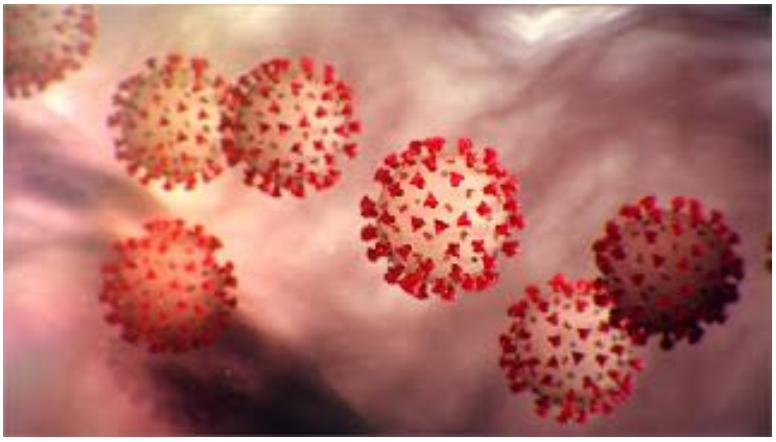


COVID-19: Mutations and Infectivity

ABSTRACT
COVID-19 is a single stranded positive mRNA virus. Such viruses are subject to significant mutations. We examine this virus and its mutability and discuss the risks such a highly mutable virus presents despite the implementation of a global vaccine provisioning.

TERRENCE MCGARTY

TGL 183, JANUARY 2021

Notice

This document represents the personal opinion of the author and is not meant to be in any way the offering of medical advice or otherwise. It represents solely an analysis by the author of certain data which is generally available. The author furthermore makes no representations that the data available in the referenced papers is free from error. The Author also does not represent in any manner or fashion that the documents and information contained herein can be used other than for expressing the opinions of the Author. Any use made and actions resulting directly or otherwise from any of the documents, information, analyses, or data or otherwise is the sole responsibility of the user and The Author expressly takes no liability for any direct or indirect losses, harm, damage or otherwise resulting from the use or reliance upon any of the Author's opinions as herein expressed. There is no representation by The Author, express or otherwise, that the materials contained herein are investment advice, business advice, legal advice, medical advice or in any way should be relied upon by anyone for any purpose. The Author does not provide any financial, investment, medical, legal or similar advice in this document or in its publications on any related Internet sites.

Furthermore, this document contains references to and quotes from papers and documents under the premise of Fair Use in order to present ideas and understandings in context. The Author has attempted to make any and all references to such material separate from those of the author per se and has referenced the source expressly in all cases. These documents are for the dissemination of ideas and have no commercial intent. The Author would appreciate any communications relating to these documents and these should be sent to:

tmcgarty@telmarc.com.

Contents

1 INTRODUCTION

The current COVID-19 virus is an elegantly engineered viral structure. We defer as to how it was engineered but one thing which makes it extremely challenging is the fact that single stranded mRNA viruses often have a high mutation rate, about one nucleotide every replication. Now given the multiple replications of the virus in every patient and the aggressive spreading of the virus, tens of thousands of mutations can occur in a short period of time[1].

COVID-19 became an official pandemic when announced in early 2020 in January in Nature and then in NEJM. As Wu et al noted January 28, 2020 in Nature:

Emerging infectious diseases, such as severe acute respiratory syndrome (SARS) and Zika virus disease, present a major threat to public health. Despite intense research efforts, how, when and where new diseases appear are still a source of considerable uncertainty. A severe respiratory disease was recently reported in Wuhan, Hubei province, China.

As of 25 January 2020, at least 1,975 cases had been reported since the first patient was hospitalized on 12 December 2019. Epidemiological investigations have suggested that the outbreak was associated with a seafood market in Wuhan.

Here we study a single patient who was a worker at the market and who was admitted to the Central Hospital of Wuhan on ***26 December 2019*** *while experiencing a severe respiratory syndrome that included fever, dizziness and a cough. Metagenomic RNA sequencing4 of a sample of bronchoalveolar lavage fluid from the patient identified a new RNA virus strain from the family Coronaviridae, which is designated here 'WH-Human 1' coronavirus (and has also been referred to as '2019-nCoV').*

Phylogenetic analysis of the complete viral genome (29,903 nucleotides) revealed that the virus was most closely related (89.1% nucleotide similarity) to a group of SARS-like coronaviruses (genus Betacoronavirus, subgenus Sarbecovirus) that had previously been found in bats in China.

This outbreak highlights the ongoing ability of viral spill-over from animals to cause severe disease in humans.

The above is the first official notice and one must read this well-crafted paragraph carefully[2]. It was received 7 January 2020 and yet the first patient was identified 12 December, a mere three weeks earlier. Isolation of the infectious element, the mRNA virus, and its complete sequencing is alleged to have been accomplished in a mere three weeks. Assertion of the source also accomplished in the same short period is open to interpretation as well.

[1] https://www.sciencemag.org/news/2021/01/viral-mutations-may-cause-another-very-very-bad-covid-19-wave-scientists-warn There is now a clear consensus that mutations are a significant issue.

[2] One should also consider the Supplementary Information as part of the publication. The detail is considerable considering the short interval to the identification of the initial infection.

One may ask what is the main driver for examining the mutants. We believe that the somewhat unique structure of COVID-19, a corona virus, and its spike protein makes it readily attackable by an mRNA vaccine. Secondly the alleged target of the ACE2 receptor matches spike with entry facilitator.

If we have spike mutations, we may then ask:

1. Does a mutated spike result in a virion deprived of an entry point?
2. Does the mutated spike result in more effective entry to the cell thus making the infectiveness greater?
3. Does a mutated spike find, via mutation, an alternative entry point? If so then does this make it more or less infective?
4. Does a mutated spike result in ineffective mRNA vaccines? Namely will this become like a typical influenza where we target multiple mRNAs on some periodic basis?
5. If there is no herd immunity, which we suspect due to mutations and the ability of the virus to evade the immune system in the nasopharynx, then are we now looking at an ongoing process of reinfection with ongoing mutant strains?
6. Is there some optimal suppression strategy, if so, what could it be?

In a recent study by Rambaut et al the authors note:

Recently a distinct phylogenetic cluster (named lineage B.1.1.7) was detected within the COG-UK surveillance dataset. This cluster has been growing rapidly over the past 4 weeks and since been observed in other UK locations, indicating further spread. Several aspects of this cluster are noteworthy for epidemiological and biological reasons and we report preliminary findings below. In summary:

The B.1.1.7 lineage accounts for an increasing proportion of cases in parts of England. The number of B.1.1.7 cases, and the number of regions reporting B.1.1.7 infections, are growing. B.1.1.7 has an unusually large number of genetic changes, particularly in the spike protein.

Three of these mutations have potential biological effects that have been described previously to varying extents:

i. *Mutation N501Y is one of six key contact residues within the receptor-binding domain (RBD) and has been identified as increasing binding affinity to human and murine ACE2.*
ii. *The spike deletion 69-70del has been described in the context of evasion to the human immune response but has also occurred a number of times in association with other RBD changes.*
iii. *Mutation P681H is immediately adjacent to the furin cleavage site, a known location of biological significance.*

The rapid growth of this lineage indicates the need for enhanced genomic and epidemiological surveillance worldwide and laboratory investigations of antigenicity and infectivity.

We will use this recent mutation and several others as a baseline for examining mutations in general. Finally, we can ask; why are COVID mutations so important? Simply, since this virus is so pandemic in its reach, the fact that so many people are infected means we have so many sources for mutation. It is thus critical to have universal immunization, not just so people are not ill, but more importantly so people do not become reservoirs for massive mutation!

The objectives of this Report are as follows:

1. To understand the dynamics of viral mutations especially as regards to the current corona virus

2. To understand the extent of the mutations and the possible negative impacts from mor aggressive forms of the virus.

3. To establish a basis for the dynamics of vaccinations as an ongoing process but one that must be addressed as quickly and efficiently as possible

4. To recognize the arguable long-term threat of mutating corona virus and the public health need for tracking

5. To recognize the initial targeting of transmissibility and virulence based on ACE2 receptor effectiveness in European, Middle Eastern, Southwest Asian, and African populations as compared to mainland Asian populations[3].

Our basic conclusion from this analysis is as follows:

THE COVID-19 VIRUS HAS SUBSTANTIAL MUTATIONS THAT RESULT IN MORE TRANSMISSIBLE AND MORE AGGRESSIVE DISEASE STATES.

DELAYS IN VACCINATING, ESPECIALLY THOSE WITH SUPPRESSED OR LOWERED IMMUNE SYSTEMS, SUCH AS CANCER PATIENTS AND THOSE OVER 75 YEARS OF AGE, WILL UNLIKELY CREATE AN EVER MORE DEADLY POOL OF VIRION VARIANTS AND THUS PRESENTS A CLEAR AND PRESENT DANGER.

THE IMPLEMENTATION OF THE VACCINE PLANS HAVE BEEN LEFT IN THE HANDS OF THE STATES WHICH THUS FAR, IN GENERAL, HAVE DEMONSTRATED AN INABILITY TO ADDRESS THE HIGH MUTATION POTENTIAL.

FAILURE TO REMEDY THIS EXECUTION QUAGMIRE WILL SURELY RESULT IN EXPLOSIVE MORBIDITY AND MORTALITY.

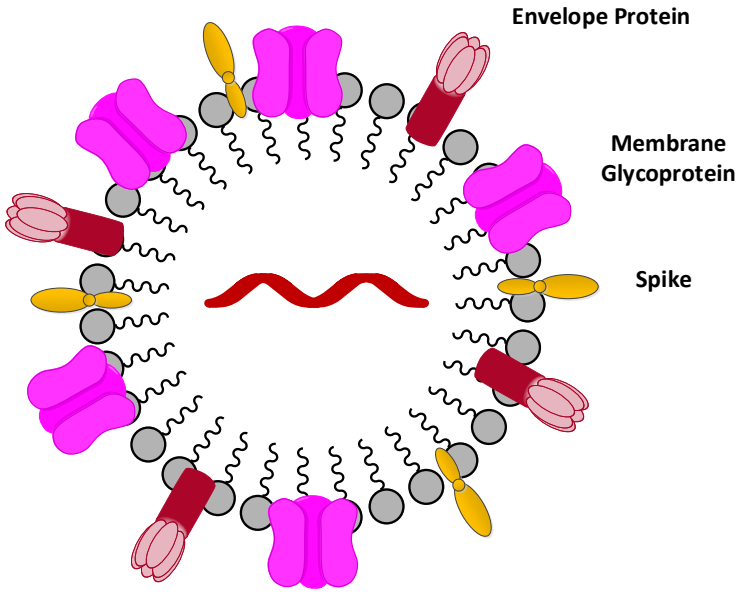
[3] This observation, referenced herein, may explain why China has such a small rate of morbidity and mortality while the US and Europe are so high. It does further beg the question of origin of the virus.

2 COVID-19 STRUCTURE

We present a brief summary of the COVID-19 corona virus which allegedly originated in some manner from Wuhan, China[4]. The actual provenance of the virus is yet to be adequately determined.

2.1 OVERVIEW

The figure below is an graphic example of the corona virus in generic terms. It is a single stranded positive RNA virus with a well-defined spike protein on the surface. The spike protein is a putative target for immunization efforts.



We now consider some detail as regards to this virus. As Artika et al note:

The schematic diagram of coronavirus life cycle.

*The coronavirus infection is initiated by the **binding of the virus particles to the cellular receptors** leading to viral entry followed by the viral and host cellular membrane fusion.*

*After the membrane fusion event, the **viral RNA is uncoated in the host cells** cytoplasm.*

[4] In March 2020 we prepared a preliminary report on the virus discussing what was known to that data and examining the possible pandemic dynamics. In February 4, 2020 we declared this a pandemic despite the WHO and CDC delaying any prophylactic actions, the delay thus resulting in the current global pandemic, https://www.telmarc.com/Documents/White%20Papers/173Corona.pdf

*The **ORF1a and ORF1ab** are translated to produce pp1a and pp1ab, which are subsequently processed by the **proteases encoded by ORF1a** to produce 16 non-structural proteins (nsps) which form the **RNA replicase–transcriptase complex (RTC)**.*

This complex localizes to modified intracellular membranes which are derived from the rough endoplasmic reticulum (ER) in the perinuclear region, and it drives the generation of negative-sense RNAs ((–) RNAs) through both replication and transcription.

*During replication, the **full-length (–)RNA copies of the genome are synthezied** and used as templates for the production of **full-length (þ)RNA genomes**.*

During transcription, a subset of 7–9 subgenomic RNAs, including those encoding all structural proteins, is produced through discontinuous transcription. In this process, subgenomic (–)RNAs are synthesized by combining varying lengths of the 30end of the genome with the 50 leader sequence necessary for translation.

These subgenomic (–)RNAs are then transcribed into subgenomic (þ)mRNAs.

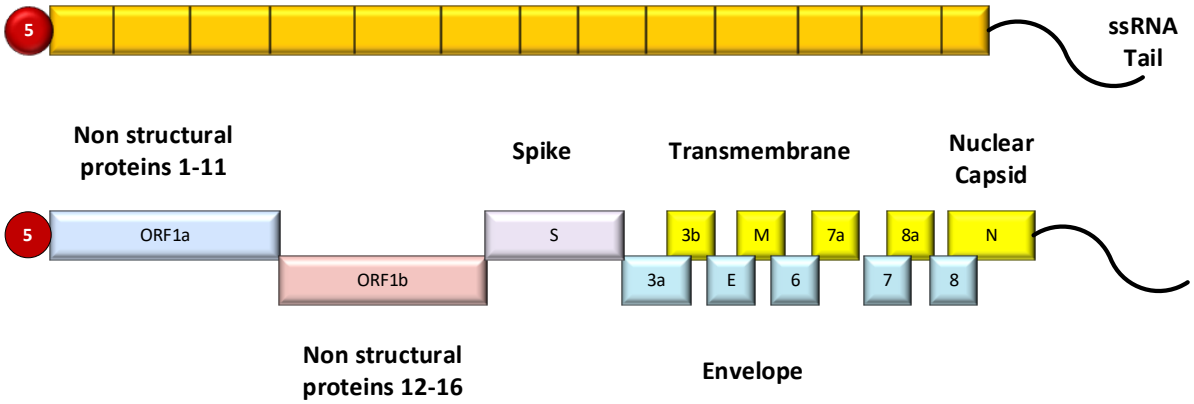
*The **subgenomic mRNAs are then translated**.*

The generated structural proteins are assembled into the ribonucleocapsid and viral envelope at the ER–Golgi intermediate compartment (ERGIC),

followed by release of the newly produced coronavirus particle from the infected cell

2.2 COVID GENE

Various authors have discussed the general structure of a corona virus gene structure and we present this below. It is a positive single stranded mRNA virus and the mRNA has a form as shown below. It is approximately 30,000 nucleotides in length and the spike protein is approximately 3,000 nucleotides in length. As with most corona viruses it contains in the mRNA, via the non-structural proteins, the ability to reconstruct itself many times over by generating the structural genes and implanting a new copy of the mRNA.



The nonstructural proteins, NSP, for the replicase transcriptase complex. There are four structure proteins:

1. S is the spike forming protein, of which we shall speak of later in detail

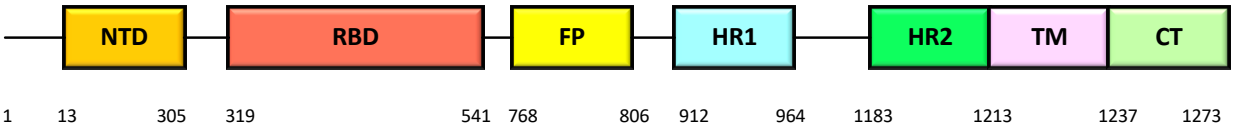
2. E is the envelope protein for the new virion

3. M is the membrane protein for the new virion

4. N is the nucleocapsid protein for the new virion

Overall, we now have the two sets; those allowing for self-reproduction and those relating to construction of the new virion.

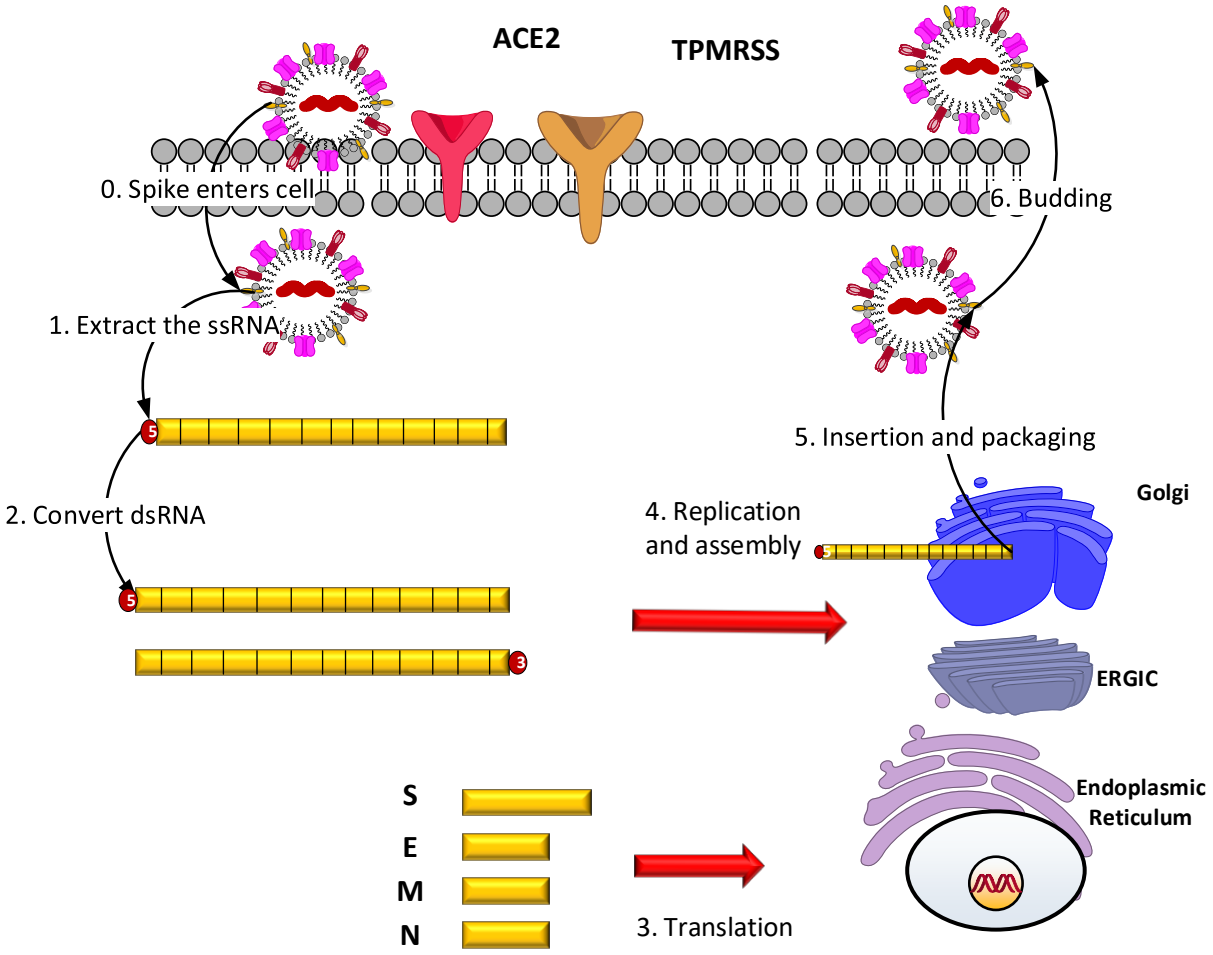
The following is the detail of the spike gene which we shall discuss later.



When examining these proteins we must note that the numbers above are those of the nucleic acids whereas the nucleotides are three times the number. Thus the 1273 nucleic acids represent almost 3900 nucleotides. Mutations in any of these nucleotides can result in a mutation of the nucleic acids and thus the protein conformation.

2.3 COVID SYSTEM

We depict the process below for the entry and reproduction of the virion. This process starts with the attachment and entry of the virion and ultimately the release of a collection of new virions until the cell is depleted and dies.



We can summarize the steps as shown above.

1. The virion spike protein attaches to the ACE2 receptor protein. As we will note later this means that there is a selective capability in this specific spike protein. As we shall note from the literature later, strangely this receptor is weak in Chinese individuals and is a strong bond in Europeans and most other ethnicities.
2. The virion then enters the cell and disassembles. The ss mRNA is extracted
3. The structure RNA sequences are then translated in the inter ER/Golgi space into the constituent proteins.
4. The ss mRNA is completed to form a complete ds mRNA which will be used to generate multiple ss mRNAs to be inserted in other new virions. This may be a point for possible mutant changes.
5. The multiple ss mRNAs and the structural proteins are assembles in the Golgi apparatus and extracted.
6. The final result is a repackaged virion which is budded outwards.

These steps are very general and there are as of yet many holes as to exactly how all this is accomplished. Yet for our purposes the processes lay out the locations of the possible mutation sites.

3 REPLICASE

One of the key and essential elements of a corona virus system is that of the non-structural proteins, the replicases. In corona viruses there are a collection of genes, in the mRNA, and their converted proteins, which result in the proteins that facilitate replication of the structure of the virion. The mRNA has structural elements that yield all of the elements including the spike protein but the background proteins including the replicase allow for the reproduction of these structural elements. Any breakdown in these genes would result in a dean end for the virus, it could not reproduce.

We examine some of the features of replicase herein. More details are in Neuman et al as well as Howley and Knipe. Our intent is not to detail replicase but to showcase its critical importance. It is not clear how much a mutation would result in with the non-structural proteins. However, it is clear that replicase proteins can greatly impact the reproduction of spike and other identifiable structural elements.

3.1 OVERVIEW

We begin with the results presented in a summary paper a few years back. As Ziebuhr notes:

Coronavirus genome replication and transcription take place at cytoplasmic membranes and involve coordinated processes of both continuous and discontinuous RNA synthesis that are mediated by the viral replicase, a huge protein complex encoded by the 20-kb replicase gene. The replicase complex is believed to be comprised of up to 16 viral subunits and a number of cellular proteins.

Besides RNA-dependent RNA polymerase, RNA helicase, and protease activities, which are common to RNA viruses, the coronavirus replicase was recently predicted to employ a variety of RNA processing enzymes that are not (or extremely rarely) found in other RNA viruses and include putative sequence-specific endoribonuclease, 30-to-50 exoribonuclease, 20-O-ribose methyltransferase, ADP ribose 100-phosphatase and, in a subset of group 2 coronaviruses, cyclic phosphodiesterase activities. …

(1) the organization of the coronavirus replicase gene, (2) the proteolytic processing of the replicase by viral proteases, (3) the available functional and structural information on individual subunits of the replicase, such as proteases, RNA helicase, and the RNA-dependent RNA polymerase, and (4) the subcellular localization of coronavirus proteins involved in RNA synthesis.

Although many molecular details of the coronavirus life cycle remain to be investigated, the available information suggests that these viruses and their distant nidovirus relatives employ a unique collection of enzymatic activities and other protein functions to synthesize a set of 50-leader-containing subgenomic mRNAs and to replicate the largest RNA virus genomes currently known.

The uniqueness of the corona virus has significant merit. It makes the virus capable of more aggressive expansion and as such becomes a driver for the mutations we are discussing. Note also from the same author:

***ERGIC**: One of the unique features of coronaviruses is the source of their membrane envelope. Differ from the other well-known enveloped viruses, coronaviruses bud into the **endoplasmic reticulum-Golgi intermediate compartment** (ERGIC), from where they obtain their **membrane envelope**. Therefore, it is not surprising to find that most of the E protein is localized to the ERGIC and Golgi complex where the E protein plays roles in the assembly, budding and trafficking of the nascent virus particle*

The ERGIC intervention is a means by which the structural components can be derived extra the nucleus and directly via the endoplasmic reticulum and Golgi apparatus, namely the path through the ERGIC. From Artika et al we have the following observations:

The replication of the coronavirus genome is viewed as the most fundamental aspect of the coronavirus biology. As the largest group of RNA virus, coronaviruses require an RNA synthesis machinery with the fidelity to faithfully replicate their RNA. Coronavirus replication is achieved by employing complex mechanisms involving various proteins encoded by both viral and host cell genomes. Evolutionary, the virus genome contains relatively constant replicative genes which are indispensable for viral replication. Despite undergoing high mutation rates, RNA viral genomes still encode proteins with arrays of conserved sequence motifs playing roles in facilitating their genome replication and expression.

***Such proteins include the RNA-dependent RNA polymerase (RdRp), RNA helicase, chymotrypsin-like proteases, papain-like proteases, and metal binding proteins**.*

*In coronavirus genomes, all of the genes encoding these proteins are located in the ORF1 strategically located at the 5-most end of the genome. **In addition, viruses also exploit cellular proteins for multiple purposes in their replication cycle, including the attachment and entry into the cells, the initiation and regulation of RNA replication and transcription, protein synthesis, and the assembly of progeny virions.***

The fact that viruses can take over the basic elements of the cell for their own purpose leads to rapid reproduction but then ultimately the demise of the cell via starvation. They continue:

For these purposes, viruses typically subvert the normal components of cellular RNA processing and translational machinery to play both integral and regulatory roles in the replication, transcription, and translation of the viral genomes …

The majority of viruses spend their entire life cycle in the cytoplasm of the host cells and have no access to the host polymerases.

Namely the virus can be totally self-sustaining in terms of its reproductive capability.

Therefore, viruses have to encode polymerases essential for their own transcription and replication. For RNA viruses, the RdRp is the most conserved viral domain and is the most fundamental component of the viral replicase machinery. The RdRp domain of coronaviruses locates in the C-terminal part of nsp12 which catalyzes the replication and transcription of the coronavirus RNA genome. The size of the coronavirus nsp12 is about 930 amino acid residues which is larger than other known viral RdRp's, commonly about 500–600 amino acid residues.

The C-terminal part, which represents about two-thirds of nsp12, has been found to align with the common viral RdRp subunit. Structure analysis of the SARS-CoV nsp12 polymerase showed that the nsp12 polymerase binds to its essential co-factors, nsp7-nsp8 heterodimer, with a second nsp8 subunit occupying a distinct binding site. The presence of nsp7 and nsp8 co-factors significantly increases the RdRp activity. The polymerase domain consists of a fingers domain, a palm domain and a thumb domain. …

The coronavirus-induced replicative structures are mostly in the form of double-membrane vesicles (DMVs) and convoluted membranes (CMs), interconnected with a reticulovesicular network of modified membranes, which seem to be continuous with the endoplasmic reticulum (ER).

The replicase proteins are localized to the DMVs and CMs.

These replicative structures together with their localized proteins, are called the ***replication-transcription complex (RTC)****. The double-stranded RNA (dsRNA) believed to function as replicative intermediate during viral RNA synthesis was detected in the interior of the DMVs. Additional small double-membrane spherule-like structures associated with "zippered" ER membranes were also observed in infectious bronchitis virus (IBV)- infected cells, but not in SARS-CoV, MHV- or MERS-CoV-infected cells*

We argue that the power that a self-contributing replicase provides the corona viruses is a significant factor in their replication. However as we examine mutations, it does seem strange that mutations are reflected in the structural elements such as the spike more often than the non-structural elements such as replicase. One would assume that a mutation in the replicase gene would have the tendency to terminate the virus propagation. At present there does not appear to be significant evidence of this assumption.

3.2 Replicase Gene and Protein

The replicase gene, in the RNA context, and resultant protein, have been studied extensively. Neuman et al provided an excellent summary of the corona replicase gene. They initially note:

The new wealth of structural and functional information revealed that the coronavirus replicase, which is but one biologically successful example of the conserved nidovirus replicative machinery, is not a patchwork amalgam of evolutionary jetsam, but an organized piece of biological machinery where proteins are generally organized into units with related functions.

The first two parts of the replicase, nsp1 and nsp2 are somewhat enigmatic, but appear to work by interfering with host defenses rather than by directly supporting virus replication.

Subunits nsp3–6 contain all the viral factors that are necessary to form viral replicative organelles, as well as two proteinases that are responsible for processing all of the viral replicase proteins.

The small subunits nsp7–11 comprise the viral primer-making activities and provide other essential support for replication.

The final part of the replicase from nsp12–16 contains the remaining RNA-modifying enzymes needed for replication, RNA capping and proofreading.

They further note:

The organization of replicase has a sort of chronological logic to it.

Nsp1–2 help to colonize the host, followed by

Nsp3–6 which lay a foundation to organize and protect the replicative machinery.

This is followed by the primer-making activities of nsp7–11 which also interact with downstream capping and RNA synthesis factors.

Finally, in the proper framework, the RNA-synthesizing enzymes from the C-terminus of the replicase are able to function. While this may be an appealing way to think of the replicase, the reality is probably much more complex. The replicase proteins are all processed from large polyproteins, and therefore are produced at the same time. Because of this, the order in which different proteins are active during the viral replication cycle remains poorly understood. The organization of the replicase also roughly follows a gradient of primary sequence conservation.

Levels of sequence conservation among the different coronaviruses are highest at the 3 end of the replicase gene, and the sequences are very divergent at the 5 end, especially in nsp1–3, which are products of nsp3 PLpro cleavage.

The DMV-making proteins and the primase group of proteins show intermediate levels of conservation with the exception of the well-conserved nsp5Mpro.

What is interesting again is the fact that given the tendency to mutate, the replicate RNA seems to be protected and preserved. As Howley and Knipe note:

The coronavirus genome, which ranges from 25 to 32 kb, is among the largest of those of all RNA viruses, including RNA viruses that have segmented genomes. This exceptional RNA molecule acts in at least three capacities152: as the initial mRNA of the infectious cycle (see Expression of the Replicase–Transcriptase Complex), as the template for RNA replication and transcription, and as the substrate for packaging into progeny viruses.

Consistent with its role as an mRNA, the coronavirus genome has a canonical eukaryotic 5′-terminal cap 1 structure and a 3′ polyadenylate tail.

The genome comprises a basic set of genes in the invariant order 5′-replicase-S-E-M-N-3′, with the huge replicase gene occupying two-thirds of the available coding capacity. The replicase–transcriptase is the only protein translated from the genome; the products of all downstream open reading frames (ORFs) are derived from subgenomic mRNAs.

The 5′-most position of the replicase gene is dictated by the requirement for expression of the replicase to set in motion all subsequent events of infection. The organization of the other basic genes, however, does not seem to reflect any underlying principle, since engineered rearrangement of the downstream gene order is completely tolerated.

Dispersed among the basic genes, there are from one to as many as eight additional ORFs, designated accessory genes, which tend to be specific to various lineages within each genus. These can fall in any of the intergenic intervals downstream of the replicase gene,456 except, curiously, never between the E and M genes. In some cases, an accessory gene can be partially or entirely embedded as an alternate reading frame within another gene, for example, the internal (I) gene of MHV or the 3b gene of SARS-CoV.

As we tend to better understand these virus elements it will be helpful to understand the long-term stability of the replicase segments.

4 SPIKE PROTEIN

We now examine the spike protein. This has become the sole target for many vaccines which is good and bad. It is good because we can really focus on the virion and reduce the load. It is bad because if we have mutations, we have issues. We will reexamine the spike protein herein. In the current environment, this spike protein is a trimer of three identical combinations of proteins. It binds to ACE2 receptor. Now when considering mutations we must consider the mutation in this gene.

4.1 OVERVIEW

We begin with a brief summary of some of the key elements of the spike protein. The work by Huang et al detail a significant amount regarding the spike protein:

With a size of 180–200 kDa, the S protein consists of an extracellular N-terminus, a ***transmembrane (TM) domain*** *anchored in the viral membrane, and a short intracellular C-terminal segment. S normally exists in a metastable, prefusion conformation; once the virus interacts with the host cell, extensive structural rearrangement of the S protein occurs, allowing the virus to fuse with the host cell membrane. The spikes are coated with polysaccharide molecules to camouflage them, evading surveillance of the host immune system during entry. The total length of SARS-CoV-2 S is 1273 aa and consists of a signal peptide (amino acids 1–13) located at the N-terminus, the S1 subunit (14–685 residues), and the S2 subunit (686–1273 residues); the last two regions are responsible for receptor binding and membrane fusion, respectively.*

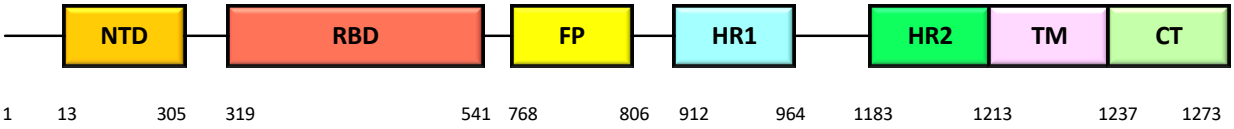
In the S1 subunit, there is an N-terminal domain (14–305 residues) and a receptor-binding domain (RBD, 319–541 residues); the fusion peptide (FP) (788–806 residues), heptapeptide repeat sequence 1 (HR1) (912–984 residues), HR2 (1163–1213 residues), TM domain (1213–1237 residues), and cytoplasm domain (1237–1273 residues) comprise the S2 subunit. S protein trimers visually form a characteristic bulbous, crown-like halo surrounding the viral particle. Based on the structure of coronavirus S protein monomers, the S1 and S2 subunits form the bulbous head and stalk region. The structure of the SARS-CoV-2 trimeric S protein has been determined by cryo-electron microscopy at the atomic level, revealing different conformations of the S RBD domain in opened and closed states and its corresponding functions.

In the native state, the CoV S protein exists as an inactive precursor. During viral infection, target cell proteases activate the S protein by cleaving it into S1 and S2 subunits, which is necessary for activating the membrane fusion domain after viral entry into target cells. Similar to other coronaviruses, the S protein of SARS-CoV-2 is cleaved into S1 and S2 subunits by cellular proteases, and the serine protease TMPRSS2 is used as a protein primer. Although the cleavage site of SARS-CoV is known, that of SARS-CoV-2 S has not yet been reported

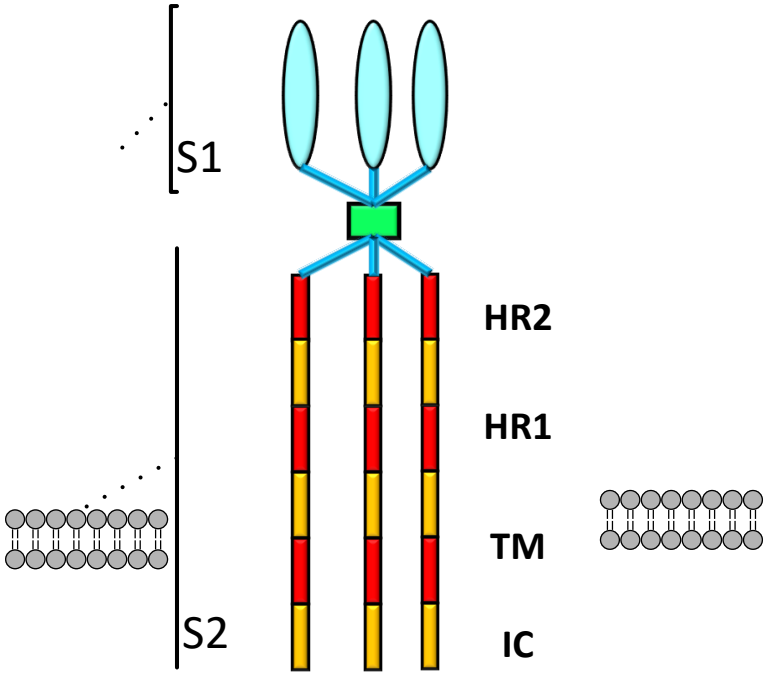
The spike protein is a key element in the invasiveness of this virus. It attaches mostly to ACE2 receptors and appears to be an aggressive attachment. It further participates along with TPMRSS in the entry of the virion.

4.2 Spike Gene and Structure

We can now examine the spike gene. We presented it previously but we now want to examine its elements and their functions in some detail. Recall that the spike "gene" is actually a segment of the mRNA and not a classic DNA construct. Further we recall that the resulting protein is created extra the nucleus and between the ER and Golgi. The structure of spike gene is detailed below:

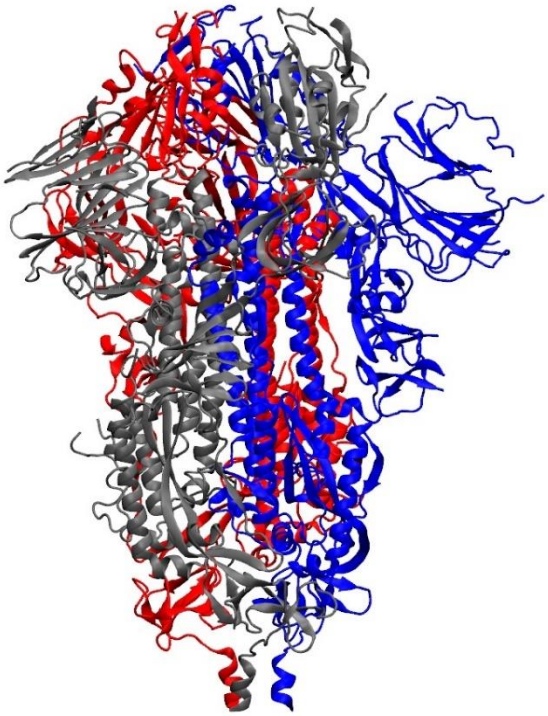


The spike trimer, three identical protein sections, is now shown below:



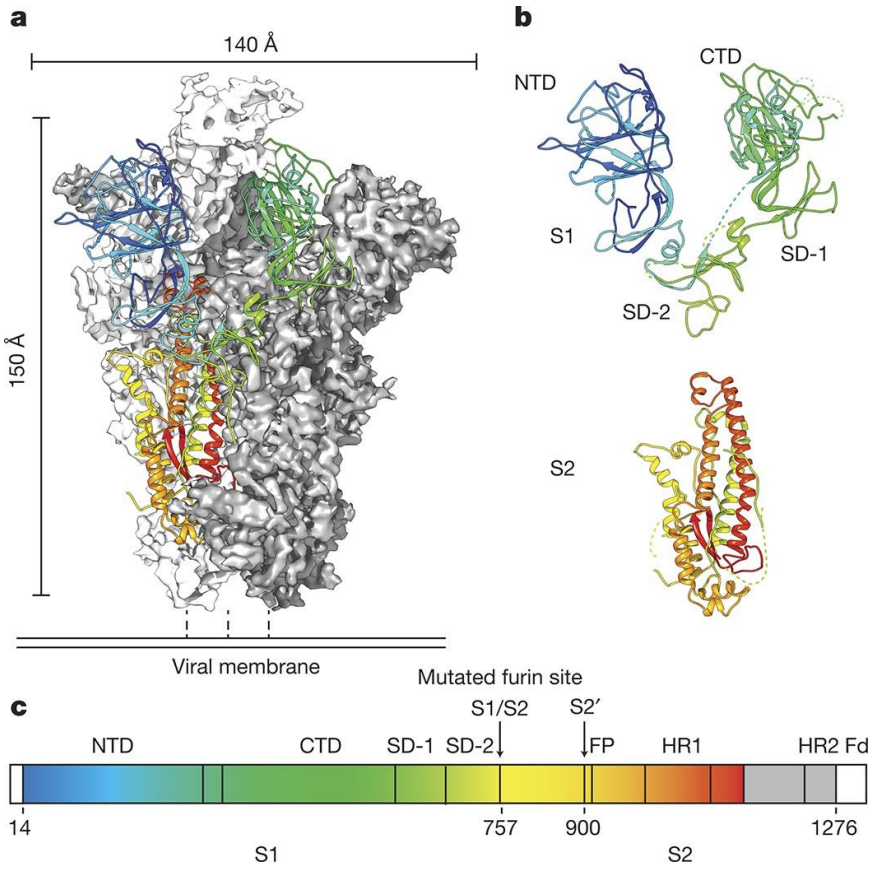
This is a complex protein complex and its binds strongly to the ACE2 receptor. From University of Arkansas, we have in the diagram below a view of this trimer[5]:

[5] https://news.uark.edu/articles/52754/chemist-developing-3d-simulations-of-coronavirus-spike-proteins and https://covid19-hpc-consortium.org/



and from Nature we have below another view of the total and the inner segments[6]:

[6] https://www.nature.com/articles/nature17200/figures/1



The spike appears to be a target for mutations. We examine a recent example in the following section.

4.3 Spike Morphing

Now we shall discuss mutant later but it worth mentioning the B.1.1.7 mutant. This mutant was first recorded in the UK and is considered more infectious. Just how one measures this we believe is problematic. However one may consider that this mutated spike can more aggressively attach and penetrate cells. This mutant is expressed by the following changes:

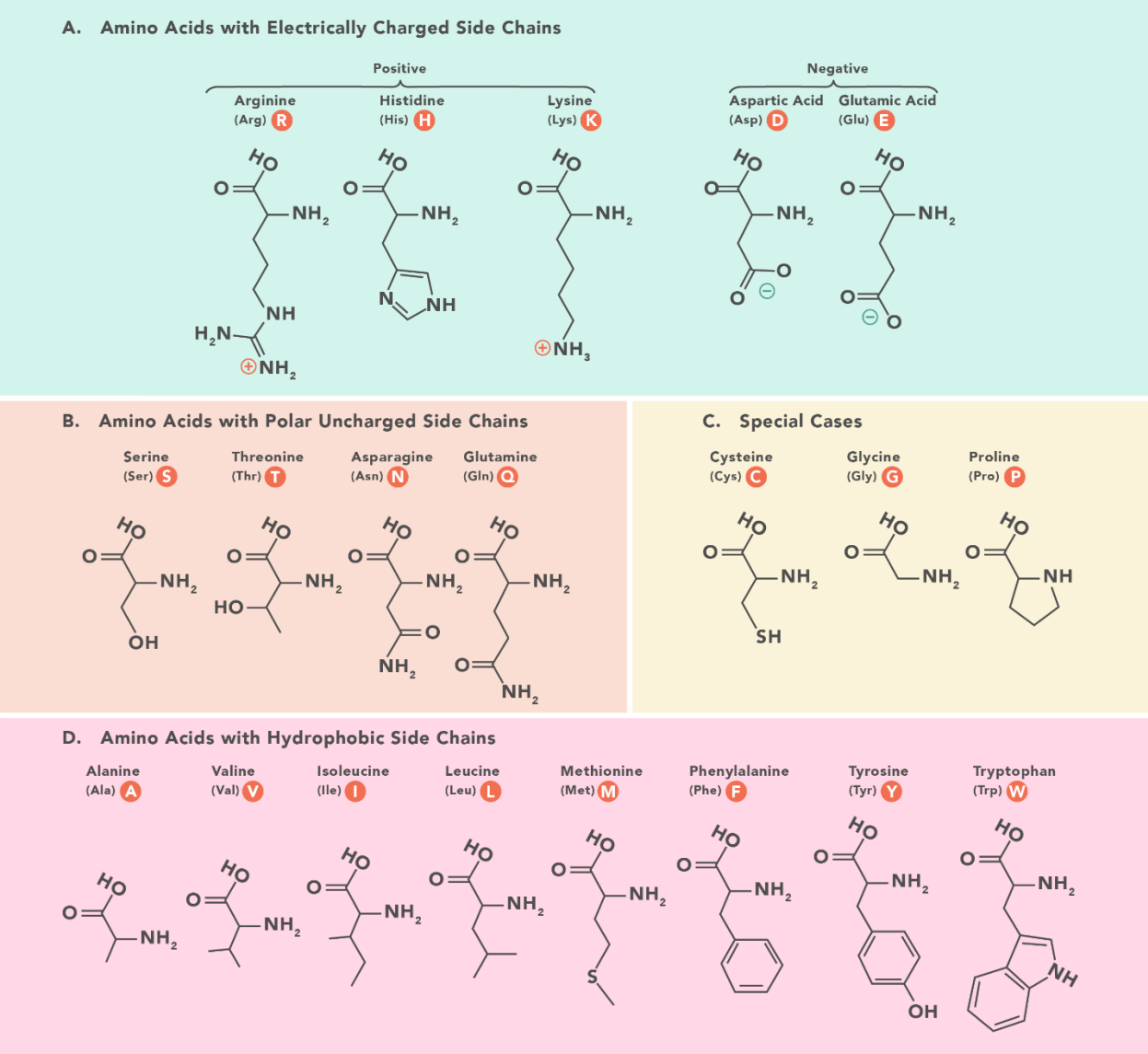
SPIKE	Nucleotide	Amino Acid
	21765-21770 deletion	HV 69-70 deletion
	21991-21993 deletion	Y144 deletion
	A23063T	N501Y
	C23271A	A570D
	C23604A	P681H
	C23709T	T716I
	T24506G	S982A
	G24914C	D1118H

Let us reinterpret this by adding the specific details of each step. In the following we have repeated the above with annotations as to the specifics as to what was changed and the resultant impacts.

SPIKE	Nucleotide	Amino Acid
	21765-21770 deletion of 6 nucleotides, thus 2 amino acids	HV 69-70 deletion Histidine-Tyrosine deletion
	21991-21993 deletion of 3 nucleotides thus one amino acid	Y144 deletion Tyrosine deletion
	A23063T change of A to T at nucleotide location 23063	N501Y, Asparagine to Tyrosine (polar uncharged to hydrophobic)
	C23271A as above change	A570D Arginine to Aspartic Acid (positive to negative side chain)
	C23604A	P681H Proline to Histidine (special to positive charge side chain)
	C23709T	T716I Threonine to Isoleucine (polar uncharged to hydrophobic side chain)
	T24506G	S982A Serine to Alanine (polar uncharged to hydrophobic side chain)
	G24914C	D1118H Aspartic Acid to Histidine (positive side chain to negative)

It should be noted that the nucleic acid changes are not insignificant the self-contributing resultant protein conformation can be substantially altered. This begs the question of how many mutants can be created and yet still target the same receptor. It also begs the question as to how many different receptors can also be targeted.

We show for reference the amino acid code below.



Thus, looking at the above, we should note the dramatic changes in this mutation. All of these are in the Spike. Now the first question should be; how does this change the ACE2 receptor binding. It may actually enhance it and thus be more effective a virus. Second, we may ask what this does to a vaccine of mRNA made to attack the original Spike? Does this now possibly weaken the vaccine while possibly strengthen the virus?

5 VIRAL MUTATIONS

Viral mutations are well known. In a sense they follow some Darwinian model. Namely there is some fitness criterion and the virus evolution sets up, a path for the most adaptive to survival. Now the COVID-19 virus is highly infectious and highly mutable. A simple back of the envelope calculation notes that there are about 30 nucleotides in a ss+mRNA virion. We can assume one nucleotide mutation per replication and about 1 million replications per infected. This is about a million mutations per person. Some of these survive and make it to another patient. The process continues.

Some of the most extreme are the species to species and back again mutations reported by Burkholz et al discuss this in the context of COVID:

A mutation analysis of a collection of SARS-CoV-2 genomes around the world via sequence, date, geographic location, and species has revealed a large number of variants from the initial reference sequence in Wuhan. It also reveals that humans infected with SARS-CoV-2 have infected mink populations in the Netherlands, Denmark, United States, and Canada. In these animals, a small set of mutations often in combination, in the spike protein receptor binding domain (RBD) has apparently transferred back into humans.

The viral genomic mutations in minks observed in the Netherlands and Denmark show the potential for new mutations on the SARS-CoV-2 spike protein RBD to be introduced into humans by zoonotic transfer. Our data suggests that close attention to viral transfer from humans to farm animals and pets will be required to prevent build-up of a viral reservoir for future zoonotic transfer.

We focus herein on the spike protein. What is important to remember is that almost all mutations are inconsequential. They most likely result in a terminal state for the virion, unable to reproduce or unable to find a host cell. Also the counterbalancing state is the few that do reproduce may not get spread. Yet so many are produced that all but one is needed to spread havoc.

5.1 GENERAL FACTORS

First we consider the generic types and causes of mutations. Simply we can look at a corona virus ss mRNA say of 30,000 nucleotides. It is divided into non-structural and structural sectors which generate proteins which are then used to replicate the virus and send it on to other cells. Thus changes in the nonstructural elements may result in a defective generation profile. If however we change the structural proteins we will then obtain a virion with varied external characteristics.

Second, if we have a workable RNA entering a cell, then we could expect that we get the correct non-structural (protein generating and replicase proteins) and structural proteins, such as spikes. This is done via the cells own checks and balances.

Third, however, when we replicate the RNA in the cell itself, a ssRNA replication is prone to mutations of various types. Thus the mRNA placed in the virion and passed on may be what is defective or just a variant.

Now we summarize some mutation processes as presented by Cann:

Process	Functions
Biochemical	Mutations due to the interference from drugs or other biochemical interferents.
Deletions	The removal of genetic material such as a nucleotide by some exogeneous factor such as radiation
Host Range	This may be the result of interference due to other biological entities changing the genetic profile.
Nonsense	These results from the change in the coding sequences due to additions/deletions
Temperature Sensitive	Temperature sensitive mutations result from mis-sense mutations in proteins.
Cold Sensitive	These are the opposite of temperature sensitive mutants
Revertant	These are the result from reverse mutations.
Suppression	These result from the suppression of a gene.

Let us begin with some basic elements of viral mutations. As Elena and Sanjuan note:

As a consequence of the lack of proofreading activity of RNA virus polymerases, new viral genetic variants are constantly created. RNA viruses readily adapt to changing environmental conditions. Therefore, the high mutation rate of RNA viruses compared with DNA organisms is responsible for their enormous adaptive capacity. The above syllogism, with some variation, is deeply rooted in the thinking of many virologists: RNA viruses mutate at the maximum error rate compatible with maintaining the integrity of genetic information (i.e., the error threshold) because this would allow them to quickly find the beneficial mutations needed for adaptation.

It is an unquestionable fact that RNA virus populations exist as swarms of mutant genotypes. Such enormous variability is an unavoidable consequence of the lack of exonuclease proofreading activity of the virus-encoded RNA polymerases with, in some cases, the added contribution of recombination. However, the argument that the more mutations are generated, the faster adaptation proceeds is flawed because it ignores the fact that the vast majority of mutations are deleterious, hence hindering adaptation, as shown by recent theoretical developments. Therefore, the adaptive value of the RNA virus extreme mutation rate has to be

carefully reconsidered, and new alternative explanations, beyond a purely mechanistic level, should be taken into consideration.

The mutation rate itself is a trait that can evolve by natural selection, provided the existence of genetic variation for the character. *Given that most mutations have deleterious fitness effects having a too high mutation rate would be prejudicial in the short term simply because (in a first approach) the population equilibrium fitness for a haploid asexual population decreases exponentially with mutation rate. In the long term, however, it can be argued that the higher the mutation rate, the more likely it is that beneficial mutations will be produced. An optimal mutation rate, which maximizes the rate of adaptation, is reached when these opposing factors are balanced. Hypotheses about the high adaptability of RNA viruses should take into account this trade-off and address why the balance between beneficial and deleterious mutational effects leads to different outcomes in RNA viruses and DNA organisms, including other viruses*

In a similar fashion Sanjuan et al also noted:

Accurate estimates of virus mutation rates are important to understand the evolution of the viruses and to combat them. *However, methods of estimation are varied and often complex. Here, we critically review over 40 original studies and establish criteria to facilitate comparative analyses.* ***The mutation rates of 23 viruses are presented as substitutions per nucleotide per cell infection (s/n/c) and corrected for selection bias where necessary, using a new statistical method. The resulting rates range from 108 to106 s/n/c for DNA viruses and from 106 to 104 s/n/c for RNA viruses.***

These calculations have been done by several authors but we shall consider this approach due to some detail.

Similar to what has been shown previously for DNA viruses, there appears to be a negative correlation between mutation rate and genome size among RNA viruses, but this result requires further experimental testing. Contrary to some suggestions, the mutation rate of retroviruses is not lower than that of other RNA viruses. We also show that nucleotide substitutions are on average four times more common than insertions/deletions (indels). Finally, we provide estimates of the mutation rate per nucleotide per strand copying, which tends to be lower than that per cell infection because some viruses undergo several rounds of copying per cell, particularly double-stranded DNA viruses. A regularly updated virus mutation rate data set will be available at www.uv.es/rsanjuan/virmut

We begin by presenting the definitions of the notations provided by the authors.

Symbol	Definition
$\mu_{s/n/c}$	Mutation rate as substitutions per nucleotide per cell infection
$\mu_{s/n/r}$	Mutation rate as substitutions per nucleotide per strand copying
f, f_s	Observed mutation frequency and observed mutation frequency for substitutions only, respectively
T, T_s	Total no. of mutations observable with a given detection method (mutational target size), subscript s is used to refer to substitutions only
L	Length (in nucleotides) of the genetic region studied
c	No. of cell infection cycles
B	No. of viral progeny particles released per infected cell (burst size, or viral yield)
N_0	No. of viral particles that start an infection (inoculum size)
N_1	No. of viral particles after growth
a	Exponential growth rate
W	Relative fitness as determined from growth rates and/or burst sizes
s	Fitness effect of a mutation, defined as 1W
E(s_v)	Average s value of single random nucleotide substitutions, excluding lethal ones
p_L	Probability that a single random nucleotide substitution is lethal
α	Selection correction factor
P_0	Fraction of cultures showing no mutants in the Luria-Delbruck fluctuation test (null class)
m	Mutation rate to a given phenotype per strand copying obtained from the Luria-Delbruck fluctuation testa
r, r_c	No. of cycles of copying and no. of cycles of copying per cell infection, respectively
δ	Ratio of indels to total mutations (indel fraction)

Some of the key relationships are shown below as used by the authors:

$$c = \frac{\log\left(\frac{N_1}{N_0}\right)}{\log B} = \text{mutation rate per cell infection}$$

and.

$$\mu = \frac{3 f_s}{T_s c \alpha} = \text{substitutions per nucleotide per cell infection (s/n/c)}$$

$$\mu = \frac{f_s}{T_s c \alpha} = \text{mutation per nucleotide per cell infection (i/n/c)}$$

and,

$$P(s) = (1 - p_L)\frac{\lambda e^{-\lambda s}}{1 - e^{-\lambda}} if\ 0 < s < 1$$

$$P(s) = -p_L \quad s = 1$$

$$P(s) = 0 \text{ otherwise}$$

and finally;

$s = 1 - \frac{a_i}{a_0}$ growth rate ratios

Using data collected we obtain:

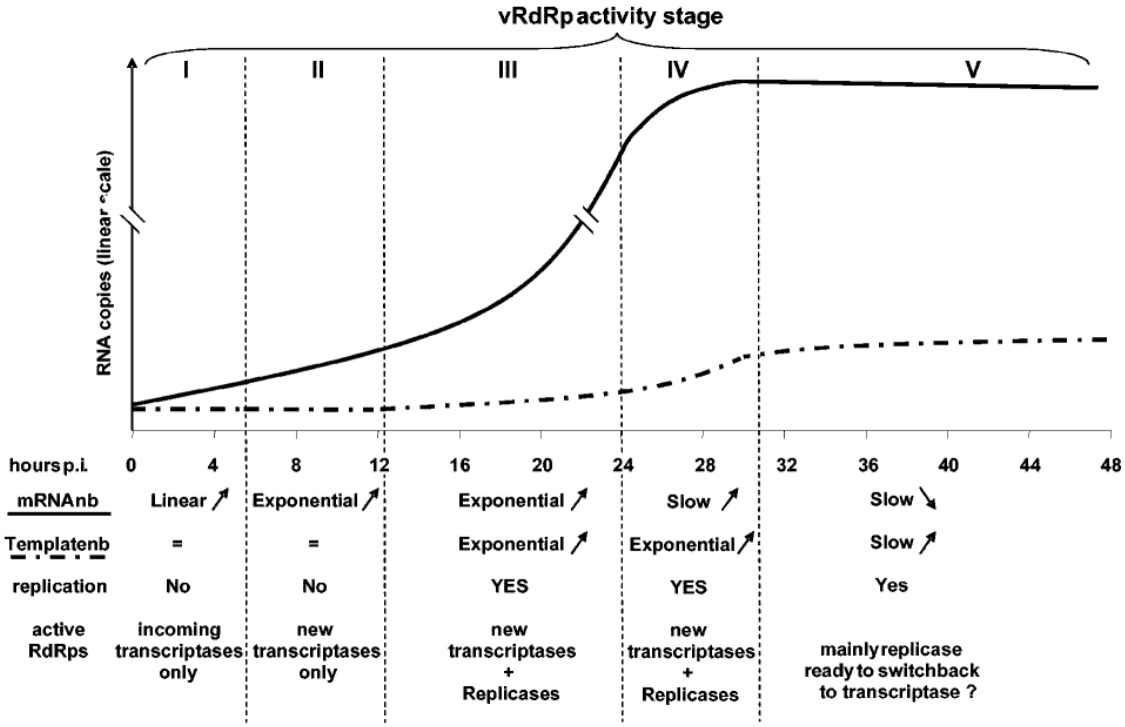
Group	Virus	Genome size (kb)	Mean μ Mutation Rate
ssRNA+	Bacteriophage Qβ	4.22	1.1 X 10^{-3}
	Tobacco mosaic virus (TMV)	6.40	8.7 X 10^{-6}
	Human rhinovirus 14 (HRV-14)	7.13	6.9 X 10^{-5}
	Poliovirus 1 (PV-1)	7.44	9.0 X 10^{-5}
	Tobacco etch virus (TEV)	9.49	1.2 X 10^{-5}
	Hepatitis C virus (HCV)	9.65	1.2 X 10^{-4}
	Murine hepatitis virus (MHV)	31.4	3.5 X 10^{-6}
ssRNA-	Vesicular stomatitis virus (VSV)	11.2	3.5 X 10^{-5}
	Influenza A virus (FLUVA)	13.6	2.3 X 10^{-5}
	Influenza B virus (FLUVB)	14.5	1.7 X 10^{-6}
dsRNA	Bacteriophage Φ6	13.4	1.6 X 10^{-6}
ssDNA	Bacteriophage ΦX174	5.39	1.1 X 10^{-6}
	Bacteriophage M13	6.41	7.9 X 10^{-7}
dsDNA	Bacteriophage X	48.5	5.4 X 10^{-7}
	Herpes simplex virus type 1 (HSV1)	152	5.9 X 10^{-8}
	Bacteriophage *T2*	169	9.8 X 10^{-8}

5.2 Replication Rates

We may ask; how quickly do the viruses replicate? As Plumet et al note:

We propose a reference model of the kinetics of a viral RNA-dependent RNA polymerase (vRdRp) activities and its regulation during infection of eucaryotic cells. After measles virus infects a cell, mRNAs from all genes immediately start to accumulate linearly over the first 5 to 6h and then exponentially until 24 h. The change from a linear to an exponential accumulation correlates with de novo synthesis of vRdRp from the incoming template.

Expression of the virus nucleoprotein (N) prior to infection shifts the balance in favor of replication. Conversely, inhibition of protein synthesis by cycloheximide favors the latter. The in vivo elongation speed of the viral polymerase is 3 nucleotides/sec. A similar profile with fivefold-slower kinetics can be obtained using a recombinant virus expressing a structurally altered polymerase. Finally, virions contain only encapsidated genomic, antigenomic, and 5-end abortive replication fragment RNAs.



The replication as described above is not representative of a single type. It appears to be a complex amalgam of linear and exponential.

6 COVID-19 MUTATIONS

We will now consider COVID mutations. These mutations effect both the structural and non-structural proteins. In turn those proteins get reflected in the viability, infectiousness, and lethality of the virus. Ultimately if death does not occur first, the virus will be attacked by the immune system. As we have remarked and as we will discuss, mutations are occurring now with COVID and as we see a larger number of infected we have a larger source of mutations. Thus understanding some of the recent work in this area is essential.

6.1 TYPES OF MUTATIONS

Not all mutations survive. In fact almost all do not. The mutations often result in incompatible RNA formats and thus non-replicable. Or they may result in ones unfit to use the resources in a cell. However, a small fraction may result in ever more virulent viruses. We first consider some of these classes. As Howley and Knipe note:

Although mutation is the ultimate source of genetic variation, the pace of evolutionary change can be measured in two rather different ways. One method is to estimate, experimentally, the rate at which mutations are generated de novo. Such rates have usually been presented as the number of mutations per nucleotide or per genome, per replication. However, because of the inherent complexities and biases in making these estimates, it has been suggested that estimates of mutation rate per nucleotide, per cell infection may be more informative.

For example, one important complicating factor is that some viruses employ so-called stamping machine replication, in which a single virus acts as the template for all progeny genomes, so that mutations accumulate linearly, while others utilize "geometric" replication, in which some of the early progeny genomes are used as templates to produce further progeny, in turn increasing the rate of mutation accumulation.

The importance of mutation rate estimates is that they reveal the intrinsic error dynamics of the RNA or DNA polymerases used in viral replication and in theory allow a count of each type of mutation—advantageous, neutral, or deleterious—before they have been shaped by natural selection, although it is always difficult to accurately count the number of lethal mutations that are rapidly removed by purifying selection.

A detailed compilation of mutation rate estimates for 27 viruses, and accounting for many of the complexities inherent in analyses of this kind, revealed that these rates vary from averages of

1.6×10^{-6} to 1.5×10^{-4} mutations/nucleotide/cell infection for RNA viruses,

1.6×10^{-5} to 4.4×10^{-3} mutations/nucleotide/cell infection for retrotranscribing viruses, and

5.9×10^{-8} to 1.1×10^{-6} mutations/nucleotide/cell infection for DNA viruses.

For RNA viruses that replicate with a RdRp, an enzyme that lacks a proofreading or repair function, this equates to mutation rates that are usually around one per genome, per replication.

This rate is key. If we assume 30,000 nucleotides per mRNA for COVID and we multiply by say 1.0×10^{-5} we obtain 3 nucleotides mutation rate. This compares to the unit number above. *Similarly, the lower mutation rates observed in large DNA viruses clearly reflect the higher fidelity of the DNA polymerases employed in their replication cycle.*

Of particular note is that mutation rate estimates in ssDNA viruses (a maximum of 1.1×10^{-6}, although only two estimates are available) are higher than those of large dsDNA viruses (which range from 5.9×10^{-8} to 5.4×10^{-7}), even though ssDNA viruses have such small genomes that they use host DNA polymerases for replication. Although it remains to be determined, it is possible that the higher mutation rates in ssDNA viruses reflect less efficient proofreading and excision repair on ssDNA and/or frequent deamination. The study of Sanjuán and Domingo-Calap226 was also of note in that it revealed that retroviruses such as HIV have error rates that often higher than those of RdRp-utilizing viruses, even though earlier studies suggested that RT exhibits higher fidelity than does RdRp[7]…

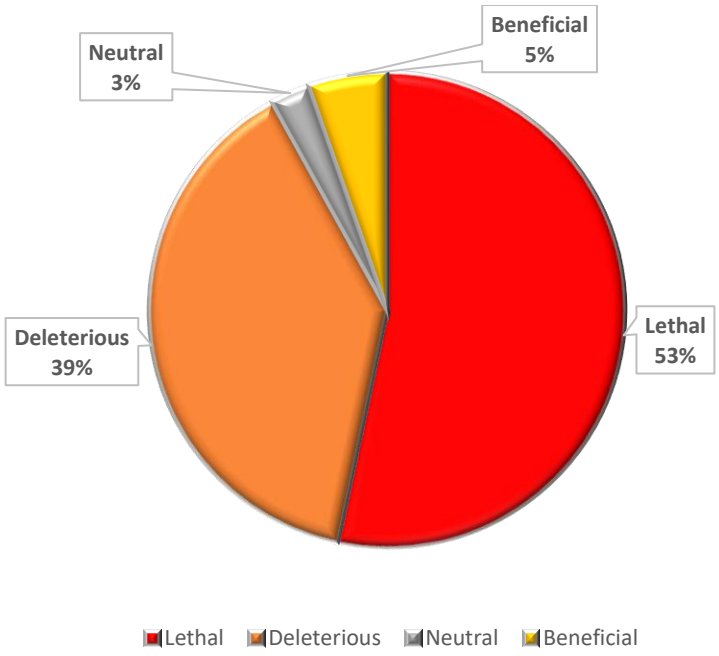
In the case of RNA viruses, and likely small DNA viruses, it is probable that few amino acid mutations are strictly neutral, with most clearly deleterious. In particular, the small genomes of RNA viruses ensure there is extensive pleiotropy, epistasis, and multifunctionality such that there is little evolutionary elbow room.

For example, mutagenesis studies of vesicular stomatitis virus (VSV) revealed that

- ***nearly 40% of random mutations are lethal,***
- ***another 29% deleterious,***
- ***a further 27% neutral,***
- ***and only 4% beneficial.***

We depict this result in the following graphic.

[7] Note the authors define RdRp as: *Although there have been attempts to infer the evolutionary history of RNA viruses based on phylogenetic analyses of the **RNA-dependent RNA polymerase (RdRp)**, the phylogenies in question are highly uncertain at the deepest branches, such as those linking the different orders of ssRNA+ and ssRNA− viruses, where there is often no more sequence similarity than expected by chance alone, greatly compromising these analyses including reliable sequence alignment.*



A similar preponderance of deleterious compared to beneficial mutations was observed in poliovirus. Importantly, however, these estimates only relate to fitness effects in a single cell and a far larger proportion of mutations are expected to be deleterious (or lethal) when considering the entirety of the virus life cycle. The same is also likely to be true of ssDNA viruses, which are also characterized by very small genome sizes. Accordingly, most estimates of the ratio of nonsynonymous (dN) to synonymous (dS) nucleotide substitutions per site (ratio dN/dS, a common measure of selection pressures, with dN/dS = 1 indicative of selective neutrality) in RNA and ssDNA viruses indicate that purifying selection is the most common evolutionary force (i.e., dN/dS < 1)….Finally, in the same way that substitution rates are time dependent in viruses, so are inferences of selection pressures based on estimates of dN/dS. In particular, the presence of transient deleterious mutations will tend to inflate estimates of dN/dS toward the present, such that sequence comparisons should again only utilize sequences sampled over the time spans.

6.2 MUTATIONS

We now consider several basic issues regarding RNA viral mutations. As Smith notes:

Highly mutable, infinitely malleable, and all-powerful: this is often the underlying assumption for how spontaneous mutations fuel RNA virus adaptation. Though essential for adaptation, mutations within RNA virus genomes can exact significant fitness costs.

Without the capacity to detect and repair mismatched or damaged nucleotides, viral RNA genomes are prone to mutations introduced by mechanisms intrinsic and extrinsic to viral replication. However, large population size, complementation, cellular chaperones, and recombination can buffer viral populations against deleterious and lethal mutations. As such,

viral replication is a rapid, tenuous dance between the generation of sufficient genetic diversity on which natural selection can act and the production of less-fit variants.

… this article is not to describe how RNA viruses evolve… rather, it aims to provide an introduction to some of the ***mechanisms by which mutations arise during RNA virus replication****, as viral mutation rates are the ultimate source of genetic diversity. What is a mutation rate? Storage and transmission of genetic information depends upon the correct formation of hydrogen bonds between nucleobases.*

Mechanisms of mutation are critical to be understood. The exact mechanisms have been examined by many authors and we attempt to use this reference as a quick summary. They continue:

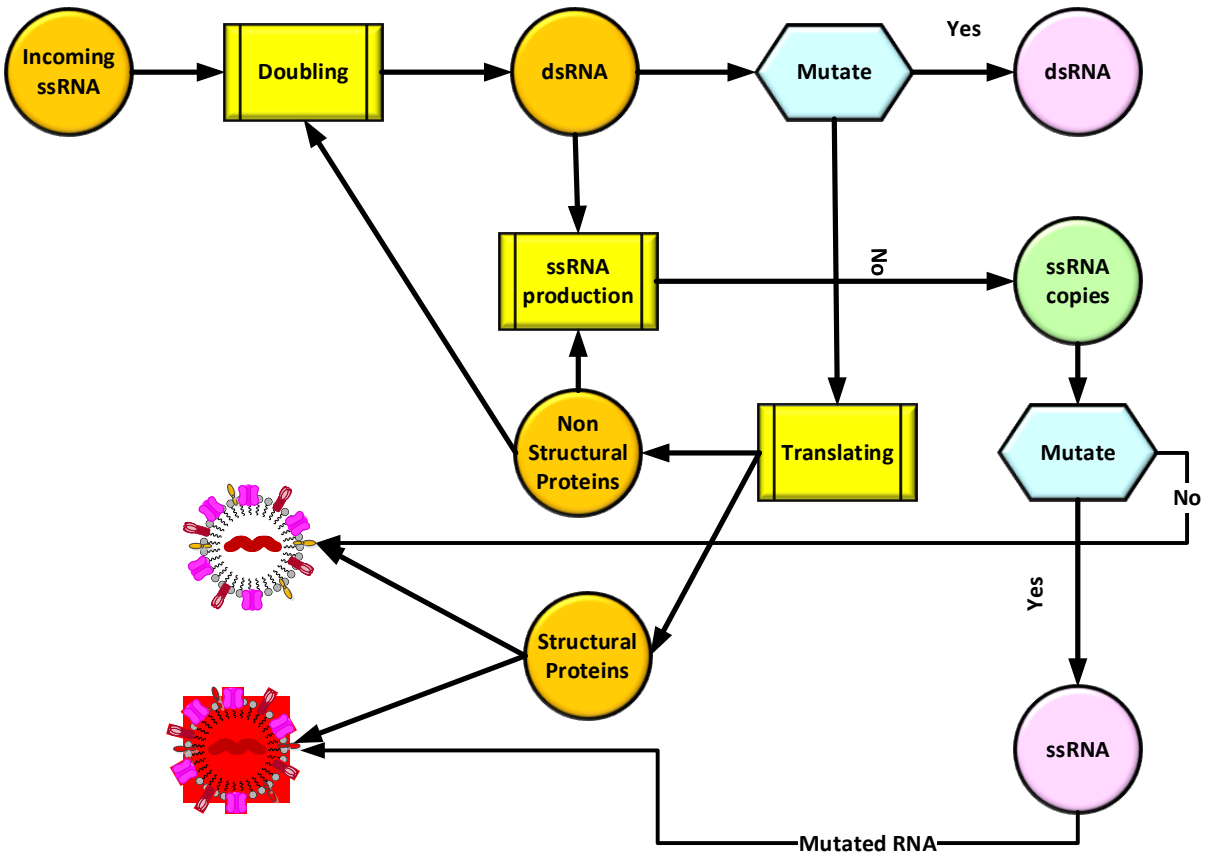
Mutations arise when mismatches are introduced during RNA virus replication or as a result of postreplicative base modification.

The RNA replication step is somewhat unique with corona viruses.

Host RNA-modifying enzymes and nitration or oxidation of nucleobases can alter hydrogen bonding and increase the probability of point mutations during subsequent rounds of replication. A mutation rate describes the rate (not frequency) at which spontaneous mutations arise during a single infection and reflects both cell- and virus-dependent mechanisms. Because most mutations are likely lethal or deleterious, natural selection and genetic drift significantly impact the observed frequency of mutations within a population.

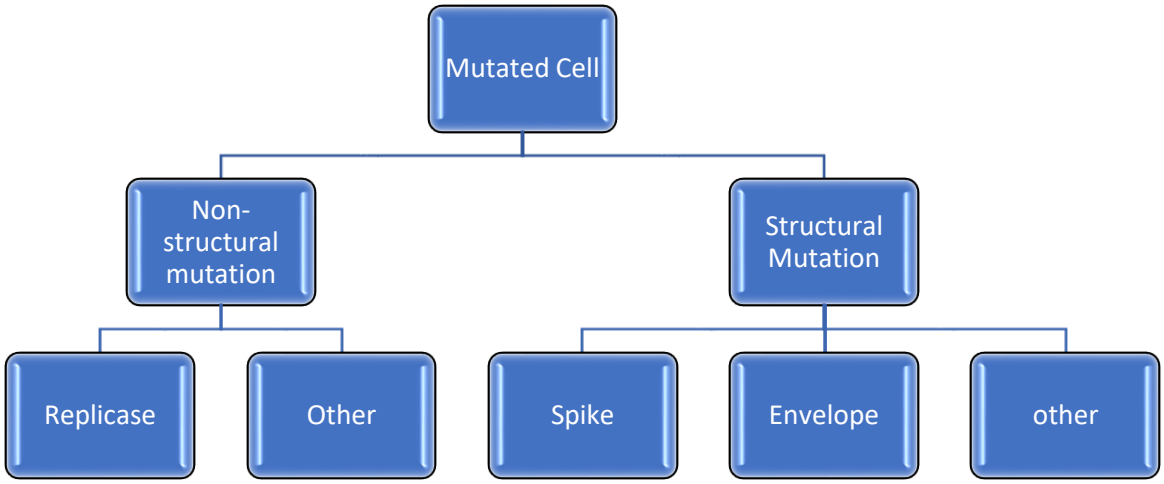
Excluding viroids, RNA viruses replicate with the highest known mutation rates, which are estimated to range between 10^{-6} and 10^{-4} substitutions per nucleotide per cell infection. Such high mutation rates enable viral populations to rapidly generate genetic diversity and thus a multitude of phenotypes on which adaptation by natural selection can occur. However, high mutation rates alone are not sufficient to drive viral adaptation. As mentioned above, many mutations are deleterious and decrease viral fitness . How a given mutation affects viral fitness is also dependent upon other characteristics of RNA virus populations, such as population size, genome size, and genome complexity

We can now consider the production process of the virus replication and see what could happen as shown below:



The above is a simplified model of the process. There can be two outcomes; a perfect copy or a mutated copy. There are in this model two places where mutation can occur; (i) in the dsRNA production process, and (ii) in the copying of the multiple ssRNA. The net result is a mutated virion. In fact the structural proteins would reproduce the same in this case but the ssRNA would be a variant whose result would not be seen until the next generation.

Now we can ask; what happens to the mutated cell? The chart below depicts some examples but it is not complete, just exemplary:



Each of these possibilities, and there are many, have differing effects. For example a spike mutation may cause the virion to not bind, or bind more strongly, or bind to another receptor, or just have no effect at all.

6.3 Causes of Mutations

Let us examine some of the causes of these mutations. As Avanzato et al note:

Long-term severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) shedding was observed from the upper respiratory tract of a female immunocompromised individual with chronic lymphocytic leukemia and acquired hypogammaglobulinemia.

Shedding of infectious SARS-CoV-2 was observed up to 70 days, and of genomic and subgenomic RNA up to 105 days, after initial diagnosis.

The length of this shedding means the virus has an exceptionally long period to reproduce and mutate. We have not found any studies to date where there is a mapping of mRNA mutations in a single patient over this period in the shedding elements. That would potentially have significant value. However as we note most mutations result in defective virions.

The infection was not cleared after the first treatment with convalescent plasma, suggesting a limited effect on SARS-CoV-2 in the upper respiratory tract of this individual. Several weeks after a second convalescent plasma transfusion, SARS-CoV-2 RNA was no longer detected. We observed marked within-host genomic evolution of SARS-CoV-2 with continuous turnover of dominant viral variants. However, replication kinetics in Vero E6 cells and primary human alveolar epithelial tissues were not affected.

Our data indicate that certain immunocompromised individuals may shed infectious virus longer than previously recognized. Detection of subgenomic RNA is recommended in persistently SARS-CoV-2-positive individuals as a proxy for shedding of infectious virus.

This is the issue we see in long-spreading immunized patients. Namely they continue to shed for long periods and thus subject the virus to substantial mutations.

As regards to these mutations Dorp et al note:

Mutations within coronaviruses, and indeed all RNA viruses, can arrive as a result of three processes.

First, mutations arise intrinsically as copying errors during viral replication, a process which may be reduced in SARS-CoV-2 relative to other RNA viruses, due to the fact that coronavirus polymerases include a proof-reading mechanism.

Second, genomic variability might arise as the result of recombination between two viral lineages co-infecting the same host.

Third, mutations can be induced by host RNA-editing systems, which form part of natural host immunity. While population genetics theory states that the majority of mutations are expected to be neutral, some may be advantageous or deleterious to the virus.

Mutations that are highly deleterious, such as those preventing virus host invasion, will be rapidly purged from the population; mutations that are only slightly deleterious may be retained, if only transiently. Conversely, neutral and in particular advantageous mutations can reach higher frequencies. Mutations in SARS-CoV-2 have already been scored as putatively adaptive using a range of population genetics methods, and there have been suggestions that specific mutations are associated with increased transmission and/or virulence.

Early flagging of such adaptive mutations could arguably be useful to control the COVID-19 pandemic. However, distinguishing neutral mutations (whose frequencies have increased through demographic processes) from adaptive mutations (which directly increase the virus' transmission) can be difficult. For this reason, the current most plausible candidate mutations under putative natural selection are those that have emerged repeatedly and independently within the global viral phylogeny. Such homoplasic sites may arise convergently as a result of the virus responding to adaptive pressures.

6.4 MUTATION RATE PRINCIPLES

As Duffy notes regarding rates of mutation we have some basic constructs:

RNA viruses have high mutation rates—up to a million times higher than their hosts—and these high rates are correlated with enhanced virulence and evolvability, traits considered beneficial for viruses. However, their mutation rates are almost disastrously high, and a small increase in mutation rate can cause RNA viruses to go locally extinct. ***Researchers often assume that natural selection has optimized the mutation rate of RNA viruses, but new data shows that, in poliovirus, selection for faster replication is stronger and faster polymerases make more mistakes.***

The fabled mutation rates of RNA viruses appear to be partially a consequence of selection on another trait, not because such a high mutation rate is optimal in and of itself*. Mutations are the building blocks of most of evolution—they are the variation upon which natural selection can act, and they are the cause of much of the novelty we see occur in evolution. However, most mutations are not beneficial for the organisms with them.*

Many mutations cause organisms to leave fewer descendants over time, so the action of natural selection on these mutations is to purge them from the population. While a small percentage of mutations are helpful and some are inconsequential (neutral or nearly neutral in effect), a large portion of mutations are harmful. While the fraction of mutations that are harmful versus beneficial may change in different organisms, in different environments, and over time, deleterious mutations are thought to always outnumber beneficial mutations. That remains true whether an organism has a low mutation rate or a high mutation rate, and biological entities differ dramatically in their per-nucleotide mutation rate

6.5 FITNESS MEASURES

Classic theories of evolution rely upon the principles of fitness. The underlying assumption is that mutations are always occurring and a few a beneficial and most are ignored and some are detrimental. Furthermore there may be some fitness driver that selects for certain types of mutations which maximize fitness for that driver. As Huang et al note regarding one of the COVID mutants:

The severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) spike (S) glycoprotein D614G mutation became the predominant globally circulating variant after its emergence in the early coronavirus disease 2019 (COVID-19) pandemic.

Studies showed that this mutation results in an open conformation of the S glycoprotein receptor-binding domain (RBD), and increased angiotensin 1-converting enzyme 2 (ACE2) binding and fusion, which result in an increase in SARS-CoV-2 transmissibility and infectivity. Dynamic tracking of SARS-CoV-2 showed that the D614G variant became predominant after emergence in Europe and North America, but not in China.

The current absence of selective pressures from antiviral treatment suggests that the driving force for viral evolution could be variations in human population genetics.

Results show that ACE2 expression is higher in Asian populations than that in European, North American, and African populations. This supports the idea that lower ACE2 expression is a driving force in the positive selection for the D614G mutation.

This is a very interesting fitness proposal. The assumption is that the virus had been mutating for long periods and then somehow jumped to a human carrier where it further mutated to adapt to the ACE2 deficiency in that carrier. This also presents a more ominous scenario of a non-nature generated viral targeting.

This study suggests that the dynamics of the SARS-CoV-2 D614G mutation during the early-to-mid pandemic is associated with enhanced transmission efficiency in populations with lower ACE2 expression. Understanding the role that human genetic diversity plays in the adaptive evolution of SARS-CoV-2 may have an important impact on public health and measures to control the pandemic.

Indeed we have examined the targeting of ACE2 in a study of therapeutic strategies as mentioned.

6.6 IMMUNOCOMPROMISED PATIENTS

Recent work has demonstrated the significant possibility that immunocompromised patients infects present a medium for extensive mutations in the virus especially the spike protein. As Kemp et al note:

SARS-CoV-2 Spike protein is critical for virus infection via engagement of ACE2, and amino acid variation in Spike is increasingly appreciated. Given both vaccines and therapeutics are designed around Wuhan-1 Spike, this raises the theoretical possibility of virus escape, particularly in immunocompromised individuals where prolonged viral replication occurs. Here we report fatal SARS-CoV-2 escape from neutralising antibodies in an ***immune suppressed individual treated with convalescent plasma, generating whole genome ultradeep sequences by both short and long read technologies over 23 time points spanning 101 days.***

Little evolutionary change was observed in the viral population over the first 65 days despite two courses of remdesivir.

However, following convalescent plasma we observed dynamic virus population shifts, with the emergence of a dominant viral strain bearing D796H in S2 and H69/V70 in the S1 NTD of the Spike protein.

As serum neutralisation waned, viruses with the escape genotype diminished in frequency, before returning during a final, unsuccessful course of convalescent plasma. In vitro, the Spike escape variant conferred decreased sensitivity to multiple units of convalescent plasma/sera from different recovered patients, whilst maintaining infectivity similar to wild type.

These data reveal strong positive selection on SARS-CoV-2 during convalescent plasma therapy and identify the combination of Spike mutations D796H and H69/V70 as a broad antibody resistance mechanism against commonly occurring antibody responses to SARS-CoV-2

Thus in immunocompromised individuals there is a mutate that evolves after a prolonged period and in this case it targets the spike protein.

A second report by Avanzato et al states:

*On February 12, 2020, a 71-year-old woman with a **10-year history of chronic lymphocytic leukemia (CLL), acquired hypogammaglobulinemia, anemia, and chronic leukocytosis** presented to the emergency department with low back and lower extremity pain.*

She underwent surgery for a spinal fracture and stenosis related to her cancer on February 14, 2020 …She could not return to her rehabilitation center because of a confirmed outbreak of COVID-19 at the facility. Chest computed tomography (CT), performed on February 28, 2020, was unremarkable. The patient had no respiratory or systemic symptoms during this time.

*Because she was residing in the rehabilitation facility around the time of the COVID-19 outbreak, **she was tested and found positive for SARS-CoV-2 on March 2, 2020** After the initial SARS-CoV-2 diagnosis, she was kept in an isolation ward in a single room with negative airflow. Attending medical staff were using full personal protective equipment comprised of powered air-purifying respirators (PAPR) or N95 respirators with goggles, gowns, and gloves.*

***Over the course of the next 15 weeks, she was tested for SARS-CoV-2 another 14 times by several diagnostic companies and remained positive through June 15, 2020, 105 days since the initial positive test.** Subsequently, she tested negative on four consecutive swabs from June 16 to July 16, indicating that her infection had cleared.*

In this report, we describe long-term SARS-CoV-2 shedding in an immunocompromised individual with CLL and acquired hypogammaglobulinemia out to 105 days after the initial positive test. …

The information available to date about SARS-CoV-2 infection in immunocompromised individuals, including those with cancers such as CLL, is limited and mostly focuses on disease severity and outcome

Although it is difficult to extrapolate from a single individual, our data suggest that long-term shedding of infectious virus may be a concern in certain immunocompromised people. Given that immunocompromised individuals could have prolonged shedding and may not have typical symptoms of COVID-19, symptom-based strategies for testing and discontinuing transmission-based precautions, as recommended by the Centers for Disease Control and Prevention (CDC), may fail to detect whether certain individuals are shedding infectious virus. The individual eventually cleared the SARS-CoV-2 infection from the upper respiratory tract after developing low neutralizing antibody titers.

How the virus was cleared and the effect of convalescent plasma on clearance of the virus is unknown. The initial administration of convalescent plasma was followed by a decreased viral load in nasal swabs, but viral loads subsequently increased, despite administration of a second dose of convalescent plasma comprising higher antibody titers. …

Throughout the course of infection, there was marked within host genomic evolution of SARS-CoV-2. Deep sequencing revealed a continuously changing virus population structure with turnover in the relative frequency of the observed genotypes over the course of infection.

These differential selective pressures may have allowed a larger genetic diversity with continuous turnover of dominant viral species throughout the course of infection. Although some sequence variants remain consistent throughout the duration of infection, we also observed variants unique to individual time points, such as the spike deletions observed on day 49 and day 70. Previously reported spike deletions, distinct from those reported here, were observed at relatively low frequency in clinical samples but were enriched upon virus isolation.

Similar to these reports, the spike deletion in the isolate on day 49 was observed as a minor variant in the individual's sample but was also selected for during passage upon virus isolation. In contrast to the previously reported deletions at the cleavage sites, both spike deletions observed on day 49 and 70 in the individual are located in the NTD of S1, a region distal from the receptor binding site.

These deleted residues are not modeled in a number of spike structures, suggesting that this region is conformationally labile. …

Many current infection control guidelines assume that persistently PCR-positive individuals are shedding residual RNA and not infectious virus, with immunocompromised people thought to remain infectious for no longer than 20 days after symptom onset.

Here we show that certain individuals may shed infectious, replication-competent virus for much longer than previously recognized. Although infectious virus could be detected up to day 70, sgRNA, a molecular marker for active SARS-CoV-2 replication, could be detected up until day 105.

An immunocompromised state has been identified as a risk factor for development of severe disease and complications from COVID-19.

A wide variety of conditions and treatments can alter the immune system and cause immunodeficiency, creating opportunities for prolonged viral replication and shedding of infectious SARS-CoV-2*. Although this report focuses on long-term shedding of one immunocompromised individual, an estimated 3 million people in the United States have some form of immunocompromising condition, including individuals with HIV infection, solid organ transplant recipients, hematopoietic stem cell transplant recipients, and individuals receiving chemotherapy and corticosteroids.*

This transient or chronic immunocompromised population is at higher risk of respiratory disease complications with respiratory infections such as influenza A virus and SARS-CoV-2*. Prolonged shedding of pH1N1 shedding was observed in immunocompromised individuals with a variety of immunocompromising conditions during the previous pandemic in 2009, such as* ***people with cancer on chemotherapy and solid organ transplant recipients***

6.7 B.1.1.7 VARIANT

The B.1.1.7 variant is currently one of the more virulent strains. As Rambaut et al report:

Recently a distinct phylogenetic cluster (named lineage B.1.1.7) was detected within the COG-UK surveillance dataset. This cluster has been growing rapidly over the past 4 weeks and since been observed in other UK locations, indicating further spread.

Several aspects of this cluster are noteworthy for epidemiological and biological reasons and we report preliminary findings below. In summary:

The B.1.1.7 lineage accounts for an increasing proportion of cases in parts of England. The number of B.1.1.7 cases, and the number of regions reporting B.1.1.7 infections, are growing. B.1.1.7 has an unusually large number of genetic changes, particularly in the spike protein. Three of these mutations have potential biological effects that have been described previously to varying extents:

- *Mutation N501Y is one of six key contact residues within the receptor-binding domain (RBD) and has been identified as increasing binding affinity to human and murine ACE2.*
- *The spike deletion 69-70del has been described in the context of evasion to the human immune response but has also occurred a number of times in association with other RBD changes.*
- *Mutation P681H is immediately adjacent to the furin cleavage site, a known location of biological significance.*

The rapid growth of this lineage indicates the need for enhanced genomic and epidemiological surveillance worldwide and laboratory investigations of antigenicity and infectivity.

They then give the mutations in this strain:

gene	nucleotide	amino acid
ORF1ab	C3267T	T1001I
	C5388A	A1708D
	T6954C	I2230T
	11288-11296 deletion	SGF 3675-3677 deletion
spike	21765-21770 deletion	HV 69-70 deletion
	21991-21993 deletion	Y144 deletion
	A23063T	N501Y
	C23271A	A570D
	C23604A	P681H
	C23709T	T716I
	T24506G	S982A
	G24914C	D1118H
Orf8	C27972T	Q27stop
	G28048T	R52I
	A28111G	Y73C
N	28280 GAT->CTA	D3L
	C28977T	S235F

6.8 Adaptive Mutability

Mutations occur frequently. Most are filtered out, repaired, or just killed off. However, we generally accept that cancer is the result of a self-sustaining mutation from a cause or causes often unknown. We know that certain cancers, such as thyroid, may be the result of radiation exposure, whereas melanoma is often the result of excess sun exposure. Viruses such as HPV cause cervical cancer and in fact such a cancer can be prevented via vaccination.

Now Russo et al argue that adaptive mutagenesis is the basis for the loss of targeting of such thinks as kinase inhibitors and growth factor inhibition. However, such is a bit of a stretch if you will see that a clear identification of the process is lacking.

The logic used is as follows (we quote from Gerlinger):

1. *Russo et al. found that human CRC cell lines that were treated with EGFR or BRAF inhibitors down-regulated the expression of high-fidelity DNA repair proteins and increased that of error-prone DNA repair proteins, which may both increase mutation rates.* The issue here is the use of first line therapeutic has a secondary effect of suppressing key DNA repair mechanisms.

2. Specifically mismatch repair, MMR, and homologous repair, HR, were impaired.

3. *Adaptive mutagenesis is a mechanism described in bacteria that increases the mutation rate in response to cell stress. This is triggered by a cell-stress signaling pathway that activates error prone DNA double-strand break repair and it is accompanied by suppression of MMR.* Namely

the stress may be reactive oxygen stress, ROS, or the equivalent, and the result is a mismatch and from there we get a mutation.

4. *Russo et al. explored whether the mammalian target of rapamycin (mTOR) pathway, a major stress signaling pathway in humans, controls drug induced mutagenesis in cancer cells. mTOR signaling was indeed inactivated by drug treatment, but inhibiting the mTOR pathway alone did not phenocopy the changes in DNA repair protein expression. The trigger of drug-induced mutagenesis in CRC cells is therefore either more complex or different from that in bacteria*. mTOR is a powerful pathway which we shall discuss. However it is not clear that this pathway can alone effect an adaptive mutagenesis.

5. *The contribution of drug-induced mutagenesis to clinically acquired resistance in patients with CRC and other cancer types is now important to assess because this remains unclear for several reasons. Mutational processes differ in the preferred DNA sequence contexts in which they occur and in the genetic variants they generate. MMR deficiency leads to high rates of deletions in nucleotide repeats and to cytosine-to-thymine base changes*. The changes are well known but the changes related to an adaptive path require a more focused driver. Thus what drives the changes, namely selecting those leading to reactivation of aggressive malignancy.

The Russo paper raises questions, presents some answers yet leaves a great deal yet to be explained. We examine adaptive mutagenesis in this section.

Let us begin with and attempt to give a definition of adaptive mutation. Before doing so let us examine the two words separately. First adaptive implies that there is some exogeneous element or process which makes the change occur in a manner which is different because of the presence of his exogeneous element. The change is somehow driven by this element. Second, look a mutation. This is a fundamental change in a gene, a change in the underlying DNA and not just some epigenetic modification. Things have been tampered with. Combined the term adaptive mutation is a tampered change in a direction compliant with the presence of the exogeneous agent. Roth et al have discussed[8]:

The term adaptive mutation has been defined as the process by which stresses that are not directly mutagenic activate mechanisms for causing mutations, even in nongrowing cells (stress-induced or stationary-phase mutagenesis). This definition assumes a mechanism that we think is unlikely to exist. In order to consider all explanations of the relevant phenomena, adaptive mutation is defined here as the process by which mutations arise under selective conditions, whether or not mutation rates increase or growth is required. The area discussed here has been reviewed previously from various points of view Darwin suggested that stress might generate the variability upon which natural selection operates.

The classic experiments of Luria & Delbruck and Lederberg demonstrated that some mutations arise without the influence of selective stress. However, the lethal selections they used could not have detected mutations induced by selective conditions. Shapiro, Cairns et al., and Hall pointed

[8] From the definition section they note: *Adaptive mutation: any process by which fitter mutations arise under selective conditions; may or may not require mutagenesis or growth.*

out this deficiency and described genetic systems in which selective conditions seemed to increase the mutation rate. The Cairns system has been analyzed in most detail and remains controversial despite this effort.

As Murray noted:

In 1943, it had long been known that bacterial cultures rapidly develop resistance to viral infection. Some biologists argued that viruses directly induced resistance mutations, while others believed the mutations arose spontaneously before exposure to the virus. But when Luria and Delbrück first attempted to distinguish between these two hypotheses, they were frustrated by what appeared to be irritatingly inconsistent mutation rates.

Then, after watching a colleague win a jackpot ($3 in dimes!) at a slot machine, Luria realized this inconsistency was telling him something: the number of mutant bacterial colonies present at the end of the experiment depended on when the mutations arose. Mutations arising in earlier generations would be present in many descendent cells (a "jackpot"), whereas mutations occurring in later generations would be present in only a few cells.

Luria passed his insight to Delbrück, who worked out the expected statistical distribution of the number of mutant cells per culture.

Their data decisively rejected the hypothesis that bacteria became resistant only after being exposed to the virus and strongly supported the prediction that the phage-resistant mutations had a constant probability of occurring in each cell division.

The Luria–Delbrück article had three important impacts beyond its direct conclusion:

it showed that elegant statistical analysis could illuminate biological processes that could not be directly observed,

it contributed to Luria and Delbrück winning the 1969 Nobel Prize in Medicine or Physiology (shared with Alfred Hershey), and

it led, indirectly, to a continuing debate about whether organisms exert physiological control over their mutation rates.

What then is the fundamental process here? As Foster notes:

Adaptive mutation is defined as a process that, during nonlethal selections, produces mutations that relieve the selective pressure whether or not other, nonselected mutations are also produced. *Examples of adaptive mutation or related phenomena have been reported in bacteria and yeast but not yet outside of microorganisms. A decade of research on adaptive mutation has revealed mechanisms that may increase mutation rates under adverse conditions. This article focuses on mechanisms that produce adaptive mutations in one strain of Escherichia coli, FC40. These mechanisms include recombination-induced DNA replication, the placement*

of genes on a conjugal plasmid, and a transient mutator state. The implications of these various phenomena for adaptive evolution in microorganisms are discussed.

In contrast as Gerlinger notes:

Adaptive mutagenesis is a mechanism described in bacteria that increases the mutation rate in response to cell stress. This is triggered by a cell-stress signaling pathway that activates error-prone DNA double-strand break repair and it is accompanied by suppression of MMR. Adaptive mutagenesis increases the probability of generating mutations that enable evolutionary adaptation of unicellular organisms to new environments.

On the basis of the pronounced similarities of drug-induced mutagenesis in CRC and adaptive mutagenesis in bacteria, Russo et al. explored whether the mammalian target of rapamycin (mTOR) pathway, a major stress signaling pathway in humans, controls drug-induced mutagenesis in cancer cells. mTOR signaling was indeed inactivated by drug treatment, but inhibiting the mTOR pathway alone did not phenocopy the changes in DNA repair protein expression. The trigger of drug-induced mutagenesis in CRC cells is therefore either more complex or different from that in bacteria.

As Hall has noted:

Adaptive mutations are spontaneous mutations that occur in microorganisms during periods of prolonged stress in non-dividing or very slowly dividing populations and that are specific to the environmental challenge that causes that stress. This article reviews the literature on adaptive mutagenesis since 1993*. The evidence that adaptive mutagenesis is both real and general is considered. The most widely used system for studying adaptive mutagenesis, reversion of an F'-borne lacI33 allele, is shown to be a special case that reflects more about F-plasmid biology than about adaptive mutagenesis in general.*

New evidence demonstrating that adaptive mutagenesis is, indeed, specific is discussed. A variety of genes whose products affect adaptive mutagenesis are discussed. A model to explain that specificity and new evidence in support of that model are considered, as are potential roles of adaptive mutagenesis in evolution and practical aspects of adaptive mutagenesis.

As Slechta et al have noted:

The term ‘‘adaptive mutation’’ refers here to the process by which selection increases the number of Lac+ revertants. This definition makes no assumptions regarding the role of growth or the contribution of general mutagenesis to reversion. Two aspects of this process are discussed here:

reversion and

mutagenesis.

*In this article, **‘‘reversion’’ means mutational correction of a particular lac allele to lac+ under selective conditions defined for this system**. By **‘‘mutagenesis,’’ we mean a genome-wide undirected increase in mutation rate occurring during the process of reversion under selection**, detected as unselected mutations carried by Lac+ revertants.*

We can now consider the types of adaptive mutations. Roth et al define three types of adaptive mutagenesis.

1. ***Functional Direction of Mutagenesis:*** *This early model proposes that nongrowing cells make only mutations that improve their fitness.*

2. ***Generally Hypermutable State****: A model proposed by Hall suggests that stress induces a general (undirected, genomewide) hypermutable state (HMS) in a subset (0.1%) of the nongrowing starved population. Cells in the subpopulation ultimately die of lethal mutations unless they first acquire a lac reversion event, which relieves the stress and terminates the HMS. Mutagenesis appears directed to lac because only Lac+ revertants survive hyper mutagenesis.*

3. ***Positional Direction of Hypermutability****: This model is a hybrid of the previous two models. It assumes that stress induces DinB in nongrowing cells and mutagenizes any region in which recombinational replication is occurring. Mutagenesis is directed to the F plasmid, whose conjugation transfer functions (tra) produce DNA ends that stimulate intense recombination between plasmid sequences. Because only the plasmid is mutagenized, associated chromosomal mutations are not expected. The observed lac revertants can be explained by a 100-fold increase in mutation rate on the plasmid (rather than a global 105-fold increase) because the whole population is affected rather than a subpopulation. The genome-wide mutagenesis seen in a minority of revertant clones (10%) is not considered central*

Now viral mutations in a sense follow the above observations. They were fundamental in the discovery and interpretation of the phenomenon.

7 OBSERVATIONS

We have previously made certain observations regarding this virus from initial understanding, putative therapeutic strategies, the impact on the immune system and secondary infections to the epidemiology and impact of the initial vaccines. Regrettably many of the issue we have examined have become significant and this current one discussed herein is most likely one of great significance. Namely mutations in viruses are common and exceptionally common in single stranded RNA which has no means of error detection.

The corona virus, AKA COVID-19, is a single stranded RNA virus. As such it is susceptible to significant mutations. There are two essential parts to the RNA virus, non-structural and structural. The non-structural produce proteins used in reproduction of the virion. The structural rely on the non-structural to produce shells, spikes and other structural elements.

The virus RNA is about 30,000 nucleotides long, and the spike is about 3,000. Mutations typically occur at one nucleotide per duplication. Considering the millions of duplication we can anticipate millions of mutations. Almost all bad mutations result in the death of the virus, but there will always be a few good ones which keep it alive and in fact can make it stronger. That is the problem.

Now where does this lead us? Well we have a vaccine based upon the spike protein. It stimulates the immune system to generate antibodies to kill off a virion attack. But if the spike mutates then the vaccine will not work. Yet a mutation in the spike may be such that the virion cannot attach to cells in which case it is harmless. However if we get a mutation which can attach and enter and is not like the one matched for in the vaccine, then we have a problem. We have this yearly with our flu shots. But flu deaths in the US are about 70,000 per year whereas COVID is now five times that. But the 70,000 is with a vaccine.

Thus, the problem is the longer we defer vaccinating people the better the virus chance to mutate to a more deadly strain. This means that since we have vaccines coming out rapidly the gross incompetence of the States will likely result in a more virulent infection and a surge later this year. The concept of herd immunity is invalid. The mutation of the virus must be stopped before it stops humanity, or at least that part outside of China.

We can now make several observations.

7.1 MUTATIONS ARE INEVITABLE

Mutations are inevitable. Once the virus enters the victim it begins the ever moving process or replications and it is that very process that enables the mutation. We believe that it is essential to better understand the mutation drivers. In so doing we should expand our understanding not only across viruses but also a wide variety of cancers.

7.2 VACCINES MUST ANTICIPATE AND THEN FOLLOW THE MUTATIONS

The vaccines that have thus far been generated are snapshots in time[9]. They are like so many military strategies a fight against the last war. The immune system can adapt and it does adapt but to what it sees as a threat. The spike protein was considered a likely antigen to kick off the immune response. However, if the spike protein changes, morphs as resulting from mutations, then we have a target which the vaccine may no longer be effective for.

As Neuman et al have noted regarding replicase:

New coronavirus protein structures continue to have important implications for virus biology and antiviral design. An unexpected benefit of the coronavirus structure boom is a wealth of previously undiscovered protein folds. The protein data bank keeps records of the discovery of new protein structures over time, and currently classifies all known protein structures into fewer than 1500 folds.

Since 2003, 18 out of 28 coronavirus proteins encompassing a total of 27 domains have been determined experimentally. Several of these have been described as "new folds" commonly defined as one with sufficiently different fold/topology based on comparison methods (DALI), fold classification schemes (CATH and SCOP) and family assignment schema of PFAM. By these criteria 16 out of the 27 domains are indeed new folds – a striking rate of fold discovery, when compared to the ~10% for model pro- and eukaryotes being reported by structural genomics centers. Why do coronaviruses possess an abundance of new folds?

One obvious reason might be that these structures have been relatively unexplored, and therefore under-represented in PDB. This is the first proteome-scale structural characterization of a coronavirus, and one with a disproportionately large number of singletons. The new folds are significantly contributed by the 16 nonstructural proteins of the replicase machinery, several of which do not have counterparts outside Nidovirales. Ideally, new folds enable us to model sequence homologues, thereby filling out the immediate neighborhood in structure space.

This is nontrivial for SARS-CoV proteins, since the new folds are (so far) either true sequence singletons (nsp1, nsp2 nsp3a, sars9b) or found exclusively in Coronaviridae (nsp4, 7, 8, 9, 10, 15, spike RBD, sars9a). Fast mutation rates in viruses may encourage divergent sampling of fold space. This, along with oligomerization has been proposed to be major facilitators of fold evolution allowing a protein to morph to a new fold analogous to structural drift

Elucidation of new folds, especially for isolated groups of divergent homologues should help in improving fold recognition and comparative modeling algorithms. These observations also have ramifications in evolution of new viral strains, a phenomenon which is the result of two antagonistic forces: greater adaptability within an ecological niche (because of intrinsically fast mutation rates) and increased evolutionary constraints due to their small genomes.

Overall, we are left with the piquant notion that proteins in viral proteomes may probably occupy a unique niche in fold space and coronaviruses, a peculiar island in this niche. Viruses are the most diverse biological entities on this planet and second only to prokaryotes in terms of

[9] https://www.researchgate.net/publication/345813274_COVID-19_Vaccine_An_Update_and_Primer

sheer biomass. While the diversity of protein structures they represent certainly defies imagination, our understanding of protein folds and their migration in tertiary fold space may well be locked up in them.

7.3 VACCINE TARGETS SHOULD BE MODELLED AND ANTICIPATED

Vaccine targets such as a spike, can be modelled in a complex manner in the context of mutations. We can anticipate the mutations using our understanding of the mutation process. This then should allow us to anticipate what changes we should be looking for and introduce those into an anticipatory vaccine. We should not have to wait to see the result. We should see the result by forecasting it, or at least a most probable set of possible mutations.

7.4 CONTAGIOUSNESS IS COMPLEX AND R0 IS USELESS AS A GENERALIZED PARAMETER

Despite the amount of work done on understanding COVID-19, we believe that there is insignificant understanding of how it is actually transmitted. There have been arguments regarding aerosols, virions, and the like but no dispositive research has clearly demonstrated the transmission path. Thus, the degrees of contagiousness is not well understood. The debate on masks, gathering etc lingers. Regrettably it is our opinion that the "science" on this issue is not in and that we are following the path taken in 1918 with the Spanish flu.

Dorp et al have noted:

COVID-19 is caused by the coronavirus SARS-CoV-2, which jumped into the human population in late 2019 from a currently uncharacterised animal reservoir. Due to this recent association with humans, SARS-CoV-2 may not yet be fully adapted to its human host. This has led to speculations that SARS-CoV-2 may be evolving towards higher transmissibility. The most plausible mutations under putative natural selection are those which have emerged repeatedly and independently (homoplasies). Here, we formally test whether any homoplasies observed in SARS-CoV-2 to date are significantly associated with increased viral transmission. To do so, we develop a phylogenetic index to quantify the relative number of descendants in sister clades with and without a specific allele.

We apply this index to a curated set of recurrent mutations identified within a dataset of 46,723 SARS-CoV-2 genomes isolated from patients worldwide. We do not identify a single recurrent mutation in this set convincingly associated with increased viral transmission. Instead, recurrent mutations currently in circulation appear to be evolutionary neutral and primarily induced by the human immune system via RNA editing, rather than being signatures of adaptation. At this stage we find no evidence for significantly more transmissible lineages of SARS-CoV-2 due to recurrent mutations.

As Musher noted:

How contagious is it?" is a question regularly asked by family, friends, and colleagues, as well as school administrators and news reporters. The purpose of this article is to analyze a large and widely scattered body of data that addresses this question, focusing on common respiratory

tract infections that are spread from person to person. Some respiratory tract infections are not at all contagious and will not be included. For example, histoplasmosis and coccidioidomycosis are caused by dimorphic fungi that need to replicate outside the human host in order to evolve into an infectious form. Agents such as Pneumocystis jiroveci (formerly carinii) and Mycobacterium avium complex are transmissible, but the occurrence of disease is so thoroughly determined by host factors that an infection is not generally regarded as contagious.

I shall also not address agents that have received recent attention principally because of their actual or potential role in biologic terrorism, and I have chosen not to include nosocomial infections, my purpose being to concentrate on community-based contagion. … Symptomatic persons are more likely than asymptomatic ones to spread infection because they discharge a greater volume of infective material, perhaps containing a greater density of infectious particles.

Furthermore, children are more likely than adults to introduce and spread infection within families, probably because of their level of personal hygiene, their contact with siblings, and their need for parental attention.

For all the respiratory tract pathogens discussed in this article, the interaction with host defenses determines whether disease results. Mucosal immune factors may block adherence and local proliferation. Humoral or cellular immune responses may contain proliferation and cell invasion, rendering infection asymptomatic.

Humans remain susceptible to organisms that have many immunologic types and to which the immune response is transient, especially if the mutation appears in antigens to which protective antibody is directed, as seen, for example, in rhinovirus and influenzavirus.

Avoidance of infections that are spread by large droplets requires avoidance of close contact with ill persons. In the case of infections that are spread by droplet nuclei, shared space should also be avoided unless ventilation is extremely good: attending a Broadway show during an outbreak of influenza is certainly not advisable for a person who has not been vaccinated. Washing hands after greeting someone with a viral infection may be appropriate, because viruses may be transmitted by direct contact, and self-inoculation is an important part of pathogenesis.

Excessive washing or boiling of eating utensils does not seem likely to prevent transmission of these organisms; with the exception of rhinovirus and RSV, they do not survive on surfaces, they are generally not present in large numbers in saliva, and with the possible exception of adenovirus, they do not infect cells of the alimentary canal

7.5 Gain of Function Issues

Gain of function, GoF, is the general concept of adding to some biological cell an addition functionality. As Casadevall and Imperiale have noted:

Given that symbolic language is the basis for much of human communication, we begin with terminology and dissect the phrase "gain of function," or GOF.

When applied to influenza virus research, ***the term GOF has taken on the meaning of something dangerous, risky, and possibly nefarious.***

However, GOF means exactly what it says, that the entity in question has gained a new property. In the case of influenza virus, the concern regarding GOF has been associated with the acquisition of a new function, such as mammalian transmissibility, increased virulence for humans, or evasion of existing host immunity.

For example, passage of H5N1 virus in ferrets allowed selection for variants with ferretto-ferret transmissibility, and the GOF was the acquisition of mammalian transmission. However, the same type of experiment can be beneficial to humanity, since the principle of passage in a nonnative host can be used to generate attenuated vaccines.

For example, some human-pathogenic viruses, such as poliovirus, were attenuated by passage in cells of another species, such as monkey cells. In those experiments, the GOF was replication in another species, and this property reduced the efficiency of replication in human cells, thus resulting in a new attenuated strain that could be used as a vaccine. Indeed, those attenuated viruses manifested a GOF, namely, attenuation. One of us recently published a GOF experiment with BK polyomavirus, in which mutation of a regulatory microRNA (miRNA) greatly enhanced replication.

Hence, GOF is a powerful experimental tool that is routinely used in biomedical research, and the concern with influenza virus research is not gain of function per se but rather the selection of variants with increased mammalian transmissibility and virulence that could affect human populations if there were deliberate or accidental release*. It is clear that GOF is a problematic phrase, and this term has acquired a particular meaning in the ongoing debate and particularly in the lay media. Unfortunately, the term GOF has come to only represent something that can be used to confer dangerous properties to a microbe.*

Despite these problems in terminology, we use the expression GOF in this essay with the understanding that we are referring to the narrow category of experiments that involve primarily changes to the virulence and transmissibility of PPP, such as influenza virus. Although influenza virus is the subject of the ongoing debate, it is important to note that these issues extend to other PPP, such as ***severe acute respiratory syndrome (SARS) coronavirus****.*

Note that this was published in 2014, six years before the current SARs pandemic. Namely GoF in viruses did and still does present a serious threat to humans. Adding an element that increases transmissibility is a complex issue but just the experimentation by careless or incompetent researchers presents such a clear and present danger,

As Sharples et al noted and a 2015 NIH panel report:

The GoF controversy began in late 2011 with the question of whether to publish the results of two experiments involving H5N1 avian influenza and continued to focus on certain research with highly pathogenic avian influenza over the next 3 years.3 The new U.S. policy expanded the

scope to include experiments with the coronaviruses that cause Severe Acute Respiratory Syndrome (SARS) and Middle East Respiratory Syndrome (MERS).

The heart of the U.S. process is an evaluation of the potential risks and benefits of certain types of GoF experiments with influenza, SARS, and MERS viruses that would "inform the development and adoption of a new U.S. Government policy governing the funding and conduct of gain-of-function research" (White House, 2014a:3). As part of the process, the government also instituted a pause in both new and current funding for some GoF research projects while the evaluation was carried out. …

The field of virology, and to some extent the broader field of microbiology, widely relies on studies that involve gain or loss of function. In order to understand the role of such studies in virology, Dr. Kanta Subbarao from the Laboratory of Infectious Disease at the National Institute of Allergy and Infectious Diseases (NIAID) at the National Institutes of Health (NIH) gave an overview of the current scientific and technical approaches to the research on pandemic strains of influenza and Severe Acute Respiratory Syndrome (SARS) and Middle East Respiratory Syndrome (MERS) coronaviruses (CoV).

As discussed in greater detail later in this chapter, many participants argued that the word choice of "gainof-function" to describe the limited type of experiments covered by the U.S. deliberative process, particularly when coupled with a pause on even a smaller number of research projects, had generated concern that the policy would affect much broader areas of virology research.

TYPES OF GAIN-OF-FUNCTION (GOF) RESEARCH

Subbarao explained that routine virological methods involve experiments that aim to produce a gain of a desired function, such as higher yields for vaccine strains, but often also lead to loss of function, such as loss of the ability for a virus to replicate well, as a consequence. In other words, any selection process involving an alteration of genotypes and their resulting phenotypes is considered a type of Gain-of-Function (GoF) research, even if the U.S. policy is intended to apply to only a small subset of such work.

Subbarao emphasized that such experiments in virology are fundamental to understanding the biology, ecology, and pathogenesis of viruses and added that much basic knowledge is still lacking for SARS-CoV and MERS-CoV.

Subbarao introduced the key questions that virologists ask at all stages of research on the emergence or re-emergence of a virus and specifically adapted these general questions to the three viruses of interest in the symposium.

To answer these questions, ***virologists use gain- and loss-of-function experiments to understand the genetic makeup of viruses and the specifics of virus-host interaction. For instance, researchers now have advanced molecular technologies, such as reverse genetics, which allow them to produce de novo recombinant viruses from cloned cDNA, and deep***

sequencing that are critical for studying how viruses escape the host immune system and antiviral controls.

Researchers also use targeted host or viral genome modification using small interfering RNA or the bacterial CRISPR-associated protein-9 nuclease as an editing tool.

7.6 Spike Morphology Changes

An added factor to understanding the spike protein is its shape in differing situations. For example we know the spike does not effect a connection at say 94F, it requires an unfolding achievable at 98F. But one suspects that we have pH issues, induction effects, charge imbalances, and a variety of other issues which affect the spike protein configuration and its ability to effect cellular entry. There are many viruses which come and go and never cause a threat to a human. However as we see mutations occur we ask if the change also will be affected by these exogeneous cell artifacts.

8 REFERENCES

1. Abbas et al, Basic Immunology, Elsevier, 4th Ed, 2014Byron et al, Translating RNA sequencing into clinical diagnostics: opportunities and challenges, Nature Review Genetics, May 2016
2. Artika et al, Molecular biology of coronaviruses: current knowledge, Heliyon 6 (2020)
3. Avanzato et al, Case Study: Prolonged Infectious SARS-CoV-2 Shedding from an Asymptomatic Immunocompromised Individual with Cancer, 2020, Cell 183, 1901–1912, December 23, 2020
4. Baer, Does Mutation Rate Depend on Itself?, PLoS Biol 6(2): e52. doi:10.1371/journal.pbio.0060052, February 2008
5. Baldo, P., et al, mTOR Pathway and mTOR inhibitors as Agents for Cancer Therapy, 2008, Curr Can Drug Targets, pp 647-668.
6. Bangaru et al, Structural analysis of full-length SARS-CoV-2 spike protein from an advanced vaccine candidate, Science, 370, 1089–1094 (2020) 27 November 2020
7. Baron S, editor. Medical Microbiology. 4th edition. Galveston (TX): University of Texas Medical Branch at Galveston; 1996.
8. Burkholz et al, Paired SARS CoV-2 Spike Protein Mutations Observed During Ongoing SARS-CoV-2 Viral Transfer from Humans to Minks and Back to Humans, bioRxiv preprint doi: https://doi.org/10.1101/2020.12.22.424003 ; this version posted December 29, 2020
9. Cairns et al, The Origin of Mutants, Nature, Vol 335 8 Sept 1988
10. Cairns, Mutation and Cancer: The Antecedents to Our Studies of Adaptive Mutation, Genetics 148: 1433–1440 (April, 1998)
11. Cann, Molecular Virology, 5th Ed, Academic, 2012
12. Cantley et al, Signal Transduction, CSHL Press, 2014
13. Carter and Saunders, Virology, 2nd Ed, Wiley, 2013
14. Casadevall and Imperiale, Risks and Benefits of Gain-of-Function Experiments with Pathogens of Pandemic Potential, Such as Influenza Virus: a Call for a Science Based Discussion, ASM, July/August 2014 Volume 5 Issue 4
15. Cavalli-Sforza and Bodmer, The Genetics of Human Populations, Dover, 1999
16. CDC, CDC 2019-Novel Coronavirus (2019-nCoV) Real-Time RT-PCR Diagnostic Panel, Feb 2, 2020
17. Chapman. J., et al, Playing the End Game: DNA Double-Strand Break Repair Pathway Choice, Molecular Cell 47, August 24, 2012
18. Charlesworth et al, The Sources of Adaptive Variation, Proc Royal Soc B 284, 2017

19. Chong et al, Transmissibility of coronavirus disease 2019 (COVID-19) in Chinese cities with 2 different transmission dynamics of imported cases, https://www.medrxiv.org/content/10.1101/2020.03.15.20036541v1
20. Ciccia, A., S. Elledge, The DNA Damage Response: Making it safe to play with knives, Mol Cell. 2010 October 22; 40(2): 179–204
21. Clark and Pazdernik, Biotechnology, Academic, 2nd Ed, 2016
22. Collier et al, Human Virology, 4th Ed, Oxford, 2011
23. Comb and Sanjuan, Variation in RNA Virus Mutation Rates across Host Cells, PLOS Pathogens, www.plospathogens.org , 1 January 2014, Volume 10, Issue 1
24. Corman et al, Diagnostic detection of 2019-nCoV by real-time RT-PCR, https://virologie-ccm.charite.de/en/ , Berlin, Jan 17, 2020
25. Cross et al, RNA regulatory processes in RNA virus biology, WIREs RNA. 2019;10:e1536.
26. Dalton et al, Pre-emptive low cost social distancing and enhanced hygiene implemented before local COVID-19 transmission could decrease the number and severity of cases, https://ssrn.com/abstract=3549276
27. Dittmar and Liberles, Evolution after Gene Duplication, Wiley-Blackwell, 2010
28. Domingo, Virus as Populations: Composition, Complexity, Dynamics, and Biological Implications, 1st Ed, Academic 2015
29. Dong et al, Epidemiological Characteristics of 2143 Pediatric Patients With 2019 Coronavirus Disease in China, Pediatrics. 2020
30. Dorp et al, No evidence for increased transmissibility from recurrent mutations in SARS-CoV-2, Nature Communications volume 11, Article number: 5986 (2020)
31. Doukas et al, Role of DNA mismatch repair genes in lung and head and neck cancer (Review), World Academy Of Sciences Journal 1: 184-191, 2019
32. Duffy, Why are RNA virus mutation rates so damn high?, PLoS Biol 16(8): e3000003. https://doi.org/10.1371/journal.pbio.3000003 2018
33. Easton and Houghton, mTOR and Cancer Therapy, Oncogene, 2006, pp 6436-6446.
34. Elena and Sanjuan, Adaptive Value of High Mutation Rates of RNA Viruses: Separating Causes from Consequences, Journal of Virology, Sept. 2005, P. 11555–11558
35. Elliott and Landomery, Molecular Biology of RNA, Oxford, 2011
36. Escherichia coli, PLOS Genetics, 11(8) 2015
37. Feldman, Mathematical Evolutionary Theory, Princeton, 1989
38. Filipe et al, Genomic epidemiology reveals multiple introductions of SARS-CoV-2 from mainland Europe into Scotland, Nature Microbiology, Vol 6, January 2021, 112–122
39. Flint et al, Principles of Virology, Vol II, 2nd Ed, ASM Press, 2009
40. Foster, Adaptive mutation: implications for evolution, Bioessays. 2000 December; 22(12): 1067–1074

41. Foster, Adaptive Mutations in E coli, Journal of Bacteriology, Aug. 2004, p. 4846–4852

42. Futuyama, Evolutionary Biology, Sinauer, 1986

43. Gerlinger, Targeted drugs ramp up cancer mutability, Science 20 Dec 2019: Vol. 366, Issue 6472, pp. 1452-1453

44. Gordon et al, A SARS-CoV-2 protein interaction map reveals targets for drug repurposing, Nature, 30 April 2020

45. Guan et al, Clinical Characteristics of Coronavirus Disease 2019 in China, NEJM, Feb 28, 2020

46. Haldane, The Causes of Evolution, Princeton, 1990

47. Hall, Adaptive mutagenesis: a process that generates almost exclusively beneficial mutations, Genetica. 1998;102-103(1-6):109-25

48. Halliwell and Gutteridge, Free Radicals in Biology and Medicine, 5th Ed, 2015Oxford,

49. Harvey and Pagel, The Comparative Method in Evolutionary Biology, Oxford, 1991

50. Hata et al, Tumor cells can follow distinct evolutionary paths to become resistant to epidermal growth factor receptor inhibition, Nat Med. 2016 March; 22(3): 262–269

51. Hay, N., The Akt-mTOR tango and its relevance to Cancer, Cancer Cell, V 8, 2005, pp 179-183.

52. Heil et al, Species-Specific Recognition of Single-Stranded RNA via Toll-like Receptor 7 and 8, Science, March 5, 2004

53. Hendrickson et al, Amplification–mutagenesis: Evidence that ‘‘directed’’ adaptive mutation and general hypermutability result from growth with a selected gene amplification, PNAS, February 19, 2002, vol. 99, no. 4

54. Herbert and Nag, A Tale of Two RNAs during Viral Infection: How Viruses Antagonize mRNAs and Small Non-Coding RNAs in The Host Cell, Viruses, 2016 June 8(6), 154

55. Hoekstra, Genetics, development and evolution of adaptive pigmentation in vertebrates, Heredity volume 97, pages222–234(2006)

56. Hofmann and Clokie, Principles and Techniques of Biochemistry and Molecular Biology, 8th Ed, Cambridge, 2018

57. Holbrook et al, Aerosol and Surface Stability of SARS-CoV-2 as Compared with SARS-CoV-1, NEJM, March 17, 2020

58. Howley and Knipe, Virology, 7th Ed, Vol 1, Wolters Kluwer, 2021

59. Hsieh and Zhang, The Devil is in the details for DNA mismatch repair, NAS, April 4, 2017, vol. 114, no. 14

60. Hsieh et al, The translational landscape of mTOR signalling steers cancer initiation and metastasis, 3 May 2012, vol 485, Nature, 55

61. Huang et al, Impact of Genetic Variability in ACE2 Expression on the Evolutionary Dynamics of SARS-CoV-2 Spike D614G Mutation, Genes, 24 December 2020

62. Huang et al, Structural and functional properties of SARS-CoV-2 spike protein: potential antivirus drug development for COVID-19, Acta Pharmacologica Sinica (2020) 41:1141–1149

63. Jackson and Loeb, The mutation rate and cancer, Genetics 148: 1483–1490 (April, 1998)

64. Jackson, S., J., Bartek, The DNA-damage response in human biology and disease, Nature Reviews, Vol 461, 22 October 2009

65. Ke et al, Structures and distributions of SARS-CoV-2 spike proteins on intact virions, Nature, 17 August 2020

66. Kemp et al, Neutralising antibodies drive Spike mediated SARS-CoV-2 evasion, medRxiv preprint doi: https://doi.org/10.1101/2020.12.05.20241927 ; this version posted December 19, 2020

67. Kinkaide, C., et al, Targeting Akt/mTOR and ERK MAPK Signalling Inhibits Hormone Refractory Prostate Cancer in Preclinical Mouse Model, Jrl Clin Inv 2008, 99 3051-3064.

68. Koma et al, How Many Adults Are at Risk of Serious Illness If Infected with Coronavirus?, KFF, March 2020

69. Kong et al, Analysis of mTOR Gene Aberrations in Melanoma Patients and Evaluation of Their Sensitivity to PI3K–AKT–mTOR Pathway Inhibitors, Clin Cancer Res; 2015

70. Korpelainen et al, RNA-seq Data Analysis, CRC Press, 2015

71. Kunkel and Erie, DNA Mismatch Repair, Annu. Rev. Biochem. 2005. 74:681–710

72. Letko et al, Functional Assessment of Cell Entry and Receptor Usage for SARS-CoV-2 and Other Lineage B Betacoronaviruses, Nature Microbiology, Feb 24, 2020

73. Li et al, Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus–Infected Pneumonia, NEJM, Jan 29, 2020

74. Li et al, Propagation analysis and prediction of the COVID-19, https://www.medrxiv.org/content/10.1101/2020.03.14.20036202v1

75. Li, Mechanisms and functions of DNA mismatch repair, Cell Research (2008) 18:85-98.

76. Liu et al, Molecular mutation characteristics of mismatch and homologous recombination repair genes in gastrointestinal cancer, Oncology Letters 18: 2789-2798, 2019

77. LoPiccolo, J., et al, Targeting the PI3K/Akt/mTOR Pathway, Drug Res Up, 2008, pp 32-50

78. Lostroh, Molecular and Cellular Biology of Viruses, CRC, 2019

79. Lund et al, Recognition of single-stranded RNA viruses by Toll-like receptor 7, PNAS, April 13, 2004, vol. 101, no. 15

80. Luria and Delbrück, Mutations of Bacteria from Virus Sensitivity to Virus Resistance. GENETICS, November 20, 1943 28: 491–511

81. Mara et al, The Genome Sequence of the SARS-Associated Coronavirus, Sciencexpress, 1 May 2003

82. McCarthy et al, Natural deletions in the SARS-CoV-2 spike glycoprotein drive antibody escape, bioRxiv preprint doi: https://doi.org/10.1101/2020.11.19.389916; this version posted November 19, 2020

83. Mizumoto et al, Estimating the asymptomatic proportion of coronavirus disease 2019 (COVID-19) cases on board the Diamond Princess cruise ship, Yokohama, Japan, 2020, www.eurosurveillance.org

84. Modrich and Lahue, Mismatch Repair in Replication Fidelity, Genetic Recombination, and Cancer Biology, Annu Rev. Biochem. 1996, 65:l0l-33

85. Mork, et al, Protein-driven inference of miRNA–disease associations, Bioinformatics, ol. 30 no. 3 2014, pages 392–397

86. Morreall et al, Evidence for Retromutagenesis as a Mechanism for Adaptive Mutation in

87. Murray, Salvador Luria and Max Delbrück on Random Mutation and Fluctuation Tests, Genetics, Vol. 202, 367–368 February 2016

88. Murray. Mathematical Biology, Springer, 1989, pp 610-650.

89. Musher, How Contagious Are Common Respiratory Tract Infections?, NEJM, 2003;348:1256-66

90. Neidhardt et al, Physiology of the Bacterial Cell, Sinauer, 1990

91. Neuman et al, Atlas of coronavirus replicase structure, Virus Research 194 (2014) 49–66

92. Nickoloff et al, Drugging the Cancers Addicted to DNA Repair, JNCI J Natl Cancer Inst (2017) 109(11)

93. Nishiura et al, Estimation of the asymptomatic ratio of novel coronavirus infections (COVID-19), International Journal of Infectious Diseases, Feb 13, 2020

94. Orr and Betancourt, Haldane's Sieve and Adaptation From the Standing Genetic Variation. Genetics 157: 875–884 (February 2001)

95. Oxford et al, Human Virology, 5th Ed, Oxford, 2016

96. Paquette, et al, mTOR Pathways in Cancer and Autophagy, Cancers 2018, 10, 18

97. Perkins, Estimating unobserved SARS-CoV-2 infections in the United States, https://www.medrxiv.org/content/10.1101/2020.03.15.20036582v1

98. Plumet et al, Dynamics of Viral RNA Synthesis during Measles Virus Infection, Journal Of Virology, June 2005, P. 6900–6908

99. Rambaut et al, Preliminary genomic characterisation of an emergent SARSCoV-2 lineage in the UK defined by a novel set of spike mutations, https://www.cogconsortium.uk , COVID-19 Genomics Consortium UK (CoG-UK)

100. Roth et al, Origin of Mutations Under Selection: The Adaptive Mutation Controversy, Annu. Rev. Microbiol. 2006. 60:477–501

101. Rothe et al, Transmission of 2019-nCoV Infection from an Asymptomatic Contact in Germany, NEJM, Jan 20, 2020

102. Russo et al, Adaptive mutability of colorectal cancers in response to targeted therapies, Science, 20 Dec 2019

103. Sanjuan et al, Viral Mutation Rates, Journal of Virology, Oct. 2010, p. 9733–9748

104. Sharples et al, Potential Risks and Benefits of Gain-of-Function Research: Summary of a Workshop, The National Academies Press at http://www.nap.edu/catalog.php?record_id=21666 , 2015

105. Shrivastav, M., et al, Regulation of DNA double-strand break repair pathway choice, Cell Research (2008) 18:134-147.

106. Simpson et al, Quantitative Zoology, Dover, 2003

107. Slechta et al, Adaptive mutation: General mutagenesis is not a programmed response to stress but results from rare coamplification of dinB with lac, PNAS October 28, 2003 100 (22) 12847-12852

108. Smith, Evolutionary Genetics, Oxford, 1989

109. Smith, The not-so-infinite malleability of RNA viruses: Viral and cellular determinants of RNA virus mutation rates, PLoS Pathog 13(4): e1006254. https://doi.org/10.1371/journal.ppat.1006254 2017.

110. Smolarz et al, Polymorphism of DNA repair genes in breast cancer, Oncotarget, 2019, Vol. 10, (No. 4), pp: 527-535

111. Sompayrac, How the Immune System Works, 5th Ed, Wiley, 2016

112. Sungnak et al, SARS-CoV-2 Entry Genes Are Most Highly Expressed in Nasal Goblet and Ciliated Cells within Human Airways, Wellcome Sanger Inst, Cambridge, UK, Feb 2020

113. Takeda and Akira, Toll-like receptors in innate immunity, International Immunology, Vol. 17, No. 1, pp. 1–14, 2005

114. Talmadge and Fidler, AACR Centennial Series: The Biology of Cancer Metastasis: Historical Perspective, AACR, Cancer Res; 70(14) July 15, 2010

115. Tan et al, Respiratory Viral Infections in Exacerbation of Chronic Airway Inflammatory Diseases: Novel Mechanisms and Insights from the Upper Airway Epithelium, Front in Cell and Dev Bio, Feb 2020

116. Torgovnick, A., B. Schumacher, DNA repair mechanisms in cancer development and therapy, Frontiers in Genetics, 1 April 2015, Volume 6, Article 157

117. Turgeon et al, DNA Damage, Repair, and Cancer Metabolism, Front in Oncology, 05 Feb 2018

118. Urganci et al, Colorectal Cancer due to Constitutional Mismatch Repair Deficiency Mimicking Neurofibromatosis I, Pediatrics, Volume 136, number 4, October 2015

119. Velic et al, DNA Damage Signalling and Repair Inhibitors: The Long-Sought-After Achilles' Heel of Cancer, Biomolecules 2015, 5, 3204-3259

120. Wang et al, Discovery of a subgenotype of human coronavirus NL63 associated with severe lower respiratory tract infection in China, 2018, in Emerging Microbes & Infections, Taylor and Francis, Jan 2020

121. Wang et al, International expansion of a novel SARS-CoV-2 mutant, medRxiv preprint doi: https://doi.org/10.1101/2020.03.15.20035204 .

122. Weinberg, R., The Biology of Cancer, Garland (New York) 2007.

123. Wolfel et al, Virological assessment of hospitalized cases of coronavirus disease 2019, medRxiv preprint doi: https://doi.org/10.1101/2020.03.05.20030502

124. Wu et al, A new coronavirus associated with human respiratory disease in China, Nature, Jan 28, 2020

125. Wu et al, Risk Factors Associated With Acute Respiratory Distress Syndrome and Death in Patients With Coronavirus Disease 2019 Pneumonia in Wuhan, China, JAMA Internal Medicine, March 13, 2020

126. Zeng, et al, Bat Severe Acute Respiratory Syndrome-Like Coronavirus WIV1 Encodes an Extra Accessory Protein, ORFX, Involved in Modulation of the Host Immune Response, Journal of Virology, July 2016

127. Zhao et al, Mismatch repair status and high expression of PD-l1 in nasopharyngeal carcinoma, Cancer Management and Research, 2019, 11 1631–1640

128. Zhou et al, A pneumonia outbreak associated with a new coronavirus of probable bat origin, Nature, 12 March 2020

129. Zhou et al, Back to the spring of Wuhan: facts and hope of COVID-19 outbreak, Frontiers Medicine, 21 Feb 2020

130. Zhu et al, A Novel Coronavirus from Patients with Pneumonia in China, 2019, NEJM Jan 24, 2020

131. Ziebuhr, The Coronavirus Replicase, CTMI (2005) 287:57—94, Springer-Verlag 2005

9 INDEX