

ORIGINAL ARTICLE / ÖZGÜN MAKALE

A Framework to Connect Viral Quasispecies, Microbiome, and Host Viral Quasispecies, Mikrobiyom ve Konak Arasında Bağlantı Kurmak İçin Bir Çerçeve

 Leman Nur NEHRİ¹,  Seher Elif KOÇOĞLU²

¹Middle East Technical University (METU), Biological Sciences, Ankara, Türkiye

²Gazi University, Institute Of Health Sciences, Biochemistry (Medicine), Ankara, Türkiye

Geliş: 04.04.2024, Kabul: 30.09.2024

Abstract

Aim: The aim of this study is to investigate the potential interactions between SARS-CoV-2 Spike protein variants and the host microbiota. While the Spike protein is known for its role in mediating viral entry into host cells, its impact on the host's microbial communities remains unclear. Given the microbiota's critical role in modulating immune responses and maintaining host homeostasis, understanding these interactions could provide new insights into disease progression and immune evasion mechanisms associated with COVID-19. By leveraging parameters extracted from the current literature and analyzing publicly available datasets, we seek to elucidate how these interactions might influence the severity of COVID-19 and the pathogenesis of emerging viral variants. This research may also highlight potential therapeutic targets for mitigating the effects of SARS-CoV-2 and its evolving forms.

Methods: This study investigates the interaction between Spike protein variants of SARS-CoV-2 and the host microbiota. To this end, the associations between various SARS-CoV-2 variants and different host factors derived from urban ecosystems have been statistically analyzed. Specifically, the influence of these host factors, which are linked to distinct microbiota compositions, on the interaction with Spike protein variants has been evaluated. A Bayesian Network approach has been employed for this analysis to model the complex relationships and dependencies among the host factors and microbiota compositions.

Results: This study investigates the interaction between Spike protein variants of SARS-CoV-2 and host factors. Hypothesis 1 (H1) posits that specific combinations of various host factors can explain the infectivity of SARS-CoV-2. The analyses reveal that 20 SARS-CoV-2 variants and mutants are significantly affected by various parameters (Table 2), indicating that H1 cannot be rejected. Additionally, it is suggested that the connections mentioned in H1 indicate the presence of a carrier within the host, potentially the microbiome. Hypothesis 2 (H2) proposes that the microbiota serves as the primary carrier of host factors, influencing the selection of specific SARS-CoV-2 mutants. To test this hypothesis, a Bayesian Network was constructed (Figure 3), which identified the probabilistic relationships between potential microbiota compositions and Spike variants.

Conclusion: As a result, it is suggested that different Spike protein variants may be present in hosts with varying microbial compositions. Additionally, the microbiota could serve as a carrier that influences the selection of viral mutants in hosts within the population, potentially impacted by external factors such as environmental conditions and human interactions.

Keywords: COVID-19, Microbiome, Spike Protein, Viral Variant, Host Factor

Corresponding Author: Leman Nur Nehri, Dr, Middle East Technical University (METU), Biological Sciences, Ankara, Türkiye. E mail: lnerkan@gmail.com

How to Cite: Nehri LN, Koçoğlu SE. A Framework to Connect Viral Quasispecies, Microbiome, and Host. *Journal of Immunology and Clinical Microbiology* 2024;9(2):73-88

©Copyright 2024 by the "International medical Education Library" The QMEL.org
Journal of Immunology and Clinical Microbiology published by Cetus Publishing.



Journal of Immunology and Clinical Microbiology 2022 Open Access (<https://dergipark.org.tr/tr/pub/jicm>)
Creative Commons Attribution Non-Commercial License: The articles in the *Journal of Immunology and Clinical Microbiology* are open access articles licensed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-sa/4.0/>) which permits unrestricted, non-commercial use, distribution and reproduction in any medium, provided the work is properly cited.

Öz

Amaç: Bu çalışmanın amacı, SARS-CoV-2 Spike proteini varyantları ile konak mikrobiotası arasındaki olası etkileşimleri araştırmaktır. Spike proteininin virüsün konak hücrelere girişini sağlamadaki rolü iyi bilinmesine rağmen, bu proteinin konak mikrobiyal topluluklar üzerindeki etkisi belirsizliğini korumaktadır. Mikrobiotanın bağışıklık yanıtlarını düzenlemede ve konak homeostazını sağlamadaki kritik rolü göz önüne alındığında, bu etkileşimlerin incelenmesi, COVID-19'un hastalık ilerleyişi ve bağışıklık kaçışı mekanizmaları hakkında yeni bilgiler sağlayabilir. Literatürdeki mevcut parametreler ve halka açık veri setleri kullanılarak bu etkileşimlerin COVID-19'un şiddeti ve ortaya çıkan virüs varyantlarının patogenezi üzerindeki etkileri araştırılmıştır. Bu araştırma aynı zamanda SARS-CoV-2 ve gelişen varyantlarının etkilerini hafifletmek için potansiyel terapötik hedef olarak mikrobiyotayı ortaya koymayı hedefler.

Yöntem: Bu çalışmada, SARS-CoV-2'nin Spike protein varyantları ile konak mikrobiota arasındaki etkileşim incelenmiştir. Bu amaçla, çeşitli SARS-CoV-2 varyantlarının kentsel ekosistemlerden elde edilen farklı konak faktörleriyle ilişkileri istatistiksel olarak analiz edilmiştir. Özellikle, bu konak faktörlerinin, farklı mikrobiota kompozisyonları ile olan etkileşimleri değerlendirilmiştir. Analiz için, konak faktörleri ile mikrobiota kompozisyonları arasındaki karmaşık ilişkileri ve bağımlılıkları modellemek amacıyla Bayesian Ağ yaklaşımı kullanılmıştır.

Bulgular: Bu çalışmada, SARS-CoV-2'nin Spike protein varyantları ile konak faktörleri arasındaki etkileşim incelenmiştir. Hipotez 1 (H1), çeşitli konak faktörlerinin belirli kombinasyonlarının SARS-CoV-2'nin enfektifliğini açıklayabileceğini öne sürmüştür. Analizler, 20 SARS-CoV-2 varyantı ve mutantının çeşitli parametrelerden önemli ölçüde etkilendiğini göstermiştir (Tablo 2). Bu sonuç, H1'in reddedilemeyeceğini ortaya koymaktadır. Ek olarak, H1'de belirtilen bağlantıların, konak içinde bir taşıyıcı olduğuna ve bunun mikrobiom olabileceğine işaret ettiği düşünülmektedir. Hipotez 2 (H2) ise, mikrobiotanın konak faktörlerini taşıyarak belirli SARS-CoV-2 mutantlarının seçimini etkileyen ana yapı olduğunu önermektedir. Bu hipotezi test etmek amacıyla oluşturulan Bayesian Ağ (Şekil 3) ile olası mikrobiota kompozisyonlarının Spike varyantları ile olasılıksal ilişkisi tespit edilmiştir.

Sonuç: Sonuç olarak, farklı Spike protein varyantlarının farklı mikrobiyal kompozisyonlara sahip konaklarda bulunabileceği önerilmektedir. Ayrıca, mikrobiota, konaklardaki viral mutantların seçimini etkileyebilecek bir taşıyıcı rolü üstlenebilir; bu etki, çevresel koşullar ve insan etkileşimleri gibi dış faktörlerden etkilenebilir.

Anahtar Kelimeler: COVID-19, Mikrobiyom, Spike Proteini, Viral Varyant, Konak Faktörler

INTRODUCTION

Coronavirus disease 2019 (COVID-19) is caused by the Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2), which is shared by many organisms such as bats, pigs, cats, and humans (World Health Organization, 2020). The SARS-CoV-2 virus belongs to the SARS-MERS viral family in the evolutionary pathway, and variants of these diseases have been seen before (Gorbalenya et al., 2020). SARS-CoV-2 is an RNA virus that belongs to the Nidovirales order and Coronaviridae family (Kahn, 2020). SARS-CoV-2 is evolutionarily related to HCV-229E, NL63, OC43, and HKU1, which are viruses that cause common colds in 15-30% of humans (Corman et al., 2019). Viruses belonging to the Nidovirales order exhibit similar structural features (Kahn, 2020). Nidoviruses have few structural

proteins and RNA as their genetic material, along with a lipid envelope that protects the genetic material from the environment (Kahn, 2020). All Nidoviruses contain a Nucleocapsid (N) protein that interacts with the Membrane protein (M); however, both structures and proteins vary among the viruses (Kahn, 2020; Wang et al., 2020). The genome sizes vary among the Nidoviruses, while the genome structures remain similar. All genomes possess two large Open Reading Frames (ORFs) that hold the genetic information of proteins responsible for transcription regulation. The parts for structural proteins (such as M and N) are located in the genome near the ORFs (Kahn, 2020; Wang et al., 2020). The life cycle of SARS-CoV-2 consists of four stages: the attachment of the virus to the cell and transfer of genetic material, processing of

genetic material, assembly of viral proteins resulting from translation, and release of the unified virions from the cell (Zhou et al., 2020). The interaction of viral proteins with various host proteins has also been the subject of many studies (Zhou et al., 2020; Wang et al., 2020). SARS-CoV-2 proteins associate with certain host proteins, forming complexes that alter the virus's effect on the host (Gorbalenya 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) exemplifies this type of relationship. Such SARS-CoV-2 virus-host protein interaction pathways can also be associated with MERS and SARS-CoV viruses, making them potential targets for drug development due to their shared patterns.

The SARS-CoV-2 genome consists of two ORF parts that encode non-structural proteins. In addition to the two ORFs, four structural gene regions carry the genetic information for the virus's structural proteins (Zhou et al., 2020). In SARS-CoV-2, 16 nonstructural proteins (derived from the cleavage of the two large ORF proteins), four structural proteins (spike (S), envelope (E), membrane (M), and nucleocapsid (N)), and eight accessory proteins are present (Wang et al., 2020). The polyproteins of Orf1a and Orf1b are cleaved into smaller non-structural proteins (NSPs). NSPs interact with each other to regulate gene expression, while the Membrane protein forms the virus's lipid membrane. The Nucleocapsid protein links to the Membrane protein and encapsidates the RNA genome. The Envelope protein is an integral membrane protein that creates an ion channel and plays a role in the virus replication process. The Spike protein is the surface glycoprotein that mediates the attachment of host cells to the virus (Wang et al., 2020).

Spike protein is one of the most important

structural proteins of SARS-CoV-2 (Walls et al., 2020). This protein recognizes and binds to the human host cell surface receptor angiotensin-converting enzyme-2 (ACE2), providing entry into the cell. The host's immune response is also triggered by the detection of the Spike protein (Walls et al., 2020). Moreover, the Spike protein determines the infectivity and transmissibility of the virus and is the major antigen inducer for the immune response (Zhou et al., 2020). Therefore, many vaccines have been designed to target the Spike protein (Wang et al., 2020). The Spike protein consists of two subunits: S1 and S2. S1 is responsible for binding to ACE2 receptors, and after this binding process, the S2 subunit facilitates fusion into the cell, allowing the virus's genetic material to enter (Zhou et al., 2020). The cleavage of the S1 subunit from S2 is critical for infection; therefore, antibodies bind to the Spike protein to prevent cleavage and inhibit the virus's fusion with the cell (Zhou et al., 2020).

SARS-CoV-2 exists as a haplotype in its host as an RNA virus, and Spike proteins can also be categorized through haplotype analysis (Korber et al., 2020). Haplotypes represent cumulative variations in genetic data on a single chromosome (Huang et al., 2020). In haplotype variations, a variant is dominant among the others, with these variants occurring at very low frequencies compared to the dominant haplotype (Huang et al., 2020). Clusters of mutants surround this main haplotype, with sequence similarities ranging between 93% and 99% among the dominant haplotype (Huang et al., 2020). In other words, the haplotype distribution in a host displays a scenario where one haplotype is central to the viral population, with some mutants present around it. Computational experiments have been conducted to verify these facts using experimental data and specific software (Huang et al., 2020; Wang et al., 2020). These findings are applicable

to RNA viruses as well. Since RNA viruses exhibit low recombination levels and lack true diversity in the conserved regions of their genome, haplotype distributions are minimal, and mutations accumulate around one or two main haplotypes (Huang et al., 2020; Wang et al., 2020). Consequently, these closely-related haplotypes in viral populations form viral quasispecies—defined as the dynamic distribution of closely related but non-identical mutant and recombinant viral genomes—or, in other words, quasispecies represent viral groups within a population composed of haplotype variations (Huang et al., 2020). These quasispecies function as a unit of selection due to their variation (Huang et al., 2020; Wang et al., 2020).

There are two main causes of variation in a viral population: recombinations and mutations (Wang et al., 2022; Simon-Loriere & Holmes, 2011; Pérez-Losada et al., 2015). Even though mutations and recombination events are high in non-conserved regions, they are rare in conserved regions. In viral quasispecies, there is a dominant haplotype that shows very low recombination events in its evolutionary history, with many mutant types surrounding these major dominant haplotypes (Töpfer et al., 2013). Specifically, they tend to identify haplotype probabilities—which represent the architecture of the viral population in a host—that shape host interactions. Spike protein is not an exception; the protein, along with SARS-CoV-2, is found in the host as a haplotype structure (Bui et al., 2022). Spike proteins are made up of small differences between different haplotypes that evolved from the same ancestor (Pérez-Losada et al., 2015). The Receptor Binding Domain (RBD) of the Spike protein, which binds to human ACE2 receptors, is not a recent acquisition by recombination but rather an ancient gain that is common to bat viruses (Boni et al., 2020).

Therefore, mutations (such as deletions and insertions), rather than recombination, have great importance in Spike protein (and SARS-CoV-2) evolution, generating Spike protein variants (Boni et al., 2020). As its evolution rate is similar across clades of SARS-CoV-2 variants, Spike protein is the major evolutionary driver, and SARS-CoV-2 variants are largely categorized according to Spike protein variants (Pérez-Losada et al., 2015). In summary, the distribution of SARS-CoV-2 variants in the host aligns with our general understanding of viral haplotype and quasispecies structure.

Microbiomes, which can be defined as the assemblage of microbes in a host, are representatives of the diseases or health condition of the host (Marchesi & Ravel, 2015). Microbiomes are key indicators of singular attributes directly related to the host (Bruijning et al., 2020), and 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 observed in the lung microbiota of terrestrial animals (Segal & Blaser, 2014), and the gut microbiota is another target for EDCs (Kumar et al., 2020). Since human microbiomes are major representatives of the host's attributes—such as diet, lifestyle, and medical record—as a whole (Scepanovic et al., 2019), changes in microbiome content can infer the evolutionary forces acting on the host (Bruijning et al., 2020). In microbiomes, ecological relations among species exist. The dominant species, also called founder species, alter the host's biological reactions by providing certain chemicals (Trosvik & de Muinck, 2015). For instance, the presence of a species can alter the host's immune response by triggering the production of more IgA, which affects the immune response, especially in respiratory areas, as the first line of immune defense

(Donaldson et al., 2018). The dominant species and other species change between health and disease conditions within a microbiome (Rinninella et al., 2019). There are many different characteristics of species within a microbiome. For example, dominant species are often in a positive relationship with other members of the microbiome, usually creating a mutualistic environment, while keystone species have a high number of both positive and negative relationships with other microbes. Keystone species are often found in low abundance but have a high number of ecological connections with other species in the microbiome (Trosvik & de Muinck, 2015). Moreover, 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, and bacteria—in the microbiota contributes to many metabolic functions of the host and plays a role in many physiological processes, especially the immune response (Zheng et al., 2020). The term dysbiosis is used to describe situations where changes in the microbiota are directly related to a host's illness (Rinninella et al., 2019). This term indicates that a microbiota community is directly related to a disease in the host, and when the host does not have this disease, the composition of the microbiota is significantly different from the disease state (Rinninella et al., 2019). To sum up, the microbiome is an area that has been studied under various conditions. The composition of the microbiota and the relative abundances of the organisms within it are related to both the disease and health conditions of the host (Shreiner et al., 2015).

Meta-community is a set of local communities that are linked by the dispersal of multiple potentially interacting species (Leibold et al., 2004), and a microbial meta-community is a variational set of

local (e.g., in some host organisms or a geographic area) microbe communities (Miller et al., 2018). Microbiomes are key indicators of certain attributes that are directly related to the host. For instance, genetic problems can be detected from microbiome content, or the host's lifestyle can be influenced by its microbiome. Human microbiota compositions show discontinuous rather than continuous variation of microbes; in other words, the microbes in the gut form 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 identified in human microbiota (Arumugam et al., 2011). Enterotypes indicate a balanced relationship between the host and its microbiota (Arumugam et al., 2011; Christensen et al., 2018; Segal & Blaser, 2014). The most important characteristic of microbiota composition is the functional relationship among microbes, rather than which specific bacterium is present (Arumugam et al., 2011). Microbiota shows phylogenetic variation at the genus and phylum levels among enterotypes and functional variation at the class level (Arumugam et al., 2011; Xiao et al., 2021; Costea et al., 2017). For instance, the Firmicutes and Bacteroides phyla are the most dominant species in the gut microbiota (Thursby & Juge, 2017). Although Bacteroides generally dominate the gut microbiome, in some enterotypes, Firmicutes can be the dominant organism (Trosvik & de Muinck, 2015; Arumugam et al., 2011; Mobeen et al., 2018). Actinobacteria, the most common phylum after Firmicutes and Bacteroidetes, is considered a keystone taxon in the gut microbiota due to its extensive ecological network with other gut microbes (Trosvik & de Muinck, 2015). Proteobacteria, another common species in the human intestinal microbiome,

METHODS

The study has 2 hypotheses:

H1: Specific combinations of various host factors can explain SARS-CoV-2 infectivity between variants and specific mutations on Spike protein.

H2: Microbiota is the main carrier of host factors inside the body which specific SARS-CoV-2 mutant is selected by the host.

From the data obtained through the literature research for this study, it has been determined that many different parameters in human life are related to both microbiota and COVID-19 (see Figure 1). To investigate these relations, a generative theoretical explanation was needed. As a result of the literature reviews, the relationship of microbiota with many parameters has been confirmed (see Table 1, Figure 1). Additionally, literature findings indicate that various parameters are associated with COVID-19, and it is known that COVID-19 causes changes in the host's microbiota composition (see Table 1, Figure 1). However, it is essential to investigate how different variants and Spike protein mutations are tolerated by various microbiota. There is no other virus like SARS-CoV-2 that provides detailed global data and the global distribution of its different mutants. Globally, the closest available data belongs to the SARS-MERS family, but even that does not come close to the breadth of data available for 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.

Table 1. Independent variables and their relations with COVID-19 and Microbiome via some examples from the literature

Independent variable:	Relatedness with COVID-19 and/or Microbiome:
Population size (in number)	[84], [100]
Urbanization percentage of the population	[100]–[103]
Deaths by indoor air pollution rates	[104]
Deaths by outdoor air pollution rates	[104]–[106]
Deaths by Covid-19 (in number)	[107], [108]
GDP per capita	[78], [109]
Gini index (income inequality)	[79]
Conflict cases	[79], [109]
Corporate Tax Rates	-
Average Household Size: Number of members	[110], [111]
Prevalence of Total Overweight Adults	[112], [113]
Consumption of the Vegetable Oil	[114], [115]
Consumption of the Animal Fat	[115]–[117]
Consumption of Sugars	[118], [119]
Prevalence of undernourishment by percentage	[120], [121]
Prevalence of Vitamin A deficiency	[122]–[124]
Vitamin D status Around the World	[125], [126]
The global prevalence of Zinc Deficiency	[127]–[129]
Iodine Levels	[69], [130]–[132]
Exposure to Solar UV Radiation	[80], [133], [134]
Average temperature	[82], [135]
Forest Area	[136], [137]
Average Precipitation	[138], [139]
Air Toxicity Levels	[81], [140]
General Toxicity Levels	[141], [142]
CO2 Emissions per capita	[143], [144]
Anemia in pregnant women	[145], [146]
CANCER (For All Types of Cancer)	[147], [148]
Lung Cancer	[149], [150]
Asthma	[151], [152]
COPD	[153], [154]
Pneumonia	[155], [156]
NDCs (Non-communicable Diseases)	[157], [158]
Diabetes	[159], [160]
Diarrheal Diseases	[161], [162]
Colorectal Cancer	[163], [164]

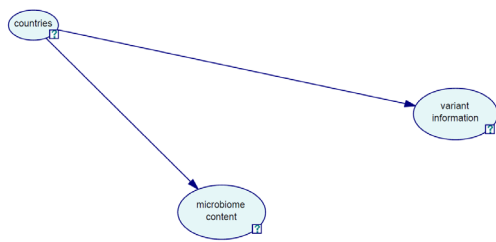


Figure 2. The Bayesian Belief Network for three nodes of countries, variant information, and microbiome content (Created by GeNIe 4.0 Academic).

The CoVariants section of the GISAID database was used to obtain data on 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. 56 of 58 countries have the relevant variant and mutant data were 32 selected for further analysis (Supplementary Material: ‘Data_mutants’ & ‘Data_parameters’).

A data file containing the data of all members of the parameter sets for the selected countries and the country names was created as a table analysis (Supplementary Material: ‘Data_parameters’). 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 (Supplementary Material: ‘Data_Sources_Variable Information’). If the data is unavailable in these sources the entry about this data was settled as NULL. These variables were used as independent variables for the analysis. The CoVariants / Per Variant section of the GISAID database was used to obtain mutant and variant data of countries. A data file containing the data of all mutants and variants on the GISAID database for the selected countries’ analysis (Supplementary Material: ‘Data_mutants’). The maximum frequency of mutants and variants for each country was used for the analysis as dependent variables. Stepwise regression analyses including all independent variables were performed to get the regression equations to describe the variance between

frequencies of the variants of the virus.

Using the significant results from this analysis (see Table 2 and Supplementary Material: ‘Data_mutants’), the relationship between gut microbiota and SARS-CoV-2 mutants was represented by a Bayesian Network. As microbiota data, the bacterial distribution of gut microbiota from Mobeen’s (2018) study was used for seven countries (Indonesia, India, Japan, Sweden, USA, Italy, and Spain) (Mobeen et al., 2018). These countries provided the frequency distribution of four types of bacteria in the gut microbiome—Bacteroides, Firmicutes, Actinobacteria, and Proteobacteria—which are common among various host microbiomes with functional effects. This distribution was used as prior probabilities in the Bayesian Belief Network, as Bayesian approaches are beneficial when data is limited, allowing the incorporation of prior knowledge (Bland & Altman, 1998). To connect the mutant data and microbiome data, the Bayesian approach was employed since the dataset is limited to the distribution of microbiomes across only seven countries. To construct the Bayesian Network, GeNIe 4.0 Academic was used (see Figure 2).

RESULTS

Hypothesis 1 (H1): Specific combinations of various host factors can explain SARS-CoV-2 infectivity between variants and specific mutations on Spike protein

In H1, it was suggested that Spike protein mutants and SARS-CoV-2 variants could be affected by selected variables. As a result of the analysis, it was found that the 20 variants and mutants were affected by various parameters (Table 2). Therefore H1 cannot be rejected.

Table 2. The results of stepwise regression analysis for different variants (predictors as the independent variables).

Spike Mutation	Predictors and Regression Results
20A.EU2	Predictors: Animal Fat. A regression equation was found ($F(1,38)=7.446, p<.010$), with an adjusted R2 of .142 and R2=.164.
20A/S:154K	No meaningful results.
20A/S:439K	Predictors: Urbanization, NDCs, Tax Rates . A regression equation was found ($F(3,36)=10.117, p<.000$), with an adjusted R2 of .412.
20A/S:478K	No meaningful results.
20A/S:484K	No meaningful results.
20A/S:98F	Predictors: Urbanization, Sunlight . A regression equation was found ($F(2,37)=5.166, p<.011$), with an adjusted R2 of .176.
20B/S:1122L	No meaningful results.
20B/S:626S	No meaningful results.
20C/S:452R	Predictors: Covid19 Mortality, Anemi, Zinc Deficiency, Conflict Rate. A regression equation was found ($F(4,35)=43.256, p<.000$), with an adjusted R2 of .813.
20C/S:484K	Predictors: Covid19 Mortality, Conflict Rate, Anemi, Diabet Air Toxicity, Tax Rates, Population Size, Sugar Consumption. A regression equation was found ($F(8,31)=51.970, p<.000$), with an adjusted R2 of .913.
20C/S:80Y	No meaningful results.
20E (EU1)	Predictors: Sunlight . A regression equation was found ($F(1,38)=8.619, p<.006$), with an adjusted R2 of .163 and R2=.185.
20H/501Y.V2	Predictors: Dierra, Household Size, GDP, Lung Cancer, Urbanization. A regression equation was found ($F(5,34)=68.963, p<.000$), with an adjusted R2 of .897.
20I/501Y.V1	Predictors: Air Toxicity, Iodine Uptake. A regression equation was found ($F(2,37)=13.647, p<.000$), with an adjusted R2 of .393.
20J/501Y.V3	Predictors: Conflict Rate, Anemi, Zinc Deficiency, Iodine Uptake, Covid19 Mortality, Animal Fat, CO2, COPD, Vegetable Oil, O2 level, Population Size, Diabet . A regression equation was found ($F(12,27)=38.240, p<.000$), with an adjusted R2 of .920.
ORF1a:S3675	No meaningful results.
S:677H.Robin1	Predictors: Covid19 Mortality, Conflict Rates, Anemi, Diabet, Air Toxicity, Tax Rates, Population Size, Sugar Consumption. A regression equation was found ($F(8,31)=51.970, p<.000$), with an adjusted R2 of .913.
S:677P.Pelican	Predictors: Covid19 Mortality Conflict Rates, Anemi, Diabet, Air Toxicity, Tax Rates, Population Size, Sugar Consumption. A regression equation was found ($F(8,31)=51.970, p<.000$), with an adjusted R2 of .913.
S:H655	Predictors: Conflict Rate, Anemi, Zinc Deficiency, Iodine Uptake, Covid19 Mortality, Animal Fat, CO2, Gini Index. A regression equation was found ($F(8,30)=31.829, p<.000$), with an adjusted R2 of .866.
S:H69-	Predictors: Rainfall, Sunlight. A regression equation was found ($F(2,36)=17.277, p<.000$), with an adjusted R2 of .461.
S:K417	Predictors: Dierra, CO2, Lung Cancer, Animal Fat. A regression equation was found ($F(4,34)=61.122, p<.000$), with an adjusted R2 of .864.
S:L18	Predictors: Gini Index, Diabet, Indoor Air Pollution Deaths, NDCs, Conflict Rates, Anemi . A regression equation was found ($F(6,32)=15.309, p<.000$), with an adjusted R2 of .693.
S:E484	Predictors: Tax Rates, Anemi, Outdoor Air Pollution Deaths, Gini Index, GDP, COPD, Temperature, Dierra, Covid19 Mortality. A regression equation was found ($F(9,29)=48.014, p<.000$), with an adjusted R2 of .918.
S:N501	No meaningful results.
S:P681	Predictors: Air Toxicity, Iodine Uptake, Anemi, Lung Cancer . A regression equation was found ($F(4,34)=9.134, p<.000$), with an adjusted R2 of .461.
S:Q677	Predictors: Population Size, Covid19 Mortality , BMI, Temperature, Vegetable Oil . A regression equation was found ($F(5,33)=17.842, p<.000$), with an adjusted R2 of .689.
S:S477	Predictors: Cancer . A regression equation was found ($F(1,37)= 10.097, p<.003$), with an adjusted R2 of .193.
S:Y144-	Predictors: Rainfall, Sunlight, Dierra, Pneume . A regression equation was found ($F(4,34)= 14.290, p<.000$), with an adjusted R2 of .583.
S:Y453F	Predictors: Lung Cancer, Vegetable Oil . A regression equation was found ($F(2,36)= 5.274, p<.010$), with an adjusted R2 of .184.

We suggest that these connections in Table 2 between host factors and viral mutants need a carrier inside of the host and it can be a microbiome:

H2: Microbiota is the main carrier of host factors inside the body which specific SARS-CoV-2 mutant is selected by the host.

To test this hypothesis, a Bayesian Network was generated (Figure 3) and some of the outputs of the Network can be represented as Figure 5.

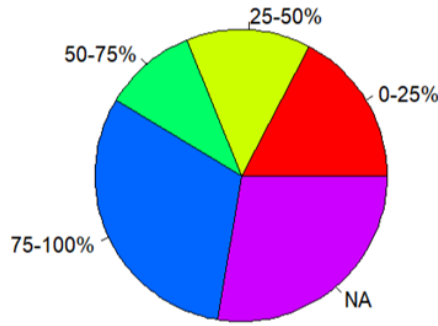


Figure 3. The proportion of variants is explained by independent variables. The separation in the proportions is based on adjusted R squares in Table 2 (The graph is created by R, on RStudio).

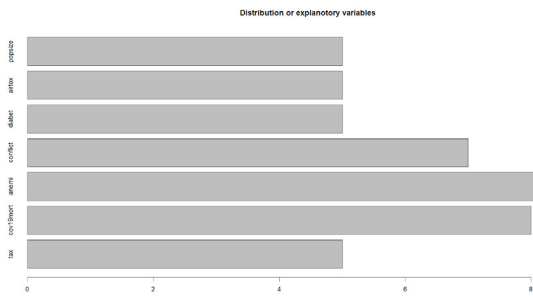


Figure 4. The independent variables of diabetes, tax rates, covid mortality rates, conflict rates, diabetes rates, air toxicity rates, and population size have the most entrants in the regression equations among the variants (The graph is created by R, on RStudio).

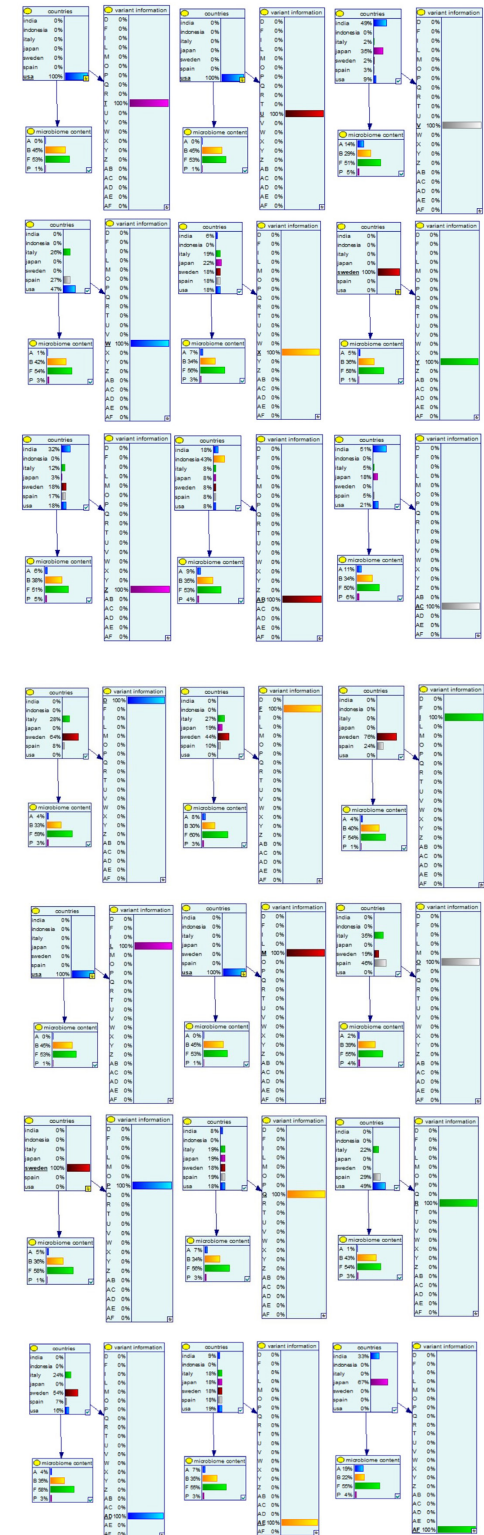


Figure 5. Some outputs of Bayesian Belief Network. *The Microbiome Content Table:* A for Actinobacteria, B for Bacteroidetes, F for Firmicutes, and P for Proteobacteria; *The Variant Information Table:* D for 20A.EU2, F for 20A/S:439K, I for 20A/S:98F, L for 20C/S:452R, M for 20C/S:484K, O for 20E (EU1), P for 20H/501Y.V2, Q for 20I/501Y.V1, R for 20J/501Y.V3, T for S:677H.Robin1, U for S:677P.Pelican, V for S: E484, W for S:H655, X for S:H69-, Y for S:K417, Z for S:L18, AB for S:P681, AC for S:Q677, AD for S:S477, AE for S:Y144-, AF for S:Y453F (The abbreviations are coherent Supplementary Material: 'Data_mutants').

DISCUSSION

Many variants (as dependent variables) were related to the independent variables at various rates (see Table 2, Figure 3). Variants also show various relationships between parameters in the literature. For instance, the 20I/501Y.V1 variant emerged in the United Kingdom and spread globally (Liu et al., 2021). This variant was predominantly found in Europe. In human reconstituted bronchial epithelium, the 20I/501Y.V1 variant replicates rapidly, contributing to its swift spread (Touret et al., 2021). This variant is also related to iodine uptake, which is linked to thyroid function. The gut microbiome plays several roles in influencing thyroid function, such as inhibiting thyroid-stimulating hormone (TSH) or modulating the immune response (Knezevic et al., 2020). Moreover, Firmicutes and Bacteroides exhibit lower abundance in inflammatory bowel disease (IBD), a condition associated with iodine malabsorption (Knezevic et al., 2020). Based on the results in Figure 5, this variant is more dependent on Actinobacteria than other variants. At this point, findings such as that polychlorinated biphenyls (PCBs), a banned air pollutant, reduce the composition of Actinobacteria in the gut microbiota (Popli et al., 2022) could be used as data to identify a link between this variant and air pollution, as shown in Table 2.

Dietary intake affects the human ACE2 receptor, the main target of the Spike protein, by influencing gene expression (Horne & Vohl, 2020; Bhattacharya et al., 2021). Therefore, changes in ACE2 structure due to dietary patterns can be linked to results such as animal fat, vegetable oil, sugar consumption, or malnutrition levels in various countries (see Table 2). Moreover, the mutation S: Y453F (see Table 2) enhances interaction with ACE, facilitating host adaptation (Ren et al., 2021). Even though studies on the relationship between

gut microbiota content and obesity are controversial, there is some evidence that Actinobacteria composition increases in the gut microbiomes of laboratory animals during obesity (Clarke et al., 2012; Kim et al., 2019). Our results show a high correlation between Actinobacteria, which has the highest relative abundance in Japan among the countries studied, and the S: Y453F mutant (see Figure 5). It is suggested that the Japanese diet promotes a healthy gut microbiome (Asano et al., 2020). Even though the adjusted R-squared value for this spike mutant is not highly descriptive (see Table 2), the strong dependence on Actinobacteria in relation to this mutant (see Figure 5) may be linked to dietary habits that influence the host microbiome. Nevertheless, all these potential links need to be explored more thoroughly.

Moreover, chronic diseases are related to SARS-CoV-2 cases and their severity (Liu et al., 2020), and our results suggest that diabetes is the most common parameter as a disease among all the variants (see Figure 4). It is well known that SARS-CoV-2 is linked to the economy (Bloem & Salemi, 2021; Elgar 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), as shown in Table 2. The economic parameters observed in Table 2 could be due to the strong relationship between economic activities and viral diseases (Adda, 2016). Many mutants can be related to different parameters. For instance, the S: Y144 mutation is another Spike protein mutation found in the 20I/501Y.V1 variant and other circulating variants, and it is associated with antibody escape (Focosi et al., 2021). This mutant has been linked to viral shedding in a patient in Washington (Avanzato et al., 2020), which is one of the largest metropolises in the United States. This city also experiences deaths due to increasing heat and excessive

ozone concentrations (Jackson et al., 2010). In this mutant's regression equation, precipitation and sunlight are included as variables (see Table 2), and it also shows a high level of association with Actinobacteria (see Figure 5). Since gut microbiome composition is influenced by both genetic and environmental factors (Qin et al., 2022), carriers of environmental factors may be related to the microbiome, especially Actinobacteria. This species is a predominant bacterium in the Italian gut microbiome compared to other nations (Rinninella et al., 2019), and infection and death rates from this Spike variant of SARS-CoV-2 are highest in Italy (Dawood et al., 2021). Another example is the S: H69 deletion in the Spike protein, which was sequenced mostly in Europe (Bal et al., 2021). This mutant occurs alongside others and is another example of immune escape, similar to S: Y144 (Meng et al., 2021). This mutant can also be associated with Actinobacteria (see Figure 5), suggesting a potential link between antibody escape and Actinobacteria (Tabib-Salazar et al., 2013) via this Spike mutant. In addition, the higher association rate with Actinobacteria may be related to the widespread use of probiotic supplements, which improve intestinal microbiota, particularly in Europe (Saxelin, 2008). Moreover, this link cannot be observed solely through Actinobacteria. Firmicutes and Bacteroides are the dominant organisms in the gut microbiome and provide the majority of ecological relations within the human gut microbiota (Bradley & Pollard, 2017). It is possible that mutants with high antibody escape rates may evade host immune defense depending on the presence of these species, as they are associated with immune responses (Donaldson et al., 2018; Peterson et al., 2015; Kosiewicz et al., 2011). Therefore, even though some connections exist in the literature, specifying the linkages between these multi-variable systems requires focused research. Additionally,

the predictors of this mutant include rainfall and sunlight (see Table 2), so the main factor linking these external factors (sunlight and rainfall) and internal factors (antibody escape and Spike mutants) needs further exploration. The composition of the microbiota, particularly Actinobacteria, may be a mediating factor for the interaction of external and internal forces on the host.

Firmicutes and Bacteroides do not vary much within a certain range, but we see that Actinobacteria shows much more variation (see Figure 5) depending on each variant. This may be related to the fact that Actinobacteria is a keystone species (Trosvik & de Muinck, 2015), and the functional relationships of keystone species shape an entire ecosystem. Therefore, changes in Actinobacteria composition may have a more decisive influence on the differentiation of variants than other microbes. It can also be argued that variants and mutants that are not associated with Proteobacteria may be independent of functional diversity in gut microbes, as they show no relation with Proteobacteria. This could be due to the fact that Proteobacteria are responsible for functional diversity in the intestine (Bradley & Pollard, 2017).

This study has limitations. The distributions shown in Figure 5 are dominated by Firmicutes and Bacteroides, while Proteobacteria and Actinobacteria are low, because these four dominant species in the human gut microbiota are present in the host at a certain interval (Mobeen et al., 2018). The small size of the data set used was accounted for by the Bayesian method, a probabilistic approach that allows for the interpretation of small data sets. Since the results presented here are the product of a probabilistic approach (see Figure 5), no significant differences are observed. However, the results obtained can help identify links between mutant variations and bacterial compositions. Another issue is that

the data used in this study covers the early days of COVID-19. Therefore, much of what this study addresses regarding virulence and spread involves mutants that emerged early in the pandemic. If a study with a broader time interval is conducted, this factor should be taken into account. It is likely that later on, the parameters relevant to COVID-19 and virus mutants may have increased, and the relevance of these parameters and the mutants at hand may have changed. The main point that this study aims to emphasize is that the host microbiota can be, or at least one of the carriers of, external factors within the host's body.

It is known that certain phyla variations are associated with various diseases, particularly in the intestinal microbiota. However, in some cases, variations that are not detected at the phylum level but are detected at the species level are also known to affect host status (Wakita et al., 2018). In this study, the geographic variations observed are at the phylum level, and two dominant phyla (Bacteroidetes and Firmicutes), one keystone phylum (Actinobacteria), and one phylum that influences the functional diversification of the microbiome (Proteobacteria) were analyzed. This is a limitation of the study, as only phylum-level analysis was possible with the available dataset. However, analyses at other levels, such as species or family, may be related to different host metabolic factors and functions. Therefore, researchers who wish to explore this topic should also consider the functional effects at different levels. The existence of a gut-lung crosstalk system (Wakita et al., 2018) may also suggest that different respiratory viral mutants could affect the transmission, virulence, and immune response of the host, as different microbiota compositions are known to influence crosstalk networks. Although the results of this study do not conclusively establish this relationship, it remains a possibility. Since lung microbiota

studies are usually conducted in laboratory environments isolated from the organism (Tulic et al., 2016), it may be necessary to perform and investigate such studies at the organismal level.

SARS-CoV-2 is in a highly advantageous position compared to other viruses in terms of both clinical data and the traceability of its mutants globally (Petrosillo et al., 2020). However, establishing a control group for this study may be necessary to study the viral mutant-microbiota relationship in detail and more meaningfully. In terms of in silico analysis, no comparable data, such as the relationship of COVID-19 with human factors, could be found for other viruses. Most comparison data remain within the axis of clinical data. Researchers who wish to investigate the viral mutant-microbiota relationship in more detail may consider establishing a comparable control group for the virus.

Additionally, there are challenges in making comprehensive comparisons among microbiome species. Understanding microbiota in terms of composition, diversity, and function is being studied, and it is thought that functional contributions are more important than species diversity in establishing microbiota composition. Ecological microbiota studies seek to understand specific gut microbiota functions in the pathways of host-microbiome interactions. When studying microbial divergence within the microbiota, it is known that there is significant species diversity among humans at the species level. Functional diversity studies, on the other hand, focus on specific genes and functions performed by particular microbial compositions, based on the concept of forming a microbiota community grounded in functional roles within the microbial ecosystem rather than species-level diversity. While microbial composition may vary greatly between individuals in terms of

species diversity, there are not significant differences in terms of functionality. In other words, the functional diversity of the human microbiome has been highly conserved among individuals since the core functions in the microbiota play crucial roles in the host's metabolic pathways (Lozupone et al., 2012).

However, despite these opposing arguments, this study aims to highlight the potential link between the macro and micro worlds that needs to be explored. One of the most effective ways to investigate this is through a combination of bioinformatics and wet lab processes—identifying indicator microbes and mutants, which can be confirmed by field studies—and conducting comprehensive studies. This approach may help answer the question: What could be the selective forces in a construct that links host factors to the survival of variants? In other words, since external elements need to be maintained inside, and a favorable environment is essential for this, a dynamic system of relationships can be constructed through the internal and external flows of the host. Developing this understanding and collecting and interpreting data in this manner require theoretical frameworks that allow different types of data to be considered on the same plane, rather than merely inferring relationships between macro and micro by combining wet lab and informatics processes.

CONCLUSION

SARS-CoV-2 has advantages in clinical data and mutation tracking compared to other viruses. However, studying the viral mutant-microbiota relationship requires a control group, as existing analyses lack experimental validation. Researchers should consider establishing such a control group. Challenges exist in comprehensively assessing microbiota types, as functional contributions are more crucial than species

diversity. Despite significant species-level diversity among individuals, the human microbiome's functional diversity is largely preserved due to key metabolic roles. This study aims to explore the link between macro and micro worlds through a combination of bioinformatics and laboratory processes. Identifying indicator microbes and mutants could clarify selective factors influencing variant survival, emphasizing the need for theoretical frameworks that integrate diverse data types.

REFERENCE

1. Avanzato, V. A., Matson, M. J., Seifert, S. N., Pryce, R., Williamson, B. N., Anzick, S. L., Barbian, K., Patel, A., Judson, S. D., Bowman, K., Martens, C., Li, D., Fischer, R. J., Munster, V. J., & Kuhn, J. H. (2020). Case study: Prolonged infectious SARS-CoV-2 shedding from an asymptomatic immunocompromised individual with cancer. *Cell*, 183(7), 1901-1912.e9. <https://doi.org/10.1016/j.cell.2020.10.049>
2. Bal, A., Destras, G., Gaymard, A., Stefic, K., Marlet, J., Eymieux, S., Billaud, G., Laurent, F., Gonzalez, C., Valette, M., & Lina, B. (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), 2100008. <https://doi.org/10.2807/1560-7917.ES.2021.26.3.2100008>
3. 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>
4. Boni, M. F., Lemey, P., Jiang, X., Lam, T. T.-Y., Perry, B. W., Castoe, T. A., Rambaut, A., & Robertson, D. L. (2020). Evolutionary origins of the SARS-CoV-2 sarbecovirus lineage responsible for the COVID-19 pandemic. *Nature Microbiology*, 5(11), 1408-1417. <https://doi.org/10.1038/s41564-020-0771-4>
5. Bradley, P. H., & Pollard, K. S. (2017). Proteobacteria explain significant functional variability in the human gut microbiome. *Microbiome*, 5(1), 36. <https://doi.org/10.1186/s40168-017-0244-z>
6. 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>
7. Bruijning, M., Henry, L. P., Forsberg, S. K. G., Metcalf, J. E., & Ayroles, J. F. (2020). When the microbiome defines the host phenotype: Selection on vertical transmission in varying environments. *bioRxiv*. <https://doi.org/10.1101/2020.04.28.066134>
 8. Bui, N.-N., Lin, Y.-T., Huang, S.-H., & Lin, C.-W. (2022). Haplotype distribution of SARS-CoV-2 variants in low and high vaccination rate countries during ongoing global COVID-19 pandemic in early 2021. *Infection, Genetics and Evolution*, 97, 105164. <https://doi.org/10.1016/j.meegid.2021.105164>
 9. Corman, V. M., Muth, D., Niemeyer, D., & Drosten, C. (2018). Hosts and sources of endemic human coronaviruses. *Advances in Virus Research*, 100, 163–188. <https://doi.org/10.1016/bs.aivir.2018.01.001>
 10. Corman, V. M., Muth, D., Niemeyer, D., & Drosten, C. (2019). Hosts and sources of endemic human coronaviruses. *Advances in Virus Research*, 100, 163–188. <https://doi.org/10.1016/bs.aivir.2018.01.001>
 11. Das, B., Nair, G. B., Pal, A., Ramamurthy, T., Ganguly, S., Majumder, K. K., & Sarkar, B. L. (2018). Gut microbiota of healthy and malnourished children in India. *Scientific Reports*, 8(1), 3215. <https://doi.org/10.1038/s41598-018-21594-4>
 12. 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>
 13. Donaldson, G. P., Lee, S. M., & Mazmanian, S. K. (2018). Gut microbiota utilize immunoglobulin A for mucosal colonization. *Science*, 360(6390), 795–800. <https://doi.org/10.1126/science.aag0926>
 14. Focosi, D., Maggi, F., Franchini, M., McConnell, S., & Casadevall, A. (2021). Analysis of immune escape variants from antibody-based therapeutics against COVID-19: A systematic review. *International Journal of Molecular Sciences*, 23(1), 29. <https://doi.org/10.3390/ijms23010029>
 15. Gorbalenya, A. E., Baker, S. C., Baric, R. S., de Groot, R. J., Drosten, C., Gulyaeva, A. A., Haagmans, B. L., Lauber, C., Leontovich, A. M., Neuman, B. W., Penzar, D., Perlman, S., Poon, L. L. M., Samborskiy, D., Sidorov, I. A., Sola, I., & Ziebuhr, J. (2020). The species severe acute respiratory syndrome-related coronavirus: Classifying 2019-nCoV and naming it SARS-CoV-2. *Nature Microbiology*, 5(4), 536–544. <https://doi.org/10.1038/s41564-020-0695-z>
 16. Huang, S., Chai, Y., Zhang, L., Wang, X., Wang, H., Zhang, L., Shi, Z., & Fu, J. (2020). Population genetic structure of SARS-CoV-2 spike gene mutations. *Genome Biology and Evolution*, 12(11), 1988–2001. <https://doi.org/10.1093/gbe/evaa240>
 17. Kahn, J. S. (2020). Nidovirales. *Encyclopedia of Virology*, 419–430. <https://doi.org/10.1016/B978-0-12-809633-8.21501-X>
 18. Knezevic, J., Starchl, C., Berisha, A. T., & Amrein, K. (2021). Thyroid-gut-axis: How does the microbiota influence thyroid function? *Nutrients*, 12(6), 1769. <https://doi.org/10.3390/nu12061769>
 19. Lazar, V., Ditu, L.-M., & Popescu, G. (2018). Aspects of gut microbiota and immune system interactions in infectious diseases, immunopathology, and cancer. *Frontiers in Immunology*, 9, 1830. <https://doi.org/10.3389/fimmu.2018.01830>
 20. 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., & Gonzalez, A. (2004). The metacommunity concept: A framework for multi-scale community ecology. *Ecology Letters*, 7(7), 601–613. <https://doi.org/10.1111/j.1461-0248.2004.00608.x>
 21. Loftus, M., Hassouneh, S. A.-D., & Yooseph, S. (2021). Bacterial associations in the healthy human gut microbiome across populations. *Scientific Reports*, 11(1), 2828. <https://doi.org/10.1038/s41598-021-82449-0>
 22. Manor, O., Levy, R., Popejoy, A. B., & McMurry, R. (2020). Health and disease markers correlate with gut microbiome composition across thousands of people. *Nature Communications*, 11(1), 5206. <https://doi.org/10.1038/s41467-020-18871-1>
 23. Marchesi, J. R., & Ravel, J. (2015). The vocabulary of microbiome research: A proposal. *Microbiome*, 3(1), 31. <https://doi.org/10.1186/s40168-015-0094-5>
 24. Miller, E. T., Svanbäck, R., & Bohannan, B. J. M. (2018). Microbiomes as metacommunities: Understanding host-associated microbes through metacommunity ecology. *Trends in Ecology & Evolution*, 33(12), 926–935. <https://doi.org/10.1016/j.tree.2018.09.002>

25. Mobeen, F., Sharma, V., & Prakash, T. (2018). Enterotype variations of the healthy human gut microbiome in different geographical regions. *Bioinformatics*, *14*(9), 560–573. <https://doi.org/10.6026/97320630014560>
26. Pérez-Losada, M., Arenas, M., Galán, J. C., Palero, F., & González-Candelas, F. (2015). Recombination in viruses: Mechanisms, methods of study, and evolutionary consequences. *Infection, Genetics and Evolution*, *30*, 296–307. <https://doi.org/10.1016/j.meegid.2014.12.022>
27. 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>
28. 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), 14. <https://doi.org/10.3390/microorganisms7010014>
29. Scanlan, P. D. (2019). Microbial evolution and ecological opportunity in the gut environment. *Proceedings of the Royal Society B: Biological Sciences*, *286*(1915), 20191964. <https://doi.org/10.1098/rspb.2019.1964>
30. Scepanovic, P., Hübenthal, M., Lauc, G., & Kreplin, D. (2019). A comprehensive assessment of demographic, environmental, and host genetic associations with gut microbiome diversity in healthy individuals. *Microbiome*, *7*(1), 130. <https://doi.org/10.1186/s40168-019-0747-x>
31. Segal, L. N., & Blaser, M. J. (2014). A brave new world: The lung microbiota in an era of change. *Annals of the American Thoracic Society*, *11*(Supplement 1), S21–S27. <https://doi.org/10.1513/AnnalsATS.201306-189MG>
32. 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>
33. Shreiner, A. B., Kao, J. Y., & Young, V. B. (2015). The gut microbiome in health and in disease. *Current Opinion in Gastroenterology*, *31*(1), 69–75. <https://doi.org/10.1097/MOG.0000000000000139>
34. Simon-Loriere, E., & Holmes, E. C. (2011). Why do RNA viruses recombine? *Nature Reviews Microbiology*, *9*(8), 617–626. <https://doi.org/10.1038/nrmicro2614>
35. Singh, B. K., Bardgett, R. D., Smith, P., & Reay, D. S. (2010). Microorganisms and climate change: Terrestrial feedbacks and mitigation options. *Nature Reviews Microbiology*, *8*(11), 779–790. <https://doi.org/10.1038/nrmicro2439>
36. Töpfer, A., Zagordi, O., Prabhakaran, S., Roth, V., Halperin, E., & Beerenwinkel, N. (2013). Probabilistic inference of viral quasispecies subject to recombination. *Journal of Computational Biology*, *20*(2), 113–123. <https://doi.org/10.1089/cmb.2012.0232>
37. Thursby, E., & Juge, N. (2017). Introduction to the human gut microbiota. *Biochemical Journal*, *474*(11), 1823–1836. <https://doi.org/10.1042/BCJ20160510>
38. Trosvik, P., & de Muinck, E. J. (2015). Ecology of bacteria in the human gastrointestinal tract—Identification of keystone and foundation taxa. *Microbiome*, *3*(1), 44. <https://doi.org/10.1186/s40168-015-0107-4>
39. Walls, A. C., Park, Y.-J., Tortorici, M. A., Wall, A., McGuire, A. T., & Velesler, D. (2020). Structure, function, and antigenicity of the SARS-CoV-2 spike glycoprotein. *Cell*, *181*(2), 281–292.e6. <https://doi.org/10.1016/j.cell.2020.02.058>
40. Yeoh, Y. K., Zuo, T., Lui, G. C.-Y., Zhang, F., Liu, Q., Li, A. Y., Chung, A. C.-K., Cheung, C. P., Tso, E. Y.-K., Fung, K. S. C., & Chan, F. K. L. (2021). Gut microbiota composition reflects disease severity and dysfunctional immune responses in patients with COVID-19. *Gut*, *70*(4), 698–706. <https://doi.org/10.1136/gutjnl-2020-323020>
41. Zheng, D., Liwinski, T., & Elinav, E. (2020). Interaction between microbiota and immunity in health and disease. *Cell Research*, *30*(6), 492–506. <https://doi.org/10.1038/s41422-020-0332-7>
42. Zhou, P., Yang, X.-L., Wang, X.-G., Hu, B., Zhang, L., Zhang, W., Si, H.-R., Zhu, Y., Li, B., Huang, C.-L., Chen, H.-D., Chen, J., Luo, Y., Guo, H., Jiang, R.-D., Liu, M.-Q., Chen, Y., Shen, X.-R., Wang, X., ... Shi, Z.-L. (2020). A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature*, *579*(7798), 270–273. <https://doi.org/10.1038/s41586-020-2012-7>