysis of average project costs illustrates projects regarding SARS-CoV-2 vaccine development received substantially larger federal grants when compared to the other ten areas of research. It is also important to note the number of fiscal years considered when comparing the total project number of each area of research. For example, when comparing total project number of 'SARS-CoV Vaccine' with that of 'SARS-CoV-2 Vaccine,' the former area of research covers projects spanning the course of fifteen fiscal years (2004-2018), whereas the latter covers projects spanning only three fiscal years (2019-2021). Therefore, the average project number per year is

thirty-one for ‘SARS-CoV Vaccine’ area of research, and 112.667 for ‘SARS-CoV-2 Vaccine’ area of research.

Table 7. Federal Funding of NIH Projects by Area of Research

Area of Research	Total Amt. Funded	Total Project Num.	Avg Project Cost
mRNA Vaccine (pre-SARS-CoV-2)	$ 528,158,597.00	1264	$ 417,846.99
Viral Vector Vaccine (pre-SARS-CoV-2)	$ 2,491,078,936.00	3126	$ 796,890.25
SARS-CoV Vaccine	$ 231,105,912.00	465	$ 497,001.96
SARS-CoV mRNA Vaccine	$ 15,786,178.00	38	$ 415,425.74
SARS-CoV Viral Vector Vaccine	$ 54,771,129.00	103	$ 531,758.53
MERS-CoV Vaccine	$ 41,981,175.00	82	$ 511,965.55
MERS-CoV mRNA Vaccine	$ 1,051,856.00	3	$ 350,618.67
MERS-CoV Viral Vector Vaccine	$ 10,666,650.00	18	$ 592,591.67
SARS-CoV-2 Vaccine	$ 722,282,359.00	338	$ 2,136,930.06
SARS-CoV-2 mRNA Vaccine	$ 14,244,613.00	29	$ 491,193.55
SARS-CoV-2 Viral Vector Vaccine	$ 451,481,743.00	39	$ 11,576,454.95

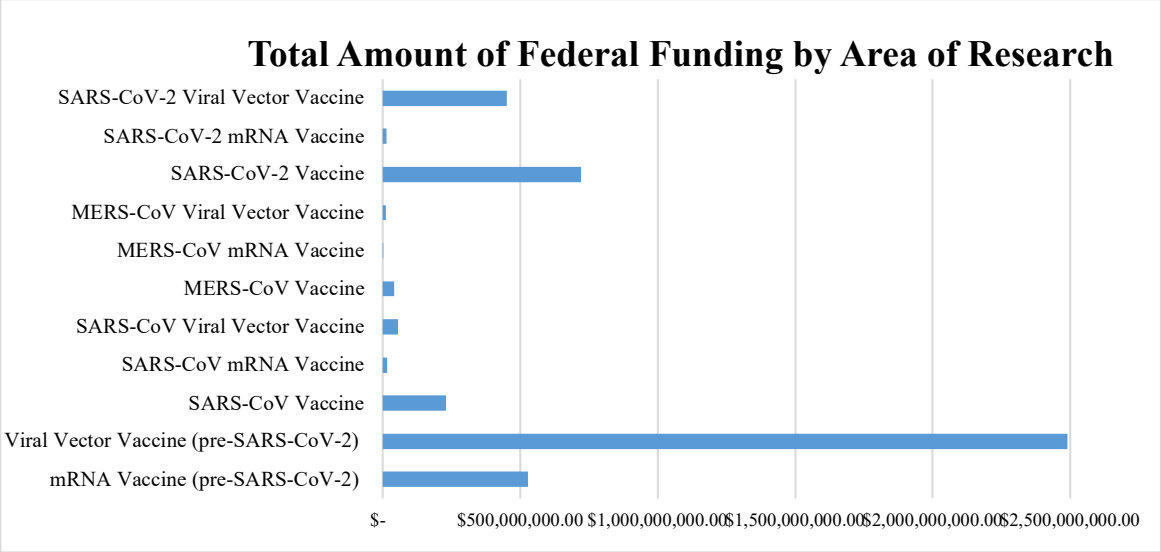

Figure 2. Total Amount of Federal Funding by Area of Research (data from NIH RePORTER)

Total Yearly Funding of Vaccine Development (pre-SARS-CoV-2)

In order to further analyze the relationship between coronavirus vaccine development and government funding, the total yearly funding was graphed for the pre-COVID-19 areas of research (Figures 3-6). The total yearly funding was not graphed for COVID-19 areas of research as SARS-CoV-2 emerged late in 2019, thus the 2020 fiscal year is the only year providing complete data.

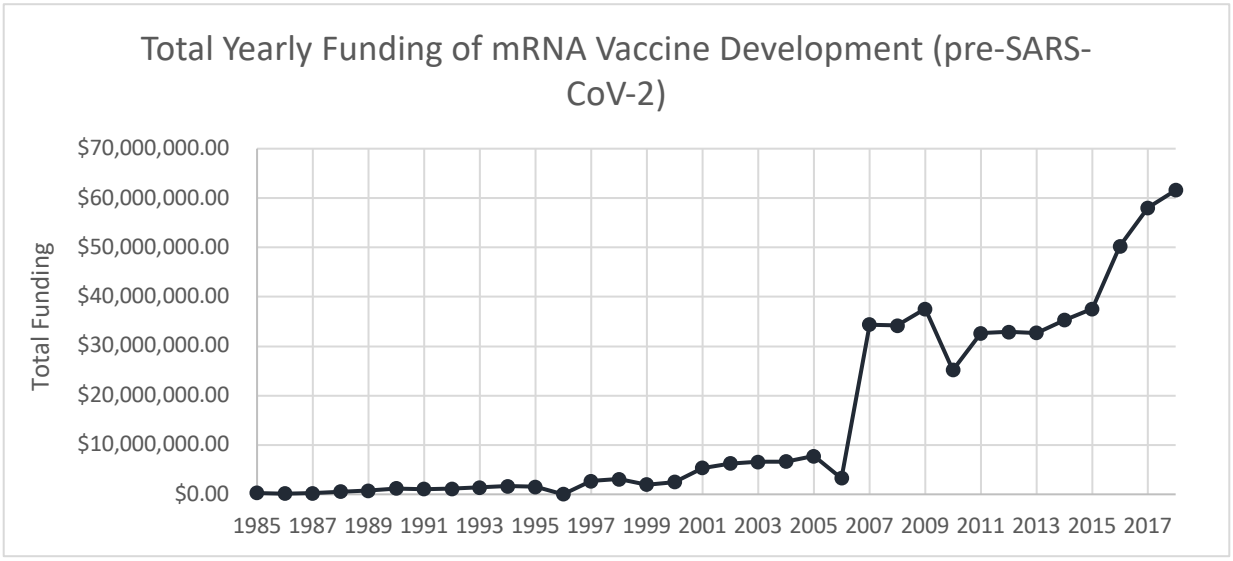

Figure 3. Line Graph Depicting Total Yearly Funding of mRNA Vaccine Development (pre-SARS-CoV-2)

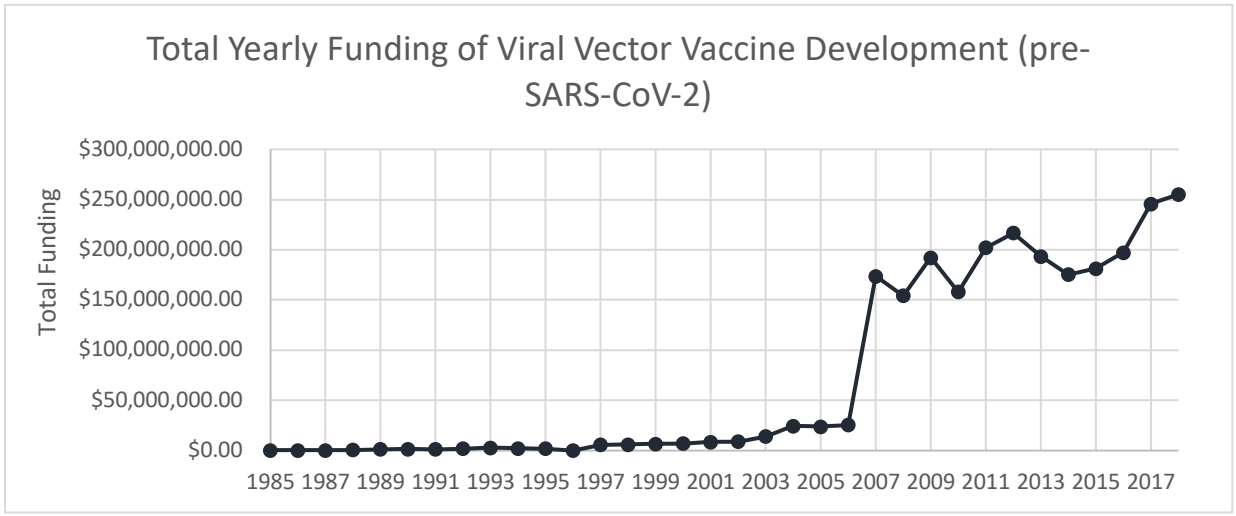

Figure 4. Line Graph Depicting Total Yearly Funding of Viral Vector Vaccine Development (pre-SARS-CoV-2)

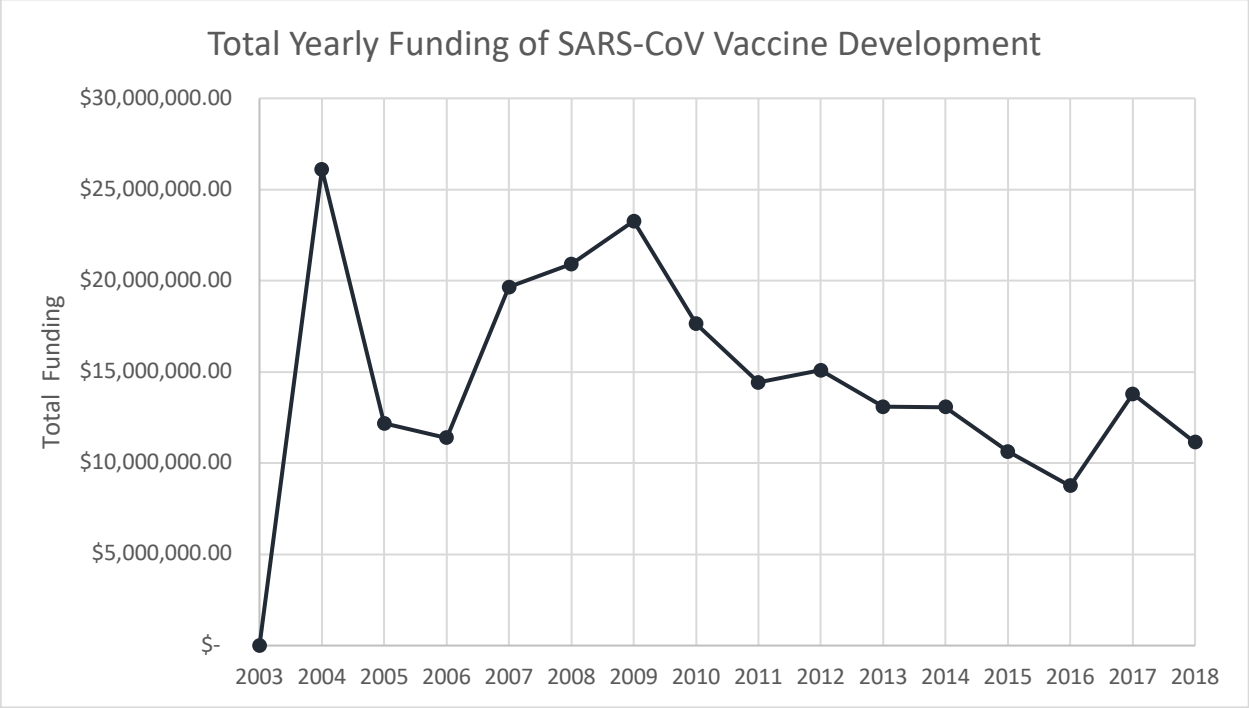

Figure 5. Line Graph Depicting Total Yearly Funding of SARS-CoV Vaccine Development

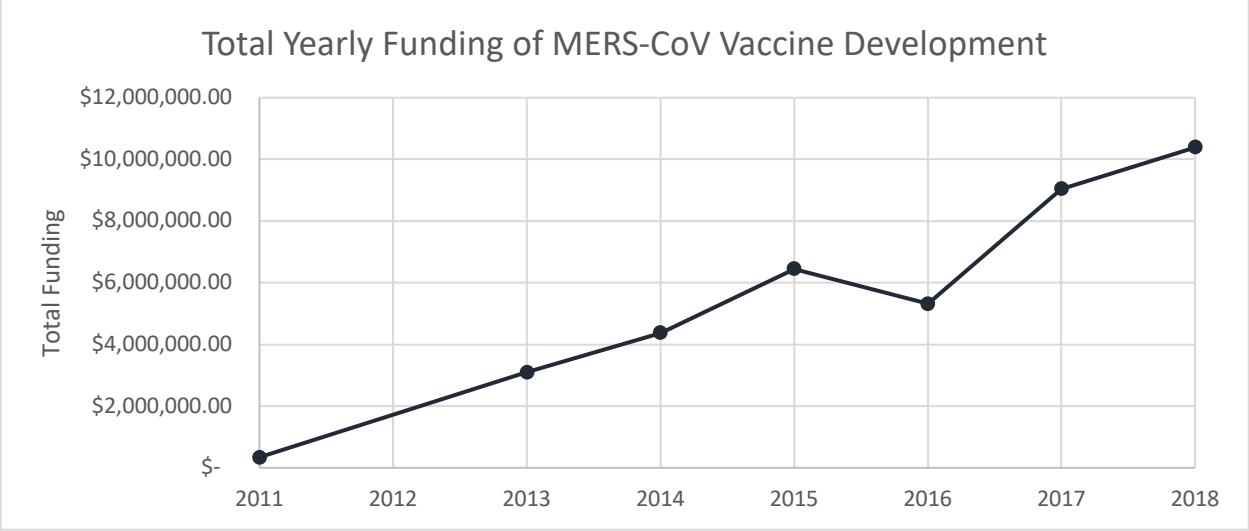

Figure 6. Line Graph Depicting Total Yearly Funding of MERS-CoV Vaccine Development

Both mRNA Vaccine Development (pre-SARS-CoV-2) and Viral Vector Vaccine Development (pre-SARS-CoV-2) research areas saw a relatively drastic increase in yearly funding from 2006 to 2007 (Figures 3, 4). This increase, although less drastic, was also observed in SARS-CoV Vaccine Development (Figure 5). Upon further research, it is likely this increase

in total yearly funding could be linked to the successful vaccines authorized by the FDA in 2006, specifically the RotaTeq (Rotavirus) vaccine, Zostavax (Shingles/Herpes Zoster) vaccine, and Gardasil (Human Papillomavirus types 6, 11, 16, and 20) vaccine.[34] The success of these vaccines could have propelled further vaccine development and increased the likelihood of receiving government funding for such development efforts. Additionally, mRNA vaccines were increasingly researched in the early 2000s, with promising studies released in 2006.[35] The upward trend observed in mRNA Vaccine Development (pre-SARS-CoV-2) and Viral Vector Vaccine Development (pre-SARS-CoV-2) following 2007 likely results from technological advancements allowing for novel research in vaccine development.

Specifically looking at total yearly funding of SARS-CoV and MERS-CoV Vaccine Development, it is interesting to compare the decline in yearly funding of SARS-CoV Vaccine Development following fiscal year 2009 with the relatively steady increase in yearly funding for MERS-CoV following the fiscal year 2011. This is likely due to the fact that there were zero recorded cases of SARS-CoV following 2004, whereas there were seven new cases of MERS-CoV, resulting in three deaths, in March 2021.[36, 37] Despite these recent cases of MERS-CoV and steady increase in total yearly funding for MERS-CoV vaccine development, the 'SARS-CoV Vaccine Development' research area received $750,423 more in the fiscal year 2018.

Chapter 4

Analysis of Government Funding of SARS-CoV-2 Vaccine Development

The severity and global impact of SARS-CoV-2 is directly related to its rapid vaccine development, and subsequently to the amount of government funding received for such rapid vaccine development. Over three years, SARS-CoV-2 Vaccine Development received $722,282,359 in government funding, $665,173,906 funded just in 2020. This year-specific total funding amount is more than double that of the next largest year-specific funding amount—$255,165,583 funded for Viral Vector Vaccine Development (pre-SARS-CoV-2) in 2018, suggesting the U.S. Government's quick identification of SARS-CoV-2 as a large threat to public health and immediate action in vaccine development.

Despite SARS-CoV and MERS-CoV having a larger fatality rate, the determined funding of SARS-CoV-2 Vaccine Development is more than three times larger than the funding of SARS-CoV Vaccine Development and ten times greater than the funding of MERS-CoV, likely due to only eight positive SARS-CoV cases and two positive MERS-CoV cases in the United States, ever (Table 1). However, the larger total number of countries with recorded cases, total number of recorded cases, total number of recorded deaths, and R_0 value associated with SARS-CoV-2 emphasizes the idea that it poses a greater threat to public health than SARS-CoV and MERS-CoV, therefore requiring immediate government action.

BIBLIOGRAPHY

1. Archived: WHO Timeline - COVID-19. https://www.who.int/news/item/27-04-2020-who-timeline---covid-19 (accessed August 12, 2020).

2. Pneumonia of unknown cause – China. https://www.who.int/csr/don/05-january-2020-pneumonia-of-unkown-cause-china/en/ (accessed August 12, 2020).

3. Novel Coronavirus – China. https://www.who.int/csr/don/12-january-2020-novel-coronavirus-china/en/ (accessed August 12, 2020).

4. Naqvi, A. A. T.; Fatima, K.; Mohammad, T.; Fatima, U.; Singh, I. K.; Singh, A.; Atif, S. M.; Hariprasad, G.; Hasan, G. M.; Hassan, M. I., Insights into SARS-CoV-2 genome, structure, evolution, pathogenesis and therapies: Structural genomics approach. *Biochimica Et Biophysica Acta-Molecular Basis of Disease* **2020,** *1866* (10). DOI: 10.1016/j.bbadis.2020.165878

5. WHO Coronavirus (COVID-19) Dashboard. https://covid19.who.int/ (accessed March 31, 2021).

6. Prompetchara, E.; Kettoy, C.; Palaga, T., Immune responses in COVID-19 and potential vaccines: Lessons learned from SARS and MERS epidemic. *Asian Pacific Journal of Allergy and Immunology* **2020,** *38* (1), 1-9. DOI: 10.12932/ap-200220-0772

7. Chams, N.; Chams, S.; Badran, R.; Shams, A.; Araji, A.; Raad, M.; Mukhopadhyay, S.; Stroberg, E.; Duval, E. J.; Barton, L. M.; Hussein, I. H., COVID-19: A Multidisciplinary Review. *Frontiers in Public Health* **2020,** *8*. DOI: 10.3389/fpubh.2020.00383

8. Zhong, N. S.; Zheng, B. J.; Li, Y. M.; Poon, L. L. M.; Xie, Z. H.; Chan, K. H.; Li, P. H.; Tan, S. Y.; Chang, Q.; Xie, J. P.; Liu, X. Q.; Xu, J.; Li, D. X.; Yuen, K. Y.; Peiris, J. S. M.; Guan, Y., Epidemiology and cause of severe acute respiratory syndrome (SARS) in Guangdong, People's Republic of China, in February, 2003. *Lancet* **2003,** *362* (9393), 1353-1358. DOI: 10.1016/s0140-6736(03)14630-2

9. Coronavirus never before seen in humans is the cause of SARS. https://www.who.int/mediacentre/news/releases/2003/pr31/en/ (accessed August 13, 2020).

10. Severe Acute Respiratory Syndrome (SARS) - multi-country outbreak - Update. https://www.who.int/csr/don/2003_03_16/en/ (accessed August 13, 2020).

11. Hui, D. S. C.; Chan, M. C. H.; Wu, A. K.; Ng, P. C., Severe acute respiratory syndrome (SARS): epidemiology and clinical features. *Postgraduate Medical Journal* **2004,** *80* (945), 373-381. DOI: 10.1136/pgmj.2004.020263

12. Ramadan, N.; Shaib, H., Middle East respiratory syndrome coronavirus (MERS-CoV): A review. *Germs* **2019,** *9* (1), 35-42. DOI: 10.18683/germs.2019.1155

13. Al-Tawfiq, J. A.; Gautret, P., Asymptomatic Middle East Respiratory Syndrome Coronavirus (MERS-CoV) infection: Extent and implications for infection control: A systematic review. *Travel Medicine and Infectious Disease* **2019,** *27*, 27-32. DOI: 10.1016/j.tmaid.2018.12.003

14. Kaur, N.; Singh, R.; Dar, Z.; Bijarnia, R. K.; Dhingra, N.; Kaur, T., Genetic comparison among various coronavirus strains for the identification of potential vaccine targets of SARS-CoV2. *Infection, genetics and evolution : journal of molecular epidemiology and evolutionary genetics in infectious diseases* **2020,** *89*, 104490-104490. DOI: 10.1016/j.meegid.2020.104490

15. Abdelghany, T. M.; Ganash, M.; Bakri, M. M.; Qanash, H.; Al-Rajhi, A. M. H.; Elhussieny, N. I., SARS-CoV-2, the other face to SARS-CoV and MERS-CoV: Future predictions. *Biomedical journal* **2020**. DOI: 10.1016/j.bj.2020.10.008

16. Krishnamoorthy, S.; Swain, B.; Verma, R. S.; Gunthe, S. S., SARS-CoV, MERS-CoV, and 2019-nCoV viruses: an overview of origin, evolution, and genetic variations. *VirusDisease* **2020,** *31* (4). DOI: 10.1007/s13337-020-00632-9

17. Dai, L.; Zheng, T.; Xu, K.; Han, Y.; Xu, L.; Huang, E.; An, Y.; Cheng, Y.; Li, S.; Liu, M.; Yang, M.; Li, Y.; Cheng, H.; Yuan, Y.; Zhang, W.; Ke, C.; Wong, G.; Qi, J.; Qin, C.; Yan, J.; Gao, G. F., A Universal Design of Betacoronavirus Vaccines against COVID-19, MERS, and SARS. *Cell* **2020,** *182* (3), 722-+. DOI: 10.1016/j.cell.2020.06.035

18. Pallesen, J.; Wang, N.; Corbett, K. S.; Wrapp, D.; Kirchdoerfer, R. N.; Turner, H. L.; Cottrell, C. A.; Becker, M. M.; Wang, L.; Shi, W.; Kong, W.-P.; Andres, E. L.; Kettenbach, A. N.; Denison, M. R.; Chappell, J. D.; Graham, B. S.; Ward, A. B.; McLellan, J. S., Immunogenicity and structures of a rationally designed prefusion MERS-CoV spike antigen. *Proceedings of the National Academy of Sciences of the United States of America* **2017,** *114* (35), E7348-E7357. DOI: 10.1073/pnas.1707304114

19. Nucleotide [Internet]. Bethesda (MD): National Library of Medicine (US), National Center for Biotechnology Information; [1988] – . Accession No. NC_004718.3, SARS coronavirus Tor2, complete genome; Available from: https://www.ncbi.nlm.nih.gov/nuccore/30271926.

20. Nucleotide [Internet]. Bethesda (MD): National Library of Medicine (US), National Center for Biotechnology Information; [1988] – . Accession No. NC_019843.3 , Middle East respiratory syndrome-related coronavirus isolate HCoV-EMC/2012, complete genome ; Available from: https://www.ncbi.nlm.nih.gov/nuccore/667489388.

21. Nucleotide [Internet]. Bethesda (MD): National Library of Medicine (US), National Center for Biotechnology Information; [1988] – . Accession No. NC_045512.2 , Severe acute respiratory syndrome coronavirus 2 isolate Wuhan-Hu-1, complete genome ; Available from: https://www-ncbi-nlm-nih-gov.ezaccess.libraries.psu.edu/nuccore/1798174254.

22. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega Sievers F, Wilm A, Dineen DG, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H,

Remmert M, Söding J, Thompson JD, Higgins D Molecular Systems Biology 7 Article number: 539. DOI: 10.1038/msb.2011.75

23. Li, Y.-D.; Chi, W.-Y.; Su, J.-H.; Ferrall, L.; Hung, C.-F.; Wu, T. C., Coronavirus vaccine development: from SARS and MERS to COVID-19. *Journal of Biomedical Science* **2020,** *27* (1). DOI: 10.1186/s12929-020-00695-2

24. Padron-Regalado, E., Vaccines for SARS-CoV-2: Lessons from Other Coronavirus Strains. *Infectious Diseases and Therapy* **2020,** *9* (2), 255-274. DOI: 10.1007/s40121-020-00300-x

25. Su, S.; Du, L. Y.; Jiang, S. B., Learning from the past: development of safe and effective COVID-19 vaccines. *Nature Reviews Microbiology* **2021,** *19* (3), 211-219. DOI: 10.1038/s41579-020-00462-y

26. Coronavirus Vaccine Tracker. https://www.nytimes.com/interactive/2020/science/coronavirus-vaccine-tracker.html (accessed March 30, 2021).

27. History. https://www.nih.gov/about-nih/who-we-are/history (accessed February 12, 2021).

28. Organization. https://www.nih.gov/about-nih/who-we-are/organization (accessed February 12, 2021).

29. Mission and Goals. https://www.nih.gov/about-nih/what-we-do/mission-goals (accessed February 12, 2021).

30. Final FY21 Appropriations: National Institutes of Health. https://www.aip.org/fyi/2021/final-fy21-appropriations-national-institutes-health#:~:text=The%20National%20Institutes%20of%20Health%20is%20receiving%20a%203%25%20funding,boost%20of%20over%20%241%20billion (accessed February 12, 2021).

31. Welcome to Research Portfolio Online Reporting Tools (RePORT). https://report.nih.gov/# (accessed February 18, 2021).

32. Frequently Asked Questions (FAQs). https://report.nih.gov/faqs (accessed February 19, 2021).

33. RePORTER. https://reporter.nih.gov/ (accessed February 22, 2021).

34. Vaccine Timeline. https://www.immunize.org/timeline/ (accessed March 26, 2021).

35. Schlake, T.; Thess, A.; Fotin-Mleczek, M.; Kallen, K. J., Developing mRNA-vaccine technologies. *Rna Biology* **2012,** *9* (11), 1319-1330. DOI: 0.4161/rna.22269

36. SARS Basics Fact Sheet. https://www.cdc.gov/sars/about/fs-sars.html (accessed March 25, 2021).

37. MERS in the U.S. https://www.cdc.gov/coronavirus/mers/us.html (accessed March 3, 2021).

CURRENT ADDRESS
State College, PA 16801

Bridget Foresman

bkf7@psu.edu

PERMANENT ADDRESS
Muncy, PA 17756

EDUCATION

The Pennsylvania State University | Schreyer Honors College ***Class of 2021***
Biochemistry and Molecular Biology (B.S.)
Minor: Mathematics Honors: Biochemistry and Molecular Biology
Requirements: 35 honors credits, academic thesis

PROFESSIONAL EXPERIENCE

Deloitte LLP **Pittsburgh, PA**
Incoming Strategy Analyst *Aug 2021*

Cigna | Express Scripts **St. Louis, MO (Remote)**
Regulated Markets Strategy Intern *June 2020 – Aug 2020*

- Accelerated bid strategy for Caremark through performance of Medicare market/competitor analysis using HealthWorksAI
- Constructed a Medicaid Enrollment Database in Excel by compiling and flattening out state-specific enrollment data available through Health Management Associates (HMA)
- Aided company-wide mental wellness effort by creating survey to identify team stressors/gauge stress levels using Adobe Acrobat, distributed to 300+ employees

RESEARCH

Dr. Andrew Belmonte Mathematics Lab **University Park, PA**
Undergraduate Research Assistant *June 2019 – Jan 2020*

- Collaborated with Melissa Rolls Biochemistry and Molecular Biology Lab, which uses genetics and live imaging in Drosophila to investigate neuronal generation of axon and dendrites
- Developed code in Python and MATLAB to construct mathematical models used in predicting the role of Stat92E in regeneration of dendrites and axons

LEADERSHIP EXPERIENCE

Schreyer Honors College Scholar Ambassador **University Park, PA**
Mental Health Task Force Member *May 2017 - Present*

- Served on service committee working to support the "Dr. Michele 'Mitch' Kirsch Fund in Support of Mental Health Initiatives" to implement training of Administrator and Staff within the Honors College
- Volunteered as a student panelist for 18 Accepted Students Programs (ASPs), speaking to 360+ students and their families
- Hosted 6 students over 3 different weekends in an SHC Overnight Program providing students with daily experience of a Schreyer Scholar

Schreyer Honors College Student Council **University Park, PA**
Service Committee Member, Social Chair (April 2018 – May 2020) *Aug 2017 - Present*

- Oversaw $18,000 budget to offer leadership, academic, service, and social opportunities to Schreyer Scholars
- Fundraised $14,000 for THON by communicating and strengthening partnerships with University Entities and Schreyer Honors College Alumni
- Planned and attended 2 Big Ten Academic Alliance Conferences in which leaders of the Big Ten Honors Colleges collaborated and exchanged various academic and professional opportunities and ideas

Penn State Global Medical Brigades **University Park, PA**
Medical Brigade Attendee—Nicaragua 2018, Executive Board (April 2018 – May 2019) *Aug 2017- Present*

- Traveled to a small community outside of San Rafael de Norte, Nicaragua on medical brigade
- Shadowed and assisted a local dentist, physician, and pharmacist for 8 hours a day to learn more of the workings of the Nicaraguan healthcare system

SKILLS AND AWARDS

Skills: Proficient in Python for Mathematical Modeling, Microsoft Office—Microsoft Excel for Data Analysis, PowerPoint, Experience with Alteryx and SQL
Awards: Schreyer Honors College Academic Excellence Scholarship, Daughters of the American Revolution Scholarship, United States Navy Academic Excellence Award