A PROBABILISTIC ASSESSMENT OF SARS-COV-2 HOST INTERACTIONS IN THE CONTEXT OF META-COMMUNITY AND URBAN ECOLOGY

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES
OF
MIDDLE EAST TECHNICAL UNIVERSITY

BY

LEMAN NUR NEHRİ

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF MASTER OF SCIENCE
IN
BIOLOGY

JUNE 2021

Approval of the thesis:

A PROBABILISTIC ASSESSMENT OF SARS-COV-2 HOST INTERACTIONS IN THE CONTEXT OF META-COMMUNITY AND URBAN ECOLOGY

submitted by **LEMAN NUR NEHRİ** in partial fulfillment of the requirements for the degree of **Master of Science** i**n Biology, Middle East Technical University** by,

Prof. Dr. Halil Kalıpçılar
Dean, Graduate School of **Natural and Applied Sciences** _____

Prof. Dr. Ayşe Gül Gözen
Head of the Department, **Biological Sciences** _____

Assist. Prof. Dr. Seçkin Eroğlu
Supervisor, **Biological Sciences, METU** _____

Examining Committee Members:

Prof. Dr. Mayda Gürsel
Biological Sciences, METU _____

Assist. Prof. Dr. Seçkin Eroğlu
Biological Sciences, METU _____

Assist. Prof. Dr. Barbaros Yet
Graduate School of Informatics, METU _____

Assoc. Prof. Dr. Arzu Karahan
Institute of Marine Sciences, METU _____

Assoc. Prof. Dr. Osman Doluca
Biomedical Engineering, Izmir University of Economics _____

Date: 29.06.2021

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name Last name : Leman Nur Nehri

Signature :

ABSTRACT

A PROBABILISTIC ASSESSMENT OF SARS-COV-2 HOST INTERACTIONS IN THE CONTEXT OF META-COMMUNITY AND URBAN ECOLOGY

Nehri, Leman Nur
Master of Science, Biology
Supervisor : Assist. Prof. Dr. Seçkin Eroğlu

June 2021, 141 pages

Coronavirus disease 2019 (COVID-19) is an ongoing pandemic that was detected in Wuhan, China, in December 2019 and spread all around the world. COVID-19 is caused by Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), which spreads through close contact. Many studies have been conducted on the transmission, virulence, and immune response of SARS-CoV-2. The intracellular mechanism of action of the virus and various host interaction pathways are also known. In addition, there are many studies on the mutant types of the virus. However, no study has been found on the microbial host interactions of the specific mutants of the Spike protein, which is one of the most important structural proteins of the virus. This protein allows the virus to enter the cell and is the main target of the ongoing vaccine studies.

In this study, interactions between Spike protein variants and bacteria of gut microbiota were analyzed with a probabilistic programming language (PPL), WebPPL. It is preferred since it is an expressive and generative language that can infer from small data sets. The relationship between the three Spike protein mutants and the two SARS-CoV-2 variants with the intestinal microbiota was also

investigated. As a result, it was found that different variants of Spike protein exist in the hosts that have dissimilar intestinal microbial compositions. Because of the fact that the microbe interplays are very dynamic systems, laboratory applications turn to be quite costly, time-consuming, and difficult in microbial interaction studies. This study is expected to be helpful for the applications of interactions between virus variants and microbiomes for laboratory environments.

Keywords: Probabilistic Programming, Spike protein, SARS-CoV-2, Microbial Meta-Community, Microbiome

ÖZ

META-TOPLULUK VE KENTSEL EKOLOJİ BAĞLAMINDA SARS-COV-2 KONAK ETKİLEŞİMLERİNİN OLASILIKSAL BİR DEĞERLENDİRMESİ

Nehri, Leman Nur
Yüksek Lisans, Biyoloji
Tez Yöneticisi: Dr. Öğr. Üyesi Seçkin Eroğlu

Haziran 2021, 141 sayfa

Koronavirüs hastalığı 2019 (COVID-19), ilk olarak Aralık 2019'da Çin'in Wuhan kentinde tespit edilen ve devam eden bir pandemidir. COVID-19, yakın temas yoluyla yayılan şiddetli akut solunum sendromu koronavirüs 2'den (SARS-CoV-2) kaynaklanır. SARS-CoV-2'nin bulaşması, virülansı ve bağışıklık tepkisi üzerine birçok çalışma yapılmıştır. Virüsün hücre içi etki mekanizması ve çeşitli konak etkileşim yolları da bilinmektedir. Ayrıca virüsün mutant tipleri ile ilgili birçok çalışma bulunmaktadır. Ancak, virüsün en önemli yapısal proteinlerinden biri olan Spike proteininin farklı mutantlarının mikrobiyal konak etkileşimleri ile ilgili herhangi bir çalışma bulunamamıştır. Bu protein, virüsün hücreye girmesini sağlar ve aşı çalışmalarının ana hedefidir.

Bu çalışmada, Spike protein varyantları ile bağırsak mikrobiyotasının bakterileri arasındaki etkileşimler, olasılıksal bir programlama dili (PPL) olan WebPPL ile analiz edilmiştir. Küçük veri setlerinden çıkarım yapabilen açıklayıcı ve generatif bir dil olduğu için bu dil tercih edilmiştir. Ayrıca, üç Spike protein mutantı ve iki SARS-CoV-2 varyantının bağırsak mikrobiyotası ile ilişkisi araştırılmıştır. Sonuç olarak, farklı bağırsak mikrobiyal kompozisyonlarına sahip konaklarda Spike proteininin

farklı varyantlarının mevcut olduğu bulunmuştur. Mikrop etkileşimleri çok dinamik sistemler olduğu için laboratuvar uygulamaları maliyetli, zaman alıcı ve mikrobiyal etkileşim çalışmaları zordur. Bu çalışmanın, laboratuvar ortamları için virüs varyantları ve mikrobiyomlar arasındaki etkileşim uygulamalarına yardımcı olması beklenmektedir.

Anahtar Kelimeler: Olasılıksal Programlama, Spike Protein, SARS-CoV-2, Mikrobiyal Meta-Komünite, Mikrobiyom

to my mother

ACKNOWLEDGMENTS

I wish to express my deepest gratitude to my supervisor Assist. Prof. Dr. Seçkin Eroğlu for his sensibility, advice, support, and insight throughout the research.

I would like to thank Assoc. Prof. Dr. Samet Bağçe for his unforgettable encouragement, and Dr. Hüseyin Enis Karahan for his generosity in sharing his knowledge. I am grateful to my dear teacher Mehmet Ali Baltaşı for his tireless teaching efforts for years.

I sincerely thank all my family, especially my lovely parents Gönül & Cesimi, my brother Mustafa, his wife Zehra, my grandmother Netice and my aunt Yadigar.

I would like to thank my dear husband Ertuğrul for every moment he added to our life, and my sweet son Cahit Asaf Nehri for giving joy to my life, with his smile everything is easy to handle.

I cannot succeed without God's help, I trust in Him and always turn to Him.

TABLE OF CONTENTS

ABSTRACT v

ÖZ vii

ACKNOWLEDGMENTS x

TABLE OF CONTENTS xi

LIST OF TABLES xiv

LIST OF FIGURES xv

CHAPTERS

1 INTRODUCTION 1

1.1 COVID-19 and SARS-CoV-2 1

1.1.1 COVID-19: Economic, sociological and technological aspects 1

1.1.2 Urbanism, climate change and SARS-CoV-2 4

1.1.3 Phylogeny, structure, and host interactions of SARS-CoV-2 5

1.1.4 Genome and proteins of SARS-CoV-2 9

1.2 Structure, function, and host interactions of Spike protein of SARS-CoV-2 10

1.2.1 Function and structure of Spike protein 10

1.2.2 Evolution and host interactions of Spike protein 11

1.3 Concept of microbiome 13

1.3.1 Microbiomes and the human microbiome 13

1.3.2 Gut microbiome 15

1.3.3 Gut microbiome and immune system 17
1.4 Research question 17
1.5 Modeling and in silico analysis 18
1.5.1 Constructing theoretical framework 19
1.5.2 Modeling the research design 21
1.5.3 Deterministic Programming vs Probabilistic Programming 24
2 MATERIALS AND METHODS 27
2.1 Parameters 27
2.2 Data 31
2.2.1 Data of countries 31
2.2.2 Data of parameters 32
2.2.3 Data of SARS-CoV-2 variants and Spike protein mutant types 32
2.2.4 Microbiome data 33
2.3 Analysis 33
2.3.1 Softwares 36
2.3.2 Set 1 analysis 36
2.3.3 Set 2 analysis 37
2.3.4 Set 3 analysis 37
2.4 Testing hypotheses 38
3 RESULTS 39
3.1 Set 1 results 39
3.2 Set 2 results 40
3.3 Set 3 results 41
4 DISCUSSION 43

5 CONCLUSION 55
REFERENCES 57
APPENDICES
A. SET 1 DATA TABLE : PARAMETERS 75
B. DATA RESOURCES 79
C. SET 2 DATA TABLE: MUTANT AND VARIANT TYPES 88
D. TIME INTERVAL FREQUENCIES 92
E. SET 3 ANALYSES 96
F. Variant and Mutant Significance Index 107
G. SARS-CoV-2 variant and mutation distributions (Global) 111
H. Regression Results 113
İ. SET-1 CORELATIONS 123
J. Data Source Tables 137

LIST OF TABLES

TABLES

Table 1.1 Hypotheses that are used in this study. 17

Table 3.1 Multi-linear regression analysis results of the dependent and independent variables. 40

LIST OF FIGURES

FIGURES

Figure 1.1. The vaccine types for COVID-19 3
Figure 1.2. A summary of the COVID-19 disease from its origin to the human disease 5
Figure 1.3. Nidovirales order 6
Figure 1.4. Structures of the nidoviruses 7
Figure 1.5. The life cycle of the SARS-CoV-2 8
Figure 1.6. CryoEM structure of Orf9b-Tom70 complex 9
Figure 1.7. Schematic diagram of the SARS coronavirus genome 9
Figure 1.8. Schematic representation of the SARS-CoV-2 Spike protein 11
Figure 1.9. SARS-CoV-2 coronavirus spike sequence variation 13
Figure 1.10. The network representation of the relations that were used in this study. 19
Figure 1.11. The schematic representation of the research design. 23
Figure 2.1. Representation of the selected parameter sets. 28
Figure 2.2. The geographical distribution of the selected countries for the analysis from the GISAID database. 32
Figure 2.3. Representation of hypotheses testing process as a flowchart. 38
Figure 4.1. The distribution of the 20I/501Y.V1 variant 45
Figure 4.2. Global distribution of S:E484 mutant. Colors represent different countries 47
Figure 4.3. Global distribution of S:Y144- mutant. Colors represent different countries 48
Figure 4.4. Global distribution of S:H69- mutant. Colors represent different countries 48

CHAPTER 1

INTRODUCTION

1.1 COVID-19 and SARS-CoV-2

1.1.1 COVID-19: Economic, sociological and technological aspects

Coronavirus disease 2019 (COVID-19) is the name of a currently well-known pandemic disease that affects hundreds of millions of people all around the world (Pal & Banerjee, 2020). There are significant variations in COVID-19 susceptibility and severity/fatality from person to person (G. Anderson & Reiter, 2020). It is known that COVID-19 vulnerability and fatality are affected by many variables: There are studied relationships between sunlight exposure (Asyary et al., 2020), dialysis, poverty, race, urbanization (Connolly et al., 2020), and COVID-19 (Bhayani et al., 2020).

COVID-19 caused many deaths around the world, and at the same time, it caused a social transformation by affecting the human population in different aspects that can be categorized as economic, social, and technological (Mofijur et al., 2021). Significant reductions in income, rising unemployment (Pal & Banerjee, 2020), technological reshapings in the healthcare area (Queen, 2021), changes in social decisions (Mofijur et al., 2021) are some examples of the different aspects of the impacts of COVID-19 on human life.

Vaccine technologies are a good example of embodying the different effects of COVID-19 on the human population because with vaccination various economic, social and technological problems and transformations regarding this disease have emerged. Since the pandemic shows its strict influence worldwide, to prevent the

virus circulation among humans and decrease the harms of the virus, many vaccines were produced and proposed to the world (Krammer, 2020). Some newly popular vaccines for the disease are based on DNA & RNA technologies currently getting approval from the FDA (Figure 1.1). Some of them are based on traditional vaccine technologies like viral-based vaccines, protein vaccines, inactivated vector vaccines, etc. (Krammer, 2020)

The vaccine development process is very costly and a long-term project in the normal standards. A vaccine can be produced in nearly 15 years, and every step in the process needs generous funding, market potential, confirmation of data, etc (Krammer, 2020). Despite the traditional vaccine development process, the COVID-19 vaccine development process is a very short-term project (Krammer, 2020). The current vaccines have been developed within nearly one year, and they have become the subject of the approval of international organizations like the FDA for immediate human use (Krammer, 2020). Since there is a big market for COVID-19 patients and also the vaccine development process is quite short, many small-scale companies are in the vaccine development process, and emerging technologies are also now accepted widely (Krammer, 2020). For instance, RNA-based vaccines have never been approved by international health organizations but in this chaotic era, they get approvals and many companies are producing and selling these vaccines (Krammer, 2020).

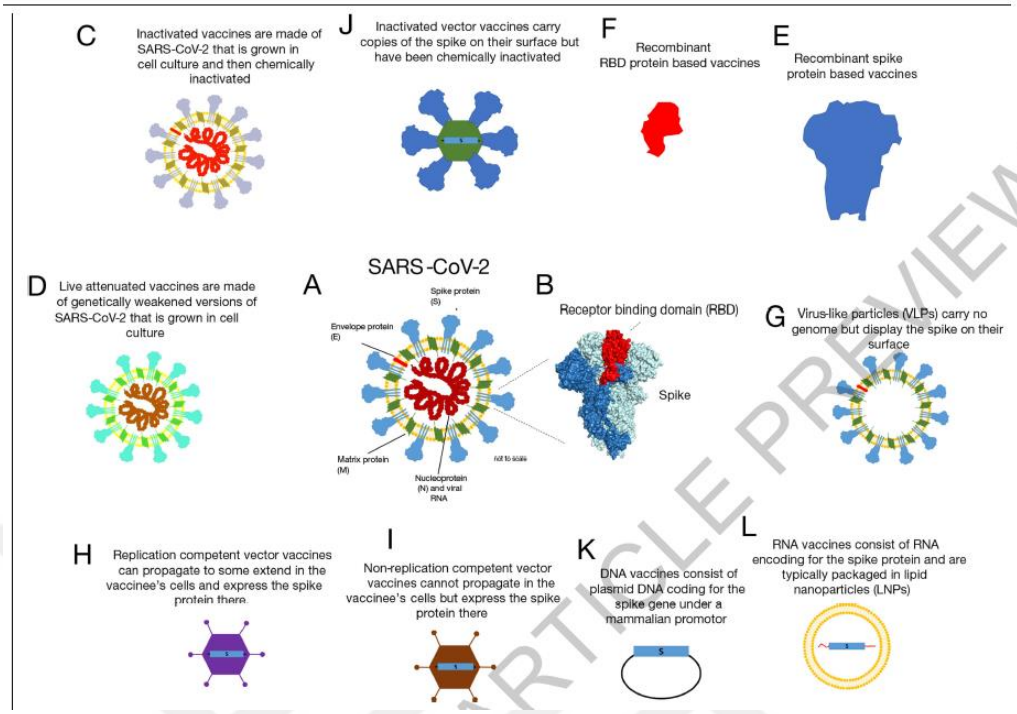


Figure 1.1. The vaccine types for COVID-19 (Krammer, 2020).

Since the companies in the market have various technologies (Figure 1.1.) and because the budget allocated by the countries is different (OECD Policy Responses to Coronavirus, 2021), there are different types of vaccines for global production and distribution. The costs of the vaccines also have variants among the Technologies; therefore, there are vaccines for countries based on the levels of income (OECD Policy Responses to Coronavirus, 2021). Just as the economic diversification of the vaccine types for countries, there is also diversification on acceptance levels for the vaccines by the people. For instance, some people are against vaccine treatment, and some people prefer different types of vaccines due to their educational background (Lazarus et al., 2021). In addition to these, since economic restrictions define the vaccine preferences, people who are citizens of different countries can access only the vaccine types supplied by the governments (OECD Policy Responses to Coronavirus, 2021).

1.1.2 Urbanism, climate change and SARS-CoV-2

Even though there are lots of microbial circulations among living organisms, modern microbial interactions have some kind of specific attributes that are caused by mainly two major parameters: Climate change and urbanism. Climate change is the name given to the change of the climate systems with chemicals, temperature, and biological processes. Many climate changes have been experienced throughout world history (Stouffer et al., 2006). Modern climate change is the main outcome of the industrial human activities which were operated especially after the 1800s (Karl & Trenberth, 2003). These activities changed the earth's ecosystem by contributing industrial chemicals to the atmosphere (Daly & Zannetti, 2007), destroying the floras and faunas, and decreasing biodiversity by human hands (Avise et al., 2009). The changes in the ecosystems can be seen firstly in microorganisms since they have an immense capacity for changing their genomes; therefore, these organisms can be counted as the major indicators of the ecosystem changes (Singh et al., 2010). In addition to climate change, urbanization is one of the significant trends that affect the ecosystem's microbial composition (Pickett, et al. 2016). Urban areas are rapidly growing worldwide, and this term is known as urbanization (Sun et al., 2020) and urbanism is the discipline that explores the relationship of these urban areas with the environment (Roggema, 2016). Due to climate change and urbanism, the microbial composition of the earth has been changed remarkably, and therefore interactions and evolution of microbial organisms are strongly affected by these processes (Reese et al., 2016).

Since the SARS-CoV-2 virus is a type of virus emerging from the areas of the interactions between rural and urban patches, this virus is an example of a microbe that emerged from human population and wild organism interactions (A. Banerjee et al., 2021). These interactions are observably strong in Wuhan, the city that SARS-CoV-2 emerged, due to rapid urbanization and composed various urban (Gui et al., 2019). In addition to urbanization processes, climate change is blamed for the emergence of SARS-CoV-2. Since urbanization and climate change have impacted

the interactions between humans and the environment, the viral interactions are affected by them and we experienced these phenomena as a pandemic in COVID-19.

1.1.3 Phylogeny, structure, and host interactions of SARS-CoV-2

COVID-19 is caused by the SARS-CoV-2 virus which is shared by many organisms like bats, pigs, cats, and humans (Figure 1.1). The SARS-CoV-2 virus belongs to the SARS-MERS viral family in the evolutionary pathway and variants of these diseases (like SARS) have been seen before (D. E. Gordon et al., 2020)

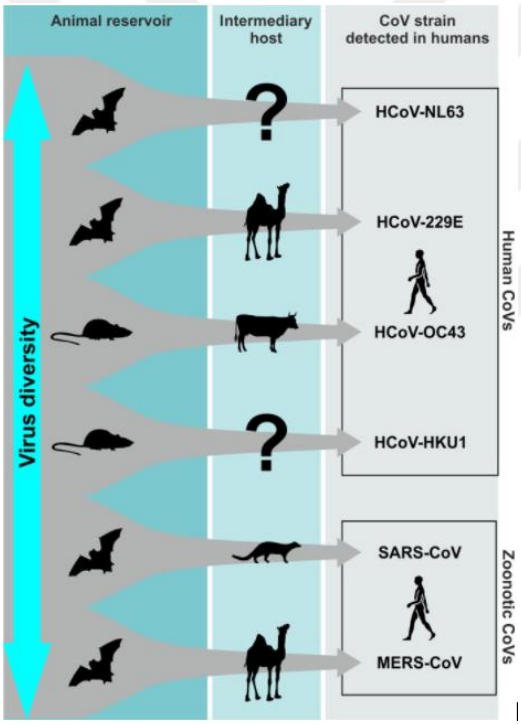


Figure 1.2. A summary of the COVID-19 disease from its origin to the human disease (Corman et al., 2018).

SARS-CoV-2 is an RNA virus belonging to the Nidovirales order Coronaviridae family (Figure 1.3) (Enjuanes et al., 2006). SARS-CoV-2 is evolutionarily related to HCV-229E, NL63, OC43, and HKU1 as belonging to the same family,

Coronaviridae, which are viruses that cause common colds of 15-30 % in humans (W. Liu et al., 2021).

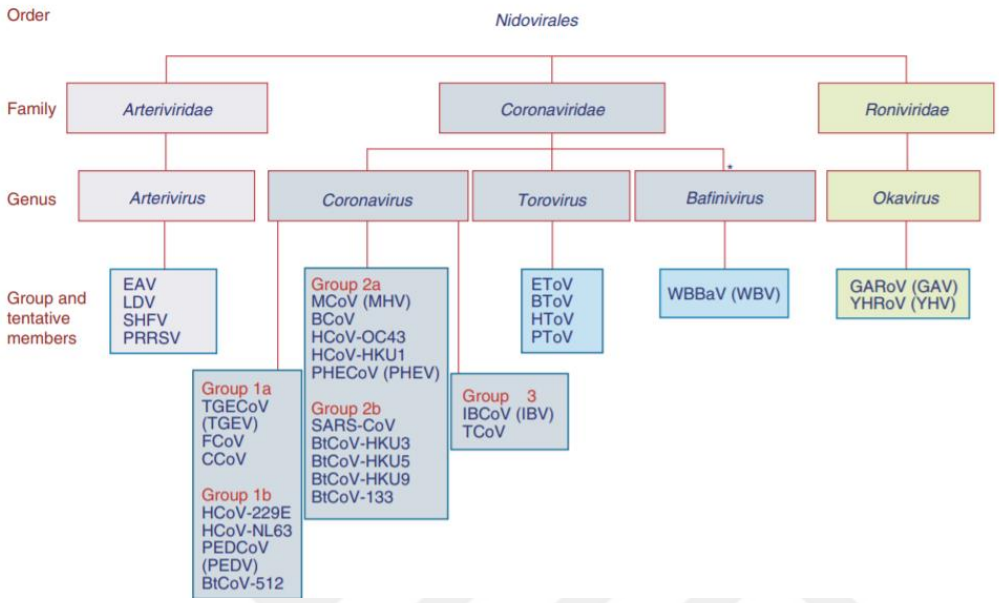


Figure 1.3. Nidovirales order (Enjuanes et al., 2006).

Viruses belonging to Nidovirales order show similar structural features (Figure 1.4). There are few structural proteins and RNA as genetic material (Enjuanes et al., 2006). Nidoviruses have a lipid envelope, and this envelope protects the genetic material from the environment. All nidoviruses have Nucleocapsid (N) protein which interacts with Membrane protein (M). Both structures and proteins vary among the viruses. The genome sizes vary among the nidoviruses, whereas the genome structures remain similar. All genomes have two big Open-reading frames (ORFs) that hold the genetic information of proteins that are responsible for regulations of transcription. The parts for structural proteins (such as M and N) stand in the genome near ORFs (Enjuanes et al., 2006).

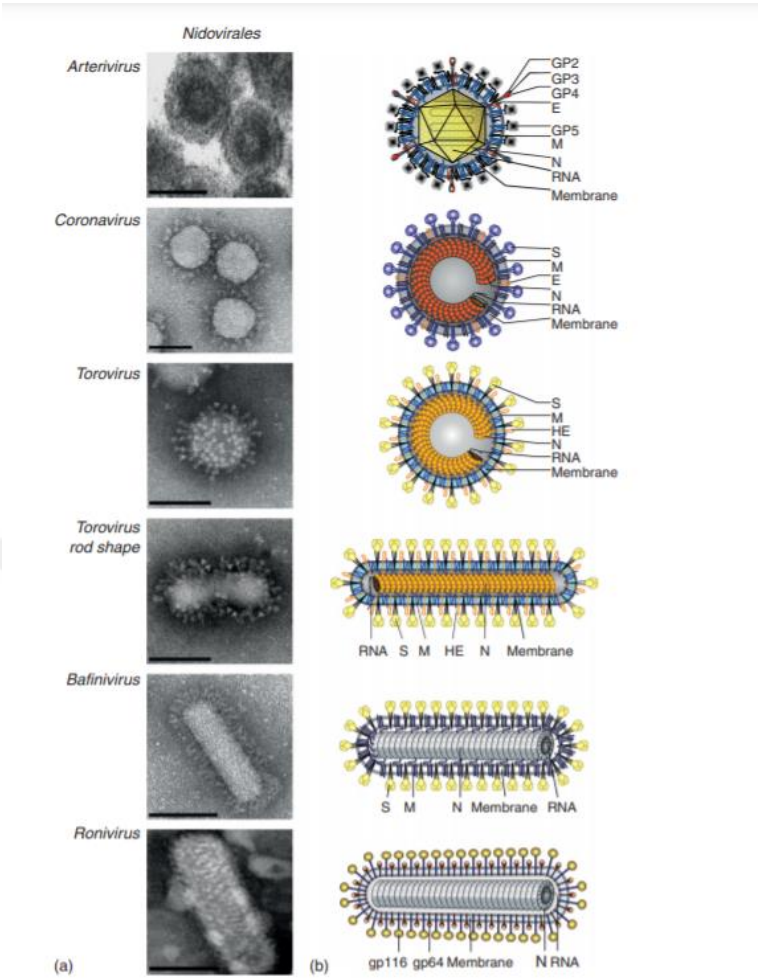


Figure 1.4. Structures of the nidoviruses (Enjuanes et al., 2006).

The life cycle of SARS-CoV-2 consists of four stages; the attachment of the virus to the cell and the transfer of genetic material, the processing of genetic material, the assembly of viral proteins resulting from translation, and the unified virions to leave the cell (Figure 1.5) (V'kovski et al., 2021).

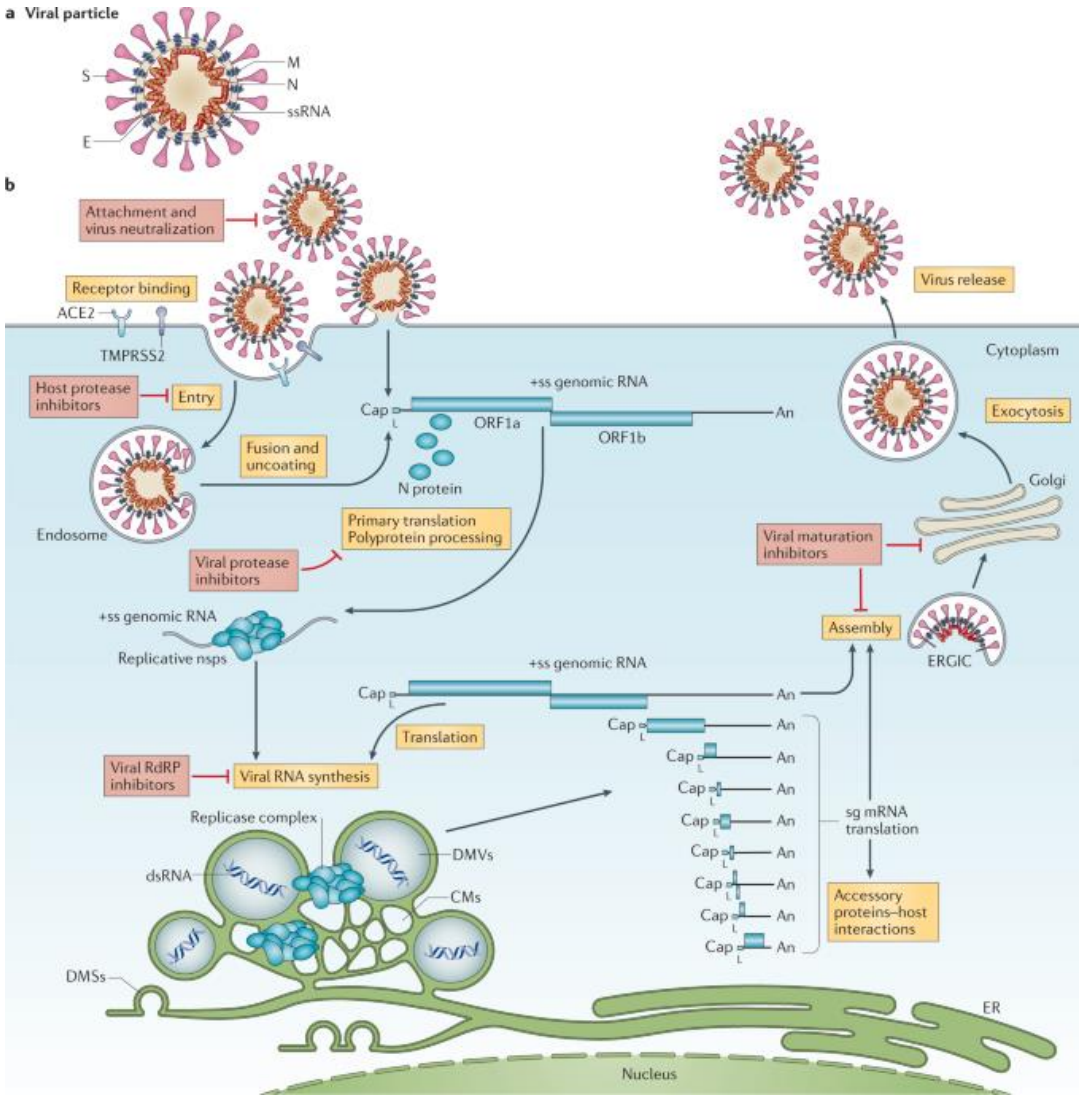


Figure 1.5. The life cycle of the SARS-CoV-2 (V'kovski et al., 2021).

SARS-CoV-2 proteins are associated with some of the host proteins and make complexes and these complexes alter the effect of the virus on the host (D. E. Gordon et al., 2020). For instance, a virus-host protein-protein interaction (PPI) formed by TOM-70 (a host cell membrane protein) and Orf-9b (a SARS-CoV-2 protein) is an example of this type of relation (Figure 1.6) (D. E. Gordon et al., 2020). Such SARS-CoV-2 virus-host protein interaction pathways can also be associated with MERS and SARS-CoV viruses and they are potential drug development targets due to their shared patterns (D. E. Gordon et al., 2020).

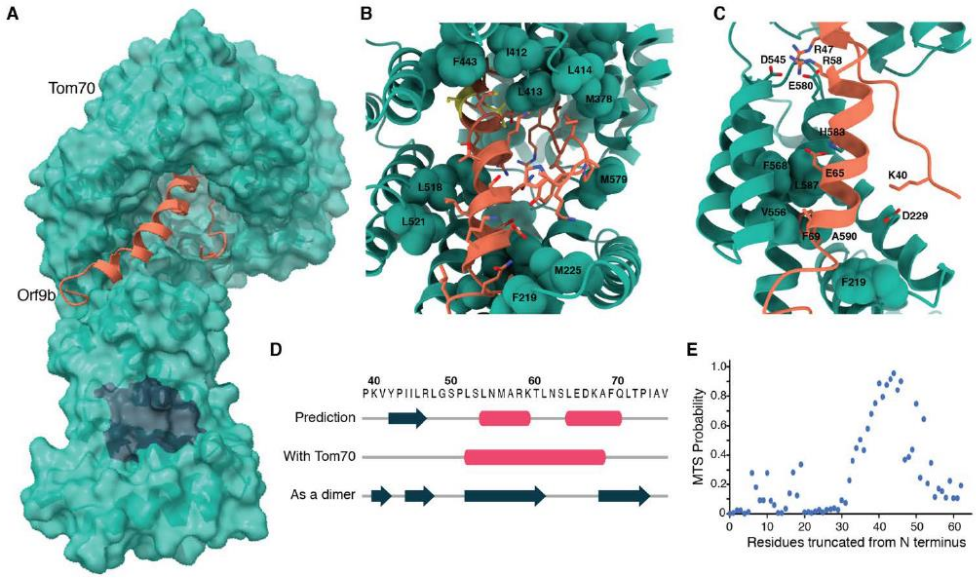


Figure 1.6. CryoEM structure of Orf9b-Tom70 complex (D. E. Gordon et al., 2020).

1.1.4 Genome and proteins of SARS-CoV-2

The SARS-CoV-2 genome consists of two ORF parts, which encode non-structural proteins. In addition to the two ORFs, four structural gene regions carry the genetic information of the structural proteins of the virus (Figure 1.7) (Wertheim et al., 2013).

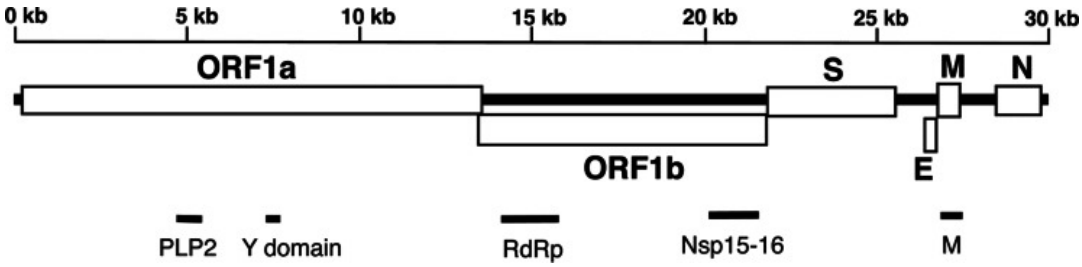


Figure 1.7. Schematic diagram of the SARS coronavirus genome (Wertheim et al., 2013).

In SARS-CoV-2, 16 nonstructural proteins (from cleavage of the two big Orf proteins), four structural proteins (spike (S), envelope (E), membrane (M), and nucleocapsid (N)), and eight accessory proteins are found (Yoshimoto, 2020). The polyproteins of Orf1a and Orf1b are cleaved to smaller non-structural proteins (NSPs). NSPs are interacting with each other and regulate gene expression (Yoshimoto, 2020). Membrane protein makes the lipid membrane of the virus, Nucleocapsid protein links via the Membrane protein and encapsidated the RNA genome. Envelope protein is an integral membrane protein and makes an ion channel and also plays a role in the virus replication process. Spike protein is the surface glycoprotein and mediates host cell attachment of the virus (Yoshimoto, 2020).

1.2 Structure, function, and host interactions of Spike protein of SARS-CoV-2

1.2.1 Function and structure of Spike protein

Spike protein is one of the most important structural proteins of SARS-CoV-2 (Guruprasad, 2021). This protein recognizes and binds to the human host cell surface receptor angiotensin-converting enzyme-2 (ACE2) receptors and provides entry into the cell (Guruprasad, 2021). The immune response of the host is also caused by the detection of the Spike protein by the host (Lu et al., 2004). Moreover, Spike protein determines the infectivity and transmissibility of the virus and it is the major antigen inducer to make an immune response (Hulswit et al., 2016). Therefore, many vaccines have been designed to target Spike protein (Du et al., 2009).

Spike protein consists of two subunits: S1 and S2 (Figure 1.8). S1 is responsible for binding to ACE2 receptors; after, this binding process, the S2 subunit performs the fusion process into the cell and allows the genetic material of the virus to enter the cell (Demers-Mathieu et al., 2020). Cleavage of the S1 subunit from S2 is important for the infection; therefore, antibodies that bind Spike protein and prevent cleavage, and inhibits the virus fusion to the cell (Demers-Mathieu et al., 2020).

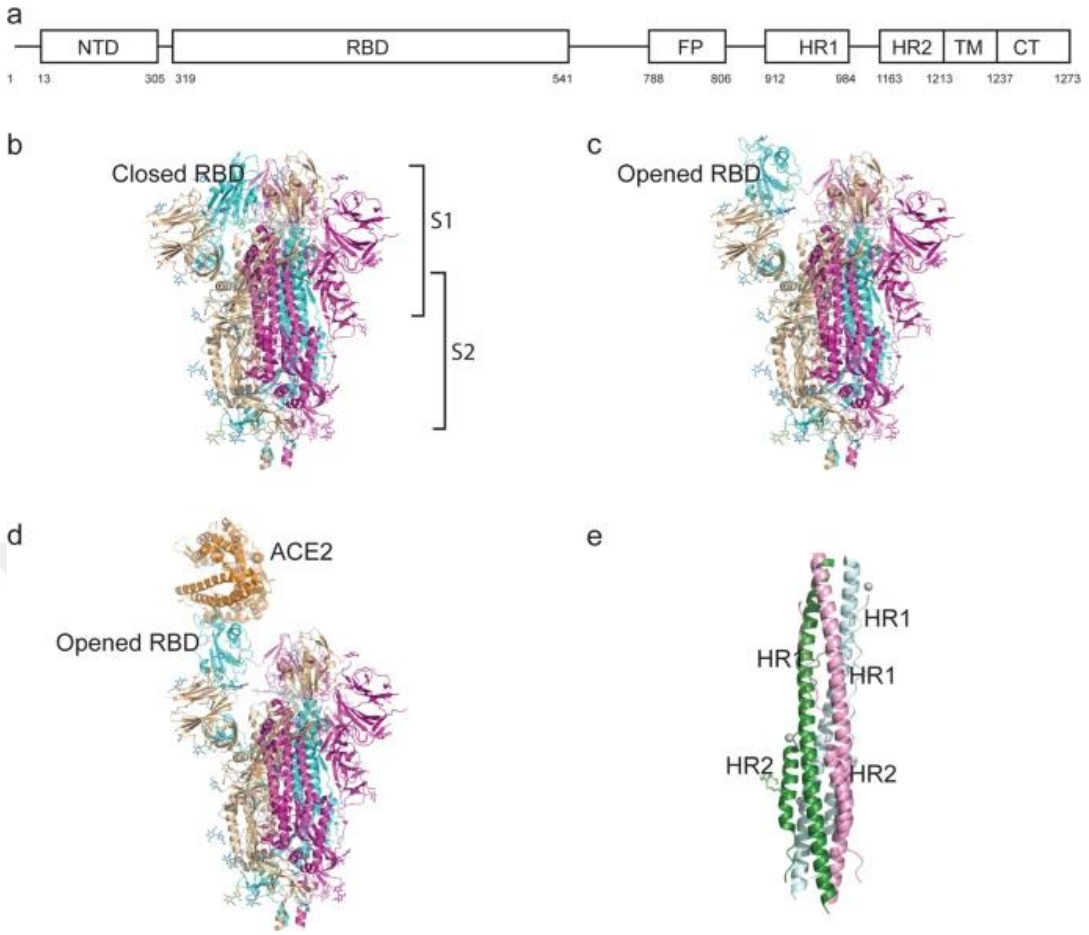


Figure 1.8. Schematic representation of the SARS-CoV-2 Spike protein (Y. Huang et al., 2020).

1.2.2 Evolution and host interactions of Spike protein

SARS-CoV-2 is also found as a haplotype in its host as an RNA virus, and Spike proteins can also be categorized by haplotype analysis. Haplotypes are cumulative variations on the genetic data in a single chromosome (Tourdot & Zhang, 2019). In haplotype variations, a variant is dominant among the other variants, and these variants are found in very low frequencies comparing to the dominant haplotype (Töpfer et al., 2013). Viruses and viral proteins are found in the host as a haplotype structure, as in the example of Spike protein. Spike proteins are made up of small

differences between different haplotypes that evolved from the same ancestor (Pereson et al., 2021).

There are two main causes of the variation in a viral population: recombinations and mutations (Töpfer et al., 2013). Even though mutations and recombination events are found in the viral genome, especially recombinations are rare (Töpfer et al., 2013). In viral quasispecies, the dominant haplotype shows very low recombination events. Quasispecies are the viral groups in a viral population composed of haplotype variations (Domingo, 2002).The Receptor Binding Domain (RBD) of Spike protein, which binds to human ACE2 receptors, is not a recent acquisition by recombination, but rather an ancient gain that is common with bat viruses (Boni et al., 2020). Therefore, mutations (such as deletion and insertion), but not recombination, have great importance in Spike protein (and SARS-CoV-2) evolution, in that they generate Spike protein variants (Figure 1.9) (Boni et al., 2020). Because its evolution rate is similar with each clade of SARS-CoV-2 variants, Spike protein is the major evolutionary driver of the cell, and therefore variants of SARS-CoV-2 are also largely categorized according to Spike protein variants (Pereson et al., 2021).

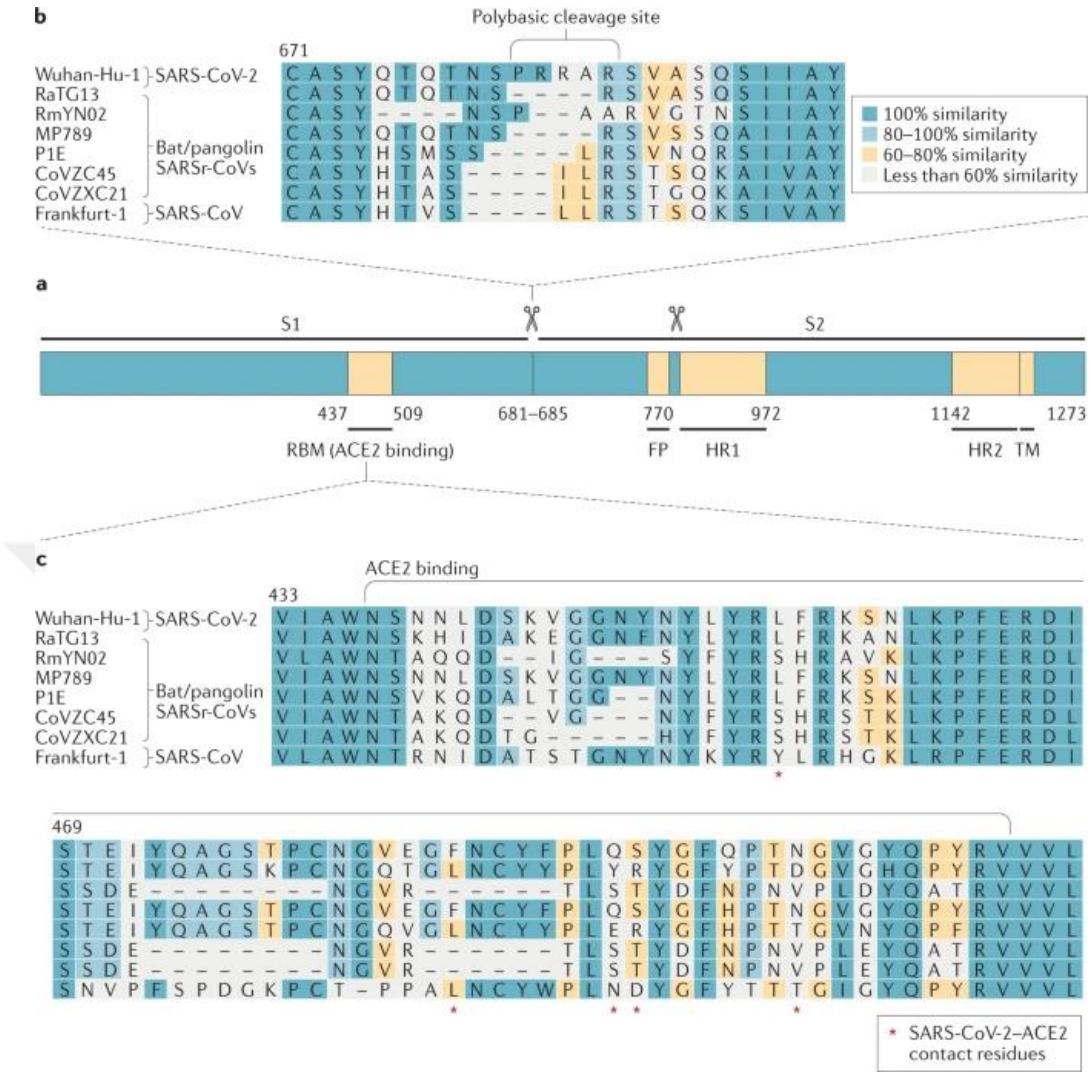


Figure 1.9. SARS-CoV-2 coronavirus spike sequence variation (V'kovski et al., 2021).

1.3 Concept of microbiome

1.3.1 Microbiomes and the human microbiome

Microbiomes, which can be defined as the assemblage of the microbes in a host, are representatives of diseases or the health condition of the host (Marchesi & Ravel, 2015). The microbiomes are the main indicators of singular attributes that are directly related to the host (Bruijning et al., 2020). The genetic problems of the host

can be detected from its microbiome content (Bresalier & Chapkin, 2020). For instance, the effects of Endocrine-Disrupting Chemicals (EDCs) in the air can be seen in the lung microbiota of the terrestrial animals easily (Segal et al., 2016). The gut microbiota is another target for the EDCs (Kumar et al., 2020). Since human microbiomes are major representative entities of the host's attributes -such as diet, lifestyle, medical record, etc.- as a whole (Scepanovic et al., 2019), the changes in microbiome content can be used for inferring the evolutionary forces that act on the host (Bruijning et al., 2020).

In microbiomes, there are ecological relations among species. The dominant species (founder species) of the microbiomes alter the host's biological reactions by providing some chemicals (Trosvik & de Muinck, 2015). For instance, the presence of a species can alter the host's immune response via triggering the host to make more IgA-Immunoglobulin A, which affects the immune response, especially in respiratory areas, as a first step reaction of the immunity (Donaldson et al., 2018). The dominant species and other species are changing in health conditions from disease conditions in a microbiome (Rinninella et al., 2019). It is known that the abundance of species in the intestinal microbiota is related to the diseases and clinical blood markers of the host organism (Manor et al., 2020). The microbial composition -viruses, fungi, bacteria- in the microbiota contributes to many metabolic functions of the host and plays a role in many physiological effects, especially the immune response. The term dysbiosis is used to describe situations where changes in the microbiota are directly related to a host's illness. This term indicates that a microbiota community is directly related to a disease of the host, in the health state where the host does not have this disease, the composition of the microbiota is significantly different from the disease state. It is an area that has been studied that these conditions, namely the composition of the microbiota and the relative abundances of the organisms in it, are related to the disease and health conditions of the host (E. Li et al., 2015). Keystone species are the species that are found in low abundance but in a very high number of ecological connections via other species in the microbiome (S. Banerjee et al., 2018). Dominant species are in a positive relationship with other

members of the microbiome via providing a usually mutualist environment, whereas keystone species have a high number of both positive and negative relations with other microbes (Trosvik & de Muinck, 2015).

1.3.2 Gut microbiome

The human gut microbiota compositions show discontinuous variation rather than a continuous variation of gut microbes; in other words, the microbes in the gut are found with certain clusters (Arumugam et al., 2011). These distinct microbial sets are called enterotypes, and three types of enterotypes (with different dominant species and different microbial compositions) have been detected in human microbiota (Arumugam et al., 2011). Enterotypes indicate a balanced relationship between the host and its microbiota (Arumugam et al., 2011). The most important characteristic of the community composition of the gut microbiota is the functional relationship, rather than which bacterium is present in the microbiome (Arumugam et al., 2011). The gut microbiota shows phylogenetic variation at the genus and phylum levels among enterotypes and represents the functional variation at the class level (Arumugam et al., 2011).

Firmicutes and Bacteroides phyla are the most dominant species in the gut microbiota (Thursby & Juge, 2017). Microbes in the gut microbiota are exposed to selective forces both by the host factors such as diets, diseases and by other microbes that are located in the gut (Hadi et al., 2020; Scanlan, 2019). This is why some low-abundance bacteria survive in the gut (Arumugam et al., 2011). Every bacterium in the gut follows different survival strategies and usually the most abundant function will be related to the most dominant type (Arumugam et al., 2011; Loftus et al., 2021; Rinninella et al., 2019). However, since the most dominant species cannot provide all functions, the functional composition of different species is important for the intestinal microbiota (Arumugam et al., 2011; S. Banerjee et al., 2018).

The human gut microbiome composition is affected by many factors. For instance, the human intestinal microbiota shows a geographical variation (Mobeen et al., 2018), which is due to different parameters (e.g. genetics, lifestyle, climate, diet, altitude etc.) that affect cumulatively (Das et al., 2018). However, even many factors affect the microbiome, enterotype variations are thought to be independent of age, gender, BMI, and geography, but they are closely related to dietary habits (Arumugam et al., 2011; Mobeen et al., 2018)

By looking at the geographical enterotype and intestinal microbiota composition variations, Firmicutes and Bacteroidetes are the most common phyla, but different countries show different abundances for these species (Mobeen et al., 2018). Bacteroides are the dominant organisms of the gut microbiome in general, but in some enterotypes, Firmicutes can be the dominant organism (Arumugam et al., 2011; Mobeen et al., 2018; Trosvik & de Muinck, 2015). Actinobacteria is the most common phyla in the gut microbiota; after, Firmicutes and Bacteroidetes, Actinobacteria is the keystone taxon of the gut microbiota and is involved in a high degree of ecological network with other gut microbes (Trosvik & de Muinck, 2015). Proteobacteria is the most common species after Bacteroides, Firmicutes, and Actinobacteria in the human intestinal microbiome (Mobeen et al., 2018). Proteobacteria is the species that represents the functional variation that occurs in the gut microbiome among different microbe compositions (Bradley & Pollard, 2017).

It is known that the mucosal immune system, which has a very important role in immunity, has a network that can be affected by various factors. This system is thought to be dysregulated due to intestinal problems. Studies have begun to show that the overall immune response of the organism is shaped by a cross-talk between the gut and the lung at the organism level (Tulic M C, Piche T, 2016). There are many studies reporting the relevance of gut-lung microbiota crosstalk to COVID-19 (Srinath et al., 2020).

1.3.3 Gut microbiome and immune system

The gut microbiome is affected by diseases and also it affects the disease conditions: Rheumatoid arthritis, type 1 diabetes, inflammatory bowel disease (IBD), allergic diseases, systemic lupus erythematosus (SLE), skin-related autoimmune pathologies, neurological inflammatory diseases, and many cancer types can be counted among diseases related to gut microbiota (Lazar et al., 2018). The composition of gut microbiota also changes during COVID-19 disease (Yeoh et al., 2021). In addition to these, viral infections in the respiratory tract and lung affect the gut microbiota by altering the function and composition of the gut microbiome (Sencio et al., 2021) since the intestinal microbiota is associated with the lung microbiota and the changes in the lung microbiome affect the gut microbial composition (Dhar & Mohanty, 2020). Moreover, gut microbiota prevents pathogen invasion by insisting on various strategies against pathogens such as killing pathogens directly, supporting the immune system of the host, or making competition for food (Pickard et al., 2017).

1.4 Research question

The main concern of this study is to try to understand whether particular gut microbiota compositions tolerate variants of the SARS-CoV-2 virus and Spike protein mutations of the virus. In other words, it aims to investigate whether the composition of gut microbiota affects the infectivity of variants of SARS-CoV-2 and variants of Spike proteins.

Table 1.1 Hypotheses that are used in this study.

H1:	*One of the following parameters can explain COVID-19 death and reproduction rates: Diet, diseases, economic parameters, environmental factors, micronutrient deficiencies, population parameters.*

H2:	*One of the following parameters can explain SARS-CoV-2 infectivity between variants and specific mutations on Spike protein: Diet, diseases, economic parameters, environmental factors, micronutrient deficiencies, population parameters.*
H3:	*Microbes residing gut can explain which SARS-CoV-2 mutant infected the host.*

1.5 Modeling and in silico analysis

From the data obtained by the literature research for his study, it has been determined that many different parameters in human life are related to both intestinal microbiota and COVID-19 (Figure 1.10). To investigate these relations (Figure 1.10), a generative theoretical explanation was needed. Moreover, to test hypotheses (Table 1.1), it was a need for modeling the in silico design and apply the model to get results with the help of probabilistic programming tools.

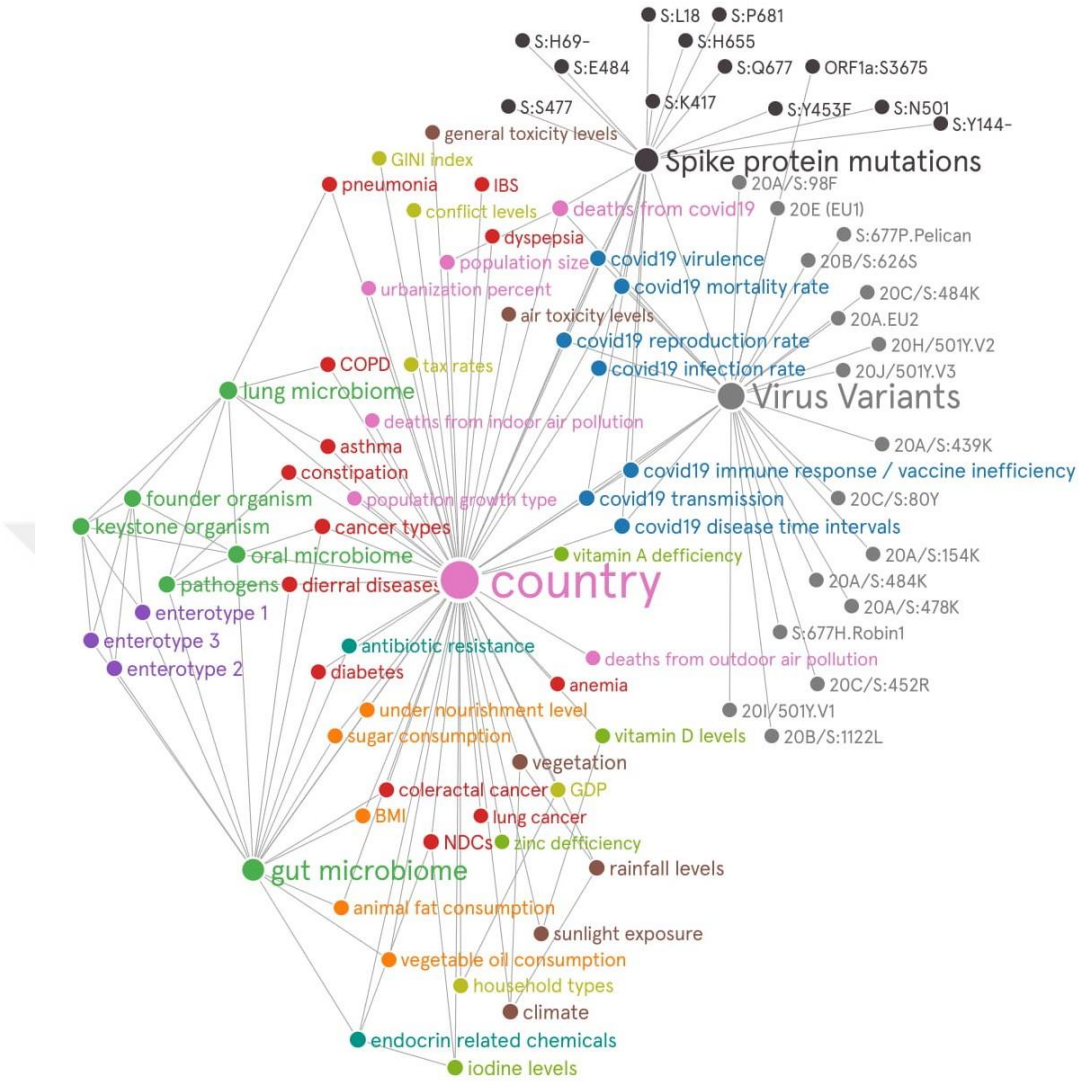


Figure 1.10. The network representation of the relations that were used in this study.

1.5.1 Constructing theoretical framework

Integrated methods are needed to wholly understand the place of humans in the ecosystem (Pickett et al., 1997). While putting people in an ecological model, it is necessary to consider human activities such as economical activities, social relations, and etc. (Grimm et al., 2008a). Therefore, clean theoretical grounds are needed to

understand the patterns that occur within urban and human-occupied systems. For this theoretical basis, it is necessary to relate economic and social aspects to biological-physical systems (Pickett et al., 1997). It is proper to use the basic theories and explanations of ecology (such as patch dynamics, spatial heterogeny) on this ground; because, in the ecological evaluation of city and city-related systems, it is sufficient to modify existing theories rather than making a new theoretical ground (Niemelä, 2000).

Different prevailing paradigms are in place for understanding different ecological relationships within city systems (Pickett et al., 2016). When different relationships are studied in different parts of the city, the methods and paradigms also change. As mentioned above, an integrated approach is required to study human relations in the city (Grimm et al., 2008a). The discipline of Urban Ecology combines the human elements of humanity with other physical, chemical, and biological factors of the biosphere in the context of ecological relations, and the discipline is maturing to provide a theoretical basis for practical studies (Grimm et al., 2008a).

However, the biosphere and ecological dynamics change and evolve as a result of evolutionary and geographical processes (Brunner et al., 2019). The impact of climate change and urbanization also manifests itself in human-occupied ecosystems from different angles: city systems are a hotspot for environmental changes (Grimm et al., 2008a). Processes such as changes in biodiversity, altering biochemical reactions are some examples of these transformations of the environment (Grimm et al., 2008a). Therefore, it is a concern that how the city systems - and therefore people - will be positioned in these changing processes (Riffat et al., 2016).

Recently, it has begun to be discovered that microbes are important not only to protect human health but also to keep city-systems sustainable (King, 2014). The importance of city microbiomes in water distribution systems and plant-microbiome relationships can be shown as an example of this issue (King, 2014). Moreover, the microbial structure of the built environment affects mental health (Hoisington et al., 2015) and also human behavior (Stamper et al., 2016). In addition, there are studies

about the effect of indoor bacteria on the built environment (Kembel et al., 2014) and studies to make the architectural design microbially sustainable (Brown et al., 2016).

The theoretical starting point of this study is based on the maturity of the studies about microbes of the city systems, human health, and human activities. In addition, it can be said that the relationship of the microbiome concept with the concept of self is questioned and includes a philosophical theoretical ground as well (Rees et al., 2018) and the impact of the microbiome concept on evaluating our self-consciousness is also discussed in the literature (Relman, 2012). Therefore, the microbiome concept creates a link both in the context of ecological relationships and health (Inkpen, 2019). In addition to all these, it has been taken into account that microbes live together as a meta-community, and theories that have already been created to understand ecological relationships at the macro level metacommunities (Leibold et al., 2004) have been adapted to the microbial level as microbial metacommunity (Miller et al., 2018). Therefore, the microbial metacommunity is suitable for the holistic evaluation of city systems and human ecosystems in terms of establishing a theoretical infrastructure.

1.5.2 Modeling the research design

The data required to know the hosts with various intestinal microbial composites infected with which mutant type have not been found in the literature. In addition, the geographic distribution analysis of the gut microbiota is limited in the existing literature. In contrast, the factors associated with COVID-19 are easier to find in the scientific literature. For this reason, a study design has been made in which relevant factors were brought together and passed through different analysis processes to reach a simulation of interaction on microbiome compositions and mutant types distribution (Figure 1.11).

As a result of the literature reviews, the relationship of the microbiota with many parameters has been confirmed. As a result of literature research different parameters are to be known as related to COVID-19. It is also known that COVID-19 causes change in the gut microbiota composition. However, no literature was found during this study as to how different variants and Spike protein mutations are tolerated by different microbiota.

In the study, the relationship between COVID-19 with some parameters from various angles was analyzed. Analysis of the relationship between different microbiota compositions and some mutants of Spike proteins -that were gathered from the previous analyses in this study- was made.

There is no second virus like SARS-CoV-2 with detailed global data and global distribution of its different mutants. Globally, the closest data belong to the SARS-MERS family, but even theirs does not come close to SARS-COV-2 (Petrosillo et al., 2020). For this reason, the results of the study could not be tested with a second virus as a control group.

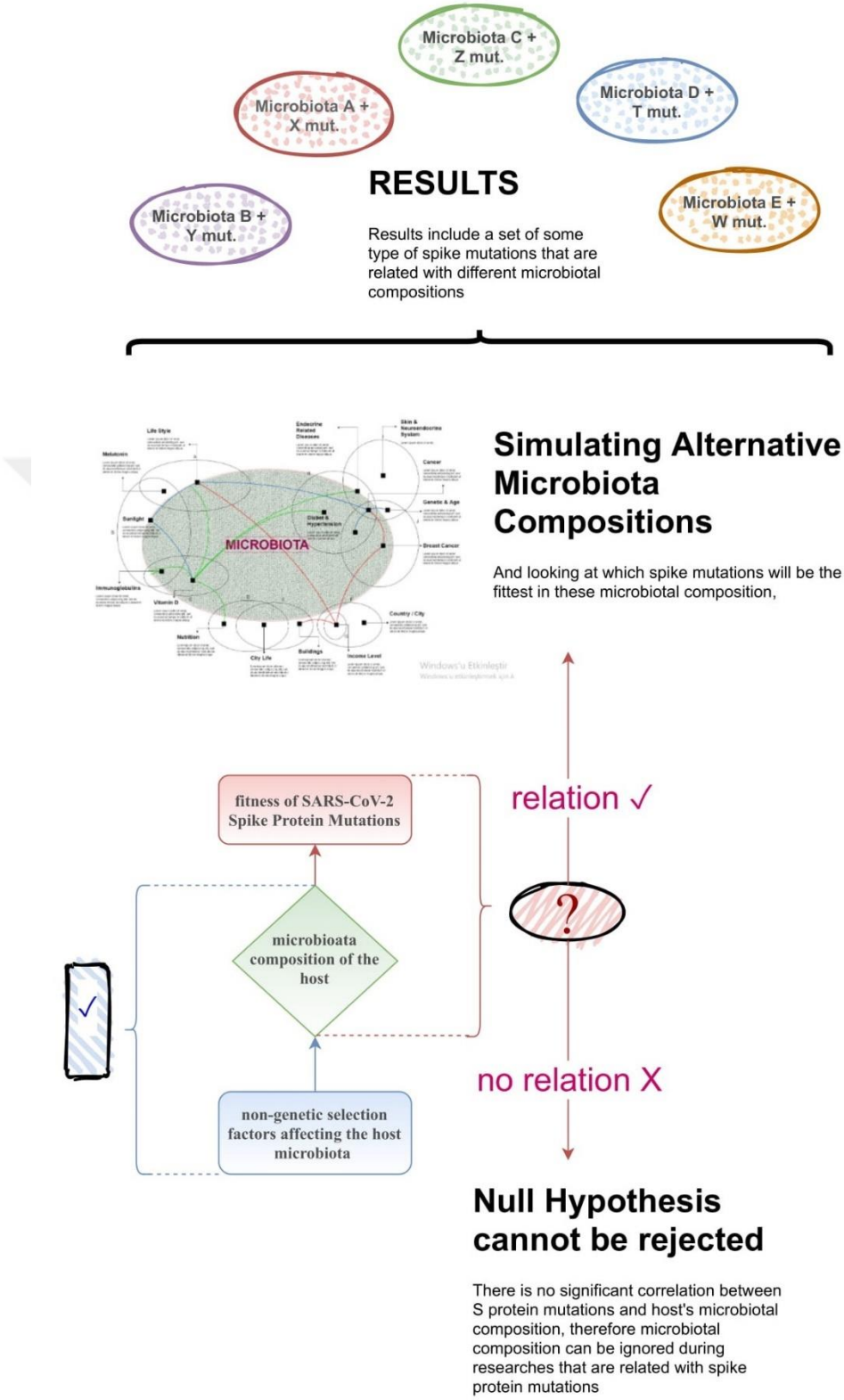


Figure 1.11. The schematic representation of the research design.

1.5.3 Deterministic Programming vs Probabilistic Programming

Although the analysis process was available to use deterministic computational tools with sufficient data at the beginning (See Appendix A), probabilistic programming tools were also included in the analysis because the data set shrank in the later parts of the analysis, and the deterministic programs made false inferences in this data set.

In deterministic programming languages, the user describes in a deterministic way what to do with the program and what processes to follow while executing the program (Mitsos et al., 2018). However, probabilistic programming languages (PPLs), such as STAN and WebPPL, do not expect the user to provide all the knowledge for execution (Ghahramani, 2015). PPLs learn by themselves over the data and they have algorithms to infer some relevant conclusions from this data in several ways and also they can learn from evidence/observation (Cornell, n.d.).

One of the features of PPLs is that they separate the model from the inference algorithm inferred from the model; therefore, it is possible to work by separating the model and the inference algorithm over PPL (Sarker, 2021). PPLs with inference algorithms, can both make inferences in the forward direction and construct causations in the backward direction and thus, causal relations between input and output can be detected by PPLs (Dimovski A.S., 2020). Therefore, it is possible to simulate the process quite efficiently.

PPLs use generative models and there may be randomness (Tavares et al., 2019). In PPLs, there is uncertainty to simulate the models; therefore, the probability is the approach to reach uncertainty (Tavares et al., 2019). PPLs use random variables that are the values that represent the uncertainties, which means that the value of the variable is a probability associated with the parameter (Obeid et al., 2018). In PPLs, parameters are unobserved (latent) random variables of interest; they are inferred from observed data (Cornell, n.d.).

There are also conditional probabilities which some variables depend on different variables. With these conditions, the PPLs can learn from data (Olmedo et al., 2018). PPLs use the subjective probability approach (Bayesian model) rather than the frequentist probability approach (Olmedo et al., 2018). By this, PPLs can capture such a pattern of reasoning; even a single model can induce complex explaining away dynamics and, many inference algorithms are available to infer from a small sample in PPLs (Cornell, n.d.). In addition to this, some PPLs can handle big data size with some algorithms such as variational inference (Cornell, n.d.).

CHAPTER 2

MATERIALS AND METHODS

In this study, a three-step process was followed. In the first stage, the relationship between COVID-19 death rates and COVID-19 reproduction rates was analyzed on a country basis with selected parameters (Figure 2.1). In the second step, the relations of the parameters on different mutants of Spike protein and variant types of the SARS-CoV-2 were examined. In the third stage, microbiota via mutant type analysis in probabilistic programming was performed by using the data obtained from the results of the first two stages. In this last stage, the possible distributions of Spike protein mutant types and SARS-CoV-2 variants in related microbial compositions were visualized.

2.1 Parameters

Six group parameter sets -diet, environmental factors, micronutrient deficiency, economic parameters, population parameters, and diseases- were created to be used in the analysis. Each parameter set was tried to be constructed in a holistic structure so that it could represent the main parameter from various angles (Figure 2.1).

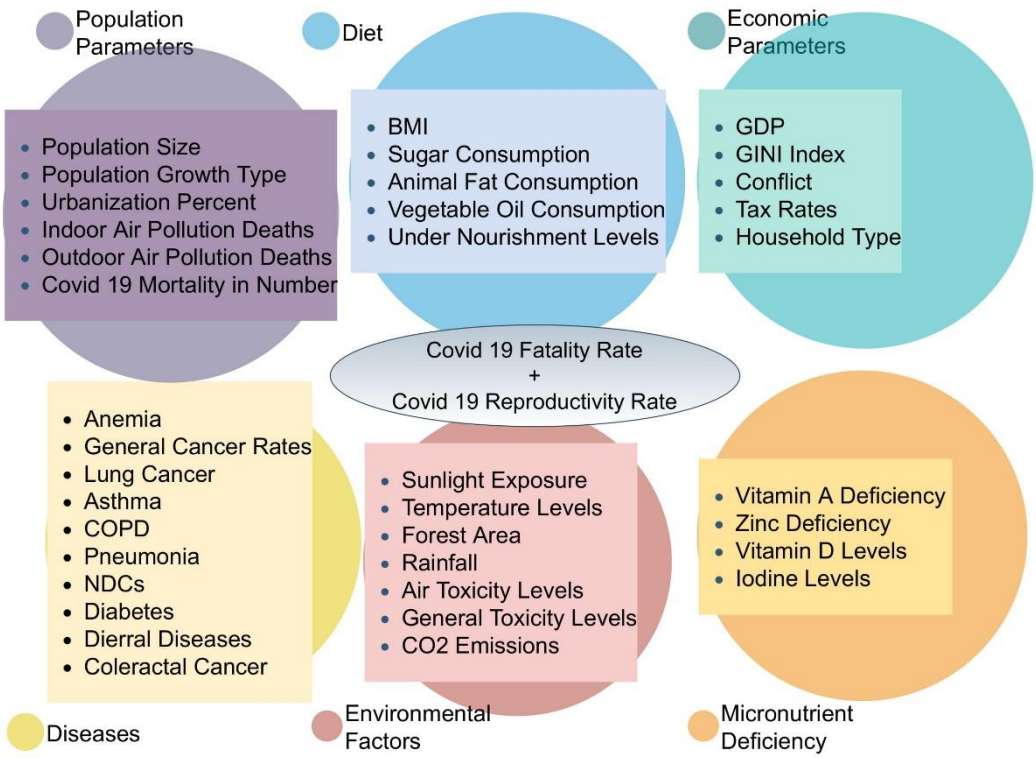


Figure 2.1. Representation of the selected parameter sets. The six sets of parameters are represented as a cluster and the dependent variables of the Set 1 Analysis (COVID-19 Fatality Rate and COVID-19 Reproductivity Rate) are represented as an ellipse. COVID19 reproduction rate shows the newly infected individuals by an infected person. If the rate is bigger than 1, the virus spread. If the rate is smaller than 1, the virus will gradually disappear from the population (See Appendix B).

Diet: Diet is one of the main parameters that are related to COVID-19 deaths (Bousquet et al., 2020). Some types of diets show a relationship with COVID-19 deaths and COVID-19 cases (Greene et al., 2021). People in malnourished countries are more prone to COVID-19 severity comparing to people living in countries with no malnutrition problems (Mertens & Peñalvo, 2021). Moreover, COVID-19 cases show a strong relationship with obesity (Ho et al., 2020). Among the cases that have Body-Mass Index (BMI) is bigger than 23 kg/m², a linear relationship of increasing COVID-19 severity (Gao et al., 2021). Since the nutrition status affects the immune system, intake of the necessary macronutrients also has a significant relation with

the COVID-19 severity condition (Chaari et al., 2020). For these reasons, the *Diet* parameter set (Figure 2.1) include the parameters of the selected countries: BMI, undernourishment levels, animal fat consumption, sugar consumption, and vegetable oil consumption (Figure 2.1). For the resources of the data of these parameters, see Appendix B.

Diseases: COVID-19 cases and COVID-19 fatality are related to many diseases: Among cancer patients, COVID-19 death rates are 13.3% higher than other patients (Moris et al., 2020), and some types of cancer patients are in the highest risk groups for COVID-19, such as lung cancers (Ömeroğlu Şimşek, 2020). Chronic obstructive pulmonary disease (COPD) and asthma are related to COVID-19 (Skevaki et al., 2021). Many of the non-communicable diseases (NDCs, diseases that cannot pass from person to person) such as diabetes or hypertension are related to COVID-19, and patients who have NDCs are the risk groups for COVID-19 (Kluge et al., 2020). Anemia is also related with COVID-19 severity (Hariyanto & Kurniawan, 2020). For these reasons, the *Diseases* parameter set (Figure 2.1) includes the parameters: Anemia, general cancer rates, lung cancer, asthma, COPD, pneumonia, NDCs, diabetes, diarrheal diseases, colorectal cancer levels of the countries (Figure 2.1). For the resources of the data of parameters, see Appendix B.

Economic Parameters: Economic statuses such as conflict, competition, and cooperation are linked with biology, especially with sociobiology, and the philosophical aspect of these connections have been constructed in literature (Hirshleifer, 1978). The applications of the economy on biological sciences are also widely studied in many disciplines such as urban ecology (Grimm et al., 2008b). The relationship of economy and biology can also be applied to the COVID-19 problem: The short-term and long-term consequences of the COVID-19 are related to household type. Moreover, the effects of COVID-19 are socially stratified (Mikolai et al., 2020). The relationship between COVID-19 cases and conflict situations is also another aspect of economic status (Bloem & Salemi, 2021). GINI index is a parameter that represents the economic inequality in a society; in other words, it shows the gap between the poor and the rich in a given country (Elgar et al., 2020),

and the relationship of the GINI index and COVID-19 deaths are also available in the literature (Elgar et al., 2020). For these reasons, the *Economic Parameters* parameter set (Figure 2.1) includes the parameters: conflict levels, GDP, GINI index, tax rates, and household type of the countries (Figure 2.1). For resources of the data of parameters, see Appendix B.

Environmental Factors: Many environmental factors like temperature (Xie & Zhu, 2020; Xiong et al., 2020), sunlight (Asyary & Veruswati, 2020), open green area (Venter et al., 2020), whether factors such as rainfall (Hariyanto & Kurniawan, 2020; Tosepu et al., 2020), air toxicity (Travaglio et al., 2021), environmental pollutants (Bashir et al., 2020) are related with COVID-19 cases and COVID-19 deaths. Also, institutional features affect COVID-19 cases (P. Li et al., 2020). For these reasons, the *Environmental Factors* parameter set (Figure 2.1) includes the parameters: sunlight exposure, rainfall, forest area, CO2 emissions, air toxicity levels, general toxicity levels, and the average temperature of the countries (Figure 2.1). For resources of the data of parameters, see Appendix B.

Micronutrient Deficiency: Micronutrients are minerals and vitamins that play important role in homeostasis by providing various functions in the body (Carr, 2020). Micronutrients are also important for the immune system to work properly and micronutrient deficiencies are related to COVID-19 also (Carr, 2020). Zinc deficiency is highly correlated with COVID-19 cases, especially in poor countries (Jothimani et al., 2020) and zinc supplementation is offered for COVID-19 patients (Wessels et al., 2020). Vitamin D is another essential micronutrient for the immune system, and COVID-19 cases and severity are strongly related to vitamin D deficiency (D. C. Anderson & Grimes, 2020). Also, provine-iodine nasal sprays protect COVID-19 cross infection (Frank et al., 2020). Vitamin A is an important micronutrient for the immune system as playing a role in immune response (Z. Huang et al., 2018). Vitamin A is important for pneumonia treatments and it can be an anti-SARS-CoV-2 regimen (R. Li et al., 2020). For these reasons, the *Micronutrient Deficiency* parameter set (Figure 2.1) includes the parameters: vitamin

D levels, vitamin A levels, zinc levels, and iodine levels of the countries (Figure 2.1). For resources of the data of parameters, see Appendix B.

Population Parameters: Population size and the median age is related to COVID-19 spread (Lulbadda et al., 2021). It is also known that air pollution increases the risk of COVID-19 fatality (Ali & Islam, 2020). Outdoor air pollution deaths (Cohen et al., 2005) and indoor air pollution deaths (Rehfuess et al., 2006) are other death rates for air pollution in a population. These three categories of death numbers - outdoor air pollution deaths, indoor air pollution deaths, and COVID-19 deaths- are counted as the population death parameter due to their strong relatedness to the pollution. Urbanization is another factor that is related to COVID-19 (Connolly et al., 2020; P. Li et al., 2020). For these reasons, the *Population Parameters* parameter set (Figure 2.1) includes population size, population growth type, urbanization percent, indoor air pollution deaths, outdoor air pollution deaths, and COVID-19 mortality deaths of the countries (Figure 2.1). For resources of the data of parameters, see Appendix B.

2.2 Data

The data used in the analysis were gathered from the data sources and collected in tables for further analysis. For the data resources and details about parameters, see Appendix B.

2.2.1 Data of countries

The CoVariants section of the GISAID database was used to obtain data of city populations that are related to different mutants of Spike protein and variants of SARS-CoV-2. In this section, 58 countries were found with related information (See Appendix A). 56 of 58 countries that have the relevant variant and mutant data were

selected for further analysis (Figure 2.2). The Caribbean island countries, Bonaire and Curacao were blinded because the information of these countries in the parameter data set was mostly missing.

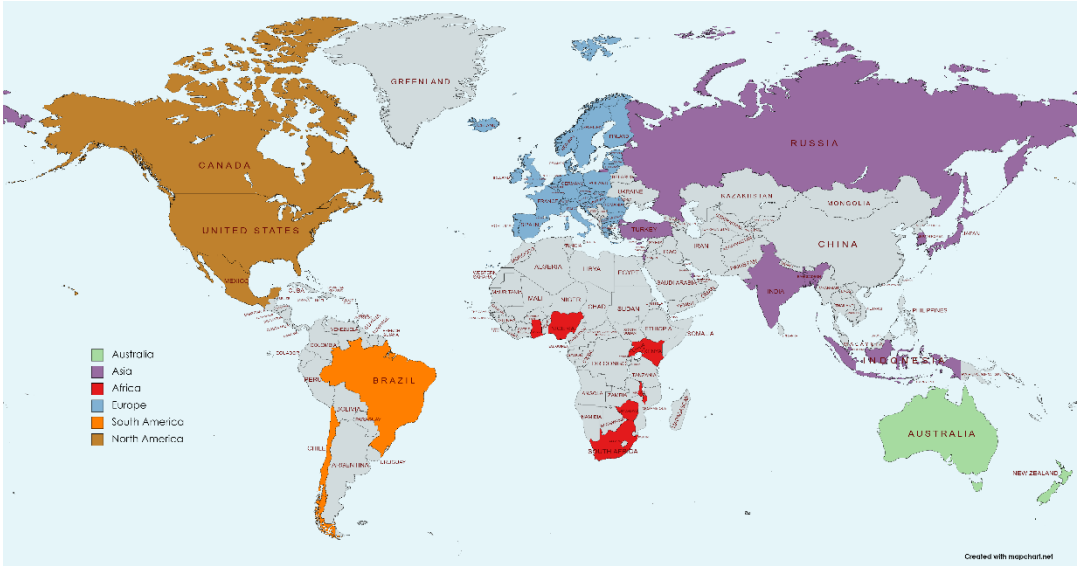


Figure 2.2. The geographical distribution of the selected countries for the analysis from the GISAID database.

2.2.2 Data of parameters

An Excel file containing the data of all members of the parameter sets for the selected countries and the country names was created as a table (see Appendix A). Each data column includes data from a single data source -only one web page or database- to provide consistency among data sets for the countries. If the data is unavailable in these sources the entry about this data was settled as NULL. For detailed information about the data resources, see Appendix B.

2.2.3 Data of SARS-CoV-2 variants and Spike protein mutant types

The CoVariants / Per Variant section of the GISAID database was used to obtain mutant and variant data of city populations (see Appendix A for more information

about the GISAID database). An Excel file containing the data of all mutants and variants on the GISAID database for the selected countries and the country names was created as a table (see Appendix C “Data Table” for the data table of mutants and variants). The maximum frequency of mutants and variants for each country was used as data in this table. Also, a table that contains time interval frequencies for each mutant and variant type was formed for the countries (see Appendix D).

2.2.4 Microbiome data

As microbiota data, bacterial distribution of gut microbiota in Mobeen's 2018 study (Mobeen et al., 2018) was used for seven countries (Indonesia, India, Japan, Sweden, USA, Italy, Spain) for Set 3 Analysis.

2.3 Analysis

Each data set was analyzed in itself. According to the results, the next data set was analyzed. Two kinds of regression analyses were used with SPSS in the first two sets of analyzes since the data was big enough and has no overfitting problem. Logistic regression analysis was performed with probabilistic programming (WebPPL) in the third set of analyses since the data was small and caused an overfitting problem in SPSS.

Math model: Relations between the variables can be measured by regression analysis and associations between the variables can be measured by correlation analysis. There are numerous regression analysis models, and the model should be selected based on the distribution of the type of the response variable. Linear regression models are based on linear equations to produce the results whereas logistic regression uses odds ratios of the independent variables to produce output (Alexopoulos EC, 2005). For a multi-linear regression model, there must be 10 cases (data entry) for each independent variable (Rodríguez del Águila & Benítez-Parejo, 2011). Confidence intervals (in this study, p-values) are used to represent the

statistical errors in the statistical analyses, in other words, they show how reliable the results are. 0.05 p-value represents the repeatability of the study is 95 % (Alexopoulos EC, 2005). The equation for multi-linear regression (Yale, n.d.)

yi = 0 + 1xi1 + 2xi2 + ... pxip + i for i = 1,2, ... n.

(0, 1, ..., p of the population regression line, n for observations, x for independent variables, and y for dependent variables).

Model selection: In the Set 3 analysis, a logistic regression analysis was performed with the parameter selection model. The parameter selection model is an important model for data analysis (Mohamad et al., 2020). Parameter selection models are used in many areas such as in global sensitivity analysis (Yuan et al., 2019) and smooth functions (Wood et al., 2016). Wrong parameter selections can cause wrong results to evaluate relations among dependent variables and independent variables (Mohamad et al., 2020). Parameter selection is also a very important task in biological modeling since the selection of model parameters is crucial for measuring biological observations (Lillacci & Khammash, 2010). Since the main question of this study is deciding whether there is a relation between SARS-CoV-2 mutants and variants (dependent variables) and bacterial microbiome contents of gut (independent variables), the parameter selection model was used as the model to investigate the relationship between selected variables. The parameter selection model was coded via WebPPL, and logistic regression analysis was embedded in the model to investigate the relationship between dependent variables and independent variables. For the code, see Appendix E.

Bayesian Approach: In set 3 analyses, the analysis was performed with a bayesian approach, not a frequentist approach. In set 3 analyses, analysis was performed with a bayesian approach, not a frequentist approach. The traditional understanding, the frequentist approach, was used in the set 1 and set 2 analyzes. Both of these approaches -Bayesian and frequentist statistics- are used to calculate probability, but there are fundamental differences in their interpretation of probability. Frequentist approaches assumes that the parameters are fixed and data is uncertain- whereas the

Bayesian approach assumes that the parameters are uncertain and the data is known (Bland & Altman, 1998). Bayesian approaches are beneficial when data is limited as they can incorporate prior knowledge about the parameters (Bland & Altman, 1998). The choice of approach is closely related to the appropriate design and parameter selection. For example, the preference for the Bayesian approach has increased recently in late-phase clinical trials (Stallard et al., 2020). The Bayesian approach is based on the principle of updating an antecedent belief with each new data or new observation, and it is not necessary to repeat the event/experiment in obtaining the probability result (as probability density) (Aitchison, 1964). However, the frequentist approach is based on the principle of repeatability of the event to obtain the result, that is, it obtains the probability result with the repeatability of the same event (Aitchison, 1964).

Bayesian Logistic Regression: Logistic regression is used as a linear classifier and has a grouping approach (Srihari, n.d.). Bayesian logistic regression uses a Bayesian approach instead of the classical maximum likelihood methods used in logistic regressions, and this inference model is important for specifying explanatory variables (X. Huang, 2010). Logistic regression analysis is an analysis that aims to get the posterior distribution about the probability of an event and is closely related to Monte Carlo diagrams (Van Erp & Van Gelder, 2013). After determining the prior regression coefficients, the most direct way to apply Bayesian logistic regression is to simulate - infer - the model with the Markov Chain Monte Carlo (MCMC) algorithm (Bayesball, n.d.-b). It is a computational technique used in MCMC Bayes inferences to generate random samples and find a sequence among these samples. In this study, the MCMC algorithm was used as an inference algorithm in the logistic regression model. Especially with very limited prior information about regression parameters, it is often preferable to include Bayesian thinking in the model (Bayesball, n.d.-a). In the case of Set 3 analyzes in this study, the Bayesian Logistic Regression model was used, since it has a limited dataset (a limited a priori microbiota-mutant type dataset for only seven countries).

Stepwise Method: In the multilinear regression method, a linear equation is created about how more than one variable can explain a dependent variable. In this study, classical multi-linear regression analysis was performed with the enter method in SPSS, and the most reasonable equation was tried to be reached with the Stepwise method. The stepwise method is a method used to reach the highest value of the regression equation. It looks at the biggest partial correlation to construct the regression equation, not the biggest correlation between the independent variable and dependent variable and tries to reach the highest regression result step by step by creating a separate equation for adding each independent variable to the previous equation (Johnsson T, 1992).

2.3.1 Softwares

IBM SPSS Statistics version 26 was used for Set 1 and Set 2 analyses for multiple linear regression analysis and bivariate correlation analysis. The missing value analysis was performed for parameters with missing data. The mean of the series was used for those whose significant value was greater than 0.05 in missing value analysis, and thus new parameter sets were created by transferring missing values. Analyzes were made with these new parameter sets.

Since the data set is too small in Set 3 analyzes, deterministic programming tools cannot give the desired results. WebPPL (http://webppl.org/), which is a JavaScript-based language and developed by cognitive scientists, suitable for small data sets with high expressive power, was used in Set 3 analyzes.

2.3.2 Set 1 analysis

First of all, missing value analysis was performed in SPSS and missing values for countries were detected in parameter values. It was checked whether the blank answers were randomly distributed. According to the results of the analysis, it was assumed that the values with significance values of the EM mean values greater than

0.05 were randomly distributed, and the null values resulting in this way were assigned with the replace missing value assignment in SPSS via the series mean method. The resulting values were used as SMEAN values in Set-1 and Set-2 analysis. Later, all variables were standardized. For this, new standardized variables (Zvariable) whose Z-scores (to represent the deviations) were obtained using the standardization method in the Descriptive option in SPSS. These standardized values were used in the analyses. Multi-linear regression and stepwise regression were used for set-1 analyses. For each subgroup parameter, both step-wise and multi-linear regression analyzes were performed. In addition, stepwise regression analysis including all independent parameters was performed. While independent variables in these analyzes were variables in parameter sets, COVID-19 fatality and COVID-19 reproduction rate values were used as dependent variables. For the data of dependent variables, see Appendix A.

2.3.3 Set 2 analysis

Stepwise regression was used for set-2 analysis. Stepwise regression analysis including all independent parameters was performed. While the independent variables in these analyzes were the variables in the parameter sets, the frequency values of each mutant type were used as the dependent variables. For the data of dependent variables, see Appendix C.

2.3.4 Set 3 analysis

In Set 3 analyzes, logistic regression analyzes for the variants and mutants (dependent variables) obtained from Set 2 analyzes were performed using WebPPL (See Appendix E). Microbiome data was used for independent variables for the selected seven countries (Indonesia, India, Japan, Sweden, USA, Italy, Spain) (Mobeen et al., 2018). These countries were selected since they were both in the GISIAD database which means the parameter data was available for them (see

Appendix A and Appendix B)- in the microbiome study (Mobeen et al., 2018). In other words, for these seven countries, the data is available for both the microbiome content of the countries and Set 1 and Set 2 analyses results.

2.4 Testing hypotheses

As each group of analysis was dependent on the previous one, the hypotheses (Table 1.1.) for each set of analyses were constructed based on data from the previous analysis' result (the result of Set-1, the result of Set-2). A small algorithm was created for hypothesis testing and the process was controlled and managed with this algorithm (Figure 2.3).

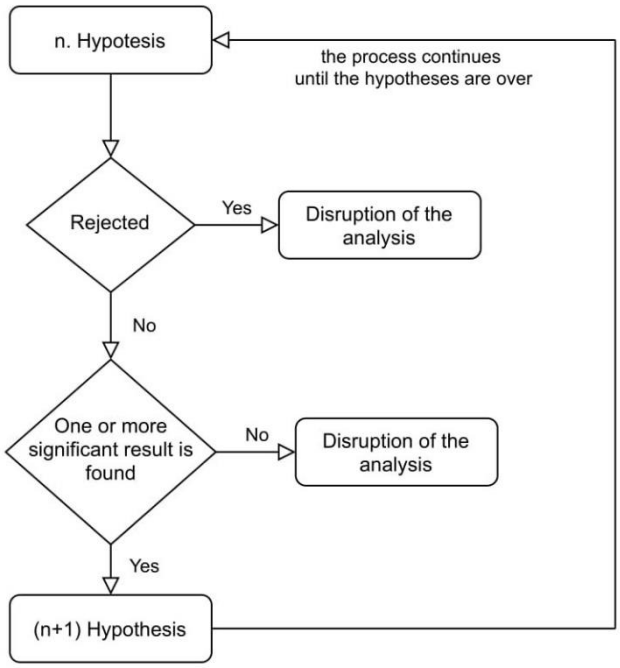


Figure 2.3. Representation of hypotheses testing process as a flowchart.

CHAPTER 3

RESULTS

The calculations were two-tailed. All calculations that were included in this study as results are accessible as the SPSS reports and WebPPL data, codes, and results in the link:

https://drive.google.com/drive/folders/1paPEhtOng3zS3Azu_TlhxzsyjnXqaasP?usp=sharing

Adjusted R2 results were used for multiple linear regression.

3.1 Set 1 results

Hypothesis 1 (H1): *One of the following parameters can explain COVID-19 death and reproduction rates: Diet, diseases, economical parameters, environmental factors, micronutrient deficiencies, population parameters.*

Set 1 analyzes were performed for Hypothesis 1. Multiple linear regression was calculated to predict fatality based on the economy. A regression equation was found (F(5,49)=6,980.,p<.000), with an adjusted R2 of .356. Therefore H1 cannot be rejected (Table 1.1). Multiple linear regression was calculated to predict reproduction based on population. A regression equation was found (F(5,49)=6,162.,p<.000), with an adjusted R2 of .323. Apart from these, no meaningful regression relationship was found. For the regression tables, See Appendix H. In addition to these, correlations between economic variables and fatality; and correlations between population parameters and reproduction were investigated. Some correlations were found between the dependent and independent variables (Table 3.1). To see the correlation matrix for each dependent variable, see Appendix İ.

Table 3.1 Multi-linear regression analysis results of the dependent and independent variables. After regression analysis, correlation analyses were done between each variable in the parameter sets and related dependent variables. Resulted in significant correlations are shown below.

Predictor values	**Predicted value**	**Multi-lineer regression**	**Correlated variables**
Population parameters	Covid-19 reproduction rate	Adjusted R Square = 0.323	**Population size vs Covid-19 reproduction rate:** Pearson = 0.307**, & Spearman = NSR; **Deaths by indoor air pollution rates vs Covid-19 reproduction rate:** Pearson = NSR & Spearman = - 0.295**
Economy parameters	Covid-19 fatality rate	Adjusted R Square = 0.356	**SMEAN(GDP) vs Covid-19 fatality rate:** Pearson = - 0.360* & Spearman = - 0.369*; **SMEAN(Gini index) vs Covid-19 fatality rate:** Pearson = 0.350* & Spearman = NSR; **SMEAN(Conflict cases) vs Covid-19 fatality rate:** Pearson = 0.483* & Spearman = NSR

NSR = Non-significant result; *Significant in 0.01 level, **Significant in 0.05 level.
SMEAN(variable) = The missing values of the data were detected by missing value analysis in SPSS, the significance of EM means was bigger than 0.05, therefore SMEAN variables were created by replacing missing values by using the series mean of the data. The created SMEAN variables were used in the multi-linear regression analysis. All variables were standardized to analyze.

3.2 Set 2 results

Hypothesis 2 (H2): One of the following parameters can explain *SARS-CoV-2 infectivity between variants and specific mutations on Spike protein: Diet, diseases, economical parameters, environmental factors, micronutrient deficiencies, population parameters.*

Set 2 analyzes were performed because it was suggested that Spike protein mutants and SARS-CoV-2 variants could also be affected by selected parameters (Figure 2.1) such as fatality and reproduction (Appendix H). As a result of the analysis, it was found that the 20 variants and mutants were affected by various parameters. Therefore H2 cannot be rejected. See the link: https://drive.google.com/drive/folders/1paPEhtOng3zS3Azu_TlhxzsyjnXqaasP?usp=sharing for the result of SET-2 analyses.

3.3 Set 3 results

Hypothesis 3 (H3): *Microbes residing gut can explain which SARS-CoV-2 mutant infected the host.*

Set 1 and Set 2 analyzes were analyzed with a program (SPSS) due to the sufficient data set size. However, in the hypothesis testing process, the data were generally eliminated at each stage and the data sets were getting smaller; therefore, SPSS and deterministic programming languages were insufficient to analyze the data of Set 3. Therefore, the analyzes were made in the WebPPL by using Bayesian Logistic Regression for Set 3.

In Set 3 analyzes, 20 variants and mutants regressed with different parameters. All the results of Set-2 can be seen in the link: https://drive.google.com/drive/folders/1paPEhtOng3zS3Azu_TlhxzsyjnXqaasP?usp=sharing. These mutants and variants were used in Set-3 analyses.

For seven countries (Indonesia, India, Japan, Sweden, USA, Italy, Spain), gut microbe (Bacteroides, Firmicutes, Actinobacteria, and Proteobacteria) relative abundances were analyzed as independent variables, and the percent frequencies of selected mutants and variants were analyzed as a dependent variable (see Appendix C) by logistic regression model in WebPPL (see Appendix E). Appendix E provides the positive results of the Set-3 analysis. To see all the analyses codes, data, and results for 20 variants and mutants check:

https://drive.google.com/drive/folders/1paPEhtOng3zS3Azu_TlhxzsyjnXqaasP?usp=sharing. The logistic regression analysis was embedded in the parameter selection model and it was examined whether the percentage of each bacterium affected the mutant and variant frequency (see Appendix E for the model and code and Appendix C for variant frequencies). For analysis code and each positive result of the combinations of dependent and independent variables, see Appendix E.

CHAPTER 4

DISCUSSION

In this study, the relationship between Spike protein mutants and intestinal bacteria was investigated by Bayesian logistic regression. It aims to investigate whether the gut microbes can be selective parameters for SARS-CoV-2 mutants and Spike protein variants in humans. To inquire this question, the hypotheses (Table 1.1) were tested with a serial algorithm (Figure 2.3).

Firstly, the relationship between COVID-19 (fatality rate and reproduction rate, Figure 2.1) and a set of parameters (Diet, Diseases, Environmental factors, Economic parameters, Micronutrient Deficiencies, and Population parameters, Figure 2.1) was tested with multi-linear regression analysis, both with the stepwise method and enter method and bivariate correlation analysis in SPSS for Set 1 analysis process (Appendix A). After Set 1 analysis, the second step of the analysis process (Set 2) was made by using SPSS with multi-linear regression analysis, both with the stepwise method and enter method, and bivariate correlation analysis for the set of parameters (Diet, Diseases, Environmental factors, Economic parameters, Micronutrient Deficiencies, Population parameters, Figure 2.1) and SARS-CoV-2 variants and Spike protein mutants that were taken in GISAID database (Appendix C). Lastly, resulted in variants of SARS-CoV-2 and resulted in mutants of Spike protein were analyzed by WebPPL with Bayesian logistic regression analysis that was embedded in a parameter selection model (Appendix E) to see whether there are relationships between SARS-CoV-2 variants, and Spike protein mutants and human gut microbes.

In Set 1 results, the COVID-19 fatality rate (dependent variable) was related to economic parameters. The effort for constructing a linkage between the biological processes with the socio-economic process is common in many disciplines such as

socio-biology (Hirshleifer, 1978) and urban ecology (Grimm et al., 2008b). The results of Set 1 analysis for COVID-19 fatality rate support these efforts by giving a piece of evidence for showing that the economic parameters can be related to COVID-19 deaths. Economic parameters set include the independent variables *of conflict levels, GDP, GINI index, tax rates, and household type* of the countries (Figure 2.1). The relatedness of COVID-19 cases and COVID-19 fatality rates with conflict cases (Bloem & Salemi, 2021), income inequality (Elgar et al., 2020), and socio-economic stratifications (Mikolai et al., 2020) were investigated in the literature. In addition to the COVID-19 fatality rate, the COVID-19 reproduction rate (dependent variable) was related to population. It is known that the spread of SARS-CoV-2 is dependent on population structure such as population size and the median age of the population (Lulbadda et al., 2021) and the results of Set 1/COVID-19 reproduction rate is relevant to these results. *Population Parameters* parameter set (Figure 2.1) includes the independent variables of population size, population growth type, urbanization percent, indoor and outdoor air pollution deaths, and COVID-19 mortality deaths of the countries (Figure 2.1). In this parameters set, the COVID-19 reproduction rate was correlated with the single parameters of population size and indoor deaths. The correlation between population size and COVID-19 reproduction rate is consistent with previous results in the literature (Lulbadda et al., 2021).

Secondly, in Set 2 results, many variants (dependent variable) were related to diet, diseases, economic parameters, environmental factors, and population parameters via various rates. Variants also show various relationships between parameters in the literature. For instance, the 20I/501Y.V1 variant emerged in the United Kingdom and spread throughout to world (To & Editor, 2021). This variant was found majorly in Europe (Figure 4.1). In human reconstituted bronchial epithelium, the 20I/501Y.V1 variant replicates furiously and due to this reason, it spreads rapidly (Touret et al., 1207). The dietary intake affects the human ACE2 receptor, the main target of the Spike protein, by affecting gene expression (Bhattacharya et al., 2021; Horne & Vohl, 2020). Therefore changing the ACE2 structure by dietary patterns can be linked with the results of Set 2. Moreover, chronic diseases are related to

SARS-CoV-2 cases and their severity (H. Liu et al., 2020). The Chronic diseases problem is a big issue for Europe since they have a long life expectancy comparing to the other countries in the world. The average age of Europe is also increasing; therefore, the financial supply of the treatments of chronic diseases is a problem for European countries (Brennan et al., 2017). Therefore, the regression relation with the results can be linked with the emergence of the variants in Europe. It is a known fact that SARS-CoV-2 is related to the economy (Bloem & Salemi, 2021; Elgar et al., 2020; Mikolai et al., 2020), environmental conditions (Asyary & Veruswati, 2020; Travaglio et al., 2021; Xie & Zhu, 2020) and population structure (Connolly et al., 2020; Lulbadda et al., 2021). Therefore, the regression relation with the results can be linked with the various factors that are found in Set 2.

Figure 4.1. The distribution of the 20I/501Y.V1 variant (GISAID database).

For the results of mutant types in Set 2 results, it can be said that environmental factors and economic parameters parameter sets are common. It is in the receptor-binding domain (RBD) of Spike protein; therefore, it is important for both antibody recognition and ACE-2 binding (GISAID, 2021c) S:E484 mutant caused a re-infection in Salvador, one of the big cities of Brazil (Nonaka et al., 2021). Salvador was the first capital of Brazil, and it promotes tourism events (Nobre, 2002) that remarkably increase the air pollution levels (Vianna et al., 2018). Salvador is a coastal city that experiences many beach pollution problems, such as pellet pollution on the beaches (Fernandino et al., 2015). Salvador also has a garbage disposal

problem in its nearby areas, which increases the susceptibility to diarrheal diseases (Rego et al., 2005). Therefore, the toxicity relation of mutants is consistent with these problems. In addition, many mutants can be related to different parameters, for instance, S:Y144 mutation is another Spike protein mutation that is found in 20I/501Y.V1 and other circulating variants, and this mutant are related to antibody escape (Figure 4.3), (GISAID, 2021e). This mutant is related to viral shedding in a patient in Washington (Avanzato et al., 2020) which is one of the biggest metropolises of the United States. This city experiences deaths from increasing heat and excessive ozone concentrations (Jackson et al., 2010). Another example is that, S:H69 is a deletion in Spike, which was sequenced mostly in Europe (Figure 4.4), (Bal et al., 2021). This mutant occurs with other mutants and also has an example of immuno-escape as S:Y144 (GISAID, 2021d). In addition, S:Y144 showed an antibody escape in a lymphoma patient (Avanzato et al., 2020), and S:H69 occurred in a chronically infected immunosuppressed patient treated with rituximab monoclonal antibodies (GISAID, 2021d). It can be pointed out that even there is a treatment with antibodies, this mutant can escape from it. Moreover, economical parameters are related to many of the mutants and variants. This situation can be caused by the relation between the economy and infectious diseases (Goenka et al., 2014). Micronutrient deficiency is another parameter that mutants and variants were related with various rates. It is known that micronutrient deficiencies are related to many diseases (Shenkin, 2006) and population parameters that are caused by demographic, social, and economic aspects of the population (Hwalla et al., 2017). This can be due to the reason that micronutrient deficiency is highly related to economic growth and population dynamics (Darnton-Hill et al., 2005). Micronutrients are essential molecules that provide many functions to the body and play roles in maintaining homeostasis (Shenkin, 2006). Economic parameters are the shared parameter set of many of the mutants and variants. This can be due to the strong relationship between economic activities and viral diseases (Adda, 2016). In this study, it is suggested that these distributions of mutants are related to host-microbiome content (Table 1.1). Therefore, the results of the Set 2 analysis are a kind

of verification that different parameters affect SARS-CoV-2 variants and Spike protein mutants. The verified mutants and the variant that are affected by various were used in Set 3 analysis to see whether there were some relations between them and gut microbes.

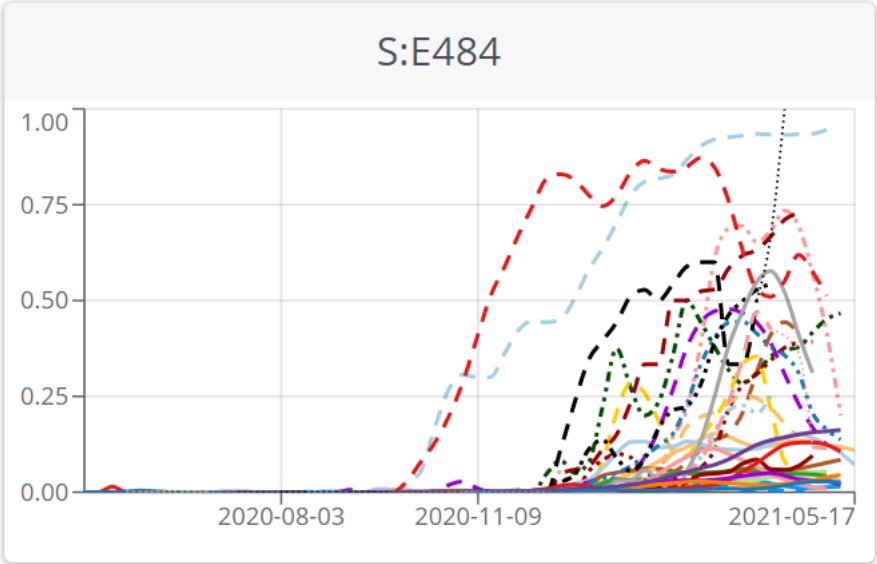


Figure 4.2. Global distribution of S:E484 mutant. Colors represent different countries (GISAID).

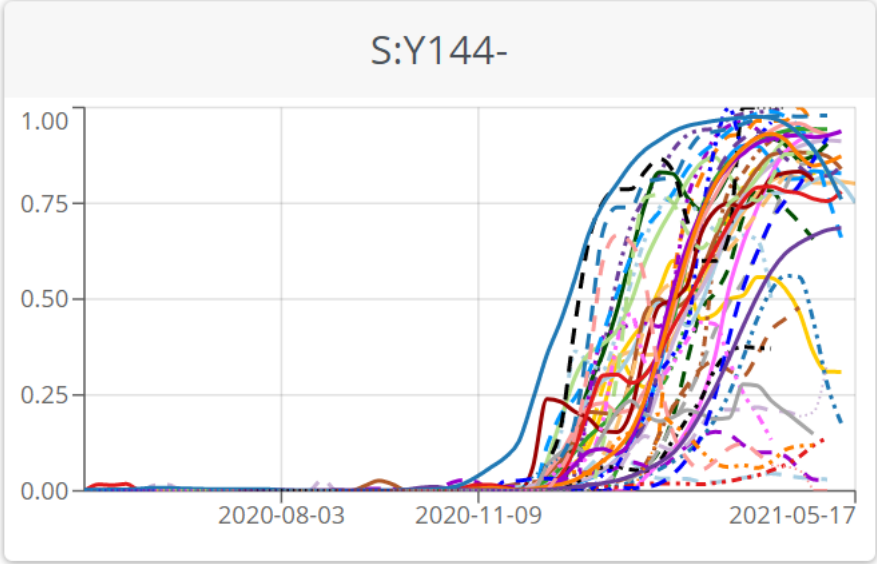


Figure 4.3. Global distribution of S:Y144- mutant. Colors represent different countries (GISAID).

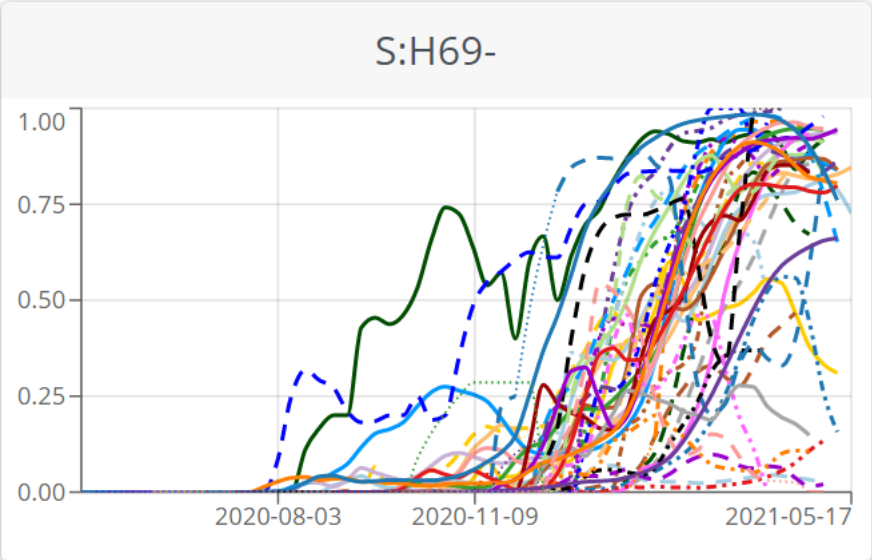


Figure 4.4. Global distribution of S:H69- mutant. Colors represent different countries (GISAID).

Set-3 probability results were very small and therefore only positive results were concerned. In Set-3 results, it was seen that bacteria gave positive regression results at different rates with variants and mutants obtained from Set-2 tests. Actinobacteria is the keystone organism in the gut microbiota (Trosvik & de Muinck, 2015). It can also be argued that variants and mutants that cannot be related to Proteobacteria can be independent of functional diversity in gut microbes, by showing no relation with Proteobacteria. This can be caused by that Proteobacteria are the bacteria responsible for the functional diversity in the intestine. (Bradley & Pollard, 2017). In addition, the higher rate of association between Actionobacter can be related to the high use of probiotic supplements, that improve the intestinal microbiota, in Europe (Nils-Gerrit Wunsch, 2021). Since Firmicutes and Bacteroides are the dominant organisms in the gut microbiome and they provide the majority of ecological relations of human gut microbiota (Bradley & Pollard, 2017), it might be possible that such mutants with high antibody escape rates (GISAID, 2021d, 2021e) would escape more from

the host immune defense depending on the contents of these species, since these species are related with an immune response (Donaldson et al., 2018; Kosiewicz et al., 2011; Peterson et al., 2015). Despite all this, the results of the set-3 analysis are too weak to establish a relationship between bacteria and mutants, and further studies are required to confirm these relationships.

Probabilistic programming languages (PPLs) are used in many areas such as cognitive science, statistics, economy, electronics, environmental modeling, and biology (A. D. Gordon et al., 2014; Krapu & Borsuk, 2019). Phylogenetic analysis is another area to use PPLs (Ronquist et al., 2021). PPLs can infer patterns from data (Gutmann et al., 2011; Merrell & Gitter, 2020). Virology is another area that PPLs are used (Töpfer et al., 2013). PPLs are diverse to make various applications in different areas; for instance, STAN is mainly used by statisticians and (Carpenter et al., 2017) WebPPL which is a feature-rich language that was generated from JavaScript (Ouyang et al., 2018). Although PPLs have some limitations (Gutmann et al., 2011), they made inferences that deterministic programming tools cannot do. In this study, WebPPL was used because of its expressive power on small data sets (Ouyang et al., 2018). Since our dataset in Set 3 analysis was quite small to infer relations between variables in deterministic programs, WebPPL was selected for analysis. The probability results of Set 3 analyses were rather small (Appendix E), but, it can be related to a very small dataset (data of seven countries). Even in a small dataset like this, WebPPL made inferences and give positive and negative relation results.

This study had some limitations. First of all, large datasets were needed to ensure integrity, as data obtained from many different grounds had to be brought together as a whole (Appendix A & Appendix B & Appendix C). This situation may have caused that parameters stand more important in the analysis process compared to others in this data set to be overlooked. Moreover, each analysis process was designed to use the data of another which means that the data was lost in each forthcoming analysis. This data loss and a large number of parameters caused overfitting problems in the outputs. Additionally, as the disciplines of urban ecology

and microbial metacommunity (Miller et al., 2018), which are the theoretical grounds on which this study is based, are the areas that newly reconstructed, there might be various gaps between theoretical and practical applications (Grimm et al., 2008b). Despite the large data sets, the missing data problem was observed intensely in some parameters, such as diseases and micronutrients in Set 1 analyses, that would affect the targeted results in the analysis. Examples such as the lack of sufficient data set to measure microbial interactions, insufficient data on geographic microbial distributions (Mobeen et al., 2018), insufficient data for the countries to be analyzed for vitamin D levels (Appendix A, Appendix B), disease data containing different missing values for each country (Appendix A), and the lack of a clinical study in the area, can be listed as an example of this situation.

SARS-CoV-2 is in a very advantageous position against other viruses in terms of both clinical data and the traceability of its mutants around the world (Petrosillo et al., 2020). However, it may be necessary to establish a control group for this study to study the viral mutant-microbiota relationship in detail and meaningfully. However, in the sense of in silico analysis, data that can be associated, such as the relationship of COVID-19 with human factors, could not be found for other viruses, and the comparison data are mostly on the axis of clinical data. Researchers who want to investigate the viral mutant-microbiota relationship in more detail may be recommended to try to establish a comparable control group for the virus.

Further analyzes were not used in the set 1 and set 2 analyzes, and after each validation, the next hypothesis was moved, as the main goal of the study was to point out a possibility in the microbiota analysis in part 3 -the gut microbiota may show different tolerances to different mutants- by making some validations in the first two sets of analyses.

Understanding microbiota in terms of composition, diversity, and function is being studied, and functional contribution rather than species is thought to be important for establishing microbiota composition. Ecological microbiota studies seek to understand specific gut microbiota functions in pathways of host-microbiome

interactions. In studies of divergence of the microbes in the microbiota, it is known that there is a great deal of species diversity at the species level among humans. Functional diversity studies look at a specific gene and the function performed by certain microbial compositions, because of the idea of forming a microbiota community based on the work performed within that microbial ecosystem rather than at the species level. Although the microbial composition varies greatly between individuals in terms of species diversity, it has been observed that there are not very serious differences between individuals in terms of functionality, in other words, the functional diversity of the human microbiome has been very conserved among people since the core functions in the microbiota have very important places in the metabolic pathways of the host (Lozupone et al., 2012). It is known that some phyla variations are associated with various diseases, especially in the intestinal microbiota (Rinninella et al., 2019).

It is known that some phyla variations are associated with various diseases, especially in the intestinal microbiota. However, in some cases, variations not detected at the phylum level but detected at the species level are also known to affect host status (Wakita et al., 2018)In this study, the geographic variations achieved are at the phylum level, and two dominant phylums (Bacteroidetes and Firmicutes), one keystone phylum (Actionobacter) and one phylum that influence the functional diversification of the microbiome (Proteobacter) were used. This is a limitation of this study because only phylum level analysis was possible, but analysis at other levels, such as species or family, may be related to different host-metabolic factors and functions. For this reason, researchers who want to work on this subject should also consider the functional effects at different levels.

The existence of a crosstalk system between the gut and lung (Srinath et al., 2020) may also suggest that different respiratory viral mutants may affect transmission, virulence, and immune response of the host, as different compositions of microbiota, are known to affect crosstalk networks, although the results of this study do not conclusively point to this relationship. Since Lung microbiota studies are usually performed in laboratory environments isolated from the organism (Tulic M C, Piche

T, 2016), it may be necessary to conduct and investigate such studies at the organismal level.

As a result of the set 3 analyzes, it was seen that some mutants and variants gave positive regression results with various bacteria in the gut microbiota, out of a total of 20 mutants/variants. Due to the very small size of the dataset, only a positive probabilistic result was taken into account when evaluating the results. Datasets, logistic regression code, and results are available in Appendix E. This study looked at whether mutants regressed positively with different bacteria and some positive results were obtained, but the dataset is very small. Therefore, studies involving more data should be conducted to support hypothesis 3. In addition, both clinical data and results from laboratory experiments are needed to study microbiota and different mutant relationships. This study proposes a method for verifying a hypothesis that is difficult to obtain from a small data set with a deterministic probabilistic approach using Bayesian logistic regression. Different datasets and different models can explore this issue better.

The data in the Set 3 analyzes are very limited, the models are too simple, and the data for the 7 countries may not be a complete representation of the countries because these data are the average of the whole country as a result of a limited study, and there is no data to show the differences between countries and within the country itself. Even if these reasons preclude constructing credibility about the set 3 results, however, there may still be a possibility that different microbiota may be a factor in the selection of different mutants and variants due to the approach of Bayesian logistic regression in interpreting this limited data set.

When the data of this study were taken from the GISAID database, there were 3-time intervals. Time intervals were thought to be related to transmission virulence and immune response of the virus, and analyzes were also performed for time intervals and parameters. However, since a significant result could not be obtained, the results were not included in this study because a meaningful regression or

correlation relationship could not be detected between any interval and any parameter set.

CHAPTER 5

CONCLUSION

In this study, it was investigated that whether there is a relationship between Spike protein mutations of SARS-CoV-2 and gut bacteria in humans. According to the results gained by using Bayesian Logistic Regression in WebPPL, a probabilistic programming language, it has been inferred that the toleration of specific human gut microbiota compositions for SARS-CoV-2 Spike proteins varies. While the results obtained are likely very small, this research suggests that there may be meaningful associations between gut bacteria and viral mutants if studied with clinical studies and larger datasets. Although it was not analyzed in detail because of the problem of small data, the results can give an insight for the researchers to develop protective treatments against pathogens by enhancing the host microbiome.

REFERENCES

Adda, J. (2016). Economic activity and the spread of viral diseases: Evidence from high frequency data. Quarterly Journal of Economics, 131(2), 891–941. https://doi.org/10.1093/qje/qjw005

Aitchison, J. (1964). Two Papers on the Comparison of Bayesian and Frequentist Approaches to Statistical Problems of Prediction: Bayesian Tolerance Regions. Journal of the Royal Statistical Society: Series B (Methodological), 26(2), 161–175. https://doi.org/10.1111/j.2517-6161.1964.tb00550.x

Alexopoulos EC. (2005). Introduction to Multivariate Regression Analysis. Hippokratia, 14(1), 23–28.

Ali, I., & Alharbi, O. (2020). COVID-19: Disease, management, treatment, and social impact. The Science of the total environment, 728, 138861. https://doi.org/10.1016/j.scitotenv.2020.138861

Ali, N., & Islam, F. (2020). The Effects of Air Pollution on COVID-19 Infection and Mortality—A Review on Recent Evidence. Frontiers in Public Health, 8(2). https://doi.org/10.3389/fpubh.2020.580057

Anderson, D. C., & Grimes, D. S. (2020). Vitamin D deficiency and COVID-19. Clinical Medicine, Journal of the Royal College of Physicians of London, 20(6), E282–E283. https://doi.org/10.7861/CLINMED.LET.20.6.9

Anderson, G., & Reiter, R. J. (2020). COVID-19 pathophysiology: interactions of gut microbiome, melatonin, vitamin D, stress, kynurenine and the alpha 7 nicotinic receptor: Treatment implications. Melatonin Research, 3(3), 322–345. https://doi.org/10.32794/mr11250066

Arumugam M, Raes J, Pelletier E, Le Paslier D, Yamada T, Mende DR, Fernandes GR, Tap J, Bruls T, Batto JM, Bertalan M, Borruel N, Casellas F, Fernandez L, Gautier L, Hansen T, Hattori M, Hayashi T, Kleerebezem M, Kurokawa K, Leclerc M, Levenez F, Manichanh C, Nielsen HB, Nielsen T, Pons N, Poulain J, Qin J, Sicheritz-Ponten T, Tims S, Torrents D, Ugarte E, Zoetendal EG, Wang J, Guarner F, Pedersen O, de Vos WM, Brunak S, Doré J; MetaHIT Consortium, Antolín M, Artiguenave F, Blottiere HM, Almeida M, Brechot C, Cara C, Chervaux C, Cultrone A, Delorme C, Denariaz G, Dervyn R, Foerstner KU, Friss C, van de Guchte M, Guedon E, Haimet F, Huber W, van Hylckama-Vlieg J, Jamet A, Juste C, Kaci G, Knol J, Lakhdari O, Layec S, Le Roux K, Maguin E, Mérieux A, Melo Minardi R, M'rini C, Muller J, Oozeer R, Parkhill J, Renault P, Rescigno M, Sanchez N, Sunagawa S, Torrejon A, Turner K, Vandemeulebrouck G, Varela E, Winogradsky Y, Zeller G, Weissenbach J, Ehrlich SD, Bork P. Enterotypes of the human gut microbiome. Nature. 2011 May 12;473(7346):174-80. doi:

10.1038/nature09944. Epub 2011 Apr 20. Erratum in: Nature. 2011 Jun 30;474(7353):666. Erratum in: Nature. 2014 Feb 27;506(7489):516. PMID: 21508958; PMCID: PMC3728647.

Asyary, A., & Veruswati, M. (2020). Sunlight exposure increased COVID-19 recovery rates: A study in the central pandemic area of Indonesia. Science of the Total Environment, 729, 139016. https://doi.org/10.1016/j.scitotenv.2020.139016

Avanzato, V. A., Matson, M. J., Seifert, S. N., Pryce, R., Williamson, B. N., Anzick, S. L., Barbian, K., Judson, S. D., Fischer, E. R., Martens, C., Bowden, T. A., de Wit, E., Riedo, F. X., & Munster, V. J. (2020). Case Study: Prolonged Infectious SARS-CoV-2 Shedding from an Asymptomatic Immunocompromised Individual with Cancer. Cell, January, 19–21.

Avise, J. C., Hubbell, S. P., & Ayala, F. J. (2009). In the light of evolution. In the Light of Evolution, 2, 1–432. https://doi.org/10.17226/12501

Bal, A., Destras, G., Gaymard, A., Stefic, K., Marlet, J., Eymieux, S., Regue, H., Semanas, Q., d'Aubarede, C., Billaud, G., Laurent, F., Gonzalez, C., Mekki, Y., Valette, M., Bouscambert, M., Gaudy-Graffin, C., Lina, B., Morfin, F., & Josset, L. (2021). Two-step strategy for the identification of SARS-CoV-2 variant of concern 202012/01 and other variants with spike deletion H69–V70, France, August to December 2020. Eurosurveillance, 26(3), 1–5. https://doi.org/10.2807/1560-7917.ES.2021.26.3.2100008

Bashir, M. F., MA, B. J., Bilal, Komal, B., Bashir, M. A., Farooq, T. H., Iqbal, N., & Bashir, M. (2020). Correlation between environmental pollution indicators and COVID-19 pandemic: A brief study in Californian context. Environmental Research, 187, 109652. https://doi.org/10.1016/j.envres.2020.109652

Banerjee, A., Mossman, K., & Baker, M. L. (2021). Zooanthroponotic potential of SARS-CoV-2 and implications of reintroduction into human populations. Cell Host and Microbe, 29(2), 160–164. https://doi.org/10.1016/j.chom.2021.01.004

Banerjee, S., Schlaeppi, K., & van der Heijden, M. G. A. (2018). Keystone taxa as drivers of microbiome structure and functioning. Nature Reviews Microbiology, 16(9), 567–576. https://doi.org/10.1038/s41579-018-0024-1

Bayesball. (n.d.). A weakly informative prior. https://bayesball.github.io/BOOK/simple-linear-regression.html#a-weakly-informative-prior

Bayesball. (n.d.). Bayesian Logistic Regression. https://bayesball.github.io/BOOK/bayesian-multiple-regression-and-logistic-models.html#bayesian-logistic-regression

Bhayani, S., Sengupta, R., Markossian, T., Tootooni, S., Luke, A., Shoham, D., Cooper, R., & Kramer, H. (2020). Dialysis, COVID-19, Poverty, and Race in Greater Chicago: An Ecological Analysis. Kidney Medicine, 2(5), 552-558.e1. https://doi.org/10.1016/j.xkme.2020.06.005

Bhattacharya, R., Gupta, A. M., Mitra, S., Mandal, S., & Biswas, S. R. (2021). A natural food preservative peptide nisin can interact with the SARS-CoV-2 spike protein receptor human ACE2. Virology, 552(October 2020), 107–111. https://doi.org/10.1016/j.virol.2020.10.002

Bland, J. M., & Altman, D. G. (1998). Statistics notes: Bayesians and frequentists. Bmj, 317(7166), 1151–1160. https://doi.org/10.1136/bmj.317.7166.1151

Bloem, J. R., & Salemi, C. (2021). COVID-19 and conflict. World Development, 140, 105294. https://doi.org/10.1016/j.worlddev.2020.105294

Bradley, P.H., Pollard, K.S. Proteobacteria explain significant functional variability in the human gut microbiome. Microbiome 5, 36 (2017). https://doi.org/10.1186/s40168-017-0244-z

Brennan, P., Perola, M., van Ommen, G. J., & Riboli, E. (2017). Chronic disease research in Europe and the need for integrated population cohorts. European Journal of Epidemiology, 32(9), 741–749. https://doi.org/10.1007/s10654-017-0315-2

Bresalier, R. S., & Chapkin, R. S. (2020). Human Microbiome in Health and Disease: The Good, the Bad, and the Bugly. Digestive Diseases and Sciences, 65(3), 671–673. https://doi.org/10.1007/s10620-020-06059-y

Brown, G.Z., Kline, J., Mhuireach, G. et al. Making microbiology of the built environment relevant to design. Microbiome 4, 6 (2016). https://doi.org/10.1186/s40168-016-0152-7

Bruijning, M., Henry, L. P., Forsberg, S. K. G., Metcalf, C. J. E., & Ayroles, J. F. (2020). When the microbiome defines the host phenotype : selection on vertical transmission in varying environments. BioRxiv Preprint, 1–13.

Brunner, F. S., Deere, J. A., Egas, M., Eizaguirre, C., & Raeymaekers, J. A. M. (2019). The diversity of eco-evolutionary dynamics: Comparing the feedbacks between ecology and evolution across scales. Functional Ecology, 33(1), 7–12. https://doi.org/10.1111/1365-2435.13268

Boni, M.F., Lemey, P., Jiang, X. et al. Evolutionary origins of the SARS-CoV-2 sarbecovirus lineage responsible for the COVID-19 pandemic. Nat Microbiol 5, 1408–1417 (2020). https://doi.org/10.1038/s41564-020-0771-4

Bousquet, J., Anto, J. M., Iaccarino, G., Czarlewski, W., Haahtela, T., Anto, A., Akdis, C. A., Blain, H., Canonica, G. W., Cardona, V., Cruz, A. A., Illario, M., Ivancevich, J. C., Jutel, M., Klimek, L., Kuna, P., Laune, D., Larenas-linnemann, D., Mullol, J., … Zubrinich, C. (2020). Is diet partly responsible

for differences in COVID-19 death rates between and within countries? Clinical and Translational Allergy, 10(1), 1–7. https://doi.org/10.1186/s13601-020-00323-0

Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M. A., Guo, J., Li, P., & Riddell, A. (2017). Stan: A probabilistic programming language. Journal of Statistical Software, 76(1). https://doi.org/10.18637/jss.v076.i01

Carr, A. C. (2020). Micronutrient status of COVID-19 patients: a critical consideration. Critical Care, 24(1), 20–21. https://doi.org/10.1186/s13054-020-03085-0

Chaari, A., Bendriss, G., Zakaria, D., & McVeigh, C. (2020). Importance of Dietary Changes During the Coronavirus Pandemic: How to Upgrade Your Immune Response. Frontiers in Public Health, 8(August), 4–12. https://doi.org/10.3389/fpubh.2020.00476

Cohen, A. J., Anderson, H. R., Ostro, B., Pandey, K. D., Krzyzanowski, M., Künzli, N., Gutschmidt, K., Pope, A., Romieu, I., Samet, J. M., & Smith, K. (2005). The global burden of disease due to outdoor air pollution. Journal of Toxicology and Environmental Health - Part A, 68(13–14), 1301–1307. https://doi.org/10.1080/15287390590936166

Connolly, C., Ali, S. H., & Keil, R. (2020). On the relationships between COVID-19 and extended urbanization. Dialogues in Human Geography, 10(2), 213–216. https://doi.org/10.1177/2043820620934209

Corman VM, Muth D, Niemeyer D, Drosten C. Hosts and Sources of Endemic Human Coronaviruses. Adv Virus Res. 2018;100:163-188. doi: 10.1016/bs.aivir.2018.01.001. Epub 2018 Feb 16. PMID: 29551135; PMCID: PMC7112090.

Cornell. (n.d.). Probabilistic Programming. https://www.cs.cornell.edu/courses/cs4110/2016fa/lectures/lecture33.html

Daly, A., & Zannetti, P. (2007). An Introduction to Air Pollution – Definitions , Classifications , and History. Science And Technology, 1–14.

Darnton-Hill, I., Webb, P., Harvey, P. W. J., Hunt, J. M., Dalmiya, N., Chopra, M., Ball, M. J., Bloem, M. W., & De Benoist, B. (2005). Micronutrient deficiencies and gender: Social and economic costs. American Journal of Clinical Nutrition, 81(5), 1198–1205. https://doi.org/10.1093/ajcn/81.5.1198

Das, B., Ghosh, T. S., Kedia, S., Rampal, R., Saxena, S., Bag, S., Mitra, R., Dayal, M., Mehta, O., Surendranath, A., Travis, S. P. L., Tripathi, P., Nair, G. B., & Ahuja, V. (2018). Analysis of the gut microbiome of rural and urban healthy indians living in sea level and high altitude areas. Scientific Reports, 8(1), 1–15. https://doi.org/10.1038/s41598-018-28550-3

Demers-Mathieu, V., Do, D.M., Mathijssen, G.B. et al. Difference in levels of SARS-CoV-2 S1 and S2 subunits- and nucleocapsid protein-reactive SIgM/IgM, IgG and SIgA/IgA antibodies in human milk. J Perinatol 41, 850–859 (2021). https://doi.org/10.1038/s41372-020-00805-w

Dhar, D., & Mohanty, A. (2020). Gut microbiota and COVID-19- possible link and implications. Virus research, 285, 198018. https://doi.org/10.1016/j.virusres.2020.198018

Dimovski A.S., L. A. (2020). Computing Program Reliability Using Forward-Backward Precondition Analysis and Model Counting. In Fundamental Approaches to Software Engineering. https://doi.org/https://doi.org/10.1007/978-3-030-45234-6_9

Domingo E. (2002). Quasispecies Theory in Virology. Journal of virology, 76(1), 463–465. https://doi.org/10.1128/JVI.76.1.463-465.2002

Donaldson GP, Ladinsky MS, Yu KB, Sanders JG, Yoo BB, Chou WC, Conner ME, Earl AM, Knight R, Bjorkman PJ, Mazmanian SK. Gut microbiota utilize immunoglobulin A for mucosal colonization. Science. 2018 May 18;360(6390):795-800. doi: 10.1126/science.aaq0926. Epub 2018 May 3. PMID: 29724905; PMCID: PMC5973787.

Du, L., He, Y., Zhou, Y., Liu, S., Zheng, B. J., & Jiang, S. (2009). The Spike protein of SARS-CoV - A target for vaccine and therapeutic development. Nature Reviews Microbiology, 7(3), 226–236. https://doi.org/10.1038/nrmicro2090

Gao, M., Piernas, C., Astbury, N. M., Hippisley-Cox, J., O'Rahilly, S., Aveyard, P., & Jebb, S. A. (2021). Associations between body-mass index and COVID-19 severity in 6·9 million people in England: a prospective, community-based, cohort study. The Lancet. Diabetes & Endocrinology, 9(6), 350–359. https://doi.org/10.1016/S2213-8587(21)00089-9

Ghahramani, Z. Probabilistic machine learning and artificial intelligence. Nature 521, 452–459 (2015). https://doi.org/10.1038/nature14541

GISAID. (2021). Variant:S:E484. https://covariants.org/variants/S.E484

GISAID. (2021). Variant:S:H69-. https://covariants.org/variants/S.H69-

GISAID. (2021). Variant:S:Y144. https://covariants.org/variants/S.Y144-

GISAID. (2021). Variant: 20E (EU1). https://covariants.org/variants/20A.EU1

GISAID. (2021). Variant: 20I/501Y.V1. https://covariants.org/variants/S.501Y.V1

Greene, M. W., Roberts, A. P., & Frugé, A. D. (2021). Negative Association Between Mediterranean Diet Adherence and COVID-19 Cases and Related Deaths in Spain and 23 OECD Countries: An Ecological Study. Frontiers in Nutrition, 8(March), 1–7. https://doi.org/10.3389/fnut.2021.591964

Grimm, N., Faeth, S., Golubiewski, N., Redman, C., Wu, J., Bai, X., & Briggs, J. (2008). Global Change and the Ecology of Cities. Science, 319, 756 - 760.

Grimm, N. B., Grove, J. M., Pickett, S. T. A., & Redman, C. L. (2008). Integrated approaches to long-term studies of urban ecological systems. Urban Ecology: An International Perspective on the Interaction Between Humans and Nature, 50(7), 123–141. https://doi.org/10.1007/978-0-387-73412-5_8

Goenka, A., Liu, L., & Nguyen, M. H. (2014). Infectious diseases and economic growth. Journal of Mathematical Economics, 50(1), 34–53. https://doi.org/10.1016/j.jmateco.2013.10.004

Gordon, A. D., Henzinger, T. A., Nori, A. V., & Rajamani, S. K. (2014). Probabilistic programming. Future of Software Engineering, FOSE 2014 - Proceedings, 167–181. https://doi.org/10.1145/2593882.2593900

Gordon DE, et al. Comparative host-coronavirus protein interaction networks reveal pan-viral disease mechanisms. Science. 2020 Oct 15:eabe9403. doi: 10.1126/science.abe9403. Epub ahead of print. PMID: 33060197.

Gui, X., Wang, L., Yao, R., Yu, D., & Li, C. (2019). Investigating the urbanization process and its impact on vegetation change and urban heat island in Wuhan, China. Environmental Science and Pollution Research, 26(30), 30808–30825. https://doi.org/10.1007/s11356-019-06273-w

Guruprasad, L. (2021). Human SARS CoV-2 Spike protein mutations. Proteins: Structure, Function and Bioinformatics, 89(5), 569–576. https://doi.org/10.1002/prot.26042

Gutmann, B., Thon, I., Kimmig, A., Bruynooghe, M., & De Raedt, L. (2011). The magic of logical inference in probabilistic programming. Theory and Practice of Logic Programming, 11(4–5), 663–680. https://doi.org/10.1017/S1471068411000238

Elgar, F. J., Stefaniak, A., & Wohl, M. J. A. (2020). Response to Lindström (2020) on "The trouble with trust: Time-series analysis of social capital, income inequality, and COVID-19 deaths in 84 countries." Social Science and Medicine, 265(October), 19–21. https://doi.org/10.1016/j.socscimed.2020.113518

Enjuanes, L., Gorbalenya, A. E., de Groot, R. J., Cowley, J. A., Ziebuhr, J., & Snijder, E. J. (2008). Nidovirales. Encyclopedia of Virology, 419–430. https://doi.org/10.1016/B978-012374410-4.00775-5

Frank, S., Capriotti, J., Brown, S. M., & Tessema, B. (2020). Povidone-Iodine Use in Sinonasal and Oral Cavities: A Review of Safety in the COVID-19 Era. Ear, Nose and Throat Journal, 99(9), 586–593. https://doi.org/10.1177/0145561320932318

Fernandino, G., Elliff, C. I., Silva, I. R., & Bittencourtc, A. C. S. P. (2015). How many pellets are too many? the pellet pollution index as a tool to assess beach pollution by plastic resin pellets in Salvador, Bahia, Brazil. Journal of Integrated Coastal Zone Management, 15(3), 325–332. https://doi.org/10.5894/rgci566

Hariyanto, T. I., & Kurniawan, A. (2020). Anemia is associated with severe coronavirus disease 2019 (COVID-19) infection. Transfusion and Apheresis Science, 59(6). https://doi.org/10.1016/j.transci.2020.102926

Hirshleifer, J. (1978). Competition, Cooperation, and Conflict in Economics and Biology. The American Economic Review, 68(2), 238–243. http://www.jstor.org/stable/1816696

Ho, J. S. Y., Fernando, D. I., Chan, M. Y., & Sia, C. H. (2020). Obesity in COVID-19: A Systematic Review and Meta-analysis. Annals of the Academy of Medicine, Singapore, 49(12), 996–1008. https://doi.org/10.47102/annals-acadmedsg.2020299

Hoisington, A.J., Brenner, L.A., Kinney, K.A. et al. The microbiome of the built environment and mental health. Microbiome 3, 60 (2015). https://doi.org/10.1186/s40168-015-0127-0

Horne, J. R., & Vohl, M. C. (2020). Biological plausibility for interactions between dietary fat, resveratrol, ACE2, and SARS-CoV illness severity. American Journal of Physiology - Endocrinology and Metabolism, 318(5), E830–E833. https://doi.org/10.1152/AJPENDO.00150.2020

Huang, X. (2010). TRACE : Tennessee Research and Creative Exchange Bayesian Logistic Regression Model for Siting Biomass-using Facilities.

Huang, Y., Yang, C., Xu, Xf. et al. Structural and functional properties of SARS-CoV-2 Spike protein: potential antivirus drug development for COVID-19. Acta Pharmacol Sin 41, 1141–1149 (2020). https://doi.org/10.1038/s41401-020-0485-4

Huang, Z., Liu, Y., Qi, G., Brand, D., & Zheng, S. (2018). Role of Vitamin A in the Immune System. Journal of Clinical Medicine, 7(9), 258. https://doi.org/10.3390/jcm7090258

Hulswit RJ, de Haan CA, Bosch BJ. Coronavirus Spike protein and Tropism Changes. Adv Virus Res. 2016;96:29-57. doi: 10.1016/bs.aivir.2016.08.004. Epub 2016 Sep 13. PMID: 27712627; PMCID: PMC7112277.

Hwalla, N., Al Dhaheri, A. S., Radwan, H., Alfawaz, H. A., Fouda, M. A., Al-Daghri, N. M., Zaghloul, S., & Blumberg, J. B. (2017). The prevalence of micronutrient deficiencies and inadequacies in the middle east and approaches to interventions. Nutrients, 9(3), 1–28. https://doi.org/10.3390/nu9030229

Inkpen S. A. (2019). Health, ecology and the microbiome. eLife, 8, e47626. https://doi.org/10.7554/eLife.47626

Jackson, J. E., Yost, M. G., Karr, C., Fitzpatrick, C., Lamb, B. K., Chung, S. H., Chen, J., Avise, J., Rosenblatt, R. A., & Fenske, R. A. (2010). Public health impacts of climate change in Washington State: Projected mortality risks due to heat events and air pollution. Climatic Change, 102(1–2), 159–186. https://doi.org/10.1007/s10584-010-9852-3

Johnsson T. (1992). A procedure for stepwise regression analysis. Statistical Papers, 33, 21–29. https://link-springer-com.ezproxy.library.yorku.ca/content/pdf/10.1007%2FBF02925308.pdf

Jothimani, D., Kailasam, E., Danielraj, S., Nallathambi, B., Ramachandran, H., Sekar, P., Manoharan, S., Ramani, V., Narasimhan, G., Kaliamoorthy, I., & Rela, M. (2020). COVID-19: Poor outcomes in patients with zinc deficiency. International Journal of Infectious Diseases, 100, 343–349. https://doi.org/10.1016/j.ijid.2020.09.014

Karl, T. R., & Trenberth, K. E. (2003). Modern Global Climate Change. Science, 302(5651), 1719–1723. https://doi.org/10.1126/science.1090228

Kembel, S. W., Meadow, J. F., O'Connor, T. K., Mhuireach, G., Northcutt, D., Kline, J., Moriyama, M., Brown, G. Z., Bohannan, B. J., & Green, J. L. (2014). Architectural design drives the biogeography of indoor bacterial communities. PloS one, 9(1), e87093. https://doi.org/10.1371/journal.pone.0087093

King, G.M. Urban microbiomes and urban ecology: How do microbes in the built environment affect human sustainability in cities?. J Microbiol. 52, 721–728 (2014). https://doi.org/10.1007/s12275-014-4364-x

Kluge, H. H. P., Wickramasinghe, K., Rippin, H. L., Mendes, R., Peters, D. H., Kontsevaya, A., & Breda, J. (2020). Prevention and control of non-communicable diseases in the COVID-19 response. The Lancet, 395(10238), 1678–1680. https://doi.org/10.1016/S0140-6736(20)31067-9

Kosiewicz, M. M., Zirnheld, A. L., & Alard, P. (2011). Gut microbiota, immunity, and disease: A complex relationship. Frontiers in Microbiology, 2(SEP), 1–11. https://doi.org/10.3389/fmicb.2011.00180

Krammer, F. SARS-CoV-2 vaccines in development. Nature 586, 516–527 (2020). https://doi.org/10.1038/s41586-020-2798-3

Krapu, C., & Borsuk, M. (2019). Probabilistic programming: A review for environmental modellers. Environmental Modelling and Software, 114, 40–48. https://doi.org/10.1016/j.envsoft.2019.01.014

Kumar, M., Sarma, D. K., Shubham, S., Kumawat, M., Verma, V., Prakash, A., & Tiwari, R. (2020). Environmental Endocrine-Disrupting Chemical Exposure:

Role in Non-Communicable Diseases. Frontiers in Public Health, 8(September), 1–28. https://doi.org/10.3389/fpubh.2020.553850Lazar V, Ditu LM, Pircalabioru GG, Gheorghe I, Curutiu C, Holban AM, Picu A, Petcu L, Chifiriuc MC. Aspects of Gut Microbiota and Immune System Interactions in Infectious Diseases, Immunopathology, and Cancer. Front Immunol. 2018 Aug 15;9:1830. doi: 10.3389/fimmu.2018.01830. PMID: 30158926; PMCID: PMC6104162.

Lazar, V., Ditu, L. M., Pircalabioru, G. G., Gheorghe, I., Curutiu, C., Holban, A. M., Picu, A., Petcu, L., & Chifiriuc, M. C. (2018). Aspects of gut microbiota and immune system interactions in infectious diseases, immunopathology, and cancer. Frontiers in Immunology, 9(AUG), 1–18. https://doi.org/10.3389/fimmu.2018.01830

Lazarus, J.V., Ratzan, S.C., Palayew, A. et al. A global survey of potential acceptance of a COVID-19 vaccine. Nat Med 27, 225–228 (2021). https://doi.org/10.1038/s41591-020-1124-9

Leibold, M.A., Holyoak, M., Mouquet, N., Amarasekare, P., Chase, J.M., Hoopes, M.F., Holt, R.D., Shurin, J.B., Law, R., Tilman, D., Loreau, M. and Gonzalez, A. (2004), The metacommunity concept: a framework for multi-scale community ecology. Ecology Letters, 7: 601-613. https://doi.org/10.1111/j.1461-0248.2004.00608.x

Li, E., Bestor, T. H., Jaenisch, R., Shreiner, A. B., Kao, J. Y., Young, V. B., Turnbaugh, P. J., Ley, R. E., Mahowald, M. A., Magrini, V., Mardis, E. R., Gordon, J. I., Wilson, R. C., & Doudna, J. A. (2015). Dicer (E-7): sc-393328. Cell, 69(1), 393328. https://doi.org/10.1097/MOG.0000000000000139.The

Li, P., Lu, M., & Zheng, Y. (2020). A Note on the Role of Cultural, Institutional, and Urbanization Features in the COVID-19 Pandemic. Asian Economic Papers, 60–70. https://doi.org/10.1162/asep_a_00798

Li, R., Wu, K., Li, Y., Liang, X., Fai Tse, W. K., Yang, L., & Lai, K. P. (2020). Revealing the targets and mechanisms of vitamin A in the treatment of COVID-19. Aging, 12(15), 15784–15796. https://doi.org/10.18632/AGING.103888

Lillacci, G., & Khammash, M. (2010). Parameter estimation and model selection in computational biology. PLoS Computational Biology, 6(3). https://doi.org/10.1371/journal.pcbi.1000696

Liu, H., Chen, S., Liu, M., Nie, H., & Lu, H. (2020). Comorbid chronic diseases are strongly correlated with disease severity among COVID-19 patients: A systematic review and meta-analysis. Aging and Disease, 11(3), 668–678. https://doi.org/10.14336/AD.2020.0502

Liu, D. X., Liang, J. Q., & Fung, T. S. (2021). Human Coronavirus-229E, -OC43, -NL63, and -HKU1 (Coronaviridae). Encyclopedia of Virology, January, 428–440. https://doi.org/10.1016/b978-0-12-809633-8.21501-x

Loftus, M., Hassouneh, S. A. D., & Yooseph, S. (2021). Bacterial associations in the healthy human gut microbiome across populations. Scientific Reports, 11(1), 1–14. https://doi.org/10.1038/s41598-021-82449-0

Lu, L., Manopo, I., Leung, B. P., Chng, H. H., Ling, A. E., Chee, L. L., Ooi, E. E., Chan, S. W., & Kwang, J. (2004). Immunological Characterization of the Spike protein of the Severe Acute Respiratory Syndrome Coronavirus. Journal of Clinical Microbiology, 42(4), 1570–1576. https://doi.org/10.1128/JCM.42.4.1570-1576.2004

Lulbadda, K. T., Kobbekaduwa, D., & Guruge, M. L. (2021). The impact of temperature, population size and median age on COVID-19 (SARS-CoV-2) outbreak. Clinical Epidemiology and Global Health, 9(June), 231–236. https://doi.org/10.1016/j.cegh.2020.09.004

Marchesi, J. R., & Ravel, J. (2015). The vocabulary of microbiome research: a proposal. Microbiome, 3(1), 1–3. https://doi.org/10.1186/s40168-015-0094-5

Manor, O., Dai, C. L., Kornilov, S. A., Smith, B., Price, N. D., Lovejoy, J. C., Gibbons, S. M., & Magis, A. T. (2020). Health and disease markers correlate with gut microbiome composition across thousands of people. Nature Communications, 11(1), 1–12. https://doi.org/10.1038/s41467-020-18871-1

Merrell, D., & Gitter, A. (2020). Inferring signaling pathways with probabilistic programming. Bioinformatics, 36(26), I822–I830. https://doi.org/10.1093/bioinformatics/btaa861

Mertens, E., & Peñalvo, J. L. (2021). The Burden of Malnutrition and Fatal COVID-19: A Global Burden of Disease Analysis. Frontiers in Nutrition, 7(January), 1–12. https://doi.org/10.3389/fnut.2020.619850

Mofijur, M., Fattah, I. M. R., Alam, M. A., Islam, A. B. M. S., Ong, H. C., Rahman, S. M. A., Najafi, G., Ahmed, S. F., Uddin, M. A., & Mahlia, T. M. I. (2021). Impact of COVID-19 on the social, economic, environmental and energy domains: Lessons learnt from a global pandemic. Sustainable Production and Consumption, 26, 343–359. https://doi.org/10.1016/j.spc.2020.10.016

Mohamad, M., Selamat, A., Krejcar, O., Fujita, H., & Wu, T. (2020). An analysis on new hybrid parameter selection model performance over big data set. Knowledge-Based Systems, 192, 105441. https://doi.org/10.1016/j.knosys.2019.105441

Moris, D., Tsilimigras, D. I., & Schizas, D. (2020). Cancer and COVID-19. The Lancet, 396(10257), 1066. https://doi.org/10.1016/S0140-6736(20)32069-9

Mikolai, J., Keenan, K., & Kulu, H. (2020). Intersecting household-level health and socio-economic vulnerabilities and the COVID-19 crisis: An analysis from the UK. SSM - Population Health, 12(June), 100628. https://doi.org/10.1016/j.ssmph.2020.100628

Miller ET, Svanbäck R, Bohannan BJM. Microbiomes as Metacommunities: Understanding Host-Associated Microbes through Metacommunity Ecology. Trends in Ecology & Evolution. 2018 Dec;33(12):926-935. DOI: 10.1016/j.tree.2018.09.002.

Mitsos, A., Najman, J., & Kevrekidis, I. G. (2018). Optimal deterministic algorithm generation. Journal of Global Optimization, 71(4), 891–913. https://doi.org/10.1007/s10898-018-0611-8

Mobeen, F., Sharma, V., & Tulika, P. (2018). Enterotype Variations of the Healthy Human Gut Microbiome in Different Geographical Regions. Bioinformation, 14(9), 560–573. https://doi.org/10.6026/97320630014560

Niemelä, J. (2000). Johnson-2000. Urban Ecosystems, 3(1), 1–9. papers3://publication/uuid/8861234A-E8E2-419A-AAFB-20E665F902E0

Nils-Gerrit Wunsch. (2021). Probiotic supplements market value in Europe 2017-2027, by application Published by Nils-Gerrit Wunsch, Jan 25, 2021 In Europe, the market value of probiotic supplements for gastrointestinal health is forecast to remain the largest, in comparison with oth. https://www.statista.com/statistics/1198594/forecast-of-the-probiotic-supplements-market-in-europe-by-application/

Nobre, E. A. C. (2002). Urban regeneration experiences in Brazil: Historical preservation, tourism development and gentrification in Salvador da Bahia. Urban Design International, 7(2), 109–124. https://doi.org/10.1057/palgrave.udi.9000066

Nonaka, C. K. V., Franco, M. M., Gräf, T., de Lorenzo Barcia, C. A., de Ávila Mendonça, R. N., de Sousa, K. A. F., Neiva, L. M. C., Fosenca, V., Mendes, A. V. A., de Aguiar, R. S., Giovanetti, M., & de Freitas Souza, B. S. (2021). Genomic Evidence of SARS-CoV-2 Reinfection Involving E484K Spike Mutation, Brazil. Emerging Infectious Diseases, 27(5), 1522–1524. https://doi.org/10.3201/eid2705.210191

OECD Policy Responses to Coronavirus. (2021). Access to COVID-19 vaccines: Global approaches in a global crisis. March, 1–30. https://www.oecd-ilibrary.org/social-issues-migration-health/access-to-covid-19-vaccines-global-approaches-in-a-global-crisis_c6a18370-en

Obeid, A., Bruza, P. D., & Wittek, P. (2018). Evaluating probabilistic programming languages for simulating quantum correlations. In arXiv.

Olmedo, F., Gretz, F., Jansen, N., Kaminski, B. L., Katoen, J. P., & McIver, A. (2018). Conditioning in probabilistic programming. ACM Transactions on Programming Languages and Systems, 40(1), 1–50. https://doi.org/10.1145/3156018

Ouyang, L., Tessler, M. H., Ly, D., & Goodman, N. D. (2018). webppl-oed: A practical optimal experiment design system. CogSci, 2192–2197.

Ömeroğlu Şimşek, G. (2020). Lung cancer management in covid-19 pandemic. Turkish Thoracic Journal, 21(5), 340–344. https://doi.org/10.5152/TurkThoracJ.2020.20061

Pak, A., Adegboye, O. A., Adekunle, A. I., Rahman, K. M., McBryde, E. S., & Eisen, D. P. (2020). Economic Consequences of the COVID-19 Outbreak: the Need for Epidemic Preparedness. Frontiers in Public Health, 8(May), 1–4. https://doi.org/10.3389/fpubh.2020.00241

Pal, R., & Banerjee, M. (2020). COVID-19 and the endocrine system: exploring the unexplored. Journal of Endocrinological Investigation, 43(7), 1027–1031. https://doi.org/10.1007/s40618-020-01276-8

Pereson, MJ, Flichman, DM, Martínez, AP, Baré, P, Garcia, GH, Di Lello, FA. Evolutionary analysis of SARS-CoV-2 Spike protein for its different clades. J Med Virol. 2021; 93: 3000– 3006. https://doi.org/10.1002/jmv.26834

Peterson, D. A., Planer, J. D., Guruge, J. L., Xue, L., Downey-Virgin, W., Goodman, A. L., Seedorf, H., & Gordon, J. I. (2015). Characterizing the interactions between a naturally primed immunoglobulin a and its conserved Bacteroides thetaiotaomicron species-specific epitope in gnotobiotic mice. Journal of Biological Chemistry, 290(20), 12630–12649. https://doi.org/10.1074/jbc.M114.633800

Petrosillo, N., Viceconte, G., Ergonul, O., Ippolito, G., & Petersen, E. (2020). COVID-19, SARS and MERS: are they closely related?. Clinical microbiology and infection : the official publication of the European Society of Clinical Microbiology and Infectious Diseases, 26(6), 729–734. https://doi.org/10.1016/j.cmi.2020.03.026

Pickard, J. M., Zeng, M. Y., Caruso, R., & Núñez, G. (2017). Gut microbiota: Role in pathogen colonization, immune responses, and inflammatory disease. Immunological reviews, 279(1), 70–89. https://doi.org/10.1111/imr.12567

Pickett, S.T.A., Cadenasso, M.L., Childers, D.L., McDonnell, M.J., and Zhou, W.. 2016. Evolution and future of urban ecological science: ecology in, of, and for the city. Ecosystem Health and Sustainability 2(7):e01229. doi: 10.1002/ehs2.1229

Pickett, S.T.A., Burch, W.R., Dalton, S.E. et al. A conceptual framework for the study of human ecosystems in urban areas. Urban Ecosystems 1, 185–199 (1997). https://doi.org/10.1023/A:1018531712889

Rees T, Bosch T, Douglas AE (2018) How the microbiome challenges our concept of self. PLOS Biology 16(2): e2005358. https://doi.org/10.1371/journal.pbio.2005358

Reese, A. T., Savage, A., Youngsteadt, E., McGuire, K. L., Koling, A., Watkins, O., Frank, S. D., & Dunn, R. R. (2016). Urban stress is associated with variation in microbial species composition-but not richness-in Manhattan. The ISME journal, 10(3), 751–760. https://doi.org/10.1038/ismej.2015.152

Rego, R. F., Moraes, L. R. S., & Dourado, I. (2005). Diarrhoea and garbage disposal in Salvador, Brazil. Transactions of the Royal Society of Tropical Medicine and Hygiene, 99(1), 48–54. https://doi.org/10.1016/j.trstmh.2004.02.008

Rehfuess, E., Corvalan, C., & Neira, M. (2006). Indoor air pollution: 4000 deaths a day must no longer be ignored. Bulletin of the World Health Organization, 84(7), 32979. https://doi.org/10.2471/BLT.06.032979

Relman, D. Learning about who we are. Nature 486, 194–195 (2012). https://doi.org/10.1038/486194a

Riffat, S., Powell, R. & Aydin, D. Future cities and environmental sustainability. Fut Cit & Env 2, 1 (2016). https://doi.org/10.1186/s40984-016-0014-2

Rinninella, E., Raoul, P., Cintoni, M., Franceschi, F., Miggiano, G. A. D., Gasbarrini, A., & Mele, M. C. (2019). What is the healthy gut microbiota composition? A changing ecosystem across age, environment, diet, and diseases. Microorganisms, 7(1). https://doi.org/10.3390/microorganisms7010014

Rodríguez del Águila, M. M., & Benítez-Parejo, N. (2011). Simple linear and multivariate regression models. Allergologia et Immunopathologia, 39(3), 159–173. https://doi.org/10.1016/j.aller.2011.02.001

Roggema, R. (2016). The future of sustainable urbanism: a redefinition. City, Territory and Architecture, 3(1). https://doi.org/10.1186/s40410-016-0052-y

Ronquist, F., Kudlicka, J., Senderov, V., Borgström, J., Lartillot, N., Lundén, D., Murray, L., Schön, T. B., & Broman, D. (2021). Universal probabilistic programming offers a powerful approach to statistical phylogenetics. Communications Biology, 4(1), 1–10. https://doi.org/10.1038/s42003-021-01753-7

Sarker, I. H. (2021). Machine Learning: Algorithms, Real-World Applications and Research Directions. SN Computer Science, 2(3), 1–21. https://doi.org/10.1007/s42979-021-00592-x

Scanlan, P. D. (2019). Microbial evolution and ecological opportunity in the gut environment. Proceedings of the Royal Society B: Biological Sciences, 286(1915), 1–9. https://doi.org/10.1098/rspb.2019.1964

Scepanovic, P., Hodel, F., Mondot, S., Partula, V., Byrd, A., Hammer, C., Alanio, C., Bergstedt, J., Patin, E., Touvier, M., Lantz, O., & Albert, M. L. (2019). A comprehensive assessment of demographic , environmental , and host genetic associations with gut microbiome diversity in healthy individuals. 1–15. Segal LN, Blaser MJ. A brave new world: the lung microbiota in an era of change. Ann Am Thorac Soc. 2014 Jan;11 Suppl 1(Suppl 1):S21-7. doi: 10.1513/AnnalsATS.201306-189MG. PMID: 24437400; PMCID: PMC3972973.

Sencio, V., Machado, M. G., & Trottein, F. (2021). The lung–gut axis during viral respiratory infections: the impact of gut dysbiosis on secondary disease outcomes. Mucosal Immunology, 14(2), 296–304. https://doi.org/10.1038/s41385-020-00361-8

Shenkin, A. (2006). Micronutrients in health and disease. Postgraduate Medical Journal, 82(971), 559–567. https://doi.org/10.1136/pgmj.2006.047670

Singh, B., Bardgett, R., Smith, P. et al. Microorganisms and climate change: terrestrial feedbacks and mitigation options. Nat Rev Microbiol 8, 779–790 (2010). https://doi.org/10.1038/nrmicro2439

Srihari, S. N. (n.d.). Linear Models for Classification: Introduction.

Srinath, B. S., Shastry, R. P., & Kumar, S. B. (2020). Role of gut-lung microbiome crosstalk in COVID-19. Research on Biomedical Engineering. https://doi.org/10.1007/s42600-020-00113-4

Skevaki, C., Karsonova, A., Karaulov, A., Fomina, D., Xie, M., Chinthrajah, S., Nadeau, K. C., & Renz, H. (2021). SARS-CoV-2 infection and COVID-19 in asthmatics: a complex relationship. Nature Reviews Immunology, 21(4), 202–203. https://doi.org/10.1038/s41577-021-00516-z

Stamper CE, Hoisington AJ, Gomez OM, Halweg-Edwards AL, Smith DG, Bates KL, Kinney KA, Postolache TT, Brenner LA, Rook GA, Lowry CA. The Microbiome of the Built Environment and Human Behavior: Implications for Emotional Health and Well-Being in Postmodern Western Societies. Int Rev Neurobiol. 2016;131:289-323. doi: 10.1016/bs.irn.2016.07.006. Epub 2016 Sep 6. PMID: 27793224.

Sun, L., Chen, J., Li, Q., & Huang, D. (2020). Dramatic uneven urbanization of large cities throughout the world in recent decades. Nature Communications, 11(1). https://doi.org/10.1038/s41467-020-19158-1

Stallard, N., Todd, S., Ryan, E. G., & Gates, S. (2020). Comparison of Bayesian and frequentist group-sequential clinical trial designs. BMC Medical Research Methodology, 20(1), 1–14. https://doi.org/10.1186/s12874-019-0892-8

Stouffer, R. J., Yin, J., Gregory, J. M., Dixon, K. W., Spelman, M. J., Hurlin, W., Weaver, A. J., Eby, M., Flato, G. M., Hasumi, H., Hu, A., Jungclaus, J. H., Kamenkovich, I. V., Levermann, A., Montoya, M., Murakami, S., Nawrath, S., Oka, A., Peltier, W. R., … Weber, S. L. (2006). Investigating the causes of the response of the thermohaline circulation to past and future climate changes. 86th AMS Annual Meeting.

Queen, D. (2021). Technological impact of COVID-19. International Wound Journal, 18(2), 129–130. https://doi.org/10.1111/iwj.13578

Tavares, Z., Zhang, X., Minaysan, E., Burroni, J., Ranganath, R., & Lezama, A. S. (2019). The random conditional distribution: For higher-order probabilistic inference. ArXiv.

Thursby, E., & Juge, N. (2017). Introduction to the human gut microbiota. Biochemical Journal, 474(11), 1823–1836. https://doi.org/10.1042/BCJ20160510

Travaglio, M., Yu, Y., Popovic, R., Selley, L., Leal, N. S., & Martins, L. M. (2021). Links between air pollution and COVID-19 in England. Environmental Pollution, 268, 115859. https://doi.org/10.1016/j.envpol.2020.115859

Trosvik, P., & de Muinck, E. J. (2015). Ecology of bacteria in the human gastrointestinal tract--identification of keystone and foundation taxa. Microbiome, 3, 44. https://doi.org/10.1186/s40168-015-0107-4 Tourdot, R. W., & Zhang, C. Z. (2019). Complete Haplotype Determination and Single-Chromosome Analysis. BioRxiv. https://doi.org/10.1101/629337

To, L., & Editor, T. H. E. (2021). The basis of a more contagious 501Y . V1 variant of. 2–4. https://doi.org/10.1038/s41422-021-00496-8

Tosepu, R., Gunawan, J., Effendy, D. S., Ahmad, L. O. A. I., Lestari, H., Bahar, H., & Asfian, P. (2020). Correlation between weather and Covid-19 pandemic in Jakarta, Indonesia. Science of the Total Environment, 725, 138436. https://doi.org/10.1016/j.scitotenv.2020.138436

Touret, F., Luciani, L., Baronti, C., Cochin, M., Driouich, J.-S., Gilles, M., Thirion, L., Nougairède, A., & De Lamballerie, X. (1207). Replicative fitness SARS-CoV-2 20I/501Y.V1 variant in a human reconstituted 1 bronchial epithelium. BioRxiv, 2021.03.22.436427. https://doi.org/10.1101/2021.03.22.436427

Töpfer A, Zagordi O, Prabhakaran S, Roth V, Halperin E, Beerenwinkel N. Probabilistic inference of viral quasispecies subject to recombination. J Comput Biol. 2013 Feb;20(2):113-23. doi: 10.1089/cmb.2012.0232. PMID: 23383997; PMCID: PMC3576916.

Tulic M C, Piche T, V. (2016). Lung-gut cross-talk: evidence, mechanisms and implications for the mucosal inflammatory diseases. Clin Exp Allergy . https://doi.org/doi: 10.1111/cea.12723.

Van Erp, N., & Van Gelder, P. (2013). Bayesian logistic regression analysis. AIP Conference Proceedings, 1553, 147–154. https://doi.org/10.1063/1.4819994

Venter, Z. S., Barton, D. N., Gundersen, V., Figari, H., & Nowell, M. (2020). Urban nature in a time of crisis: Recreational use of green space increases during the COVID-19 outbreak in Oslo, Norway. Environmental Research Letters, 15(10). https://doi.org/10.1088/1748-9326/abb396

V'kovski, P., Kratzel, A., Steiner, S. et al. Coronavirus biology and replication: implications for SARS-CoV-2. Nat Rev Microbiol 19, 155–170 (2021). https://doi.org/10.1038/s41579-020-00468-6

Vianna, N. A., Novaes, P., Gnoatto, N., Miraglia, S. G. E. K., Saldiva, P. H. N., & Andrade, L. R. (2018). Diesel Emissions Increase Air Pollution during the Carnival Festival in Salvador , Bahia-Brazil. Preprints, 2, 1–14. https://doi.org/10.20944/preprints201804.0346.v1

Xie, J., & Zhu, Y. (2020). Association between ambient temperature and COVID-19 infection in 122 cities from China. *Science of the Total Environment*, *724*, 138201. https://doi.org/10.1016/j.scitotenv.2020.138201

Yale. (n.d.). Multiple Linear Regression. http://www.stat.yale.edu/Courses/1997-98/101/linmult.htm

Yan Yam, E. Lo. (2021). Climate change and the origin of SARS-cov-2. Journal of Travel Medicine, 27(8), 1–2. https://doi.org/10.1093/JTM/TAAA224

Yeoh YK, Zuo T, Lui GC, et alGut microbiota composition reflects disease severity and dysfunctional immune responses in patients with COVID-19Gut 2021;70:698-706.

Yoshimoto, F.K. The Proteins of Severe Acute Respiratory Syndrome Coronavirus-2 (SARS CoV-2 or n-COV19), the Cause of COVID-19. Protein J 39, 198–216 (2020). https://doi.org/10.1007/s10930-020-09901-4

Yuan, Z., Liang, P., Silva, T., Yu, K., & Mottershead, J. E. (2019). Parameter selection for model updating with global sensitivity analysis. Mechanical Systems and Signal Processing, 115, 483–496. https://doi.org/10.1016/j.ymssp.2018.05.048

Wakita, Y., Shimomura, Y., Kitada, Y., Yamamoto, H., Ohashi, Y., & Matsumoto, M. (2018). Taxonomic classification for microbiome analysis, which correlates well with the metabolite milieu of the gut. BMC Microbiology, 18(1), 1–11. https://doi.org/10.1186/s12866-018-1311-8

Wertheim, J. O., Chu, D. K., Peiris, J. S., Kosakovsky Pond, S. L., & Poon, L. L. (2013). A case for the ancient origin of coronaviruses. Journal of virology, 87(12), 7039–7045. https://doi.org/10.1128/JVI.03273-12

Wessels, I., Rolles, B., & Rink, L. (2020). The Potential Impact of Zinc Supplementation on COVID-19 Pathogenesis. Frontiers in Immunology, 11(July), 1–11. https://doi.org/10.3389/fimmu.2020.01712

Wood, S. N., Pya, N., & Säfken, B. (2016). Smoothing Parameter and Model Selection for General Smooth Models. Journal of the American Statistical Association, 111(516), 1548–1563. https://doi.org/10.1080/01621459.2016.1180986

APPENDICES

A. SET 1 DATA TABLE : PARAMETERS

COUNTRIES	GDP	GINI INDEX	CONFLICT	TAX RATES	HOUSEHOLD	BMI	VEGETABLE OIL CONSUMPTION	ANIMAL FAT CONSUMPTION	SUGAR CONSUMPTION	UNDERNOURHSMENT
ARUBA	29008	null	null	25	2.9	null	null	null	null	null
AUSTRALIA	55057	34.40	null	30	2.5	64.50	23.40	5.70	60.40	3
AUSTRIA	50122	30.80	17	25	2.3	54.30	15.50	15.80	46.90	3
BANGLADESH	1856	32.40	634	25	4.5	20.00	7.00	0.20	9.30	13
BELGIUM	46345	27.20	26	25	2.3	59.50	11.30	17.10	58.20	3
BRAZIL	8717	53.40	6658	34	3.3	56.50	19.90	4.20	42.80	3
BULGARIA	9828	41.30	31	10	2.3	61.70	10.00	3.40	34.30	3
CANADA	46190	33.30	null	26	2.4	64.10	25.90	15.20	90.30	3
CHILE	14897	44.40	777	25	3.6	63.10	7.90	2.50	48.00	4
CROATIA	14944	29.70	12	18	2.8	59.60	8.10	4.00	56.10	3
CYPRUS	27858	32.70	20	13	2.8	59.10	15.80	0.40	59.30	7
CZECHIA	23490	25.00	6	19	2.3	62.30	14.80	15.20	50.10	3
DENMARK	60213	28.20	30	22	null	55.40	1.70	22.90	55.00	3
ESTONIA	23718	30.30	1	20	2.3	55.80	7.70	9.20	52.30	3
FINLAND	48771	27.30	8	20	2.1	57.90	6.20	11.90	40.30	3
FRANCE	40496	32.40	252	32	2.2	59.50	17.40	14.10	38.30	3
GERMANY	46468	31.90	230	30	2.1	56.80	14.40	12.60	48.10	3
GHANA	2202	43.50	115	25	3.5	32.00	6.30	0.20	14.10	7
GREECE	19581	32.90	173	24	2.6	62.30	29.60	3.10	30.30	3
HUNGARY	16730	29.60	1	9	2.4	61.60	10.40	15.10	40.90	3
ICELAND	67084	26.10	1	20	null	59.10	9.00	18.60	57.10	3
INDIA	2100	35.70	3413	30	4.6	19.70	7.70	0.00	22.20	14
INDONESIA	4136	38.20	451	25	3.9	28.20	9.80	0.40	16.90	9
IRELAND	78799	31.40	13	13	2.8	60.60	13.80	15.30	83.90	3
ISRAEL	43589	39.00	182	23	3.1	64.30	28.60	1.80	31.00	3
ITALY	33226	35.90	173	28	2.4	58.50	26.80	5.50	32.50	3
JAPAN	40247	32.90	null	30	2.3	27.20	15.50	1.00	26.90	3
KENYA	1817	40.80	210	30	3.6	25.50	5.40	1.10	17.50	23
LATVIA	17819	35.10	2	20	2.4	57.80	12.90	22.20	51.80	3
LITHUANIA	19551	35.70	null	15	2.4	59.60	9.60	10.70	92.00	3
LUXEMBOURG	114685	35.40	null	25	2.4	58.70	11.30	9.70	162.20	3
MALAWI	412	44.70	139	30	4.5	23.40	2.90	1.00	10.80	19
MEXICO	9946	45.40	7650	30	3.7	64.90	10.30	2.50	49.60	7
NETHERLANDS	52295	28.10	101	25	2.2	57.80	15.80	7.40	44.90	3
NEW ZEALAND	41558	NULL	null	28	2.7	65.60	7.80	11.10	56.20	3
NIGERIA	2230	35.10	2714	30	4.9	28.90	10.80	0.30	11.10	13
NORTH MACEDONIA	6022	33.00	4	10	null	58.10	17.00	6.60	49.70	3
NORWAY	75420	27.60	13	22	2.2	58.30	15.80	11.10	43.90	3
POLAND	15695	30.20	22	19	2.8	58.30	7.00	17.20	44.30	3
PORTUGUAL	23214	33.50	10	32	2.6	57.50	15.20	8.30	37.80	3
QATAR	62088	NULL	null	10	null	71.70	null	null	null	null
ROMANIA	12913	35.80	null	16	2.7	57.70	14.70	7.60	28.40	3
RUSSIA	11585	37.50	55	20	2.6	57.10	14.90	4.30	76.70	3
SINGAPORE	65233	NULL	null	17	3.3	31.80	null	null	null	null
SLOVAKIA	19266	25.00	4	21	2.9	56.20	16.10	17.10	70.20	6
SLOVENIA	25941	24.60	2	19	2.5	56.10	8.10	19.10	48.10	3
SOUTH AFRICA	6001	63.00	751	28	3.4	53.80	13.00	0.60	43.80	6
SOUTH KOREA	31846	31.40	42	28	2.5	30.30	18.60	3.90	38.60	3
SPAIN	29565	34.70	232	25	2.6	61.60	28.20	4.60	33.20	3
SWEDEN	51648	30.00	26	21	null	56.40	7.50	20.50	48.70	3
SWITZERLAND	81989	33.10	6	21	2.2	54.30	20.30	9.20	49.20	3
TURKEY	9127	41.90	670	22	4.1	66.80	18.00	2.70	31.80	3
UGANDA	794	42.80	547	30	4.5	22.40	8.20	0.40	12.10	null
UK	42329	35.10	134	19	2.3	63.70	13.10	5.60	37.80	3
USA	65298	41.40	1015	26	2.5	67.90	19.50	3.30	66.20	3
ZIMBABWE	1464	50.30	69	25	4.1	38.20	11.70	0.50	33.30	null

ANTIBIOTIC RESISTANCE	CANCER	LUNG_CANCER	ASTHMA	COPD	PHEUMONIA	NDC	DIABET	DIERRAL DISEASES	COLREACTAL CANCER	DYSPENSIA	CONSTIPATION	IBS	ANEMIA
null	null	null	null	34.5	null	null	11.62	null	null	null	null	null	null
35	452.4	34.00	10.71	143.2	9.07	15281	5.07	0.43	null	null	30.7	8.9	20.10
32	255.7	11.77	5.33	141.9	4.59	15492	6.35	0.49	128.6	null	null	null	24.50
null	106.2	3.30	2.75	null	32.50	21148	8.38	30.02	null	null	null	8.5	45.70
43	349.2	10.10	4.66	190.6	19.69	15967	4.29	2.38	125.7	null	null	6.7	23.50
null	215.4	6.02	4.88	273.8	40.05	19291	8.11	3.46	23.6	null	null	null	37.30
48	247.1	6.28	3.43	100.8	12.95	21449	5.81	0.42	77.1	null	20.6	null	28.00
10	248.0	25.05	5.77	164.3	11.58	15798	7.37	1.95	86.6	null	16.7	17.5	17.40
null	180.9	7.32	5.03	153.3	19.76	17125	8.46	1.75	31.1	null	null	null	27.00
41	290.8	10.21	3.91	131.4	6.43	17557	5.59	0.42	117.6	null	null	null	28.70
null	256.7	5.30	5.55	55.5	9.35	15129	9.24	1.25	null	null	null	null	29.00
38	292.6	12.32	3.26	106.7	15.66	16824	6.82	1.25	149.8	19.0	13.0	20.0	26.60
15	351.1	27.80	5.70	289.9	18.23	16824	6.41	2.61	136.2	3.4	null	null	22.80
34	278.5	7.65	2.65	71.2	9.19	18362	4.02	0.12	null	null	null	null	25.50
18	271.2	8.07	6.19	96.2	5.06	16031	5.76	0.34	84.2	null	null	5.1	21.90
42	341.9	8.22	6.55	52.0	10.98	14645	4.77	0.69	117.8	null	22.4	4.7	25.10
27	313.2	11.13	6.55	125.5	11.42	16561	8.31	1.66	156.2	20.4	null	22.0	23.30
null	115.9	0.60	2.71	null	114.91	22158	4.97	34.78	null	null	null	null	54.30
47	264.7	7.27	5.28	null	14.41	16276	4.55	0.14	71.1	null	null	null	23.70
39	338.2	22.69	3.28	220.4	5.65	19716	7.55	1.28	152.8	null	null	null	27.30
null	265.1	22.95	8.83	159.7	14.80	14841	5.31	0.94	null	null	null	null	22.10
71	97.1	1.70	2.48	null	52.50	22628	10.39	85.52	6.3	null	null	null	50.10
null	141.1	6.30	5.26	null	26.57	22613	6.32	46.02	17.8	null	null	null	42.00
49	372.8	16.81	7.92	191.7	17.59	15873	3.28	0.53	97.6	null	null	15.0	21.50
null	240.7	9.11	5.54	107.1	17.85	14465	6.74	2.15	93.3	null	null	null	23.80
47	292.6	8.75	3.67	115.9	7.11	14507	4.78	0.45	132.1	13.4	34.1	15.0	24.60
null	285.1	8.04	4.60	40.9	22.16	12718	5.72	0.71	151.2	null	null	20.0	34.10
null	149.2	2.10	3.37	null	93.62	20462	2.92	76.94	null	null	null	null	38.20
51	301.5	6.04	3.93	59.5	11.20	21024	4.91	0.18	null	null	null	null	25.40
46	293.4	4.71	3.35	150.3	13.88	20867	3.67	0.27	null	null	null	null	25.60
44	291.9	7.97	7.24	138.1	13.36	16122	4.42	1.49	null	null	null	null	23.30
null	154.2	0.40	4.16	null	91.62	21872	3.94	70.11	null	null	null	null	41.80
49	140.4	4.96	3.58	242.8	19.82	18853	13.06	4.15	11.1	8.5	19.0	35.0	19.60
34	349.6	18.10	7.24	194.6	16.40	15985	5.29	0.96	119.5	13.8	22.0	6.2	23.20
null	422.9	18.06	8.03	238.7	8.14	16037	8.08	0.86	null	null	19.9	null	20.20
null	110.4	0.40	2.99	null	92.93	19727	2.42	73.67	null	null	null	null	57.80
null	277.0	5.48	4.29	112.0	5.63	19003	10.08	0.41	null	null	null	null	28.10
9	327.5	15.80	8.17	178.5	15.28	15495	5.31	2.67	144.1	null	null	null	21.50
40	267.3	13.59	4.78	114.3	15.88	17977	5.91	0.40	81.1	null	null	null	26.20
44	261.8	5.35	8.71	104.4	24.59	15662	9.85	0.63	100.1	null	null	null	25.20
null	107.2	4.20	4.59	56.6	18.80	17278	16.52	0.35	null	null	null	null	33.40
57	263.1	7.81	4.45	164.9	21.69	20520	9.74	0.56	56.3	null	null	10.0	27.40
null	234.3	5.77	2.98	null	17.73	23454	6.18	0.28	66.8	null	null	19.0	24.00
null	233.0	11.50	4.56	null	41.85	11776	10.99	0.23	null	7.9	3.9	8.6	31.80
48	296.8	7.34	3.02	93.3	20.23	18561	7.29	0.36	110.8	null	null	null	27.30
39	309.0	11.83	4.52	122.0	12.24	15229	7.25	0.15	114.7	null	null	null	25.60
54	209.5	7.09	2.24	331.0	62.16	19374	5.52	32.75	13.Eyl	null	null	null	28.10
null	242.7	10.37	4.59	104.3	17.68	13334	6.80	1.23	null	null	null	5.7	25.80
48	277.2	5.08	5.24	139.9	10.17	14312	7.17	0.75	113.4	24.0	29.0	7.3	24.50
6	288.6	14.04	8.57	113.6	11.21	15069	4.79	2.30	122.7	25.0	6.5	null	21.80
33	317.6	10.71	6.18	null	7.89	14186	5.59	1.03	129.4	null	null	8.4	25.10
55	231.5	4.90	5.19	null	13.36	18402	12.13	0.70	15.0	28.4	24.5	8.6	34.40
null	153.8	2.10	4.66	null	78.81	20819	2.50	57.79	null	null	null	null	34.30
45	319.9	19.67	9.11	210.7	24.11	16908	4.28	0.83	121.0	41.0	null	12.0	21.90
13	362.2	26.15	5.45	248.2	15.88	19743	10.79	1.72	61.1	12.0	18.0	10.0	16.20
null	200.4	5.80	2.98	null	137.63	24788	1.82	44.94	null	null	null	null	34.00

VITAMIN A DEFFICIENCY	ZINC DEFFICIENCY	VITAMIN D LEVELS	IODINE UPTAKE	SUNLIGHT EXPOSURE	TEMPERATURE	RAINFALL	AIR TOXICTY LEVELS	GENERAL TOXICTY LEVELS	FOREST AREA
null	null	null	null	null	null	null	null	null	2.3
null	3.80	70.2	104	3206	21.65	534	71	7.60	17.4
null	7.40	9.5	111	1888	6.35	1110	null	10.90	47.2
21.70	29.70	null	126	4029	25.00	2666	154	77.10	14.5
null	6.80	56.4	80	1645	9.55	847	null	8.90	22.8
13.30	7.30	52.4	360	4552	24.95	1761	124	14.20	59.7
18.30	15.30	null	198	2331	10.55	608	11	27.50	35.6
null	8.00	67.7	null	1887	5.35	537	34	7.30	38.7
7.90	5.70	null	984	3982	8.45	1522	63	19.30	24.2
9.20	12.40	null	140	1976	10.9	1113	65	21.20	34.2
null	6.10	null	null	3439	18.45	498	null	15.80	18.7
5.80	11.00	58.2	119	1707	7.55	677	12	12.30	34.6
null	6.20	25.5	61	1691	7.50	703	null	9.40	15.7
8.70	10.50	43.7	65	1781	5.10	626	null	5.90	56.1
null	4.60	42.9	164	1494	1.70	536	26	5.00	73.7
null	3.90	61.0	85	1907	10.70	867	38	11.10	31.2
null	9.00	45.2	148	1812	8.50	700	28	10.10	32.7
75.80	21.60	null	54	5166	27.20	1187	null	26.90	35.0
null	7.20	42.9	null	2753	15.40	652	null	18.40	30.3
7.00	8.40	null	80	1932	9.85	589	33	14.30	22.5
null	3.10	46.1	150	957	1.75	1940	null	7.20	0.5
62.00	31.20	36.4	133	4514	23.65	1083	149	51.90	24.1
19.60	31.20	null	229	5220	25.85	2702	70	40.70	49.7
null	4.00	37.1	82	1509	9.30	1118	null	8.60	11.2
null	5.50	55.1	null	3682	19.20	435	78	16.90	6.5
null	5.80	39.9	94	2444	13.45	832	53	18.50	31.8
null	16.40	59.1	null	2521	11.15	1668	70	9.80	68.4
84.40	25.30	null	118	5803	24.75	630	null	14.20	6.3
13.00	11.60	null	59	1671	5.60	667	null	11.30	54.8
11.10	7.50	null	75	1801	6.20	656	null	11.70	35.1
null	3.60	null	148	1687	8.65	934	null	9.00	36.5
59.20	40.60	null	null	5019	21.90	1181	null	null	24.7
26.80	16.90	null	235	4974	21.00	758	97	18.90	33.9
null	5.00	53.2	154	1662	9.25	778	34	9.70	10.9
null	4.40	39.8	66	2487	10.55	1732	null	7.00	37.4
29.50	20.60	52.7	130	5251	26.80	1150	null	null	24.1
29.70	14.10	null	228	2403	null	619	22	30.60	39.7
null	6.20	67.2	null	1439	1.50	1414	16	5.70	33.3
9.30	10.30	33.5	84	1749	7.85	600	35	16.90	30.9
null	6.30	null	null	2585	15.15	854	null	9.10	36.2
null	33.40	null	203	4905	27.15	74	null	44.30	0.0
16.20	7.50	null	102	2071	8.8	637	null	15.80	30.1
14.10	8.30	29.1	93	1795	5.1	460	75	9.30	49.8
null	15.40	null	null	3979	26.45	2497	60	11.80	22.5
8.30	16.30	null	183	1795	6.80	824	37	15.30	40.1
null	9.80	null	null	2256	8.90	1162	null	null	61.7
16.90	20.00	37.0	177	4111	17.75	495	112	18.00	14.1
null	13.00	46.1	null	2535	11.5	1274	63	19.50	64.7
null	5.40	52.7	109	2705	13.30	636	34	10.40	37.2
null	6.10	95.0	null	1587	2.10	624	null	5.00	68.7
null	4.90	50.0	141	2158	5.50	1537	17	9.00	31.9
12.40	21.70	null	75	2924	11.10	593	null	18.70	28.5
27.90	20.50	null	464	5499	22.80	1180	null	26.10	12.1
null	4.60	56.2	null	1576	8.45	1220	20	8.30	13.1
null	5.00	70.4	249	2736	8.55	715	75	9.60	33.9
35.80	48.40	null	245	4918	21.00	657	null	null	45.3

COVID19 FATALITY RATE	COVID19 REPRODUCTION RATE	POPULATION SIZE	POPULATION GROWTH TYPE	URBANIZATION PERCENT	COVID19 MORTALITY	INDOOR AIR POLLUTION DEATHS	OUTDOOR AIR POLLUTION DEATHS	CO2 EMISSIONS	FOOD INSECURTY
null	null	106310	3	44	100	null	null	8.72	null
3.05	1.03	25203199	2	86	910	0.02	2.86	16.88	null
1.65	0.86	8955108	3	59	10311	0.07	4.31	7.89	null
1.53	0.61	163046173	1	37	11755	7.93	6.32	0.50	null
2.44	0.90	11539326	2	98	24367	0.09	4.65	8.71	null
2.77	0.97	211049518	2	87	411854	0.90	4.07	2.33	null
4.09	0.71	7000116	3	75	16609	1.43	5.78	6.69	null
1.94	0.97	37411038	3	81	24445	0.01	2.84	15.59	null
2.19	0.94	18952035	3	88	26726	0.61	4.49	4.55	null
2.16	0.86	4130299	3	57	7315	0.84	5.30	4.48	null
0.48	0.91	1198573	3	67	326	0.05	6.55	6.37	null
1.80	0.80	10689213	3	74	29479	0.19	5.74	9.93	null
0.98	1.14	5771876	3	88	2492	0.02	4.43	6.06	null
0.95	0.84	1325649	3	69	1183	0.75	2.22	14.13	null
1.05	0.81	5532159	2	85	918	0.03	2.03	8.11	null
1.84	0.80	65129730	2	81	105631	0.03	3.31	5.33	null
2.42	0.99	83517046	3	77	84482	0.06	4.44	9.52	null
0.84	0.91	30417857	1	57	780	4.85	2.66	0.47	0.49
3.05	0.80	10473452	3	79	10764	0.06	5.84	7.08	null
3.57	0.62	9684679	3	72	28173	1.32	5.17	5.11	null
0.45	1.01	339037	2	94	29	0.04	2.67	10.82	null
1.09	1.27	1366417755	2	34	230010	4.86	8.26	1.84	null
2.74	0.99	270625567	2	56	46349	4.15	3.46	2.01	null
1.96	1.06	4882498	3	63	4915	0.06	3.59	8.19	null
0.76	0.60	8519373	1	93	6370	0.04	5.76	7.58	null
3.00	0.86	60550092	3	71	122005	0.06	4.78	5.79	null
1.70	1.17	126860299	3	92	10470	0.01	3.67	9.31	null
1.74	0.80	52573967	1	28	2825	4.85	1.73	0.33	2.99
1.79	1.09	1906740	2	68	2166	0.63	4.58	3.70	null
1.58	1.10	2759631	3	68	3993	0.15	4.90	4.76	null
1.18	0.99	615726	3	91	800	0.03	3.85	15.63	null
3.37	0.81	18628748	1	17	1151	5.90	0.94	0.08	2.99
9.25	0.92	127575529	2	80	217740	1.65	5.25	3.70	null
1.13	0.86	17097123	3	92	17245	0.02	4.49	9.66	null
0.99	1.09	4783061	2	87	26	0.03	2.11	7.69	null
1.25	0.86	200963603	1	51	2063	4.06	3.19	0.68	9.99
3.27	0.57	2083457	3	58	5016	2.15	6.64	3.61	null
0.66	0.84	5378858	3	83	767	0.05	2.78	8.23	null
2.43	0.59	37887771	3	60	68482	0.61	5.71	8.89	null
2.03	0.89	10226178	3	66	16983	0.15	3.72	5.32	null
0.23	0.89	2832071	3	99	489	0.00	9.29	38.74	null
2.69	0.66	19364557	3	54	28616	1.37	4.47	3.97	null
2.29	1.00	145872259	3	75	111895	0.02	5.43	11.31	null
0.05	1.14	5804343	3	100	31	0.01	6.69	6.84	null
3.09	0.81	5457011	3	54	11886	0.27	5.74	6.62	null
1.76	0.97	2078654	3	55	4279	0.37	4.31	6.87	null
3.44	1.10	58558267	2	67	54511	0.92	4.09	8.18	10.00
1.48	0.95	51225320	3	81	1847	0.00	5.90	12.15	null
2.21	1.04	46736782	3	81	78566	0.13	4.11	5.89	null
1.43	1.06	10036391	2	88	14151	0.05	2.44	4.27	null
1.60	1.04	8591361	3	74	10676	0.06	3.60	4.52	null
0.84	0.73	83429607	2	76	41527	0.09	10.12	5.24	null
0.81	1.06	44269587	1	24	343	4.35	1.54	0.13	2.99
2.88	0.91	67530161	3	84	127570	0.06	4.08	5.82	null
1.78	0.89	329064916	2	82	592537	0.02	3.84	16.16	null
4.11	0.83	14645473	1	32	1574	5.93	1.79	0.72	4.99

B. DATA RESOURCES

SUMMARY

These data were gathered from different sources such as literature sources or publicly online databases. This document provides the references that provide the data of the table content.

A wide range of the aspects of the countries (such as sociological, economic, e.g.) that can have any impact on the survival of the specific variants, were included as the attributes of the table.

Countries (city_name)

https://covariants.org/per-country

Number of total countries that are included in this database is 58.

Bonaire and Curacao were excluded in the study due to insufficient information on the selected references.

56 countries with common Sars-CoV-2 variant data were included in the study.

These countries were selected for further analysis.

Countries that were used in this study by continents:

AFRICA: 1. Ghana 2. Kenya 3. Malawi 4. Nigeria 5. South Africa 6. Uganda 7. Zimbabwe ASIA: 1. Cyprus 2. Turkey 3. Russia 4. Bangladesh 5. India 6. Indonesia 7. Israel 8. Japan 9. Qatar 10. Singapore 11. South Korea AUSTRALIA: 1. Australia 2. New Zeland EUROPE: 1. Austria 2. Belgium 3. Bulgaria 4. Croatia 5. Czech Republic 6. Denmark 7. Estonia 8. Finland 9. France 10. Germany 11. Greece 12. Hungary 13. Iceland 14. Ireland 15. Italy 16. Latvia 17. Lıthuania 18. Luxembourg 19. Netherlands 20. North Macedonia 21. Norway 22. Poland 23. Portugual 24. Romania 25. Slovakia 26. Slovenia 27. Spain 28. Sweden 29.

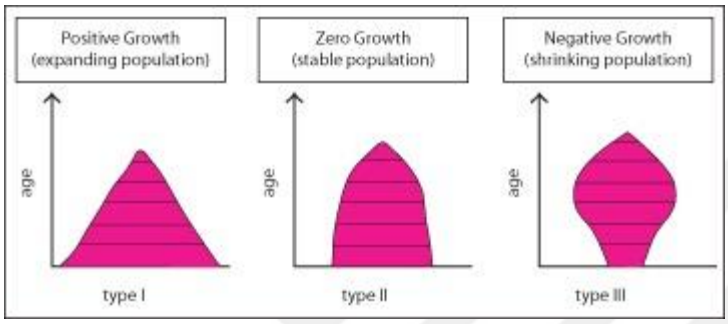
Switzerland 30. United Kingdom NORTH AMERICA: 1. Canada 2. USA 3. Mexico SOUTH AMERICA: 1. Brazil 2. Chile 3. Aruba

Population Size (population_number)

https://www.populationpyramid.net/

Population pyramid (population_growth)

https://www.populationpyramid.net/



(figure source: https://www.ck12.org/biology/population-structure-1501903452.12/lesson/Age-Sex-Structure-of-Populations-Advanced-BIO-ADV/ last access time: 27.04.2021, 17:03)

GDP per capita (GDP)

https://data.worldbank.org/indicator/NY.GDP.PCAP.CD

Last entry (current) data was used.

The fractional numbers rounded to whole numbers.

Exposure to Solar UV Radiation (Sunligh_exposure)

https://apps.who.int/gho/data/view.main.35300

For the countries that have not any information about sunlight exposure in this application, the information of the nearest country was used (for Aruba, Venezuela used.)

Climate (temperature)

https://worldpopulationreview.com/country-rankings/hottest-countries-in-the-world

Average temperature was used for representing climate.

Drug Resistance Index (antibiotic_use_freq)

https://resistancemap.cddep.org/DRI.php

Gini Index (income_inequality)

https://data.worldbank.org/indicator/SI.POV.GINI?name_desc=false&view=map&year=2019

To estimate income inequality.

The fractional numbers rounded to whole numbers.

Corporate Tax Rates (tax_rates)

https://taxfounhttps://data.worldbank.org/indicator/SI.POV.GINI?name_desc=false&view=map&year=2019dation.org/publications/corporate-tax-rates-around-the-world/

The fractional numbers rounded to whole numbers.

Total Deaths due to COVID-19 (covid_19_mortality)

https://www.worldometers.info/coronavirus/?utm_campaign=homeAdvegas1?

COVID-19 recovery cases in number (covid_19_mortality)

https://www.worldometers.info/coronavirus/?utm_campaign=homeAdvegas1?

COVID-19 Case Fatality Rate by % (covid_19_mortality_freq)

https://ourworldindata.org/explorers/coronavirus-data-explorer?tab=table&zoomToSelection=true&time=2020-03-01..latest&pickerSort=asc&pickerMetric=location&Metric=Case+fatality+rate&Interval=Cumulative&Relative+to+Population=true&Align+outbreaks=false

COVID-19 Reproduction Rate (covid_19_reproduction)

https://ourworldindata.org/explorers/coronavirus-data-explorer?tab=table&zoomToSelection=true&time=2020-03-01..latest&pickerSort=asc&pickerMetric=location&Metric=Reproduction+rate&Interval=7-day+rolling+average&Relative+to+Population=true&Align+outbreaks=false&country=USA~GBR~CAN~DEU~ITA~IND

Prevalence of Total Overweight Adults (overweight_adults)

https://apps.who.int/gho/data/view.main.CTRY2430A?lang=en

Last entry (current) data was used (2016).

The fractional numbers rounded to whole numbers.

Consumption of the Vegetable Oil (consumption_veg_oil)

https://data.worldobesity.org/maps-obesity-day/?mapid=62

This database uses the data of the FAO (Food and Agriculture of the United Nations : http://www.fao.org/faostat/en/#data/FBS) and visualize the data.

The fractional numbers rounded to whole numbers.

Consumption of the Animal Fat (consumption_animal_fat)

https://data.worldobesity.org/maps-obesity-day/?mapid=61

This database uses the data of the FAO (Food and Agriculture of the United Nations : http://www.fao.org/faostat/en/#data/FBS) and visualize the data.

The fractional numbers rounded to whole numbers.

Consumption of Sugars (consumption_sugar)

https://data.worldobesity.org/maps-obesity-day/?mapid=67

This database uses the data of the FAO (Food and Agriculture of the United Nations : http://www.fao.org/faostat/en/#data/FBS) and visualize the data.

The fractional numbers rounded to whole numbers.

Prevalence of undernourishment by percentage (under_nourishment)

https://data.worldbank.org/indicator/SN.ITK.DEFC.ZS

Conflict Cases (conflict)

https://acleddata.com/dashboard/#/dashboard

Total events (reported) were used.

Vegetation Index (forest_area)

https://data.worldbank.org/indicator/AG.LND.FRST.ZS

Forest Area is used for representing vegetation.

Average Precipitation (rainfall)

https://data.worldbank.org/indicator/AG.LND.PRCP.MM

IPC/CH (IPC/CH)

https://hungermap.wfp.org/

To class https://hungermap.wfp.org/ify food insecurity.

The upper bound of the color scheme is used as an integer in the DB.

The index in this database was used as a percentage (e.g. index 1 in this database was used as 100 in the study).

The resulting fractional numbers rounded to whole numbers.

Anemia in pregnant women (anemia)

https://ourworldindata.org/grapher/anemia-pregnant-women-vs-children?tab=table

The fractional numbers rounded to whole numbers.

Global prevalence of Zinc Deficiency (zinc_defficiency)

https://ourworldindata.org/grapher/global-prevalence-of-zinc-deficiency

Most recent data (2005) was used.

The fractional numbers rounded to whole numbers.

Prevalence of Vitamin A deficiency (vit_A_defficiency)

https://ourworldindata.org/grapher/prevalence-of-vitamin-a-deficiency-in-children?tab=table

The fractional numbers rounded to whole numbers.

Vitamin D status Around the World (vit_D_defficiency)

https://www.osteoporosis.foundation/educational-hub/topic/vitamin-d

The fractional numbers rounded to whole numbers.

Average Household Size: Number of members (household_type)

https://population.un.org/Household/index.html#/countries/533

Urban Population as % of Total Population (open_closed_index)

https://data.worldbank.org/indicator/SP.URB.TOTL.IN.ZS

Indoor Air Pollution Deaths (indoor_deaths)

https://ourworldindata.org/indoor-air-pollution?country= (From table "Share of Deaths From Indoor Air Pollution Percent")

Outdoor Air Pollution Deaths (outdoor_deaths)

https://ourworldindata.org/outdoor-air-pollution (From table "Share of Deaths From Outdoor Air Pollution Percent")

CO2 Emissions (co2)

https://ourworldindata.org/co2-emissions (From "CO2 emissions per capita")

Variant Name (variant_name)

https://covariants.org/per-country

In the DB of https://covariants.org/per-country , there are 27 variants of the virus that are related with the countries.

In our tables, the variants that are related with the selected countries were used.

Frequency (frequency)

https://covariants.org/per-country

The variant frequency of the countries.

The upper bound of the frequency data is used for each time interval.

Interpolated data that has no frequency was ignored.

Time Interval (time_interval)

https://covariants.org/per-country

There are three time intervals for dividing the viral circulation.

Categorization of the time intervals: interval_1: 0 - 2020/07/28, interval_2: 2020/07/28 - 2020/10/29, interval_3: 2020/10/29 - 2021/04/19.

Related Spike Mutations of the Variants (variant_has_mutations)

https://covariants.org/variants

Disease Statistics (disease_frequency)

CANCER (For All Types of Cancer): https://gco.iarc.fr/today/online-analysis-map?v=2020&mode=population&mode_population=continents&population=900&populations=900&key=asr&sex=0&cancer=39&type=0&statistic=5&prevalence=0&population_group=0&ages_group%5B%5D=0&ages_group%5B%5D=17&nb_items=10&group_cancer=1&include_nmsc=1&include_nmsc_other=1&projection=natural-earth&color_palette=default&map_scale=quantile&map_nb_colors=5&continent=0&show_ranking=0&rotate=%255B10%252C0%255D

Lung Cancer: https://ourworldindata.org/grapher/lung-cancer-deaths-per-100000-by-sex-1950-2002?tab=table

Asthma: https://ourworldindata.org/grapher/asthma-prevalence

COPD: https://statistics.blf.org.uk/copd (Number of Deaths by COPD per million section).

Pneumonia: https://ourworldindata.org/grapher/pneumonia-death-rates-age-standardized

NDCs (Non-communicable Diseases): https://ourworldindata.org/grapher/burden-of-disease-rates-from-ncds?tab=table (to get more information about NDCs: https://ourworldindata.org/burden-of-disease)

Diabetes: https://ourworldindata.org/grapher/diabetes-prevalence

Thyroid Diseases (Thyroid Diseases Relatedness via Iodine Levels): https://www.who.int/vmnis/iodine/status/summary/IDD_estimates_table_2007.pdf?ua=1 (in this section, the iodine levels were used to indicate thyroid diseases. "Median urinary iodine concentration" data of the reference table was used).

Diarrheal Diseases: https://ourworldindata.org/grapher/diarrheal-disease-death-rates

Colorectal Cancer: https://www.worldgastroenterology.org/UserFiles/file/wdhd-2008-map-of-digestive-disorders.pdf (The data of "Global Colorectal Cancer Incidence" section was used.) (The sum of female and male incidence rates was used.)

Dyspepsia: https://www.worldgastroenterology.org/UserFiles/file/wdhd-2008-map-of-digestive-disorders.pdf (The data of "Global Functional Dyspepsia Prevalence" section was used.)

Constipation: https://www.worldgastroenterology.org/UserFiles/file/wdhd-2008-map-of-digestive-disorders.pdf (The data of “Global Functional Constipation Prevalence” section was used.) (Upper bound of the prevalence statistics was used.)

Irritable Bowel Syndrome: https://www.worldgastroenterology.org/UserFiles/file/wdhd-2008-map-of-digestive-disorders.pdf (The data of “Global Irritable Bowel Syndrome Prevalence” section was used.)(The data that has the latest survey date was used.)(If there was no other data, the data that is related with children was used .)

Air Toxicity Levels (air_toxicity_levels)

https://www.iqair.com/world-air-quality-ranking

For countries that have more than one entry, the most toxic city data was used.

General Toxicity Levels (general_toxicity_levels)

https://www.iqair.com/world-most-polluted-countries

C. SET 2 DATA TABLE: MUTANT AND VARIANT TYPES

COUNTRIES	GENERAL DOMINANT VARIANT // MUTATION	GENERAL DOMINANT VARIANT // MUTATION Freq	GENERAL 20A.EU2 Freq	GENERAL 20A/S:154K Freq	GENERAL 20A/S:439K Freq	GENERAL 20A/S:478K Freq	GENERAL 20A/S:484K Freq
ARUBA	ORF1a:S3675	0.89	0	0	0	0	0
AUSTRALIA	S:S477	0.91	0	0	0	0	0
AUSTRIA	ORF1a:S3675	0.91	0.36	0	0.24	0	0
BANGLADESH	ORF1a:S3675	0.98	0	0	0	0	0
BELGIUM	ORF1a:S3675	0.97	0.40	0	0.18	0	0
BRAZIL	S:E484	0.95	0	0	0	0	0
BULGARIA	more than one	0.98	0	0	0	0	0
CANADA	more than one	0.78	0.16	0	0	0	0
CHILE	ORF1a:S3675	0.80	0	0	0	0	0
CROATIA	S:H69-	0.92	0	0	0.37	0	0
CYPRUS	20A/S:439K	0.91	0	0	0.91	0	0
CZECHIA	ORF1a:S3675	0.97	0	0	0.74	0	0
DENMARK	ORF1a:S3675	0.97	0.18	0	0.15	0	0.03
ESTONIA	S:H69-	0.91	0	0	0	0	0
FINLAND	ORF1a:S3675	0.92	0.24	0	0	0	0
FRANCE	ORF1a:S3675	0.91	0.68	0	0.08	0	0.03
GERMANY	more than one	0.96	0.11	0	0.11	0	0
GHANA	S:P681	0.80	0	0	0	0	0
GREECE	ORF1a:S3675	0.97	0	0	0	0	0
HUNGARY	S:S477	0.93	0.93	0	0	0	0
ICELAND	20E (EU1)	0.90	0	0	0.08	0	0
INDIA	S:P681	0.74	0	0.37	0	0.29	0
INDONESIA	S:P681	0.75	0	0	0	0	0
IRELAND	S:P681	0.98	0	0	0.74	0	0
ISRAEL	S:P681	0.86	0	0	0	0	0
ITALY	20I/501Y.V1	0.89	0.11	0	0.18	0	0.01
JAPAN	S:N501	0.94	0	0	0	0	0
KENYA	ORF1a:S3675	0.79	0	0	0	0	0
LATVIA	S:Y144-	0.64	0	0	0	0	0
LITHUANIA	S:H69-	0.89	0	0	0	0	0
LUXEMBOURG	ORF1a:S3675	0.95	0.40	0	0	0	0
MALAWI	more than one	1.00	0	0	0	0	0
MEXICO	S:P681	0.90	0	0	0	0	0
NETHERLANDS	20I/501Y.V1	0.94	0.14	0	0.11	0	0
NEW ZEALAND	S:P681	1.00	0	0	0	0	0
NIGERIA	more than one	1.00	0	0	0	0	0.43
NORTH MACEDONIA	S:H69-	1.00	0	0	0	0	0
NORWAY	ORF1a:S3675	0 0.99	0.34	0	0.15	0	0
POLAND	more than one	0.99	0	0	0.24	0	0
PORTUGUAL	ORF1a:S3675	0.97	0.15	0	0	0	0
QATAR	ORF1a:S3675	1.00	0	0	0	0	0
ROMANIA	more than one	0.87	0	0	0	0	0
RUSSIA	S:S477	0.29	0	0	0	0	0
SINGAPORE	ORF1a:S3675	0.73	0	0	0	0	0
SLOVAKIA	more than one	0.98	0	0	0	0	0
SLOVENIA	S:H69-	0.92	0.41	0	0.75	0	0
SOUTH AFRICA	S:N501	1.00	0	0	0	0	0
SOUTH KOREA	more than one	0.20	0	0	0	0	0
SPAIN	ORF1a:S3675	0.93	0.03	0	0.06	0	0
SWEDEN	more than one	0.96	0.27	0	0.31	0	0
SWITZERLAND	ORF1a:S3675	0.96	0.33	0	0.07	0	0
TURKEY	S:N501	0.92	0	0	0	0	0
UGANDA	S:P681	1.00	0	0	0	0	0
UK	more than one	0.99	0.04	0	0.06	0.02	0
USA	ORF1a:S3675	0.83	0	0	0	0	0
ZIMBABWE	more than one	0.94	0	0	0	0	0

GENERAL 20A/S:98F Freq	GENERAL 20B/S:1122L Freq	GENERAL 20B/S:626S Freq	GENERAL 20C/S:452R Freq	GENERAL 20C/S:484K Freq	GENERAL 20C/S:80Y Freq	GENERAL 20E (EU1) Freq	GENERAL 20H/501Y.V2 Freq	GENERAL 20I/501Y.V1 Freq	GENERAL 20J/501Y.V3 Freq
0	0	0	0	0	0	0	0	0.79	0
0	0	0	0	0	0	0	0	0.51	0
0	0	0	0	0	0	0.16	0.27	0.83	0
0	0	0	0	0	0	0	0.68	0	0
0.48	0	0	0	0	0	0.23	0.10	0.83	0.09
0	0	0	0	0	0	0	0	0	0.93
0	0	0	0	0	0	0	0	0.98	0
0	0	0	0	0	0	0	0	0.73	0
0	0	0	0	0	0	0	0	0	0.33
0	0	0	0	0	0	0	0	0.73	0
0	0	0	0	0	0	0.17	0	0.42	0
0	0	0	0	0	0	0	0	0.90	0
0.06	0	0.08	0	0	0	0.57	0	0.94	0
0	0	0	0	0	0	0.50	0	0.90	0
0	0	0	0	0	0	0.08	0	0.71	0
0.03	0	0	0	0	0.13	0.16	0.11	0.77	0
0.12	0	0	0	0	0	0.35	0.02	0.93	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0.89	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0.90	0	0	0
0	0	0	0	0	0	0	0	0.13	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0.84	0	0.91	0
0	0	0	0	0	0	0.04	0.03	0.78	0
0	0	0	0	0	0	0.67	0	0.89	0.04
0	0	0	0	0	0	0	0	0.50	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	1.00	0	0.62	0
0	0	0	0	0	0	0.67	0	0.85	0
0.27	0	0	0	0	0	0.27	0.23	0.69	0
0	0	0	0	0	0	0	1.00	0	0
0	0	0	0.07	0	0	0	0	0	0
0.35	0	0	0	0	0	0.46	0.03	0.94	0.02
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0.30	0	0	0	0.43	0.10	0.96	0
0.22	0	0	0	0	0	0	0	0.98	0
0	0	0	0	0	0	0.71	0	0.87	0
0	0	0	0	0	0	0	0	0.46	0
0	0	0	0	0	0	0	0	0.83	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0.43	0.28	0
0	0	0	0	0	0	0	0	0.97	0
0	0	0	0	0	0	0	0	0.76	0
0	0	0	0	0	0	0	0.88	0	0
0	0	0	0	0	0	0	0	0.19	0
0.03	0	0	0	0	0	0.84	0	0.83	0.05
0.11	0.48	0	0	0	0	0.40	0.04	0.92	0
0.06	0	0	0	0	0	0.35	0.02	0.90	0
0	0	0	0	0	0	0	0.20	0.54	0
0	0	0	0	0	0	0	0	0	0
0.02	0	0.01	0	0	0.02	0	0	0.98	0
0	0	0	0.19	0.14	0	0	0	0.56	0.06
0	0	0	0	0	0	0	0	0	0

GENERAL ORF1a:S3675 Freq	GENERAL S:677H.Robin1 Freq	GENERAL S:677P.Pelican Freq	GENERAL S:E484 Freq	GENERAL S:H655 Freq	GENERAL S:H69- Freq	GENERAL S:K417 Freq
0.89	0	0	0	0	0.83	0
0.71	0	0	0	0	0.60	0
0.91	0	0	0.29	0	0.85	0.29
0.98	0	0	0.73	0	0	0.83
0.97	0	0	0.15	0.09	0.83	0.10
0.95	0	0	0.95	0.93	0	0
0.98	0	0	0	0	0.98	0
0.78	0	0	0	0	0.78	0
0.80	0	0	0.37	0.34	0	0
0.74	0	0	0	0	0.92	0
null	0	0	null	null	null	null
0.97	0	0	0	0	0.95	0
0.97	0	0	0.03	0	0.95	0
0.92	0	0	0	0	0.91	0
0.92	0	0	0	0	0.76	0
0.91	0	0	0.14	0.05	0.79	0.12
0.96	0	0	0.04	0	0.93	0.02
0	0	0	0	0	0	0
0.97	0	0	0	0	0.90	0
0	0	0	0	0	0	0
0	0	0	0	0	0	0
0.11	0	0	0.50	0	0.10	0
0	0	0	0	0	0	0
0.97	0	0	0.06	0	0.91	0
0.81	0	0	0.04	0	0.83	0.03
0.60	0	0	0.06	0.05	0.90	0
0.50	0	0	0.56	0	0.60	0
0.79	0	0	0	0	0	0
0.26	0	0	0	0	0.62	0
0.88	0	0	0	0	0.89	0
0.95	0	0	0.26	0	0.69	0.24
0.93	0	0	1.00	0	0	1.00
0.09	0	0	0	0	0	0
0.99	0	0	0.06	0.02	0.95	0.03
0.83	0	0	0	0	0	0
1.00	0	0	0.50	0	1.00	0
0	0	0	0	0	1.00	0
0.99	0	0	0.11	0	0.98	0.10
0.96	0	0	0	0	0.89	0
0.97	0	0	0.09	0	0.86	0
1.00	0	0	0	0	0.46	0.62
0.87	0	0	0	0	0.83	0
0	0	0	0	0	0	0
0.73	0	0	0.58	0	0.28	0.44
0.81	0	0	0	0	0.87	0
0.81	0	0	0	0	0.92	0
0.94	0	0	0.91	0	0	0.99
0.20	0	0	0	0	0.20	0
0.93	0	0	0.09	0.05	0.85	0
0.96	0	0	0.05	0	0.92	0.04
0.96	0	0	0.03	0	0.91	0.01
0.73	0	0	0.38	0	0.53	0.28
0	0	0	0	0	0	0
0.99	0	0	0.02	0.01	0.98	0
0.83	0.03	0.02	0.16	0.06	0.59	0
0.94	0	0	0	0	0	0.94

GENERAL S:L18 Freq	GENERAL S:N501 Freq	GENERAL S:P681 Freq	GENERAL S:Q677 Freq	GENERAL S:S477 Freq	GENERAL S:Y144- Freq	GENERAL S:Y453F Freq
0	0.81	0.81	0	0	0.79	0
0	0.67	0.81	0	0.91	0.60	0
0.13	0.89	0.89	0	0.36	0.82	0
0.76	0.80	0.68	0	0	0	0
0.12	0.94	0.84	0	0.40	0.83	0
0.93	0.96	0	0	0	0.11	0
0	0.98	0.98	0	0	0.98	0
0.19	0.73	0.78	0.10	0.16	0.76	0
0.34	0.44	0	0	0	0	0
0	0.73	0.73	0	0	0.73	0
null	null	null	null	null	null	null
0	0.96	0.93	0	0	0.92	0
0.08	0.95	0.96	0.03	0.18	0.90	0.52
0.50	0.90	0.90	0	0	0.90	0
0.12	0.88	0.78	0	0.24	0.77	0
0.11	0.88	0.81	0.03	0.70	0.79	0
0.16	0.96	0.94	0.09	0.13	0.91	0
0	0	0.80	0	0	0	0
0	0.91	0.95	0	0	0.95	0
0	0	0	0	0.93	0	0
0	0	0	0	0	0	0
0.06	0.20	0.74	0.11	0	0.17	0
0	0	0.75	0	0	0	0
0.38	0.93	0.98	0	0	0.94	0
0.03	0.81	0.86	0.03	0	0.80	0
0.06	0.94	0.91	0.03	0.11	0.89	0
0.01	0.50	0.50	0.06	0	0.50	0
0	0.78	0.57	0	0	0	0
0.48	0.63	0.62	0	0	0.64	0
0.75	0.86	0.88	0	0	0.87	0
0.30	0.93	0.70	0	0.45	0.70	0
0	1.00	0	0	0	0	0
0	0.07	0.90	0	0	0.07	0
0.11	0.98	0.95	0.01	0.14	0.94	0
0	0.75	1.00	0	0	0	0
0	0.44	0.47	0.81	0	0.88	0
0	0	0	0	0	0	0
0.15	0.98	0.97	0	0.34	0.97	0
0	0.99	0.99	0	0	0.93	0
0.05	0.93	0.92	0	0.15	0.83	0
0	0.79	0.46	0	0	0.46	0
0	0.87	0.87	0	0	0.87	0
0	0	0.21	0	0.29	0	0
0	0.69	0.68	0	0	0.28	0
0	0.97	0.98	0	0	0.98	0
0	0.77	0.80	0	0.41	0.76	0
0.50	1.00	0	0	0	0	0
0	0.19	0.20	0	0	0.20	0
0.08	0.89	0.87	0.03	0.03	0.84	0
0.10	0.96	0.94	0	0.27	0.93	0
0.02	0.92	0.95	0.03	0.34	0.94	0
0	0.92	0.64	0	0	0.67	0
0	0	1.00	0	0	0	0
0.40	0.99	0.99	0.02	0.04	0.98	0
0.06	0.64	0.64	0.08	0.05	0.62	0
0	0	0	0	0	0	0

D. TIME INTERVAL FREQUENCIES

COUNTRIES	20E (EU1)_time_1	20E (EU1)_time_2	20E (EU1)_time_3	20I/501Y.V1_time_1	20I/501Y.V1_time_2	20I/501Y.V1_time_3
ARUBA	0	0	0	0	0	0.79
AUSTRALIA	0	0	0	0	0	0.52
AUSTRIA	0	0.09	0.16	0	0	0.83
BANGLADESH	0	0	0	0	0	0
BELGIUM	0.04	0.23	0.22	0	0	0.83
BRAZIL	0	0	0	0	0	0
BULGARIA	0	0	0	0	0	0.98
CANADA	0	0	0	0	0	0.73
CHILE	0	0	0	0	0	0
CROATIA	0	0	0	0	0	0.73
CYPRUS	0	0	0	0	0	0.90
CZECHIA	0	0	0	0	0	0.90
DENMARK	0	0.29	0.57	0	0	0.94
ESTONIA	0	0	0.50	0	0	0.90
FINLAND	0	0	0.08	0	0	0.71
FRANCE	0	0.14	0.16	0	0	0.77
GERMANY	0	0.18	0.35	0	0	0.93
GHANA	0	0	0	0	0	0
GREECE	0	0	0	0	0	0.89
HUNGARY	0	0	0	0	0	0
ICELAND	0	0.90	0.79	0	0	0
INDIA	0	0	0	0	0	0.13
INDONESIA	0	0	0	0	0	0
IRELAND	0.05	0.74	0.84	0	0	0.91
ISRAEL	0	0	0.04	0	0	0.78
ITALY	0	0.64	0.67	0	0	0.89
JAPAN	0	0	0	0	0	0.50
KENYA	0	0	0	0	0	0
LATVIA	0	0.40	1.00	0	0	0.62
LITHUANIA	0	0.33	0.67	0	0	0.85
LUXEMBOURG	0	0.25	0.27	0	0	0.69
MALAWI	0	0	0	0	0	0
MEXICO	0	0	0	0	0	0
NETHERLANDS	0.02	0.35	0.46	0	0	0.94
NEW ZEALAND	0	0	0	0	0	0
NIGERIA	0	0	0	0	0	0
NORTH MACEDONIA	0	0	0	0	0	0
NORWAY	0.15	0.43	0.26	0	0	0.96
POLAND	0	0	0	0	0	0.98
PORTUGUAL	0	0.64	0.71	0	0	0.87
QATAR	0	0	0	0	0	0.46
ROMANIA	0	0	0	0	0	0.87
RUSSIA	0	0	0	0	0	0
SINGAPORE	0	0	0	0	0	0.28
SLOVAKIA	0	0	0	0	0	0.97
SLOVENIA	0	0	0	0	0	0.76
SOUTH AFRICA	0	0	0	0	0	0
SOUTH KOREA	0	0	0	0	0	0.19
SPAIN	0.58	0.81	0.84	0	0	0.83
SWEDEN	0	0.40	0.39	0	0	0.92
SWITZERLAND	0.03	0.35	0.31	0	0	0.90
TURKEY	0	0	0	0	0	0.54
UGANDA	0	0	0	0	0	0.56
UK	0.01	0.65	0.67	0	0	0.98
USA	0	0	0	0	0	0.56
ZIMBABWE	0	0	0	0	0	0

ORF1a:S3675_time_1	ORF1a:S3675_time_2	ORF1a:S3675_time_3	S:E484_time_1	S:E484_time_2	S:E484_time_3
0	0	0.89	0	0	0
0	0	0.71	0	0	0
0	0	0.91	0	0	0.29
0	0	0.98	0	0	0.73
0	0	0.97	0	0	0.15
0	0	0.95	0	0.27	0.95
0	0	0.98	0	0	0
0	0.10	0.78	0	0	0
0	0	0.80	0	0	0.37
0	0	0.74	0	0	0
0	0	0.97	0	0	0
0	0	0.97	0	0	0
0	0	0.97	0	0	0.03
0	0	0.92	0	0	0
0	0	0.92	0	0	0
0	0	0.91	0	0	0.14
0.05	0	0.96	0	0	0.04
0	0	0	0	0	0
0	0	0.97	0	0	0
0	0	0	0	0	0
0	0	0	0	0	0
0	0	0.11	0	0.03	0.50
0	0	0	0	0	0
0	0	0.97	0	0	0.06
0	0	0.81	0	0	0.04
0	0	0.60	0	0	0.07
0	0	0.50	0	0	0.56
0	0	0.79	0	0	0
0	0	0.26	0	0	0
0	0	0.88	0	0	0
0	0	0.95	0	0	0.26
0	0	0.93	0	0	1.00
0	0	0.09	0	0	0
0	0	0.99	0	0	0.06
0	0	0.83	0	0	0
0	0	1.00	0	0	0.50
0	0	0	0	0	0
0	0	0.99	0	0	0.11
0	0	0.96	0	0	0
0	0	0.97	0	0	0.09
0	0	1.00	0	0	0
0	0	0.87	0	0	0
0	0	0	0	0	0
0	0	0.71	0	0	0.58
0	0	0.81	0	0	0
0	0	0.81	0	0	0
0	0.07	0.94	0	0.18	0.91
0	0	0.20	0	0	0
0	0	0.93	0.01	0	0.09
0	0	0.96	0	0	0.05
0	0	0.96	0	0	0.03
0	0	0.73	0	0	0.38
0	0	0.83	0	0	0.16
0	0	0.99	0	0	0.02
0	0	0.83	0	0	0.16
0	0	0.91	0	0	0

S:H69-_time_1	S:H69-_time_2	S:H69-_time_3	S:L18_time_1	S:L18_time_2	S:L18_time_3
0	0	0.83	0	0	0
0	0	0.60	0	0	0
0	0.07	0.85	0	0	0.13
0	0	0	0	0	0.76
0	0.02	0.83	0	0	0.12
0	0	0	0	0	0.93
0	0	0.98	0	0	0
0	0	0.78	0	0.09	0.19
0	0	0	0	0	0.34
0	0.21	0.92	0	0	0
0	0.74	0.94	0	0	0
0	0.74	0.95	0	0	0
0	0.27	0.95	0	0.02	0.08
0	0	0.91	0	0	0.50
0	0	0.79	0	0	0.12
0	0.04	0.79	0	0	0.11
0	0.07	0.93	0	0.05	0.16
0	0	0	0	0	0
0	0	0.90	0	0	0
0	0	0	0	0	0
0	0	0	0	0	0
0	0	0.10	0	0.02	0.06
0	0	0	0	0	0
0	0	0.91	0	0.24	0.38
0	0	0.83	0	0	0.03
0	0	0.90	0	0	0.07
0	0	0.60	0	0	0.01
0	0	0	0	0	0
0	0	0.62	0	0.11	0.47
0	0	0.89	0	0.33	0.75
0	0.01	0.69	0	0	0.30
0	0	0	0	0	0
0	0	0	0	0	0
0	0.03	0.95	0	0.03	0.11
0	0	0	0	0	0
0	0	1.00	0	0	0
0	0	1.00	0	0	0
0	0.04	0.98	0	0.12	0.18
0	0	0.89	0	0	0
0	0	0.86	0	0	0.08
0	0	0.46	0	0	0
0	0	0.83	0	0	0
0	0	0	0	0	0
0	0	0.28	0	0	0
0	0	0.89	0	0	0
0	0.32	0.92	0	0	0
0	0	0	0	0.04	0.50
0	0	0.20	0	0	0
0	0.06	0.85	0	0.01	0.08
0	0	0.92	0	0.05	0.10
0.01	0.04	0.91	0	0	0.02
0	0	0.53	0	0	0
0	0	0.59	0	0	0.05
0	0.04	0.98	0.02	0.34	0.40
0	0	0.59	0	0	0.06
0	0	0	0	0	0

S:P681_time_1	S:P681_time_2	S:P681_time_3	S:Y144-_time_1	S:Y144-_time_2	S:Y144-_time_3
0	0	0.81	0	0	0.79
0	0	0.81	0	0	0.60
0	0	0.89	0	0	0.82
0.03	0.25	0.68	0	0	0
0	0	0.84	0	0	0.83
0	0	0	0	0	0.11
0	0	0.98	0	0	0.98
0	0	0.78	0	0	0.76
0	0	0	0	0	0
0	0	0.73	0	0	0.73
0	0	0.93	0	0	0.92
0	0	0.93	0	0	0.92
0	0	0.96	0	0	0.90
0	0	0.90	0	0	0.90
0	0	0.78	0	0	0.77
0	0	0.81	0	0	0.79
0	0	0.94	0	0	0.91
0	0	0.80	0	0	0
0	0	0.95	0	0	0.95
0	0	0	0	0	0
0	0	0	0	0	0
0.01	0.09	0.74	0	0.03	0.17
0	0	0.75	0	0	0
0	0	0.98	0	0	0.94
0	0	0.86	0	0	0.80
0	0	0.91	0	0.03	0.89
0	0.02	0.50	0	0	0.50
0	0	0.57	0	0	0
0	0	0.62	0	0	0.64
0	0	0.88	0	0	0.87
0	0	0.70	0	0	0.70
0	0	0	0	0	0
0	0	0.90	0	0	0.07
0	0	0.95	0	0	0.94
0	0.21	1.00	0	0	0
0	0	0.47	0	0	0.88
0	0	0	0	0	0
0	0	0.97	0	0	0.97
0	0	0.98	0	0	0.93
0	0	0.92	0	0	0.83
0	0	0.46	0	0	0.46
0	0	0.87	0	0	0.87
0	0.01	0.21	0	0	0
0	0	0.68	0	0	0.28
0	0	0.98	0	0	0.98
0	0	0.80	0	0	0.76
0	0	0	0	0	0
0.01	0	0.20	0	0	0.20
0	0	0.87	0.02	0	0.84
0	0	0.94	0	0	0.93
0	0	0.95	0	0	0.94
0	0	0.64	0	0	0.67
0	0.05	0.64	0	0.01	0.61
0	0	0.99	0	0	0.98
0	0.05	0.64	0	0.01	0.62
0	0	0	0	0	0

E. SET 3 ANALYSES

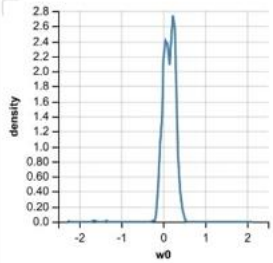
```
//Bacteriodetes == freq
//n == percentage
// k == percentage of the mutant X
var data = [
  {freq: 343, n:100, k:10}, //India
  {freq: 354 , n:100, k:0}, // Indonesia
   {freq: 202 , n:100, k:90}, // Italy
  {freq: 157 , n:100, k:60}, // Japan
   {freq: 346 , n:100, k:92}, //Sweden
  {freq: 549 , n:100, k:85}, //Spain
  {freq: 455 , n:100, k:59} //USA

]


var logistic = function(x) {
  return 1/( 1 + Math.exp(-x))
}

var model = function(){
  var w0 = gaussian(0,1)
  var w1 = gaussian(0,1)
map(
function(dt){
observe(Binomial({p: logistic (w0 + w1 * dt.freq), n: dt.n}) , dt.k)
 }, data)
```

WebPPL - probabilistic programming for the web

```
  return {w0: w0, w1: w1}
}
var options = {method : "MCMC", samples: 50000}
var dist= Infer(options,model)
viz.marginals(dist)
print("expectation : "+ expectation(marginalize(dist,'w1')))
print("positive expectation : " + expectation(marginalize(dist,'w1'),function(p){p>0}))
```

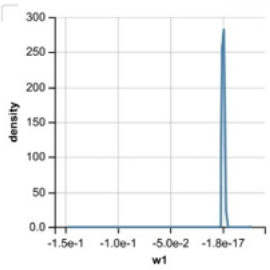
run

w0:



w1:

WebPPL - probabilistic programming for the web




```
expectation : 0.0003370870744748967
positive expectation : 0.9392400000000001
```

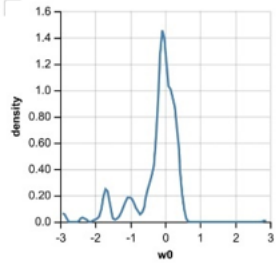
```
//Bacteriodetes == freq
//n == percentage
// k == percentage of the mutant AB
var data = [
  {freq: 343, n:100, k:74}, //India
  {freq: 354 , n:100, k:75}, // Indonesia
   {freq: 202 , n:100, k:91}, // Italy
  {freq: 157 , n:100, k:50}, // Japan
   {freq: 346 , n:100, k:94}, //Sweden
  {freq: 549 , n:100, k:87}, //Spain
  {freq: 455 , n:100, k:64} //USA

]


var logistic = function(x) {
  return 1/( 1 + Math.exp(-x))
}

var model = function(){
  var w0 = gaussian(0,1)
  var w1 = gaussian(0,1)
map(
function(dt){
observe(Binomial({p: logistic (w0 + w1 * dt.freq), n: dt.n}) , dt.k)
 }, data)
```

WebPPL - probabilistic programming for the web

```
  return {w0: w0, w1: w1}
}
var options = {method : "MCMC", samples: 50000}
var dist= Infer(options,model)
viz.marginals(dist)
print("expectation : "+ expectation(marginalize(dist,'w1')))
print("positive expectation : " + expectation(marginalize(dist,'w1'),function(p){p>0}))
```

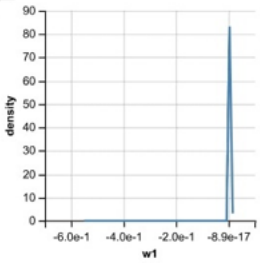
run

w0:



w1:

WebPPL - probabilistic programming for the web




```
expectation : 0.0040919021419380626
positive expectation : 0.9987800000000002
```

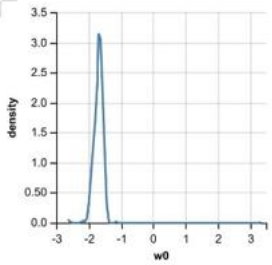
```
//Firmicutes == freq
//n == percentage
// k == percentage of the mutant Q
var data = [
  {freq: 434, n:100, k:13}, //India
  {freq: 520 , n:100, k:0}, // Indonesia
   {freq: 672 , n:100, k:89}, // Italy
  {freq: 610 , n:100, k:50}, // Japan
   {freq: 556 , n:100, k:92}, //Sweden
  {freq: 437 , n:100, k:83}, //Spain
  {freq: 534 , n:100, k:56} //USA

]


var logistic = function(x) {
  return 1/( 1 + Math.exp(-x))
}

var model = function(){
  var w0 = gaussian(0,1)
  var w1 = gaussian(0,1)
map(
function(dt){
observe(Binomial({p: logistic (w0 + w1 * dt.freq), n: dt.n}) , dt.k)
 }, data)
```

WebPPL - probabilistic programming for the web

```
  return {w0: w0, w1: w1}
}
var options = {method : "MCMC", samples: 50000}
var dist= Infer(options,model)
viz.marginals(dist)
print("expectation : "+ expectation(marginalize(dist,'w1')))
print("positive expectation : " + expectation(marginalize(dist,'w1'),function(p){p>0}))
```

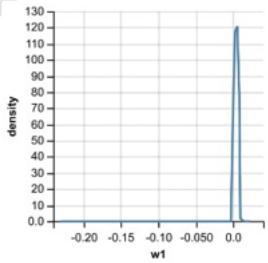
run

w0:



w1:

WebPPL - probabilistic programming for the web




```
expectation : 0.0031657830212496888
positive expectation : 0.99794
```

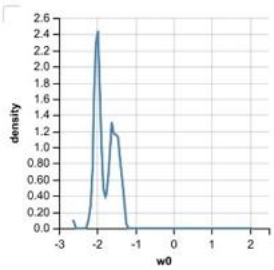
```
//Firmicutes == freq
//n == percentage
// k == percentage of the mutant AE
var data = [
  {freq: 434, n:100, k:17}, //India
  {freq: 520 , n:100, k:0}, // Indonesia
   {freq: 672 , n:100, k:89}, // Italy
  {freq: 610 , n:100, k:50}, // Japan
   {freq: 556 , n:100, k:93}, //Sweden
  {freq: 437 , n:100, k:84}, //Spain
  {freq: 534 , n:100, k:62} //USA

]


var logistic = function(x) {
  return 1/( 1 + Math.exp(-x))
}

var model = function(){
  var w0 = gaussian(0,1)
  var w1 = gaussian(0,1)
map(
function(dt){
observe(Binomial({p: logistic (w0 + w1 * dt.freq), n: dt.n}) , dt.k)
 }, data)
```

WebPPL - probabilistic programming for the web

```
  return {w0: w0, w1: w1}
}
var options = {method : "MCMC", samples: 50000}
var dist= Infer(options,model)
viz.marginals(dist)
print("expectation : "+ expectation(marginalize(dist,'w1')))
print("positive expectation : " + expectation(marginalize(dist,'w1'),function(p){p>0}))
```

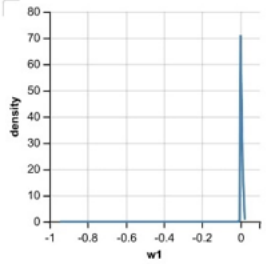
run

w0:



w1:

WebPPL - probabilistic programming for the web




```
expectation : 0.003637278731872112
positive expectation : 0.99904
```

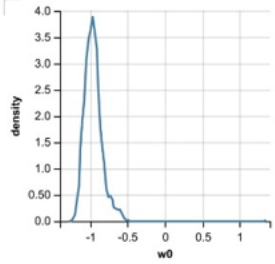
```
//Firmicutes == freq
//n == percentage
// k == percentage of the mutant AB
var data = [
  {freq: 434, n:100, k:74}, //India
  {freq: 520 , n:100, k:75}, // Indonesia
   {freq: 672 , n:100, k:91}, // Italy
  {freq: 610 , n:100, k:50}, // Japan
   {freq: 556 , n:100, k:94}, //Sweden
  {freq: 437 , n:100, k:87}, //Spain
  {freq: 534 , n:100, k:64} //USA

]


var logistic = function(x) {
  return 1/( 1 + Math.exp(-x))
}

var model = function(){
  var w0 = gaussian(0,1)
  var w1 = gaussian(0,1)
map(
function(dt){
observe(Binomial({p: logistic (w0 + w1 * dt.freq), n: dt.n}) , dt.k)
 }, data)
```

WebPPL - probabilistic programming for the web

```
  return {w0: w0, w1: w1}
}
var options = {method : "MCMC", samples: 50000}
var dist= Infer(options,model)
viz.marginals(dist)
print("expectation : "+ expectation(marginalize(dist,'w1')))
print("positive expectation : " + expectation(marginalize(dist,'w1'),function(p){p>0}))
```

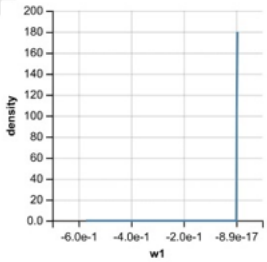
run

w0:



w1:

WebPPL - probabilistic programming for the web




```
expectation : 0.003904307736800334
positive expectation : 0.9991399999999999
```

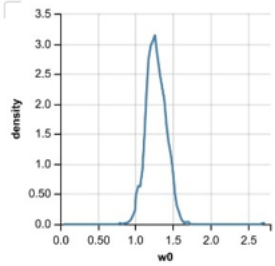
```
//Proteobacteria == freq
//n == percentage
// k == percentage of the mutant X
var data = [
  {freq: 91, n:100, k:10}, //India
  {freq: 24 , n:100, k:0}, // Indonesia
   {freq: 21 , n:100, k:90}, // Italy
  {freq: 9 , n:100, k:60}, // Japan
   {freq: 14 , n:100, k:92}, //Sweden
  {freq: 10 , n:100, k:85}, //Spain
  {freq: 10 , n:100, k:59} //USA

]


var logistic = function(x) {
  return 1/( 1 + Math.exp(-x))
}

var model = function(){
  var w0 = gaussian(0,1)
  var w1 = gaussian(0,1)
map(
function(dt){
observe(Binomial({p: logistic (w0 + w1 * dt.freq), n: dt.n}) , dt.k)
 }, data)
```

WebPPL - probabilistic programming for the web

```
  return {w0: w0, w1: w1}
}
var options = {method : "MCMC", samples: 50000}
var dist= Infer(options,model)
viz.marginals(dist)
print("expectation : "+ expectation(marginalize(dist,'w1')))
print("positive expectation : " + expectation(marginalize(dist,'w1'),function(p){p>0}))
```

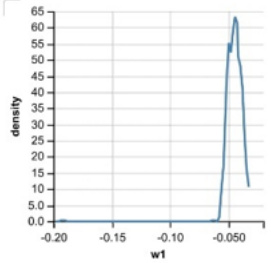
run

w0:



w1:

WebPPL - probabilistic programming for the web




```
expectation : -0.044871291447590736
positive expectation : 0
```

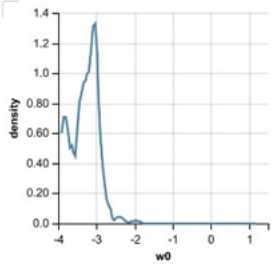
```
//Proteobacteria == freq
//n == percentage
// k == percentage of the mutant AC
var data = [
  {freq: 91, n:100, k:11}, //India
  {freq: 24 , n:100, k:0}, // Indonesia
   {freq: 21 , n:100, k:3}, // Italy
  {freq: 9 , n:100, k:6}, // Japan
   {freq: 14 , n:100, k:0}, //Sweden
  {freq: 10 , n:100, k:3}, //Spain
  {freq: 10 , n:100, k:8} //USA

]


var logistic = function(x) {
  return 1/( 1 + Math.exp(-x))
}

var model = function(){
  var w0 = gaussian(0,1)
  var w1 = gaussian(0,1)
map(
function(dt){
observe(Binomial({p: logistic (w0 + w1 * dt.freq), n: dt.n}) , dt.k)
 }, data)
```

WebPPL - probabilistic programming for the web

```
  return {w0: w0, w1: w1}
}
var options = {method : "MCMC", samples: 50000}
var dist= Infer(options,model)
viz.marginals(dist)
print("expectation : "+ expectation(marginalize(dist,'w1')))
print("positive expectation : " + expectation(marginalize(dist,'w1'),function(p){p>0}))
```

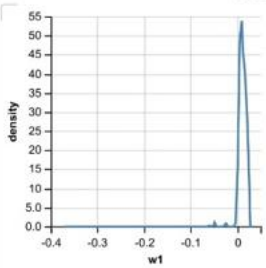
run

w0:



w1:

WebPPL - probabilistic programming for the web




```
expectation : 0.010637415293605344
positive expectation : 0.9727600000000004
```

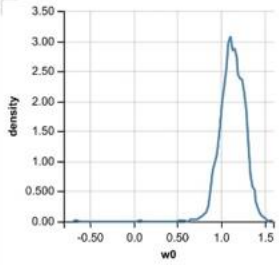
```
//Proteobacteria == freq
//n == percentage
// k == percentage of the mutant AB
var data = [
  {freq: 91, n:100, k:74}, //India
  {freq: 24 , n:100, k:75}, // Indonesia
   {freq: 21 , n:100, k:91}, // Italy
  {freq: 9 , n:100, k:50}, // Japan
   {freq: 14 , n:100, k:94}, //Sweden
  {freq: 10 , n:100, k:87}, //Spain
  {freq: 10 , n:100, k:64} //USA

]


var logistic = function(x) {
  return 1/( 1 + Math.exp(-x))
}

var model = function(){
  var w0 = gaussian(0,1)
  var w1 = gaussian(0,1)
map(
function(dt){
observe(Binomial({p: logistic (w0 + w1 * dt.freq), n: dt.n}) , dt.k)
 }, data)
```

WebPPL - probabilistic programming for the web

```
  return {w0: w0, w1: w1}
}
var options = {method : "MCMC", samples: 50000}
var dist= Infer(options,model)
viz.marginals(dist)
print("expectation : "+ expectation(marginalize(dist,'w1')))
print("positive expectation : " + expectation(marginalize(dist,'w1'),function(p){p>0}))
```

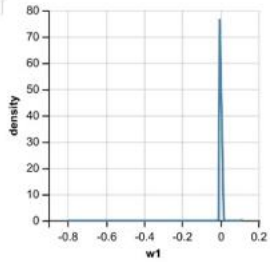
run

w0:



w1:

WebPPL - probabilistic programming for the web




```
expectation : 0.0016791682658504746
positive expectation : 0.5366000000000004
```

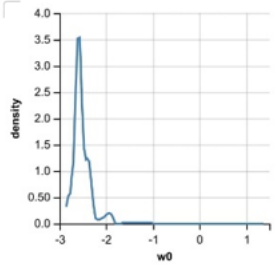
```
//Actinobacteria == freq
//n == percentage
// k == percentage of the mutant S:E484 / V
var data = [
  {freq: 128, n:100, k:50}, //India
  {freq:90 , n:100, k:0}, // Indonesia
   {freq:41 , n:100, k:6}, // Italy
  {freq:221 , n:100, k:56}, // Japan
   {freq:49 , n:100, k:5}, //Sweden
  {freq:1 , n:100, k:9}, //Spain
  {freq:1 , n:100, k:16} //USA

]


var logistic = function(x) {
  return 1/( 1 + Math.exp(-x))
}

var model = function(){
  var w0 = gaussian(0,1)
  var w1 = gaussian(0,1)
map(
function(dt){
observe(Binomial({p: logistic (w0 + w1 * dt.freq), n: dt.n}) , dt.k)
 }, data)
```

WebPPL - probabilistic programming for the web

```
  return {w0: w0, w1: w1}
}
var options = {method : "MCMC", samples: 50000}
var dist= Infer(options,model)
viz.marginals(dist)
print("expectation : "+ expectation(marginalize(dist,'w1')))
print("positive expectation : " + expectation(marginalize(dist,'w1'),function(p){p>0}))
```

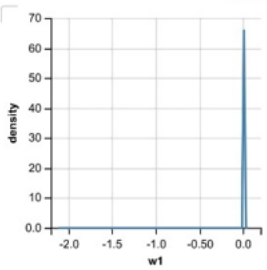
run

w0:



w1:

WebPPL - probabilistic programming for the web




```
expectation : 0.012470003161642571
positive expectation : 0.9887600000000002
```

Features

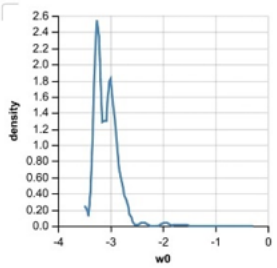
```
//Actinobacteria == freq
//n == percentage
// k == percentage of the mutant S:Q677 / AC
var data = [
  {freq: 128, n:100, k:11}, //India
  {freq:90 , n:100, k:0}, // Indonesia
   {freq:41 , n:100, k:3}, // Italy
  {freq:221 , n:100, k:6}, // Japan
   {freq:49 , n:100, k:0}, //Sweden
  {freq:1 , n:100, k:3}, //Spain
  {freq:1 , n:100, k:8} //USA

]


var logistic = function(x) {
  return 1/( 1 + Math.exp(-x))
}

var model = function(){
  var w0 = gaussian(0,1)
  var w1 = gaussian(0,1)
map(
function(dt){
observe(Binomial({p: logistic (w0 + w1 * dt.freq), n: dt.n}) , dt.k)
 }, data)
```

WebPPL - probabilistic programming for the web

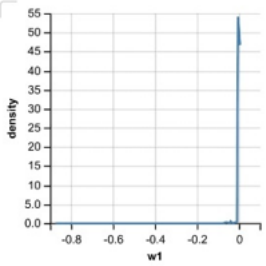
```
  return {w0: w0, w1: w1}
}
var options = {method : "MCMC", samples: 50000}
var dist= Infer(options,model)
viz.marginals(dist)
print("expectation : "+ expectation(marginalize(dist,'w1')))
print("positive expectation : " + expectation(marginalize(dist,'w1'),function(p){p>0}))
```

run

w0:



w1:




```
expectation : 0.0004964056495771601
positive expectation : 0.7725400000000004
```

F. Variant and Mutant Significance Index

Variants and VSI (Variant Significance Index)

Variant Name	Transmissibility	Virulence	Immune Evasion	VSI
20E (EU1)	2	4	2	8[1]
20A.EU2	4[2]	4[3]	5[4]	13
20H/501Y.V2	5[5]	2[6]	4	11
20J/501Y.V3	4[7]	2[8]	4[9]	10
20C/S:452R	5[10]	0[11]	4[12]	9

[1] Hodcroft EB, Zuber M, Nadeau S, Crawford KHD, Bloom JD, Veesler D, Vaughan TG, Comas I, Candelas FG, Stadler T, Neher RA. Emergence and spread of a SARS-CoV-2 variant through Europe in the summer of 2020. medRxiv [Preprint]. 2020 Nov 27:2020.10.25.20219063. doi: 10.1101/2020.10.25.20219063. PMID: 33269368; PMCID: PMC7709189.
[2] Jiahui Chen, Rui Wang, Menglun Wang, Guo-Wei Wei,
Mutations Strengthened SARS-CoV-2 Infectivity,
Journal of Molecular Biology,
Volume 432, Issue 19,
2020,
Pages 5212-5226,
ISSN 0022-2836,
https://doi.org/10.1016/j.jmb.2020.07.009.
(https://www.sciencedirect.com/science/article/pii/S0022283620304563)
[3] Liu Z, VanBlargan LA, Bloyet LM, Rothlauf PW, Chen RE, Stumpf S, Zhao H, Errico JM, Theel ES, Liebeskind MJ, Alford B, Buchser WJ, Ellebedy AH, Fremont DH, Diamond MS, Whelan SPJ. Landscape analysis of escape variants identifies SARS-CoV-2 spike mutations that attenuate monoclonal and serum antibody neutralization. bioRxiv [Preprint]. 2020 Nov 8:2020.11.06.372037. doi: 10.1101/2020.11.06.372037. Update in: Cell Host Microbe. 2021 Jan 27;: PMID: 33442690; PMCID: PMC7805447.
[4] Liu Z, VanBlargan LA, Bloyet LM, Rothlauf PW, Chen RE, Stumpf S, Zhao H, Errico JM, Theel ES, Liebeskind MJ, Alford B, Buchser WJ, Ellebedy AH, Fremont DH, Diamond MS, Whelan SPJ. Landscape analysis of escape variants identifies SARS-CoV-2 spike mutations that attenuate monoclonal and serum antibody neutralization. bioRxiv [Preprint]. 2020 Nov 8:2020.11.06.372037. doi: 10.1101/2020.11.06.372037. Update in: Cell Host Microbe. 2021 Jan 27;: PMID: 33442690; PMCID: PMC7805447.
[5] https://www.cdc.gov/coronavirus/2019-ncov/cases-updates/variant-surveillance/variant-info.html
[6] https://www.cdc.gov/coronavirus/2019-ncov/science/science-briefs/scientific-brief-emerging-variants.html
[7] https://www.cdc.gov/coronavirus/2019-ncov/science/science-briefs/scientific-brief-emerging-variants.html
[8] https://www.the-scientist.com/news-opinion/a-guide-to-emerging-sars-cov-2-variants-68387
[9] https://www.cdc.gov/coronavirus/2019-ncov/science/science-briefs/scientific-brief-emerging-variants.html
[10] https://www.cdc.gov/coronavirus/2019-ncov/cases-updates/variant-surveillance/variant-info.html
[11] https://covariants.org/variants/S.L452R
[12] https://www.cdc.gov/coronavirus/2019-ncov/cases-updates/variant-surveillance/variant-info.html

20C/S:484K	5[13]	0	5[14]	10
20A/S:484K	5	5	2	12[15]
20I/501Y.V1	5	5	2	12[16]
20A/S:154K	5[17]	5[18]	5[19]	15
20A/S:478K	5	5	5	15[20]
20A/S:439K	5	0	5	10[21]
S:677H.Robin1	0	4[22]	4[23]	8
S:677P.Pelican	0	4[24]	4[25]	8

20A/S:98F	0	0	0	0[26]
20C/S:80Y	0	0	0	0
20B/S:626S	0	0	0	0

[13] Zhou H, Dcosta BM, Samanovic MI, Mulligan MJ, Landau NR, Tada T. B.1.526 SARS-CoV-2 variants identified in New York City are neutralized by vaccine-elicited and therapeutic monoclonal antibodies. bioRxiv [Preprint]. 2021 Mar 24:2021.03.24.436620. doi: 10.1101/2021.03.24.436620. PMID: 33791698; PMCID: PMC8010725.

[14] Zhou H, Dcosta BM, Samanovic MI, Mulligan MJ, Landau NR, Tada T. B.1.526 SARS-CoV-2 variants identified in New York City are neutralized by vaccine-elicited and therapeutic monoclonal antibodies. bioRxiv [Preprint]. 2021 Mar 24:2021.03.24.436620. doi: 10.1101/2021.03.24.436620. PMID: 33791698; PMCID: PMC8010725.

[15] https://nccid.ca/covid-19-variants/ (This variant exhibits similar mutations to other SARS-CoV-2 variants including B.1.1.7)

[16] https://www.cdc.gov/coronavirus/2019-ncov/cases-updates/variant-surveillance/variant-info.html

[17] https://cov-lineages.org/lineages/lineage_B.1.html

[18]Convergent evolution of SARS-CoV-2 spike mutations, L452R, E484Q and P681R, in the second wave of COVID-19 in Maharashtra, India

Sarah Cherian, Varsha Potdar, Santosh Jadhav, Pragya Yadav, Nivedita Gupta, Mousmi Das, Soumitra Das,

doi: https://doi.org/10.1101/2021.04.22.440932;

Starr TN, Greaney AJ, Dingens AS, Bloom JD. Complete map of SARS-CoV-2 RBD mutations that escape the monoclonal antibody LY-CoV555 and its cocktail with LY-CoV016. bioRxiv [Preprint]. 2021 Feb 22:2021.02.17.431683. doi: 10.1101/2021.02.17.431683. Update in: Cell Rep Med. 2021 Apr 5;:100255. PMID: 33655250; PMCID: PMC7924270.

[19] McCallum M, Marco A, Lempp F, Tortorici MA, Pinto D, Walls AC, Beltramello M, Chen A, Liu Z, Zatta F, Zepeda S, di Iulio J, Bowen JE, Montiel-Ruiz M, Zhou J, Rosen LE, Bianchi S, Guarino B, Fregni CS, Abdelnabi R, Caroline Foo SY, Rothlauf PW, Bloyet LM, Benigni F, Cameroni E, Neyts J, Riva A, Snell G, Telenti A, Whelan SPJ, Virgin HW, Corti D, Pizzuto MS, Veesler D. N-terminal domain antigenic mapping reveals a site of vulnerability for SARS-CoV-2. bioRxiv [Preprint]. 2021 Jan 14:2021.01.14.426475. doi: 10.1101/2021.01.14.426475. Update in: Cell. 2021 Mar 16;: PMID: 33469588; PMCID: PMC7814825.

[20] https://covariants.org/variants/20A.S.478K

[21] https://covariants.org/variants/S.N439K

[22] https://covariants.org/variants/S.Q677H.Robin1

[23] https://www.cdc.gov/coronavirus/2019-ncov/cases-updates/variant-surveillance/variant-info.html

[24] https://covariants.org/variants/S.Q677P.Pelican

[25] https://www.cdc.gov/coronavirus/2019-ncov/cases-updates/variant-surveillance/variant-info.html

[26] https://covariants.org/variants/S.S98F

20B/S:1122L	0	0	0	0
S:N501	5	5	4	14[27]
S:E484	5	4	4	13[28]
S:H69-	4	4	4	12[29]
S:Q677	4	5	0	9[30]
S:Y453F	4	5	4	13[31]
S:S477	4	5	5	14[32]
S:L18	0	0	4	4[33]
S:Y144-	0	5	5	10[34]
S:K417	0	2	5	7[35]
S:H655	0	0	0	0[36]
S:P681	0	0	4	4[37]
ORF1a:S3675	0	0	0	0[38]

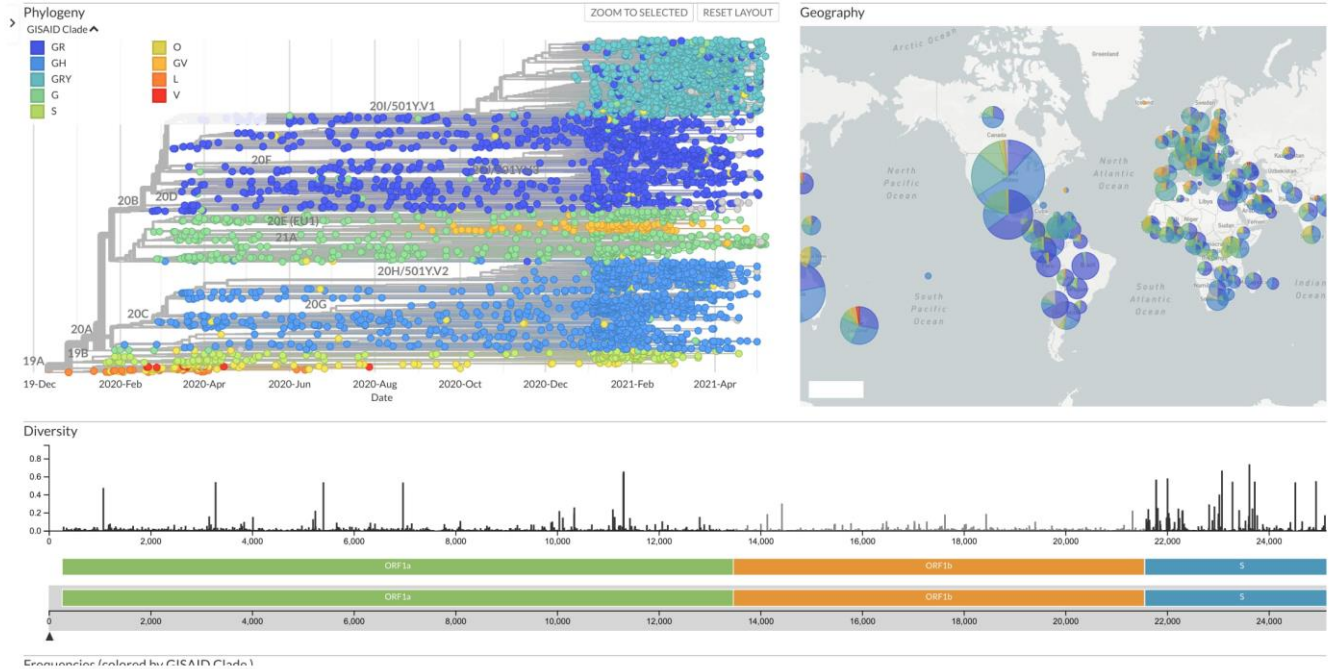
[27] https://covariants.org/variants/S.N501
[28] https://covariants.org/variants/S.E484
[29] https://covariants.org/variants/S.H69-
[30] https://covariants.org/variants/S.Q677
[31] https://covariants.org/variants/S.Y453F
[32] https://covariants.org/variants/S.S477
[33] https://covariants.org/variants/S.L18
[34] https://covariants.org/variants/S.Y144-
[35] https://covariants.org/variants/S.K417
[36] https://covariants.org/variants/S.H655
[37] https://covariants.org/variants/S.P681
[38] https://covariants.org/variants/ORF1a.S3675

Table A. The virulence index levels.

Virulence Index	***Represents disease severity of the variant.***
+2	no evidence of increased disease severity
0	effect on disease severity is unknown
+4	potentially more virulent
+5	increased virulent
Immune Evasion Index	***Represents immune evasion or vaccine efficacy of the variant.***
+2	no evidence of increased propensity of reinfection
0	effect on vaccine efficacy or immune response is unknown
+4	potential for immune escape
+5	increased propensity of reinfection

G. SARS-CoV-2 variant and mutation distributions (Global)

The screen shot was gathered from https://www.gisaid.org/phylodynamics/global/nextstrain/ (Last Access time: 15.05.2021, 17:25)


Phylogeny
GISAID Clade
GR
GH
GRY
G
S
O
GV
L
V
ZOOM TO SELECTED
RESET LAYOUT
20I/501Y.V1
20F
20B
20D
20E (EU1)
21A
20H/501Y.V2
20C
20G
20A
19B
19A
19-Dec
2020-Feb
2020-Apr
2020-Jun
2020-Aug
2020-Oct
2020-Dec
2021-Feb
2021-Apr
Date
Geography
Diversity
ORF1a
ORF1b
S
Frequencies (colored by GISAID Clade)

H. Regression Results

```
REGRESSION
  /MISSING LISTWISE
  /STATISTICS COEFF OUTS R ANOVA
  /CRITERIA=PIN(.05) POUT(.10)
  /NOORIGIN
  /DEPENDENT Zcov19fatality
  /METHOD=ENTER Zgdp_1 Zginiindex_1 Zconflict_1 Ztax_1 Zhousehold_1.
```

Regression

Variables Entered/Removed[a]

Model	Variables Entered	Variables Removed	Method
1	Zscore: SMEAN (household), Zscore: SMEAN(tax), Zscore: SMEAN (conflict), Zscore: SMEAN(gdp), Zscore: SMEAN (giniindex)[b]	.	Enter

a. Dependent Variable: Zscore(cov19fatality)

b. All requested variables entered.

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,645[a]	,416	,356	,80226322

a. Predictors: (Constant), Zscore: SMEAN(household), Zscore: SMEAN(tax), Zscore: SMEAN(conflict), Zscore: SMEAN(gdp), Zscore: SMEAN(giniindex)

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	22,462	5	4,492	6,980	,000[b]
	Residual	31,538	49	,644		
	Total	54,000	54			

a. Dependent Variable: Zscore(cov19fatality)

b. Predictors: (Constant), Zscore: SMEAN(household), Zscore: SMEAN(tax), Zscore: SMEAN(conflict), Zscore: SMEAN(gdp), Zscore: SMEAN(giniindex)

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t
		B	Std. Error	Beta	
1	(Constant)	,000	,108		,004
	Zscore: SMEAN(gdp)	-,406	,134	-,410	-3,040
	Zscore: SMEAN(giniindex)	,200	,135	,202	1,482
	Zscore: SMEAN(conflict)	,482	,127	,486	3,796
	Zscore: SMEAN(tax)	-,070	,121	-,071	-,578
	Zscore: SMEAN (household)	-,417	,144	-,421	-2,899

Coefficients[a]

Model		Sig.
1	(Constant)	,997
	Zscore: SMEAN(gdp)	,004
	Zscore: SMEAN(giniindex)	,145
	Zscore: SMEAN(conflict)	,000
	Zscore: SMEAN(tax)	,566
	Zscore: SMEAN (household)	,006

a. Dependent Variable: Zscore(cov19fatality)

```
REGRESSION
  /DESCRIPTIVES MEAN STDDEV CORR SIG N
  /MISSING LISTWISE
  /STATISTICS COEFF OUTS R ANOVA ZPP
  /CRITERIA=PIN(.05) POUT(.10)
  /NOORIGIN
  /DEPENDENT Zcov19reprodcution
  /METHOD=ENTER Zpopsize Zurbanization Zcov19mortality Zindoordeath Zoutdoo
rdeath.
```

Regression

Descriptive Statistics

	Mean	Std. Deviation	N
Zscore(cov19reprodcution)	,0000000	1,00000000	55
Zscore(popsize)	,0067019	1,00794676	55
Zscore(urbanization)	,0241139	,99264930	55
Zscore(cov19mortality)	,0083045	1,00726610	55
Zscore(indoordeath)	,0000000	1,00000000	55
Zscore(outdoordeath)	,0000000	1,00000000	55

Correlations

		Zscore (cov19reprodcution)	Zscore (popsize)	Zscore (urbanization)
Pearson Correlation	Zscore(cov19reprodcution)	1,000	,307	,185
	Zscore(popsize)	,307	1,000	-,263
	Zscore(urbanization)	,185	-,263	1,000
	Zscore(cov19mortality)	,085	,481	,099
	Zscore(indoordeath)	-,178	,340	-,796
	Zscore(outdoordeath)	-,180	,263	,196
Sig. (1-tailed)	Zscore(cov19reprodcution)	.	,011	,088
	Zscore(popsize)	,011	.	,026
	Zscore(urbanization)	,088	,026	.
	Zscore(cov19mortality)	,269	,000	,237
	Zscore(indoordeath)	,096	,006	,000
	Zscore(outdoordeath)	,095	,026	,076
N	Zscore(cov19reprodcution)	55	55	55
	Zscore(popsize)	55	55	55
	Zscore(urbanization)	55	55	55

Page 1

Correlations

		Zscore (cov19mortality)	Zscore (indoordeath)	Zscore (outdoordeath)
Pearson Correlation	Zscore(cov19reprodcution)	,085	-,178	-,180
	Zscore(popsize)	,481	,340	,263
	Zscore(urbanization)	,099	-,796	,196
	Zscore(cov19mortality)	1,000	-,048	,087
	Zscore(indoordeath)	-,048	1,000	-,186
	Zscore(outdoordeath)	,087	-,186	1,000
Sig. (1-tailed)	Zscore(cov19reprodcution)	,269	,096	,095
	Zscore(popsize)	,000	,006	,026
	Zscore(urbanization)	,237	,000	,076
	Zscore(cov19mortality)	.	,365	,264
	Zscore(indoordeath)	,365	.	,087
	Zscore(outdoordeath)	,264	,087	.
N	Zscore(cov19reprodcution)	55	55	55
	Zscore(popsize)	55	55	55
	Zscore(urbanization)	55	55	55

Correlations

		Zscore (cov19reprodcution)	Zscore (popsize)	Zscore (urbanization)
	Zscore(cov19mortality)	55	55	55
	Zscore(indoordeath)	55	55	55
	Zscore(outdoordeath)	55	55	55

Correlations

		Zscore (cov19mortality)	Zscore (indoordeath)	Zscore (outdoordeath)
	Zscore(cov19mortality)	55	55	55
	Zscore(indoordeath)	55	55	55
	Zscore(outdoordeath)	55	55	55

Page 2

Variables Entered/Removed[a]

Model	Variables Entered	Variables Removed	Method
1	Zscore (outdoordeath), Zscore (cov19mortality), Zscore (indoordeath), Zscore (popsize), Zscore (urbanization)[b]	.	Enter

a. Dependent Variable: Zscore(cov19reprodcution)

b. All requested variables entered.

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,621[a]	,386	,323	,82256390

a. Predictors: (Constant), Zscore(outdoordeath), Zscore(cov19mortality), Zscore(indoordeath), Zscore (popsize), Zscore(urbanization)

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	20,846	5	4,169	6,162	,000[b]
	Residual	33,154	49	,677		
	Total	54,000	54			

a. Dependent Variable: Zscore(cov19reprodcution)

b. Predictors: (Constant), Zscore(outdoordeath), Zscore(cov19mortality), Zscore(indoordeath), Zscore (popsize), Zscore(urbanization)

Coefficients[a]

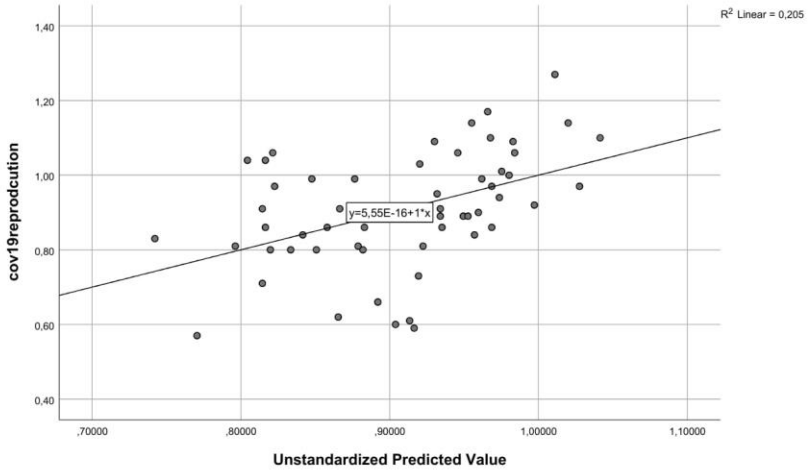
Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.
1	(Constant)	-,008	,111		-,069	,945
	Zscore(popsize)	,725	,150	,730	4,827	,000
	Zscore(urbanization)	,208	,188	,207	1,105	,275
	Zscore(cov19mortality)	-,262	,133	-,264	-1,966	,055
	Zscore(indoordeath)	-,359	,193	-,359	-1,867	,068
	Zscore(outdoordeath)	-,456	,123	-,456	-3,696	,001

Coefficients[a]

Model		Correlations Zero-order	Correlations Partial	Correlations Part
1	(Constant)			
	Zscore(popsize)	,307	,568	,540
	Zscore(urbanization)	,185	,156	,124
	Zscore(cov19mortality)	,085	-,270	-,220
	Zscore(indoordeath)	-,178	-,258	-,209
	Zscore(outdoordeath)	-,180	-,467	-,414

a. Dependent Variable: Zscore(cov19reprodcution)

Page 4


R2 Linear = 0,205
1,40
1,20
1,00
0,80
0,60
0,40
cov19reprodcution
y=5,55E-16+1*x
,70000
,80000
,90000
1,00000
1,10000
Unstandardized Predicted Value

Page 1

```
CORRELATIONS
  /VARIABLES=cov19reprodcution PRE_2
  /PRINT=TWOTAIL NOSIG
  /MISSING=PAIRWISE.
```

Correlations

Correlations

		cov19reprodcution	Unstandardized Predicted Value
cov19reprodcution	Pearson Correlation	1	,001
	Sig. (2-tailed)		,996
	N	55	55
Unstandardized Predicted Value	Pearson Correlation	,001	1
	Sig. (2-tailed)	,996	
	N	55	55

```
NONPAR CORR
  /VARIABLES=cov19reprodcution PRE_2
  /PRINT=SPEARMAN TWOTAIL NOSIG
  /MISSING=PAIRWISE.
```

Nonparametric Correlations

Correlations

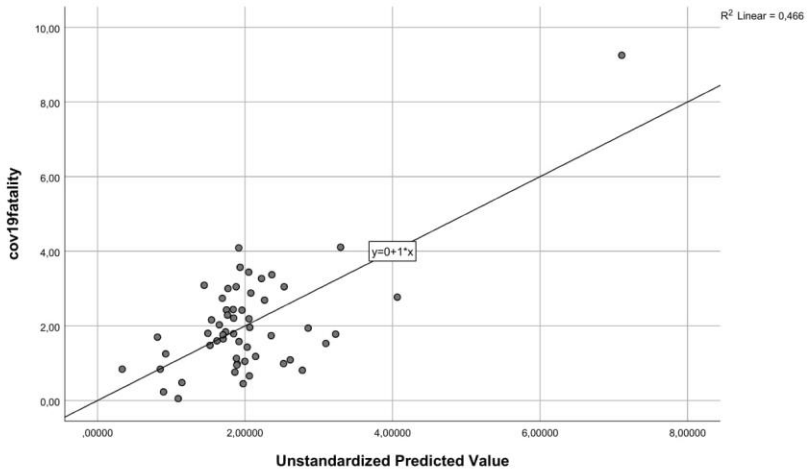
			cov19reprodcution
Spearman's rho	cov19reprodcution	Correlation Coefficient	1,000
		Sig. (2-tailed)	.
		N	55
	Unstandardized Predicted Value	Correlation Coefficient	,037
		Sig. (2-tailed)	,786
		N	55

Correlations

			Unstandardized Predicted Value
Spearman's rho	cov19reprodcution	Correlation Coefficient	,037
		Sig. (2-tailed)	,786
		N	55
	Unstandardized Predicted Value	Correlation Coefficient	1,000
		Sig. (2-tailed)	.
		N	55

```
GRAPH
  /SCATTERPLOT(BIVAR)=PRE_2 WITH cov19fatality
  /MISSING=LISTWISE.
```

Graph

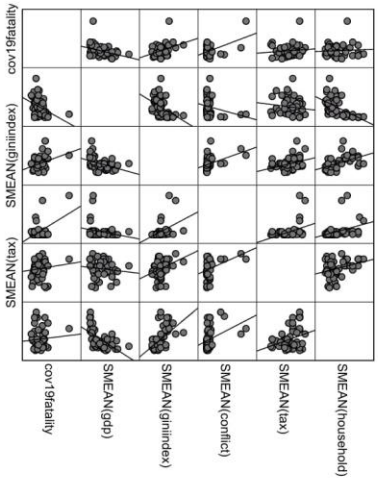


Page 2

İ. SET-1 CORELATIONS

```
GRAPH
  /SCATTERPLOT(MATRIX)=cov19fatality gdp_1 giniindex_1 conflict_1 tax_1 hou
sehold_1
  /MISSING=LISTWISE.
```

Graph



Page 1

```
CORRELATIONS
  /VARIABLES=cov19fatality gdp_1 giniindex_1 conflict_1 tax_1 household_1
  /PRINT=TWOTAIL NOSIG
  /MISSING=PAIRWISE.
```

Correlations

Correlations

		cov19fatality	SMEAN(gdp)	SMEAN (giniindex)
cov19fatality	Pearson Correlation	1	-,360**	,350**
	Sig. (2-tailed)		,007	,009
	N	55	55	55
SMEAN(gdp)	Pearson Correlation	-,360**	1	-,413**
	Sig. (2-tailed)	,007		,002
	N	55	56	56
SMEAN(giniindex)	Pearson Correlation	,350**	-,413**	1
	Sig. (2-tailed)	,009	,002	
	N	55	56	56
SMEAN(conflict)	Pearson Correlation	,483**	-,229	,434**
	Sig. (2-tailed)	,000	,090	,001
	N	55	56	56
SMEAN(tax)	Pearson Correlation	,110	-,125	,337*
	Sig. (2-tailed)	,423	,358	,011
	N	55	56	56
SMEAN(household)	Pearson Correlation	,070	-,560**	,496**
	Sig. (2-tailed)	,610	,000	,000
	N	55	56	56

Correlations

		SMEAN (conflict)	SMEAN(tax)	SMEAN (household)
cov19fatality	Pearson Correlation	,483**	,110	,070
	Sig. (2-tailed)	,000	,423	,610
	N	55	55	55
SMEAN(gdp)	Pearson Correlation	-,229	-,125	-,560**
	Sig. (2-tailed)	,090	,358	,000
	N	56	56	56
SMEAN(giniindex)	Pearson Correlation	,434**	,337*	,496**
	Sig. (2-tailed)	,001	,011	,000
	N	56	56	56
SMEAN(conflict)	Pearson Correlation	1	,386**	,375**
	Sig. (2-tailed)		,003	,004
	N	56	56	56
SMEAN(tax)	Pearson Correlation	,386**	1	,300*
	Sig. (2-tailed)	,003		,025
	N	56	56	56
SMEAN(household)	Pearson Correlation	,375**	,300*	1
	Sig. (2-tailed)	,004	,025	
	N	56	56	56

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

```
NONPAR CORR
  /VARIABLES=cov19fatality gdp_1 giniindex_1 conflict_1 tax_1 household_1
  /PRINT=SPEARMAN TWOTAIL NOSIG
  /MISSING=PAIRWISE.
```

Nonparametric Correlations

Page 2

Correlations

			cov19fatality	SMEAN(gdp)
Spearman's rho	cov19fatality	Correlation Coefficient	1,000	-,369**
		Sig. (2-tailed)	.	,006
		N	55	55
	SMEAN(gdp)	Correlation Coefficient	-,369**	1,000
		Sig. (2-tailed)	,006	.
		N	55	56
	SMEAN(giniindex)	Correlation Coefficient	,203	-,495**
		Sig. (2-tailed)	,138	,000
		N	55	56
	SMEAN(conflict)	Correlation Coefficient	,003	-,210
		Sig. (2-tailed)	,981	,121
		N	55	56
	SMEAN(tax)	Correlation Coefficient	,095	-,186
		Sig. (2-tailed)	,491	,170
		N	55	56
	SMEAN(household)	Correlation Coefficient	-,034	-,601**
		Sig. (2-tailed)	,806	,000
		N	55	56

Correlations

			SMEAN (giniindex)	SMEAN (conflict)
Spearman's rho	cov19fatality	Correlation Coefficient	,203	,003
		Sig. (2-tailed)	,138	,981
		N	55	55
	SMEAN(gdp)	Correlation Coefficient	-,495**	-,210
		Sig. (2-tailed)	,000	,121
		N	56	56
	SMEAN(giniindex)	Correlation Coefficient	1,000	,669**
		Sig. (2-tailed)	.	,000
		N	56	56
	SMEAN(conflict)	Correlation Coefficient	,669**	1,000
		Sig. (2-tailed)	,000	.
		N	56	56
	SMEAN(tax)	Correlation Coefficient	,353**	,528**
		Sig. (2-tailed)	,008	,000
		N	56	56
	SMEAN(household)	Correlation Coefficient	,521**	,422**
		Sig. (2-tailed)	,000	,001
		N	56	56

Correlations

			SMEAN(tax)	SMEAN (household)
Spearman's rho	cov19fatality	Correlation Coefficient	,095	-,034
		Sig. (2-tailed)	,491	,806
		N	55	55
	SMEAN(gdp)	Correlation Coefficient	-,186	-,601**
		Sig. (2-tailed)	,170	,000
		N	56	56
	SMEAN(giniindex)	Correlation Coefficient	,353**	,521**
		Sig. (2-tailed)	,008	,000
		N	56	56
	SMEAN(conflict)	Correlation Coefficient	,528**	,422**
		Sig. (2-tailed)	,000	,001
		N	56	56
	SMEAN(tax)	Correlation Coefficient	1,000	,188
		Sig. (2-tailed)	.	,165
		N	56	56
	SMEAN(household)	Correlation Coefficient	,188	1,000
		Sig. (2-tailed)	,165	.
		N	56	56

**. Correlation is significant at the 0.01 level (2-tailed).

```
CORRELATIONS
  /VARIABLES=cov19reprodcution popsize urbanization cov19mortality indoorde
ath outdoordeath
  /PRINT=TWOTAIL NOSIG
  /MISSING=PAIRWISE.
```

Correlations

Correlations

		cov19reprodcution	popsize	urbanization	cov19mortality
cov19reprodcution	Pearson Correlation	1	,307*	,185	,085
	Sig. (2-tailed)		,023	,175	,537
	N	55	55	55	55
popsize	Pearson Correlation	,307*	1	-,249	,482**
	Sig. (2-tailed)	,023		,064	,000
	N	55	56	56	56
urbanization	Pearson Correlation	,185	-,249	1	,108
	Sig. (2-tailed)	,175	,064		,429
	N	55	56	56	56
cov19mortality	Pearson Correlation	,085	,482**	,108	1
	Sig. (2-tailed)	,537	,000	,429	
	N	55	56	56	56
indoordeath	Pearson Correlation	-,178	,340*	-,796**	-,048
	Sig. (2-tailed)	,193	,011	,000	,729
	N	55	55	55	55
outdoordeath	Pearson Correlation	-,180	,263	,196	,087
	Sig. (2-tailed)	,189	,052	,152	,528
	N	55	55	55	55

Correlations

		indoordeath	outdoordeath
cov19reprodcution	Pearson Correlation	-,178	-,180
	Sig. (2-tailed)	,193	,189
	N	55	55
popsize	Pearson Correlation	,340*	,263
	Sig. (2-tailed)	,011	,052
	N	55	55
urbanization	Pearson Correlation	-,796**	,196
	Sig. (2-tailed)	,000	,152
	N	55	55
cov19mortality	Pearson Correlation	-,048	,087
	Sig. (2-tailed)	,729	,528
	N	55	55
indoordeath	Pearson Correlation	1	-,186
	Sig. (2-tailed)		,175
	N	55	55
outdoordeath	Pearson Correlation	-,186	1
	Sig. (2-tailed)	,175	
	N	55	55

*. Correlation is significant at the 0.05 level (2-tailed).

**. Correlation is significant at the 0.01 level (2-tailed).

```
NONPAR CORR
  /VARIABLES=cov19reprodcution popsize urbanization cov19mortality indoorde
ath outdoordeath
  /PRINT=SPEARMAN TWOTAIL NOSIG
  /MISSING=PAIRWISE.
```

Nonparametric Correlations

Page 2

Correlations

			cov19reprodcution	popsize
Spearman's rho	cov19reprodcution	Correlation Coefficient	1,000	,006
		Sig. (2-tailed)	.	,968
		N	55	55
	popsize	Correlation Coefficient	,006	1,000
		Sig. (2-tailed)	,968	.
		N	55	56
	urbanization	Correlation Coefficient	,255	-,104
		Sig. (2-tailed)	,060	,445
		N	55	56
	cov19mortality	Correlation Coefficient	-,123	,678**
		Sig. (2-tailed)	,372	,000
		N	55	56
	indoordeath	Correlation Coefficient	-,295*	,195
		Sig. (2-tailed)	,029	,155
		N	55	55
	outdoordeath	Correlation Coefficient	-,203	-,019
		Sig. (2-tailed)	,138	,888
		N	55	55

Correlations

			urbanization	cov19mortality
Spearman's rho	cov19reprodcution	Correlation Coefficient	,255	-,123
		Sig. (2-tailed)	,060	,372
		N	55	55
	popsize	Correlation Coefficient	-,104	,678**
		Sig. (2-tailed)	,445	,000
		N	56	56
	urbanization	Correlation Coefficient	1,000	,003
		Sig. (2-tailed)	.	,980
		N	56	56
	cov19mortality	Correlation Coefficient	,003	1,000
		Sig. (2-tailed)	,980	.
		N	56	56
	indoordeath	Correlation Coefficient	-,774**	,141
		Sig. (2-tailed)	,000	,304
		N	55	55
	outdoordeath	Correlation Coefficient	,086	,279*
		Sig. (2-tailed)	,533	,039
		N	55	55

Correlations

			indoordeath	outdoordeath
Spearman's rho	cov19reprodcution	Correlation Coefficient	-,295*	-,203
		Sig. (2-tailed)	,029	,138
		N	55	55
	popsize	Correlation Coefficient	,195	-,019
		Sig. (2-tailed)	,155	,888
		N	55	55
	urbanization	Correlation Coefficient	-,774**	,086
		Sig. (2-tailed)	,000	,533
		N	55	55
	cov19mortality	Correlation Coefficient	,141	,279*
		Sig. (2-tailed)	,304	,039
		N	55	55
	indoordeath	Correlation Coefficient	1,000	-,069
		Sig. (2-tailed)	.	,616
		N	55	55
	outdoordeath	Correlation Coefficient	-,069	1,000
		Sig. (2-tailed)	,616	.
		N	55	55

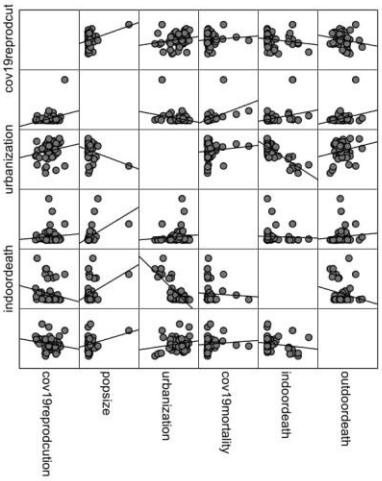
*. Correlation is significant at the 0.05 level (2-tailed).

**. Correlation is significant at the 0.01 level (2-tailed).

```
GRAPH
  /SCATTERPLOT(MATRIX)=cov19reprodcution popsize urbanization cov19mortalit
y indoordeath
    outdoordeath
  /MISSING=LISTWISE.
```

Graph

Page 5


cov19reproducut
urbanization
indoordeath
cov19reproduction
popsize
urbanization
cov19mortality
indoordeath
outdoordeath

Page 6

J. Data Source Tables

All databases and webpages were accessed in 15.05.2021 (Last Access time).

These data were gathered from different sources such as literature sources or publicly online databases. This document provides the references that provide the data of the table content.

A wide range of the aspects of the countries (such as sociological, economic, e.g.) that can have any impact on the survival of the SARS-CoV-2 were included as the attributes of the table.

- Countires were selected from https://covariants.org/per-country database. Number of total countries that are included in this database is 58. Bonaire and Curacao were excluded in the study due to insufficient information on the selected references. 56 countries with common Sars-CoV-2 variant data were included in the study.

Variable name	Variable type	Used in	Data source
Population Parameters			
Population size (in number)	Independent variable	Multi-Linear Regression Analysis, Bi-variate Corelation Analysis	https://www.populationpyramid.net/
Urbanization percent of the population	Independent variable	Multi-Linear Regression Analysis, Bi-variate Corelation Analysis	https://data.worldbank.org/indicator/SP.URB.TOTL.IN.ZS
Deaths by indoor air pollution rates	Independent variable	Multi-Linear Regression Analysis, Bi-variate Corelation Analysis	https://ourworldindata.org/indoor-air-pollution?country=
Deaths by outdoor air pollution rates	Independent variable	Multi-Linear Regression Analysis, Bi-variate Corelation Analysis	https://ourworldindata.org/outdoor-air-pollution

Deaths by Covid-19 (in number)	Independent variable	Multi-Linear Regression Analysis, Bi-variate Corelation Analysis	https://www.worldometers.info/coronavirus/?utm_campaign=homeAdvegas1?
Economic Parameters			
GDP per capita	Independent variable	Multi-Linear Regression Analysis, Bi-variate Corelation Analysis	https://data.worldbank.org/indicator/NY.GDP.PCAP.CD The fractional numbers rounded to whole numbers, Last entry (current) data was used.
Gini index (income inequality)	Independent variable	Multi-Linear Regression Analysis, Bi-variate Corelation Analysis	https://data.worldbank.org/indicator/SI.POV.GINI?name_desc=false&view=map&year=2019 The fractional numbers rounded to whole numbers
Conflict cases	Independent variable	Multi-Linear Regression Analysis, Bi-variate Corelation Analysis	https://acleddata.com/dashboard/#/dashboard Total events (reported) were used
Corporate Tax Rates	Independent variable	Multi-Linear Regression Analysis, Bi-variate Corelation Analysis	https://taxfounhttps://data.worldbank.org/indicator/SI.POV.GINI?name_desc=false&view=map&year=2019dation.org/publications/corporate-tax-rates-around-the-world/ The fractional numbers rounded to whole numbers.
Average Household Size: Number of members	Independent variable	Multi-Linear Regression Analysis, Bi-variate Corelation Analysis	https://population.un.org/Household/index.html#/countries/533
Diet Parameters			
Prevalence of Total Overweight Adults	Independent variable	Multi-Linear Regression Analysis, Bi-variate Corelation Analysis	https://apps.who.int/gho/data/view.main.CTRY2430A?lang=en Last entry (current) data was used (2016), The fractional numbers rounded to whole numbers.

Consumption of the Vegetable Oil	Independent variable	Multi-Linear Regression Analysis, Bi-variate Corelation Analysis	https://data.worldobesity.org/maps-obesity-day/?mapid=62 This database uses the data of the FAO (Food and Agriculture of the United Nations : http://www.fao.org/faostat/en/#data/FBS) and visualize the data. The fractional numbers rounded to whole numbers.
Consumption of the Animal Fat	Independent variable	Multi-Linear Regression Analysis, Bi-variate Corelation Analysis	https://data.worldobesity.org/maps-obesity-day/?mapid=62 This database uses the data of the FAO (Food and Agriculture of the United Nations : http://www.fao.org/faostat/en/#data/FBS) and visualize the data. The fractional numbers rounded to whole numbers.
Consumption of Sugars	Independent variable	Multi-Linear Regression Analysis, Bi-variate Corelation Analysis	https://data.worldobesity.org/maps-obesity-day/?mapid=62 This database uses the data of the FAO (Food and Agriculture of the United Nations : http://www.fao.org/faostat/en/#data/FBS) and visualize the data. The fractional numbers rounded to whole numbers.
Prevalence of undernourishment by percentage	Independent variable	Multi-Linear Regression Analysis, Bi-variate Corelation Analysis	https://data.worldbank.org/indicator/SN.ITK.DEFC.ZS
Micronutrient Deficiency Parameters			
Prevalence of Vitamin A deficiency	Independent variable	Multi-Linear Regression Analysis, Bi-variate Corelation Analysis	https://ourworldindata.org/grapher/prevalence-of-vitamin-a-deficiency-in-children?tab=table The fractional numbers rounded to whole numbers.
Vitamin D status Around the World	Independent variable	Multi-Linear Regression Analysis, Bi-variate Corelation Analysis	https://www.osteoporosis.foundation/educational-hub/topic/vitamin-d The fractional numbers rounded to whole numbers.
Global prevalence of Zinc Deficiency	Independent variable	Multi-Linear Regression Analysis, Bi-variate Corelation Analysis	https://ourworldindata.org/grapher/global-prevalence-of-zinc-deficiency Most recent data was used, The fractional numbers rounded to whole numbers.
Iodine Levels	Independent variable	Multi-Linear Regression Analysis, Bi-variate Corelation Analysis	https://www.who.int/vmnis/iodine/status/summary/IDD_estimates_table_2007.pdf?ua=1

Environmental Parameters			
Exposure to Solar UV Radiation	Independent variable	Multi-Linear Regression Analysis, Bi-variate Corelation Analysis	https://apps.who.int/gho/data/view.main.35300 For the countries that have not any information about sunlight exposure in this application, the information of the nearest country was used (for Aruba, Venezuela used.)
Avarage temperature	Independent variable	Multi-Linear Regression Analysis, Bi-variate Corelation Analysis	https://worldpopulationreview.com/country-rankings/hottest-countries-in-the-world
Forest Area	Independent variable	Multi-Linear Regression Analysis, Bi-variate Corelation Analysis	https://data.worldbank.org/indicator/AG.LND.FRST.ZS
Average Precipitation	Independent variable	Multi-Linear Regression Analysis, Bi-variate Corelation Analysis	https://data.worldbank.org/indicator/AG.LND.PRCP.MM
Air Toxicity Levels	Independent variable	Multi-Linear Regression Analysis, Bi-variate Corelation Analysis	https://www.iqair.com/world-air-quality-ranking For countries that have more than one entry, the most toxic city data was used.
General Toxicity Levels	Independent variable	Multi-Linear Regression Analysis, Bi-variate Corelation Analysis	https://www.iqair.com/world-most-polluted-countries
CO2 Emissions per capita	Independent variable	Multi-Linear Regression Analysis, Bi-variate Corelation Analysis	https://ourworldindata.org/co2-emissions
Diseases Parameters			
Anemia in pregnant women	Independent variable	Multi-Linear Regression Analysis, Bi-variate Corelation Analysis	https://ourworldindata.org/grapher/anemia-pregnant-women-vs-children?tab=table The fractional numbers rounded to whole numbers.
CANCER (For All Types of Cancer)	Independent variable	Multi-Linear Regression Analysis, Bi-variate Corelation Analysis	https://gco.iarc.fr/today/online-analysis-map?v=2020&mode=population&mode_population=continents&population=900&populations=900&key=asr&sex=0&cancer=39&type=0&statistic=5&prevalence=0&population_group=0&ages_group%5B%5D=0&ages_group%5B%5D=17&nb_items=10&group_cancer=1&include_nmsc=1&include_nmsc_other=1&projection=natural-

			earth&color_palette=default&map_scale=quantile&map_nb_colors=5&continent=0&show_ranking=0&rotate=%255B10%252C0%255D
Lung Cancer	Independent variable	Multi-Linear Regression Analysis, Bi-variate Corelation Analysis	https://ourworldindata.org/grapher/lung-cancer-deaths-per-100000-by-sex-1950-2002?tab=table
Asthma	Independent variable	Multi-Linear Regression Analysis, Bi-variate Corelation Analysis	https://ourworldindata.org/grapher/asthma-prevalence
COPD	Independent variable	Multi-Linear Regression Analysis, Bi-variate Corelation Analysis	Data used from the Number of Deaths by COPD per million section in this database.
Pneumonia	Independent variable	Multi-Linear Regression Analysis, Bi-variate Corelation Analysis	https://ourworldindata.org/grapher/pneumonia-death-rates-age-standardized
NDCs (Non-communicable Diseases)	Independent variable	Multi-Linear Regression Analysis, Bi-variate Corelation Analysis	https://ourworldindata.org/grapher/burden-of-disease-rates-from-ncds?tab=table To get more information about NDCs: https://ourworldindata.org/burden-of-disease
Diabetes	Independent variable	Multi-Linear Regression Analysis, Bi-variate Corelation Analysis	https://ourworldindata.org/grapher/diabetes-prevalence
Diarrheal Diseases	Independent variable	Multi-Linear Regression Analysis, Bi-variate Corelation Analysis	https://ourworldindata.org/grapher/diarrheal-disease-death-rates
Colorectal Cancer	Independent variable	Multi-Linear Regression Analysis, Bi-variate Corelation Analysis	https://www.worldgastroenterology.org/UserFiles/file/wdhd-2008-map-of-digestive-disorders.pdf The data of "Global Colorectal Cancer Incidence" section was used, The sum of female and male incidence rates was used.