

A Multi-Pronged Investigation of the SARS-CoV-2 Coronavirus

L. Van Warren
UALR Bioinformatics
April 6, 2020

INTRODUCTION	4
COMPARISON OF PANDEMIC VIRUSES	4
WHERE IS CV IN THE PHYLOGENETIC TREE OF RNA VIRUSES?	4
COMPARATIVE GENOMICS VIA BLAST	8
14-SEQUENCE ANALYSIS	8
41-SEQUENCE ANALYSIS	9
142-SEQUENCE ANALYSIS	9
CONSTRUCTING A REFERENCE SEQUENCE USING AN INTERACTIVE SIEVE	10
SIEVE MOP UP	11
516-SEQUENCE ANALYSIS	11
SPECIFIC ACCOMPLISHMENT: SIMILARITY TO SARS AND MERS-COV	11
MOP UP: COMPARISON WITH 96% SCRIPPS BAT VIRUS RATG13	12
PROTEOMICS	12
TARGETING THE ACE2 RECEPTOR	14
INTERDICTION OF VIRAL INFECTION AT THE ACE2 BINDING SITE	14
PHYRE2 <i>IN-SILICO</i> STRUCTURE PREDICTION AND SEQUENCE COMPARISON	15
INTERACTION OF THE SPIKE (S) PROTEIN AND THE ACE2 RECEPTOR	16
IMPORTANT LESSONS FROM THE WUHAN REPORT	18
SPECIFIC ACCOMPLISHMENT: ANNOTATION OF THE REPORT	18
THE CYTOKINE STORM SYNDROME	18
BLOOD PRESSURE DYSREGULATION	18
COMPETING WITH THE VIRUS VIA ACE2 INHIBITORS AND AGONISTS	19
STATISTICAL ANALYSIS OF CV INFECTION RATES AND MORTALITY	20
APPARENT VS. ACTUAL INFECTION AND MORTALITY	20
SPECIFIC ACCOMPLISHMENT: FAST ESTIMATION OF INFECTION AND MORTALITY	20
KINSA ATYPICAL ILLNESS FEVER MAP OF APRIL 2, 2020	22
CYCLICITY	23
DECISION-MAKING IN A PANDEMIC IS NOT LIKE TRADITIONAL POLITICAL DECISION MAKING	25
PLEASING THE MEAN	26
TECHNICAL COMMUNICATION IN CRISIS	26
LATENCY IN TRADITIONAL COMMUNICATION PATHWAYS	26
FUTURE WORK	27

APPLICATION OF MACHINE LEARNING	27
COMMAND AND CONTROL FOR VIRAL ANALYSIS	27
REFERENCES AND BIBLIOGRAPHIC ENTRIES	27
ACKNOWLEDGEMENTS	27
APPENDIX 1: DATA SOURCES FOR EPIDEMIOLOGICAL DATA	28
ABBREVIATIONS / GLOSSARY / TOOLS	28
Figure 1 - Comparison of Pandemic Viruses	4
Figure 2 - Root of Viral Phylogenetic Tree	5
Figure 3 - Just RNA Viruses	5
Figure 4 - In Nidovirales Bats Appear	6
Figure 5 - For Coronavirineae Chickens & Pigs Trade Places and More Bats Appear	6
Figure 6 – Descending to Betacoronavirus Humans Retake the Lead and More Bats Appear	7
Figure 7 - List of 14 FASTA files and NCBI Accession numbers for Coronavirus SARS-CoV-2	8
Figure 8 - Visualization of Precision Alignment of 14 Initial CV Sequences	9
Figure 9 - Visualization of Rapid Alignment of 41 CV Sequences	9
Figure 10 - Visualization of Rapid Alignment of 142 CV Sequences	9
Figure 11 - A Canonical 50 Sequence Reference List	10
Figure 12 - Similarity in Structural Proteins of Bat and Human Coronaviruses	11
Figure 13 - Gene Components of SARS-CoV-2 derived from GenBank: MN908947.3	12
Figure 14 - Nidovirus Life-Cycle from Ziebuhr, et. al.	13
Figure 15 - Organization of SARS-CoV Genome from Fehr and Perlman	13
Figure 16 - Betacoronavirus Receptors after Fehr and Stanley	14
Figure 17 - Analysis 1 Recovering a papain-like protease	15
Figure 18 - Analysis 2 - Recovering nsp12 Sars-Cov-2 Protein	16
Figure 19 - Chimera Rendering of 6ACG Coronavirus Spike Protein Interacting with ACE2	17
Figure 20 - Renin-Angiotensin-Aldosterone Organ Level Pathway for BP Regulation	19
Figure 21 - Derivation of Fast Estimator for Exponential Growth	21
Figure 22 - CV Infection and Mortality Forecasts for April 14 Made March 30, 2020	22
Figure 23 - CV Infection and Mortality Forecasts for April 14 Made April 4, 2020	22
Figure 24 - Contrast Enhanced Kinsa Fever Map of April 2, 2020	23
Figure 25 - CDC.gov - 1918 Pandemic Influenza: Three Waves	24
Figure 26 - Singh, and Adhikari – Policy Dependent Forecast of India's Covid-19 Epidemic	25

Introduction

With the advent of the COVID-19 pandemic our department undertook a multi-pronged attack to investigate the virus, its rapid spread, and a review of clinical lessons learned. Each thread of this attack is enumerated below.

Comparison of Pandemic Viruses

The figure below shows the date of a few significant pandemics, the viral formal name, genus, Baltimore Type Classification, Transmission (zoonotic except for smallpox), the size of the sequenced virus in nucleotides, and the geometric size of the virus in nanometers

Date	Viral Pandemic	Formal Name	Genus	Type	Transmission	Size nucleotides	Size nm
1519	Smallpox Yellow Fever	Variola	Orthopoxvirus	dsDNA	human→human	188251	292
1647	Fever	YFV	Flavivirus	(+)-ssRNA	mosquito→human	11564	50
1889	Russian Flu	H2N2	Influenza A	(-)-ssRNA	bird→pig→human	13460	130
1918	Spanish Flu	H1N1	Influenza A	(-)-ssRNA	bird→pig→human	13158	130
1957	Asian Flu	H2N2	Influenza A	(-)-ssRNA	bird→pig→human	13460	130
1968	Hong Kong Flu	H3N2	Influenza A	(-)-ssRNA	bird→pig→human	13498	130-100-
1981	AIDS	HIV-1	Lentivirus	(+)-ssRNA	simian→human	15524	140
2002	SARS	SARS-CoV	Betacoronavirus	(+)-ssRNA	bats→human	29874	80-400
2009	Swine Flu	H1N1	Influenza D	(-)-ssRNA	bird→pig→human	12906	130
2012	MERS	MERS-CoV	Betacoronavirus	(+)-ssRNA	camel→human	30484	60-140
2014	Ebola	EBOV	Ebolavirus	(-)-ssRNA	bats→human	19897	80
2019	COVID-19	SARS-COV-2	Betacoronavirus	(+)-ssRNA	bats→human	29945	60-140

Figure 1 - Comparison of Pandemic Viruses

As RNA viruses go, SARS-Cov-2 (CV hereafter), is relatively large at ~30,000 bases over twice the size of influenza viruses in nucleotide bases. The smallpox variola virus is an order of magnitude larger than the other pandemic viruses listed.

Where is CV in the [phylogenetic tree of RNA viruses](#)?

In the great scheme of viral taxonomy CV lives here:

Viral Taxa □ RNA Viruses □ Nidovirales □ Coronidovirineae □ Coronaviridae □
Orthocoronavirinae □ Betacoronavirus □ Sarbecovirus □ [SARS-CoV-2](#)

An interesting set of hosts pivots take place as we descend down the tree. At the beginning humans are the leading candidate for all viral hosts:

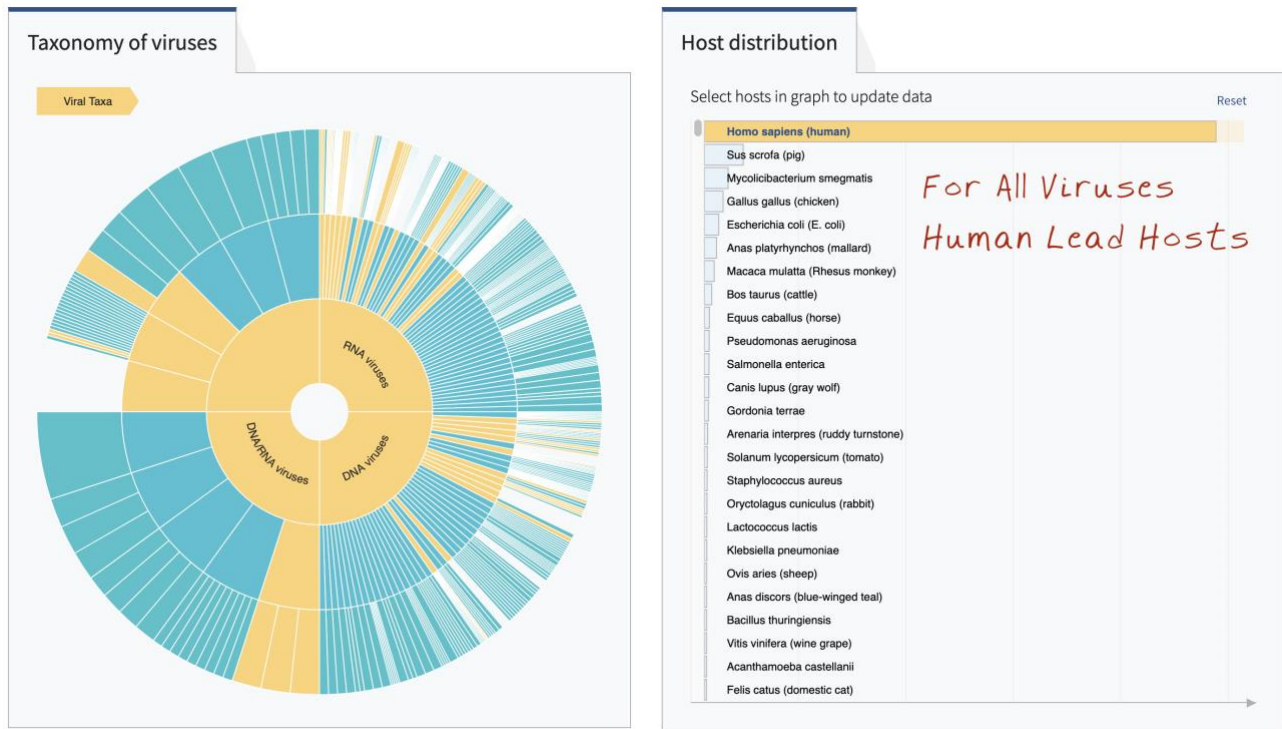


Figure 2 - Root of Viral Phylogenetic Tree

If we then limit ourselves to just RNA viruses, humans remain the principal host.

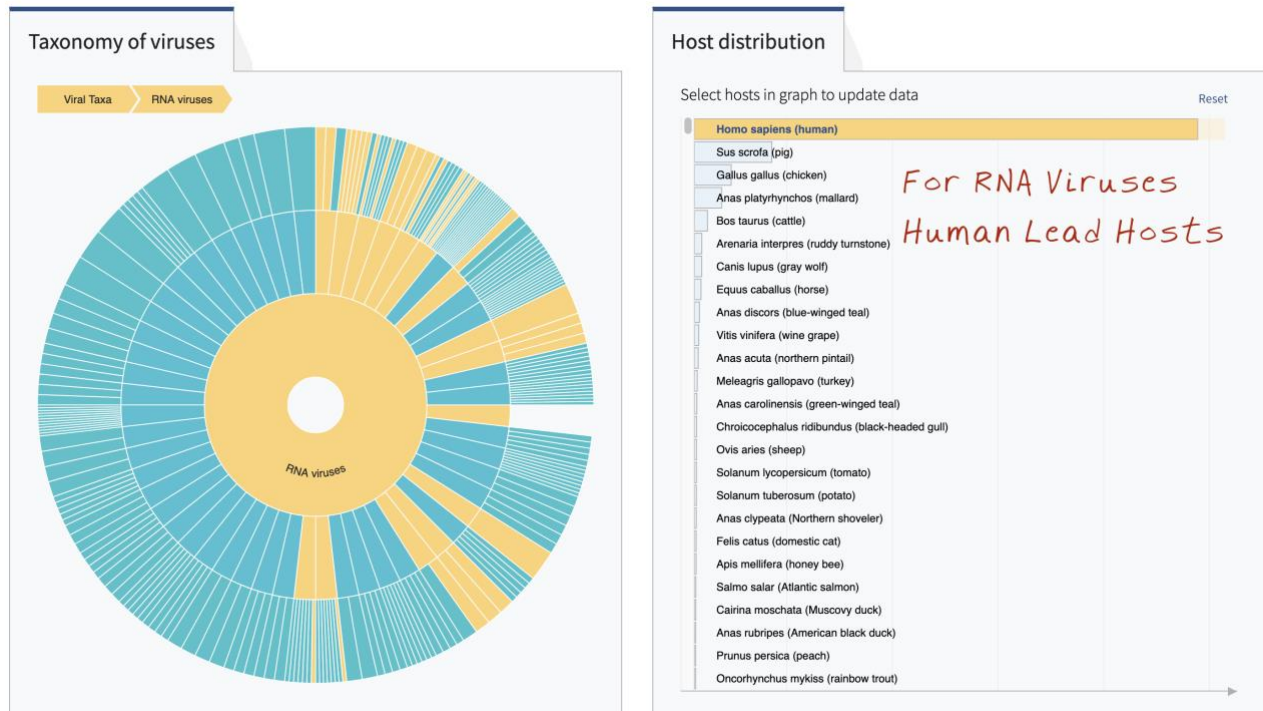


Figure 3 - Just RNA Viruses

As we enter the sector Nidovirales, human hosts are eclipsed by pigs and chickens, and we see the appearance of bats in the list of hosts.

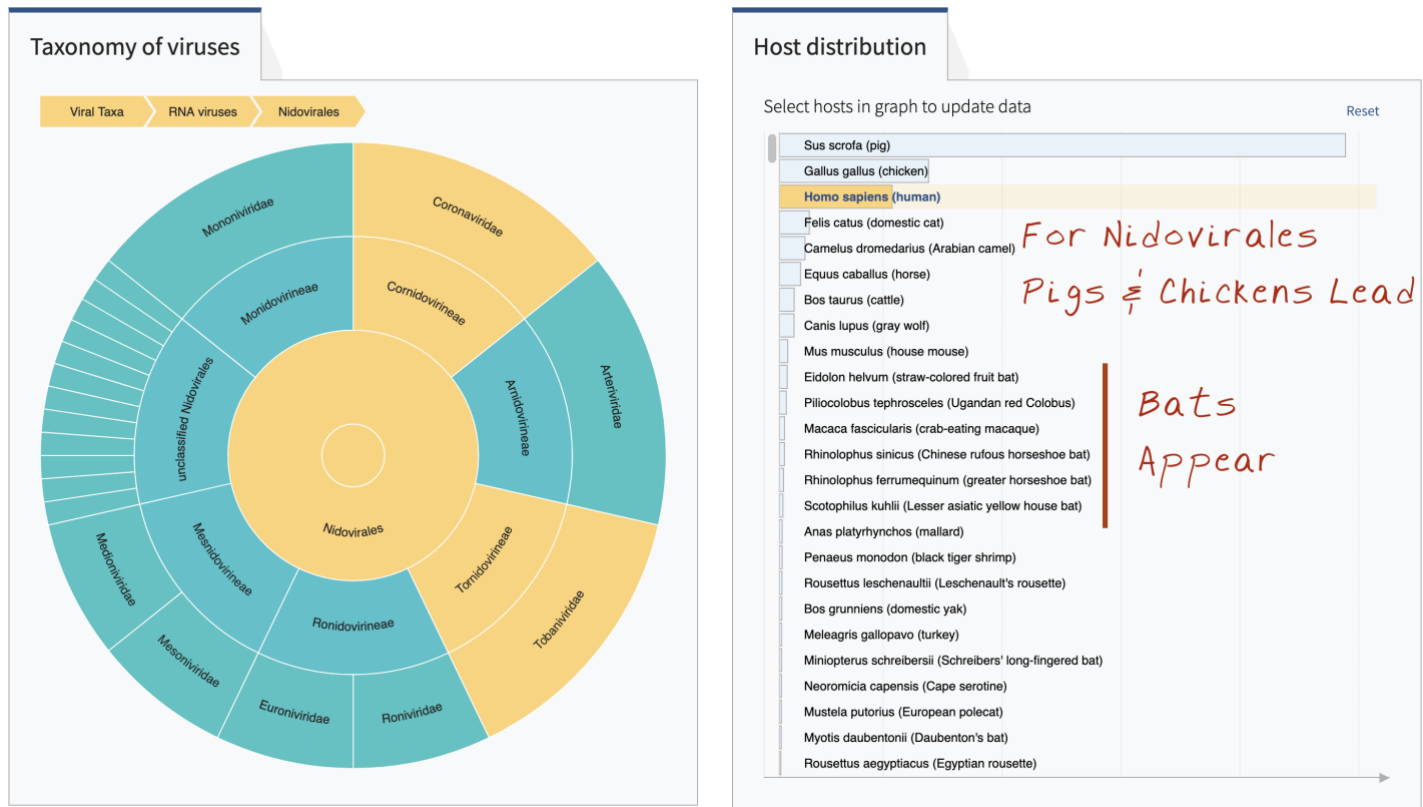


Figure 4 - In Nidovirales Bats Appear

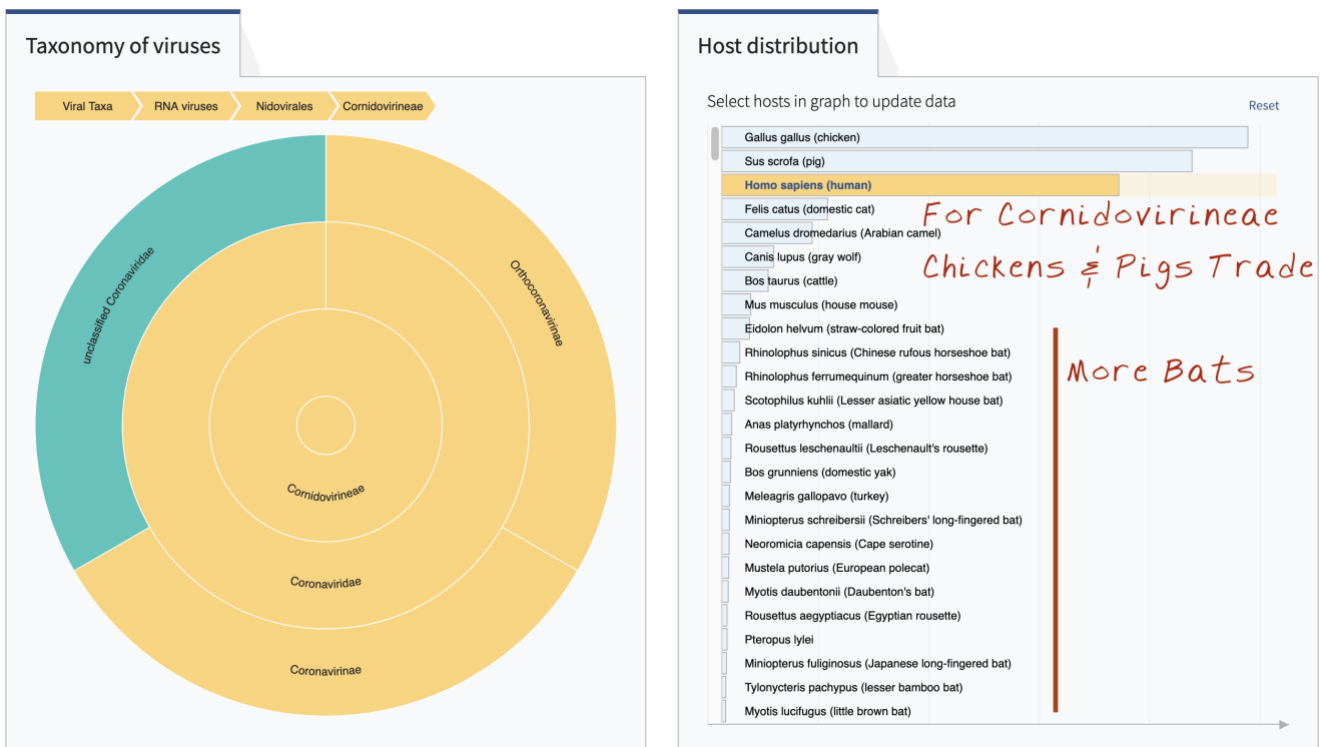


Figure 5 - For Cornidovirineae Chickens & Pigs Trade Places and More Bats Appear

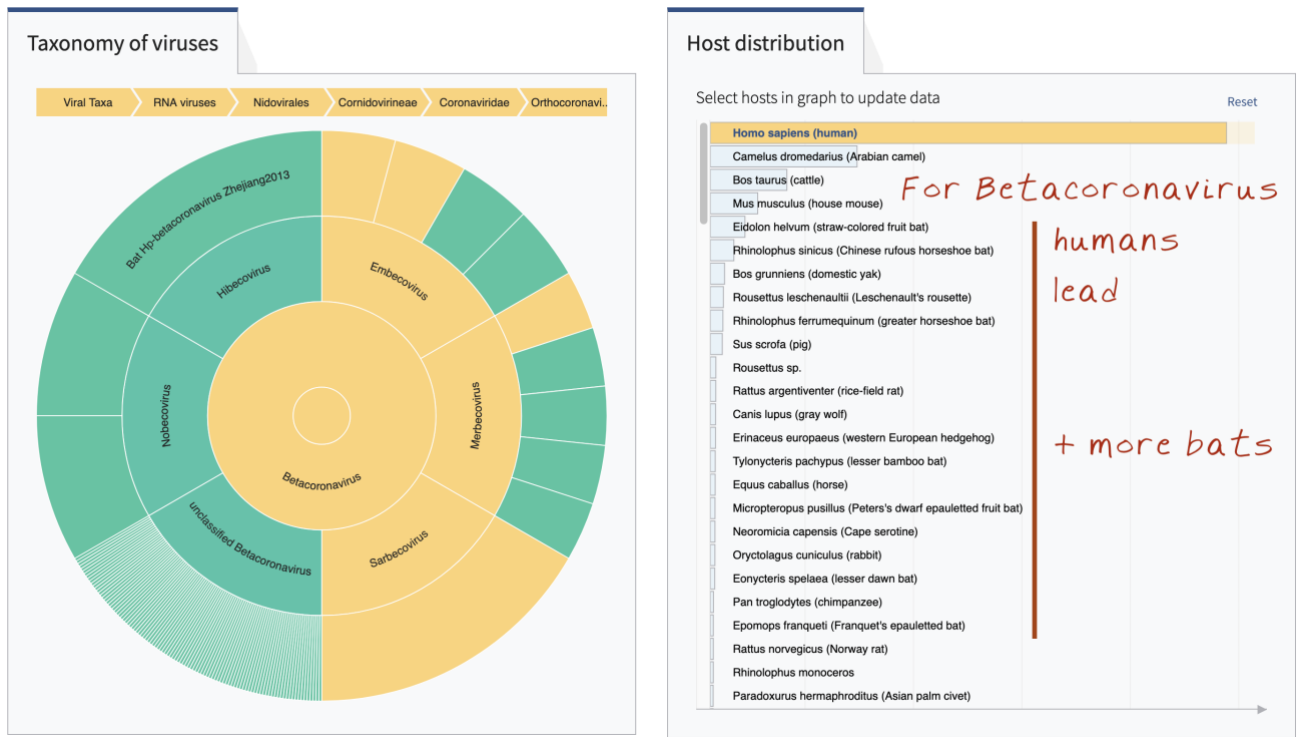


Figure 6 – Descending to Betacoronavirus Humans Retake the Lead and More Bats Appear

Now that we have located CV in the tree, let's analyze some specific viral sequences.

Comparative Genomics via BLAST

14-Sequence Analysis

Using BLAST-nucleotide (BLASTn) fourteen initially delivered CV sequences were analyzed. Ten of these were provided as accession numbers with companion FASTA files. Four carried WHO labeling. The size field is the actual size of the fasta file that contains the nucleotides. They are encoded as CTAG even though RNA bases are CUAG. These initial files use uppercase for all nucleotide bases, an issue that emerged later on in larger datasets that contain amino acids. All of these files were either identical or similar in both Coverage and Identity except for the WHO2 file that only had 65% coverage. The list of files is sorted by a nearly exact indicator of the size of this positive-sense single-stranded virus. For these accession numbers the file size is typically 17 units larger than the actual number of nucleotides, the reason being the name of the sequence and a terminal newline are embedded in the file.

Size	Date	Fasta file	NCBI Accession	Q-Cvr	%Ident.	Locale	
30912	Mar 13	NMDC60013002-09.fasta	NC_045512.2	100%	99.99%	Wuhan	SARS CV 2
30907	Mar 13	NMDC60013002-10.fasta	MN996528.1	100%	99.99%	Wuhan	SARS CV 2
30907	Mar 13	NMDC60013002-08.fasta	MN996528.1	100%	100.00%	Wuhan	SARS CV 2
30907	Mar 13	NMDC60013002-06.fasta	MN996528.1	100%	100.00%	Wuhan	SARS CV 2
30906	Mar 13	NMDC60013002-07.fasta	MN996528.1	100%	99.99%	Wuhan	SARS CV 2
30865	Mar 13	NMDC60013002-03.fasta	LR757996.1	100%	100.00%	Wuhan	seafood market
30857	Mar 13	NMDC60013002-04.fasta	LR757995.1	99%	100.00%	Wuhan	seafood market
30848	Mar 13	NMDC60013002-01.fasta	LR757998.1	100%	100.00%	Wuhan	seafood market
30187	Mar 13	2019-nCoV_WH04.fasta	LR757995.1	99%	100.00%	Wuhan	seafood market
30183	Mar 13	2019-nCoV_WH03.fasta	MT019532.1	100%	100.00%	Wuhan	seafood market
30181	Mar 13	2019-nCoV_WH01.fasta	LR757998.1	100%	100.00%	Wuhan	seafood market
29698	Mar 13	2019-nCoV_WH02.fasta	LR757997.1	65%	100.00%	Wuhan	seafood market
27891	Mar 13	NMDC60013002-05.fasta	MT020781.2	99%	99.08%	Finland	SARS CV 2
19925	Mar 13	NMDC60013002-02.fasta	LR757997.1	100%	100.00%	Wuhan	seafood market

Figure 7 - List of 14 FASTA files and NCBI Accession numbers for Coronavirus SARS-CoV-2

The initial alignment was performed using a local version of the well-known *mafft* program which is available in both online and desktop versions. There are two options, one to do a fast, but rough alignment, the other to do an exhaustive alignment. The fast but rough alignment runs in just a few seconds on a typical desktop PC. The exhaustive alignment took 18 minutes and 47 seconds running in a single thread and utilized 10.1 Gigabytes of random-access memory in processing these 14 sequences. Each sequence was compared with every other sequence. This image is hot linked for viewing at higher resolution.

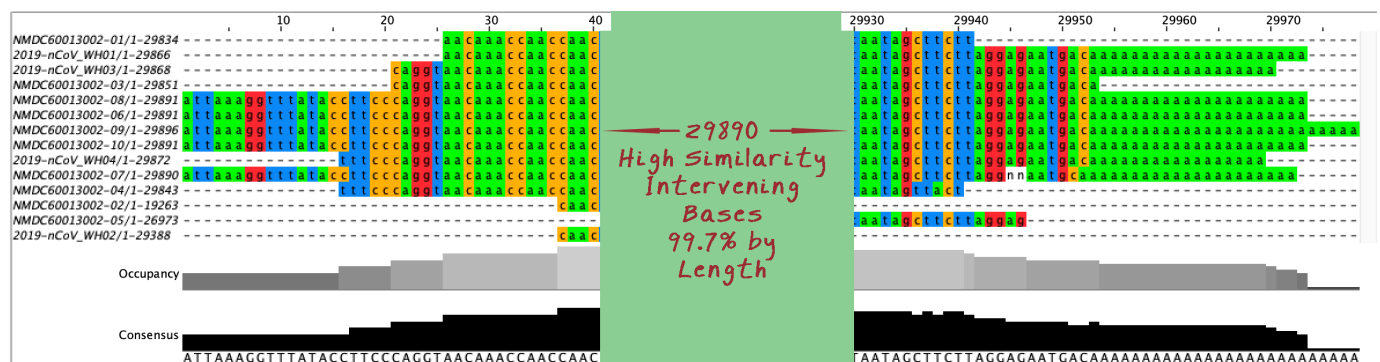


Figure 8 - Visualization of Precision Alignment of 14 Initial CV Sequences

The sequences differ mostly at their 5' leaders and 3' polyadenylated tails with relatively few insertions or deletions. The only exception was the WHO2 sequence, a noisy scaffold sequence that contained many unspecified bases. This pattern of leader and tail being the primary variation persisted as the number of sequences was increased as seen below.

41-Sequence Analysis

Subsequent to this more real time data was becoming available from worldwide sources. A second analysis was performed in a manner similar to the first but using the faster alignment method. The results showed the same significant leader and tail variation with little variation in the internal sequence. This image is also hot linked for higher resolution examination.

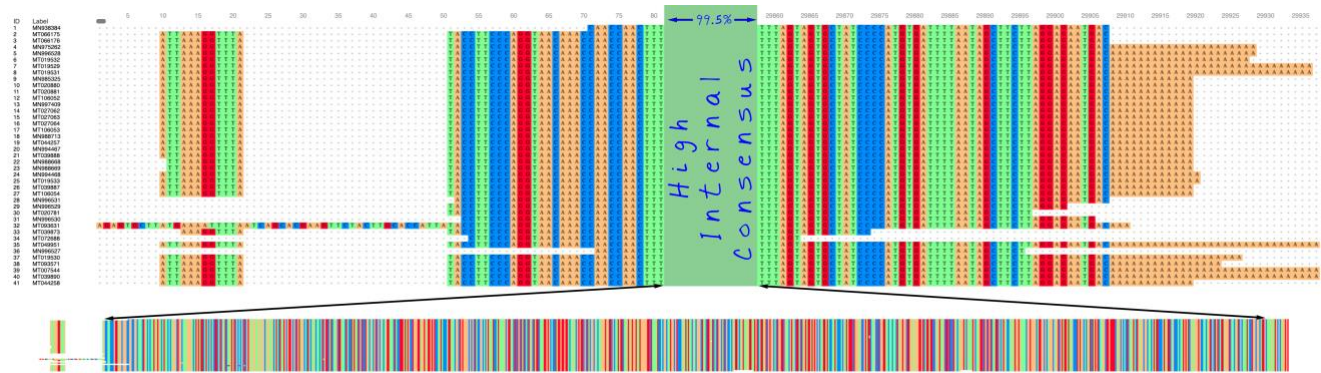


Figure 9 - Visualization of Rapid Alignment of 41 CV Sequences

142-Sequence Analysis

At this point the variation in the leaders and tails, on sorting, was becoming consistent as can be seen below. Sequences were arriving so rapidly that between the time the analysis was begun and ended, more had arrived.

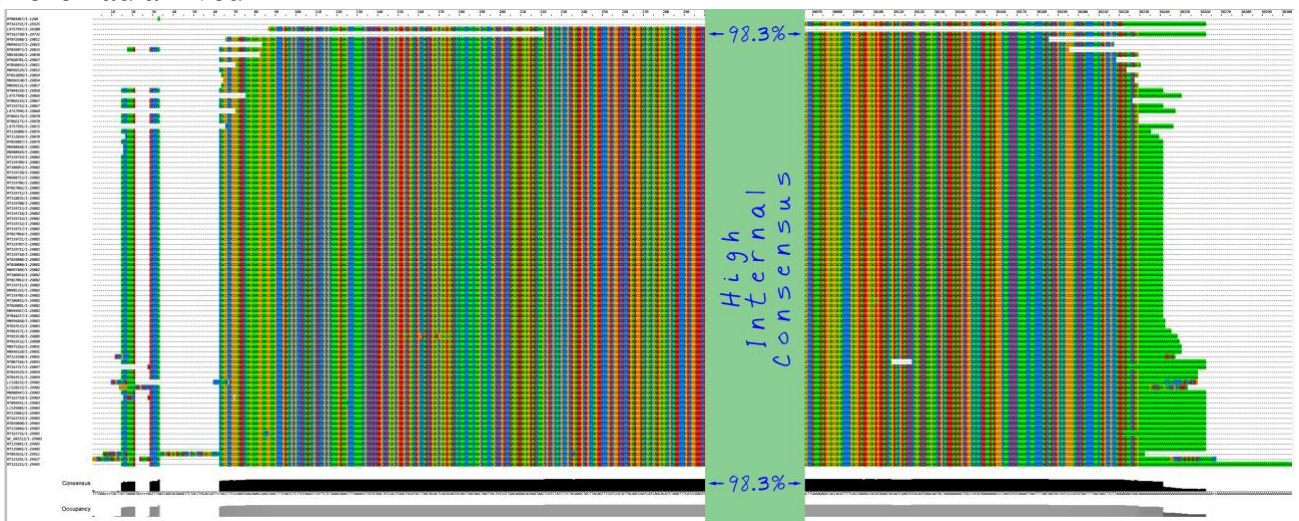


Figure 10 - Visualization of Rapid Alignment of 142 CV Sequences

As these sequences arrived it became desirable to construct a *reference sequence* that could be used for proteomic analysis. Understanding the significant mutations and separating strains by

virulence and disease impact will be important tasks for follow-on work. Given the explosion of a world-wide pandemic of historic proportion it seemed useful to 'freeze' a version of the virus that would lend itself to more detailed examination.

Constructing a Reference Sequence Using an Interactive Sieve

A sieve procedure devised to accomplish this was performed as follows. Using the interactive JalView the ragged ends of the 142 sequences were trimmed, and outlier sequences were eliminated. This was a completely 'human umpired' activity, but the interactive approach rapidly produced a representative/canonical set of 50 sequences of very high similarity which presumably representative of the most dangerous form of the virus.

1. Create 142Alignment.pir using MAFFT
2. Open 142Alignment.pir using JalView
3. Sort the sequences by length
4. Cull partial sequences
5. Cull sequences missing > 200 bases
6. Cull 2280-2285 deadbox (2 cases)
7. Cull 1695-1961 deadbox (89 cases)
8. Cull those whose leader deletion is > 60
9. Sort by bases 60-90 (undo this)
10. Cull those with 5' deletions > 60 bases
11. Cull those with 3' polyAtail < 12 A's
12. Trim polyAtails to 12 bases
13. Cull those with outlier tails (8 cases)
14. Cull those with outlier leaders
15. Cull those with missing internal regions
16. Cull missing header regions insertion alignment artifact
17. Cull those with missing internal regions
18. Cull 7387-7391 with internal insertion alignment artifact
19. Cull 1 atypical, noisy, mutated sequence
20. Cull 15804-15804 with internal insertion alignment artifact
21. Cull 15686-15641 with internal insertion alignment artifact
22. Cull 13039-13040 with internal insertion alignment artifact
23. Cull 14605-14608 with internal insertion alignment artifact
24. Cull 16578-16588 with AA
25. Cull 20658-20662 with AA
26. Cull one with Y base

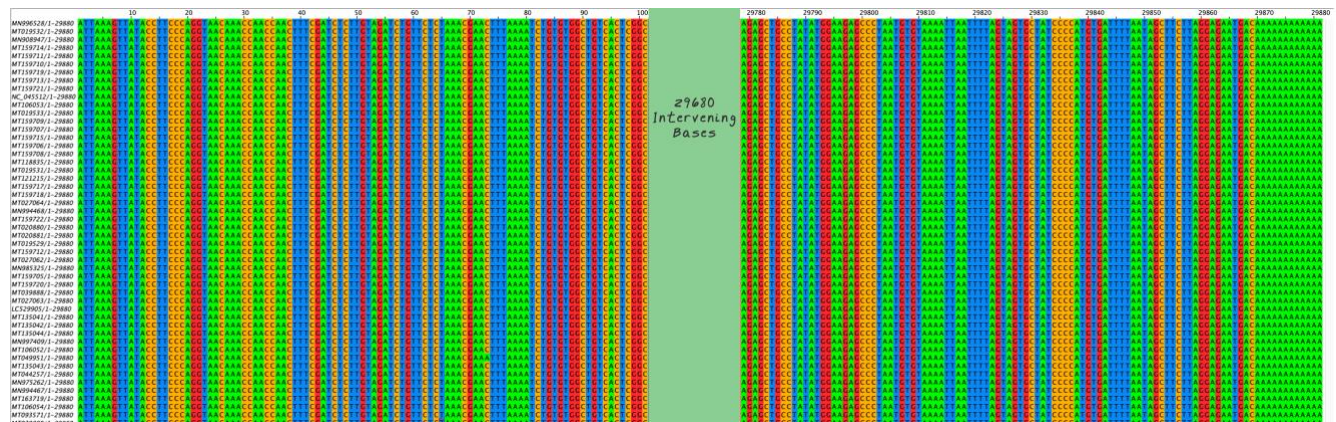


Figure 11 - A Canonical 50 Sequence Reference List

Sieve Mop Up

A quality control follow up or 'mop up' activity will be to realign the 50-sequence subset to confirm the accuracy and completeness of the interactive sieve.

516-Sequence Analysis

Around this time in our investigation we discovered a German sequence sharing database called [gisaid](#) that at that time (March 14, 2020) contained 516 sequences. Previous to this time gisaid had allowed batch downloads with a single button, but this feature was disabled, and we really needed the data. It was downloaded by hand one click at a time in an arduous process that had an unexpected windfall. Each time a sequence was downloaded, the meta data for that sequence was displayed. This created a sense of touring the world where CV had become entrenched and also showed those players who were on the front lines collecting and sequencing the samples. A number of different models of sequencers were used, with degrees of coverage that ranged widely from 0.9x to over 1000x. We also confirmed that the running time grows rapidly with the number of sequences being aligned and were not able to get a full 516 sequence run to terminate in a reasonable amount of time. Nonetheless we have not only the 516 sequences, but also metadata which can be used in later analysis.

Specific Accomplishment: Similarity to SARS and MERS-COV

As an incidental exercise we compared our CV sequences with SARS and MERS-COV. Running NCBI BLASTn comparisons showed 71% similarity of CV to bat SARS and 68% to MERS-COV. Downstream structural proteins were conserved between CV, SARS and MERS-COV. An interesting feature was that in our preliminary analysis the similar region of the human and bat coronaviruses were identical but contained in both cases a dissimilar leader region of about 29% of the length of the virus. We wondered by what process a virus could have taken this big a step.

Descriptions

Graphic Summary

Alignments

Taxonomy

Sequences producing significant alignments

Download Manage Columns Show 100

select all

16 sequences selected

GenPept

Graphics

	Description	Max Score	Total Score	Query Cover	E value	Per. Ident	Accession
☒	P1a [Bat SARS CoV Rf1/2004]	7142	7142	44%	0.0	79.71%	ABD75321.1
☒	ORF1 [Bat CoV 273/2005]	7122	12579	71%	0.0	79.74%	ABG47059.1
☒	P1b [Bat SARS CoV Rf1/2004]	5273	5273	26%	0.0	95.17%	ABD75322.1
☒	spike protein [Bat SARS CoV Rf1/2004]	1883	1883	12%	0.0	74.42%	ABD75323.1
☒	spike protein [Bat CoV 273/2005]	1880	1880	12%	0.0	74.27%	ABG47060.1
☒	nucleocapsid protein [Bat CoV 273/2005]	586	586	4%	0.0	89.79%	ABG47067.1
☒	nucleocapsid protein [Bat SARS CoV Rf1/2004]	580	580	4%	0.0	89.07%	ABD75315.1
☒	membrane protein [Bat SARS CoV Rf1/2004]	413	413	2%	2e-134	89.95%	ABD75314.1
☒	ORF3a [Bat SARS CoV Rf1/2004]	405	405	2%	8e-131	68.73%	ABD75316.1
☒	ORF7a [Bat SARS CoV Rf1/2004]	195	195	1%	6e-60	86.07%	ABD75319.1
☒	ORF9a [Bat SARS CoV Rf1/2004]	132	132	0%	2e-38	70.10%	AAZ67037.1
☒	ORF6 [Bat SARS CoV Rf1/2004]	93.6	93.6	0%	5e-25	68.85%	ABD75318.1
☒	envelope protein [Bat SARS CoV Rf1/2004]	87.8	87.8	0%	9e-23	93.42%	ABD75313.1
☒	ORF9b [Bat SARS CoV Rf1/2004]	81.3	81.3	0%	1e-20	78.00%	AAZ67038.1
☒	putative ORF8 [Bat CoV 273/2005]	50.4	50.4	1%	5e-09	30.91%	ABG47066.1
☒	ORF8 [Bat SARS CoV Rf1/2004]	50.1	50.1	1%	5e-09	30.91%	AAZ67036.1

Figure 12 - Similarity in Structural Proteins of Bat and Human Coronaviruses

Mop Up: Comparison with 96% Scripps Bat Virus RaTG13

The [Scripps paper in Nature](#) on March 17, 2020 contained the assertion: "Although RaTG13, sampled from a *Rhinolophus affinis* bat, is ~96% identical overall to SARS-CoV-2, its spike diverges in the RBD, which suggests that it may not bind efficiently to human ACE2." This meant that for transfer to humans only a 4% change, probably in the spike protein was required. This seemed to limit the size of the step necessary from our initial analysis which increased the plausibility of incremental zoonotic transfer. This whole process underscores the importance of secure curation of sequences along with their timestamps for entry into the database. A nefarious actor could engineer a virus, enter it into the database after the fact and then claim it was a naturally occurring one. This is not being suggested here, but the possibility cannot be ruled out without careful curation, possibly involving blockchain or other secure distributed ledger technologies used by cryptocurrencies.

Proteomics

Components of SARS-CoV-2:

gene/UTR	start	stop	product	protein id	notes
5' UTR	1	265	start of transcription		
ORF1ab	266	21555	replicase polyprotein1a polyprotein1b	QHD43415.1	pp1ab; translated by -1 ribosomal frameshift
S	21563	25384	Spike entry point glycoprotein	QHD43416.1	structural protein; binds to host ACE2 receptor
ORF3a	25393	26220	ORF3a protein	QHD43417.1	
E	26245	26472	Envelope protein	QHD43418.1	structural protein; ion channel activity
M	26523	27191	Membrane glycoprotein	QHD43419.1	structural protein
ORF6	27202	27387	ORF6 protein	QHD43420.1	
ORF7a	27394	27759	ORF7a protein	QHD43421.1	
ORF8	27894	28259	ORF8 protein	QHD43422.1	
N	28274	29533	nucleocapsid phosphoprotein	QHD43423.2	structural protein; Binds nsp3 and M protein, tethers viral genome to replicase-transcriptase complex, changes structure when phosphorylated.
ORF10	29558	29674	ORF10 protein	QHI42199.1	
3' UTR	29675	29903	end of transcription		includes poly-(A) tail

Figure 13 - Gene Components of SARS-CoV-2 derived from GenBank: [MN908947.3](#)

In the paper, ' [Virus-encoded proteinases and proteolytic processing in the Nidovirales](#)', Ziebuhr et. al. enumerate replication strategies in the life-cycle of an entire class of related viruses. Quoting from their paper:

both the virus, and of vaccine-related efforts:

Betacoronavirus	Receptor	References
MHV	mCEACAM	[121,122]
BCoV	N-acetyl-9-O-acetylneuraminic acid	[123]
SARS-CoV	ACE2	[124]
MERS-CoV	DPP4	[100]

Figure 16 - Betacoronavirus Receptors after Fehr and Stanley

There is a specific, [high-affinity association](#) between the S Spike Entry Protein of the SARS-CoV and ACE2. Syncytia or, globs of infected host cells, can be formed when viral particles enable cell-to-cell fusion via ACE2 receptors. There is 100,000-fold viral replication within infected cells in 48 hours in culture. We can use our mathematical derivation below to compute virus doubling time based on this to be 2.9 hours and the confirmation of this is that $2^{(48/2.89)}$ is $\sim 100,000$.

ACE2 is an essential regulator of cardiac function which gives rise to our request for blood pressure data in the second below.

Targeting the ACE2 Receptor

In their [letter to Nature of 2003](#) following the SARS outbreak, Li et. al. determined that 'Angiotensin-converting enzyme 2 is a functional receptor for the SARS coronavirus'. They noted that Anti-ACE2 antibody inhibited cell pathology in a dose-dependent manner. Not only was their clever methodology noteworthy, but for all players in the vaccine chain, they provide a direct path for vaccine development by identifying unambiguously optimal first targets that include the S1 component of the spike protein.

Interdiction of Viral Infection at the ACE2 Binding Site

This passage written 17 years ago remains especially important today. Quoting from their letter:

A number of antibodies, peptides and small compounds have been shown to bind to ACE2 (refs 29, 30). It is possible that some of these may be useful in the treatment of SARS, either by blocking the S-protein-binding site, or by inducing a conformation in ACE2 that is unfavorable to binding or fusion. Alternatively, a soluble form of the receptor itself may slow viral replication in an infected individual. Identification of ACE2 as a SARS-CoV receptor will facilitate description of the receptor-binding domain of the S protein, presumably the most effective target epitope of an S1-protein-based subunit vaccine. Also, it is likely that a cell line approved for vaccine production and made permissive for viral replication by ACE2 expression will be the most efficient large-scale producer of a whole-killed or attenuated virus for use as a vaccine. A mouse transgenic for human ACE2 may be useful as an animal model of SARS. Finally, study of the interaction between the SARS-CoV S protein and ACE2 of other animals may

provide insights into the origins of the virus. Thus, if SARS returns as a threat to human health, these studies may contribute to its control.

Phyre2 *in-silico* Structure Prediction and Sequence Comparison

The Phyre2 protein folding server allows simulations of plausible protein-foldings to occur *in-silico* by comparing nucleotide base files with existing structure entries. The service places a limit is 3600 base. To address this, we partitioned a section of the whole viral genome into two overlapping pieces, a [3600-base leader](#), [a 3600-base tail](#), The links embedded in the text dispatch the reader into two complete analyses. We recovered from the first analysis a SARS coronavirus papain-like protease with the structure of a viral deubiquitinating enzyme.

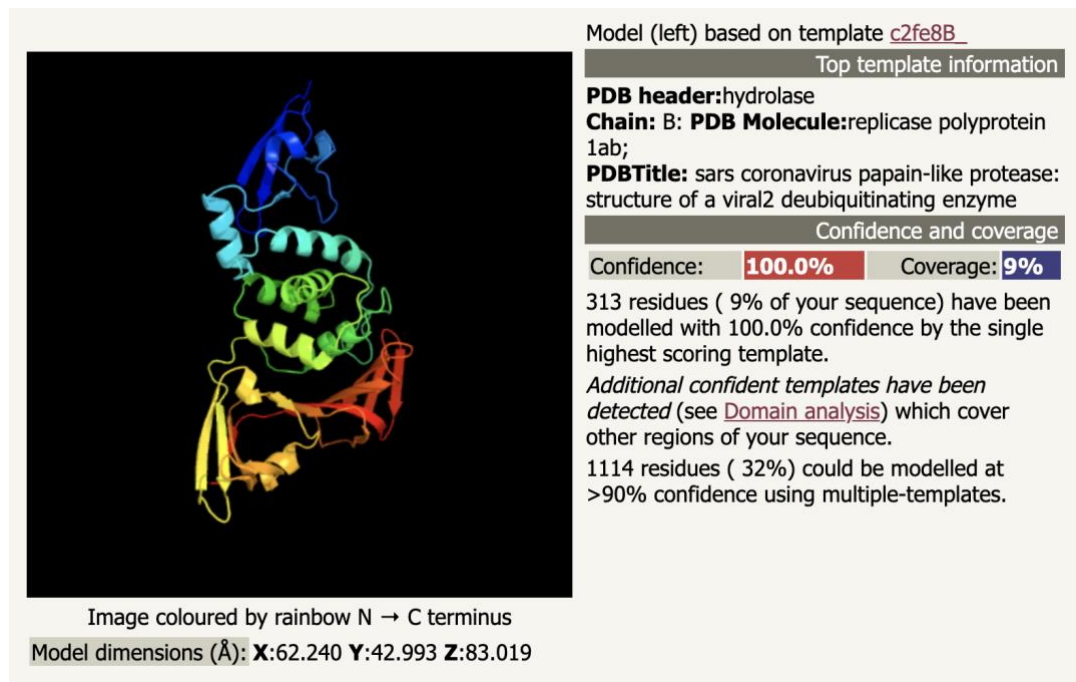


Figure 17 - Analysis 1 Recovering a papain-like protease

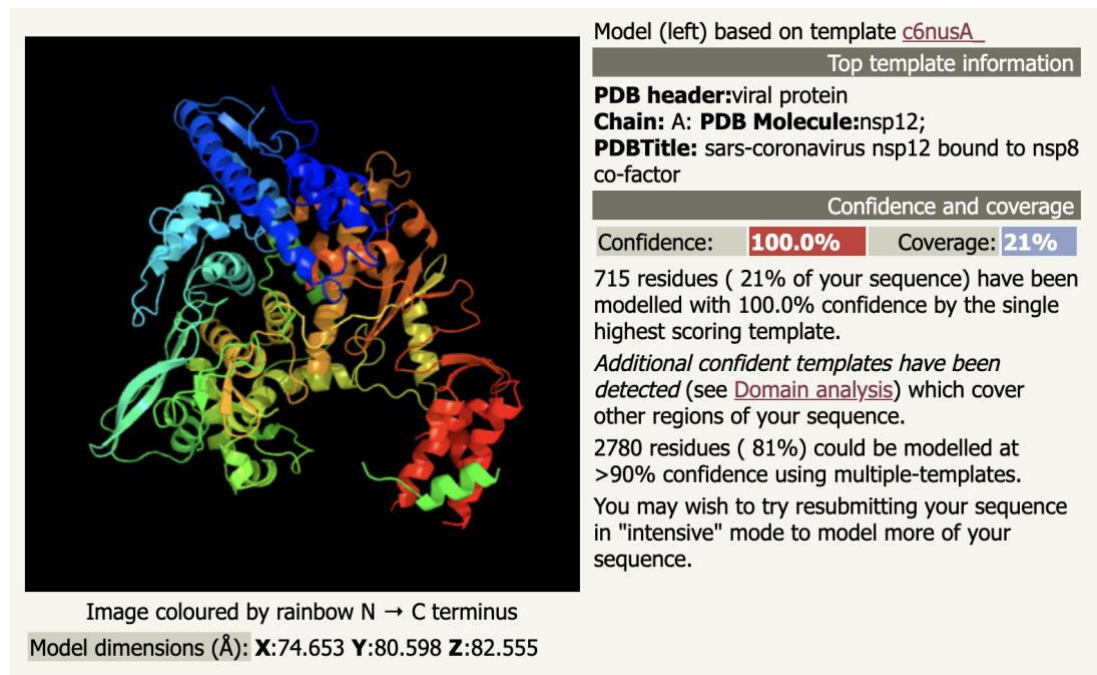


Figure 18 - Analysis 2 - Recovering nsp12 Sars-Cov-2 Protein

Interaction of the Spike (S) Protein and the ACE2 Receptor

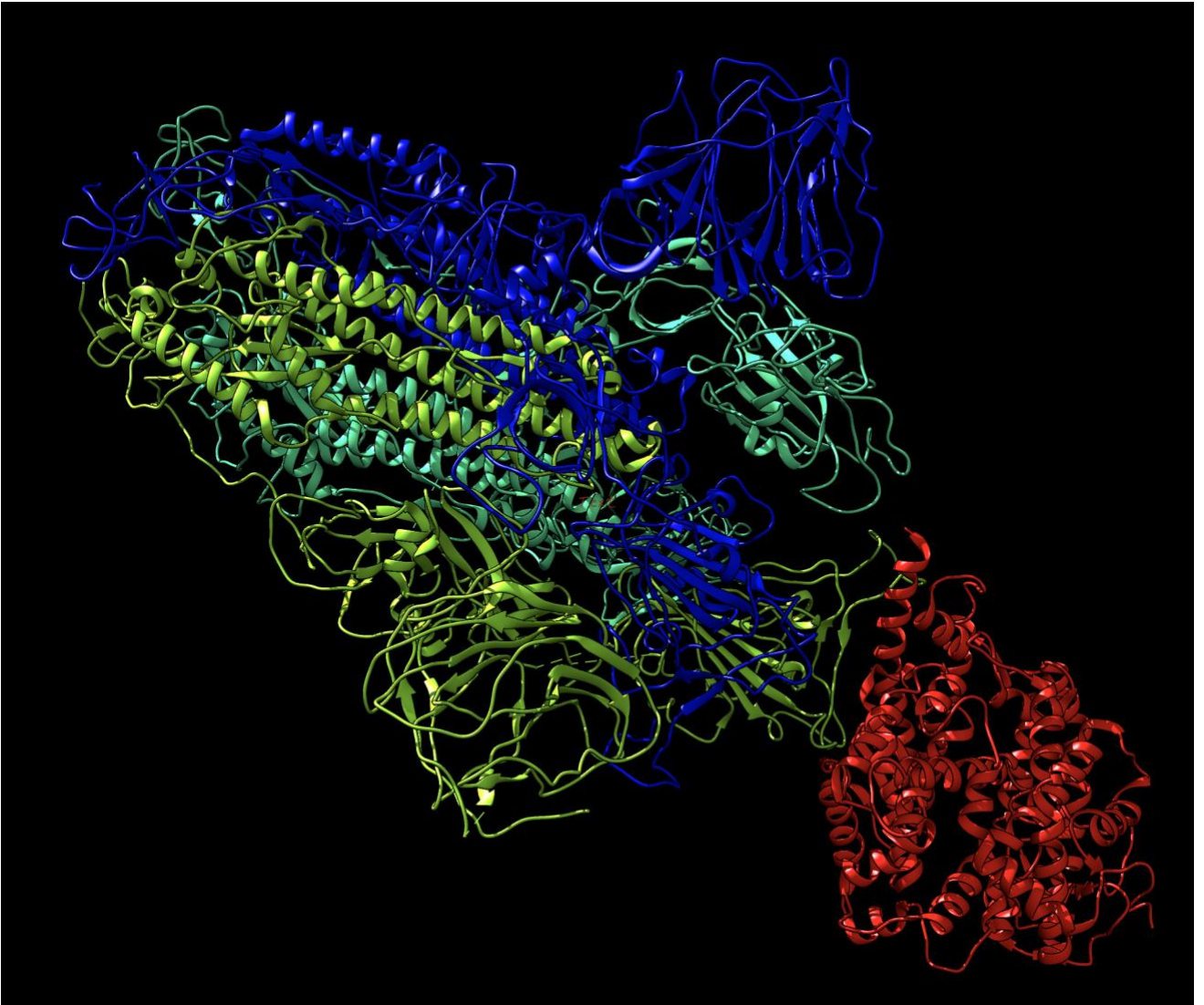


Figure 19 - Chimera Rendering of [6ACG](#) Coronavirus Spike Protein Interacting with ACE2

Important Lessons from the Wuhan Report

Specific Accomplishment: Annotation of the Report

The full name of the Wuhan Report is the Chinese Clinical Guidance for COVID-19 Pneumonia, Diagnosis and Treatment (7th Edition) published by China National Health Commission on March 4, 2020.

For readability the report was converted to Microsoft Word and Adobe PDF formats. To improve accessibility for the general audience each term of art in these two versions was hot-linked to curated definitions. The report is an excellent resource for BINF, health-care providers, and epidemiologists.

The Cytokine Storm Syndrome

Section 7 of the report contained warning signals for severe and critically severe types of patients, citing from this section:

7. Warning signals for severe and critically severe types

a) Adults

- (1) Progressive decline in the number of peripheral lymphocytes;
- (2) Progressive increase in the levels of peripheral inflammatory biomarkers, such as [IL-6](#) and CRP;
- (1) Progressive increase in [lactic acid concentration](#);
- (2) Pulmonary lesions progress rapidly in a short time.

b) Children

- (1) Increased respiration rate;
- (2) Poor mental responsiveness and drowsiness;
- (3) Progressive increase in lactic acid concentration;
- (4) Imaging showed bilateral or multilobes infiltration and pleural effusion; or pulmonary lesions progress rapidly in a short time;

Puja Mehta, in [their Lancet correspondence](#), have used the term 'cytokine storm' to characterize a hyperinflammatory syndrome that is associated with a turn for the worse in Covid-19, MERS, and SARS patients. Quoting from the linked reference, "Such an episode is seen as an under-recognized condition leading to multi-organ failure. [...] In such conditions of hyperinflammation, immunosuppression can be beneficial using tocilizumab, an IL-6 receptor blocker."

The Wuhan report also cites the use and dosages of: [a-interferon](#), [lopinavir](#)/[ritonavir](#), [Ribavirin](#), [chloroquine phosphate](#), [Abidol](#), [methylprednisolone](#), and tocilizumab.

Blood Pressure Dysregulation

The pandemic virus SARS-CoV-2 uses its spike protein uniquely to enter respiratory epithelial and other somatic cells by binding to the ACE2 receptor. The long name of this receptor is the Angiotensin Converting Enzyme 2 and it is part of the renin-angiotensin-aldosterone

pathway that regulates blood pressure. ACE2 is [an essential regulator](#) of cardiac function. This pathway involves multiple organs including the lungs, liver, kidney and the arteriolar vasculature that are key players in a exquisitely balanced system that insures that a blood pressure remains within viable limits.

Renin-angiotensin-aldosterone system

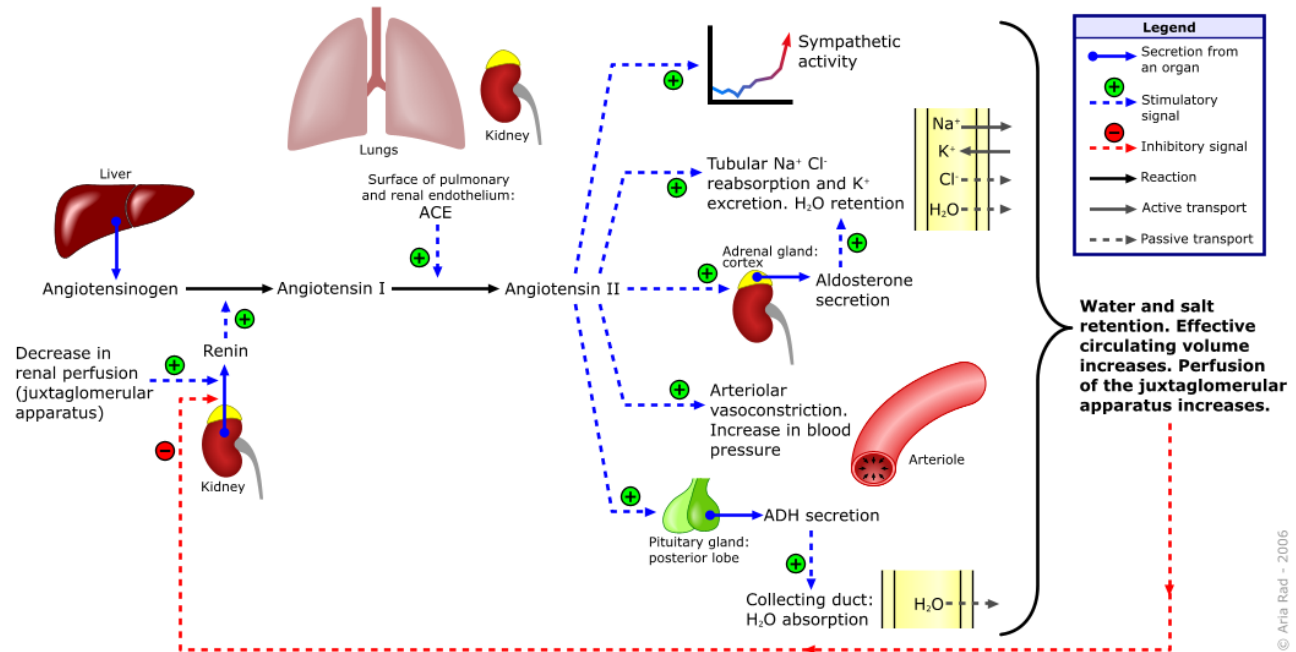


Figure 20 - Renin-Angiotensin-Aldosterone Organ Level Pathway for BP Regulation

One of us wondered if, under high viral loads, that the presence of the virus could present a dysregulating force that would be a contributing factor to precipitous blood pressure spikes or drops, that would make maintaining patients in critical care more difficult.

This hypothesis could be easily tested by obtaining blood pressure data versus time for Covid-19 patients. A solicitation was made through the university channels in the form of an administrative request for the acquisition of this data following proper protocols that would respect HIPAA, and institutional review board policies.

Competing with the Virus via ACE2 Inhibitors and Agonists

If it did turn out that blood pressure regulation was a factor in Covid-19 morbidity, this could help in the treatment of intensive care patients. It is also possible that strategies of angiotensin inhibitors and agonists could not only regulate problematic blood pressure pharmacologically but could also compete with the virus for occupation of the receptor site. Angiotensin II, a potent [vasoconstrictor](#), has a [circulation half-life](#) in rats of 16 +/- 1 seconds in peripheral blood and 15-30 minutes in tissues [Ref]. With these rapid response times an automatic feedback control system could be used based on instantaneous blood pressure readings. Such a strategy would have to be carried out with the utmost care in a fully qualified clinical setting.

Statistical Analysis of CV Infection Rates and Mortality

Apparent vs. Actual Infection and Mortality

When one computes mortality, one is computing it against APPARENT infections versus ACTUAL infections. According to Prof. Neil Ferguson of Imperial College there can be a threefold difference between these numbers. This has the effect of deflating the ACTUAL mortality rate and inflating the APPARENT mortality rate. Say that a calculation of 4.5% mortality rate in a certain region may too high by a factor of 3. If the actual number is 1.5% the ACTUAL number of infected people would be three times higher due to incomplete testing. It appears that more complete testing was being done earlier in the outbreak. Some of these determined true 'Infection Prevalence', when they occurred on closed groups like the Diamond Princess Cruise Ship. Informal results here were 1.5% to 2.0% in the early reports. With a paucity of test kits at this writing (April 5, 2020), it is a certainty that positives are being underreported. On a given day when the Johns Hopkins console reports 1,252,265 worldwide infects, on the order of 3.8 million could actually be infected. The mortality rate APPEARS to be 4.5% only because when people die, they get counted more reliably than when they get infected, show few symptoms, or are never tested.

Specific Accomplishment: Fast Estimation of Infection and Mortality

While monitoring daily statistics it was desired to predict both infection rates and mortality into the near future to assess the effectiveness of stay-at-home, social-distancing and prophylactic measures such as masks and gloves. The degree of exponentiality of the infection growth curve can be determined by looking at the logarithm of the number of infections versus time. When the slope of the log curve is a straight line, the exponential growth curve and doubling time are good approximations for what is coming next. Knowing this allows the following estimator to be developed that chooses two times, t_1 and t_2 , along with two observations o_1 and o_2 along a portion of an exponential growth curve where the slope of the log curve is known to be straight. Then for some minor interval of time, an extrapolation or forecast can be made that assumes no change in doubling time or amplitude. Here is the math, apologies for handwriting, but we are after all in the middle of a pandemic. The steps are:

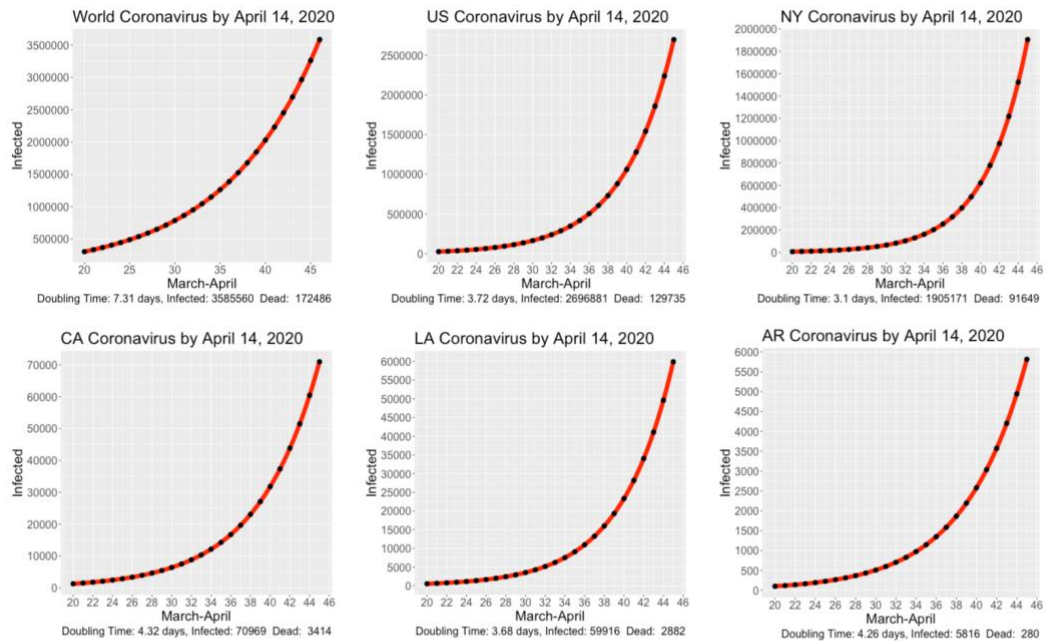
- Assume that doubling law of unknown period is occurring
- Call the leading constant coefficient the a for amplitude
- Call the inverse of the doubling time b
- Assume that a and b remain constant for the interval of interest
- Given observations o_i at times t_i solve for a and b to discover the shape of the curve.

$$\begin{aligned}
 a 2^{bt_1} &= O_1 \\
 a 2^{bt_2} &= O_2 \\
 \log_2(a 2^{bt_1}) &= \log_2(O_1) \\
 \log_2(a 2^{bt_2}) &= \log_2(O_2) \\
 \log_2(a) + \log_2(2^{bt_1}) &= \log_2(O_1) \\
 \log_2(a) + \log_2(2^{bt_2}) &= \log_2(O_2) \\
 \log_2(a) + bt_1 &= \log_2(O_1) \\
 \log_2(a) + bt_2 &= \log_2(O_2) \\
 \text{minus} \quad & \\
 \hline
 b(t_1 - t_2) &= \log_2(O_1) - \log_2(O_2) \\
 b &= \frac{\log_2(O_1) - \log_2(O_2)}{t_1 - t_2} \\
 b &= \frac{\log_2(O_2) - \log_2(O_1)}{t_2 - t_1} \\
 \boxed{b = \frac{\log_2(O_2/O_1)}{t_2 - t_1}} &\rightarrow \boxed{a = \frac{O_1}{2^{bt_1}}}
 \end{aligned}$$

Figure 21 - Derivation of Fast Estimator for Exponential Growth

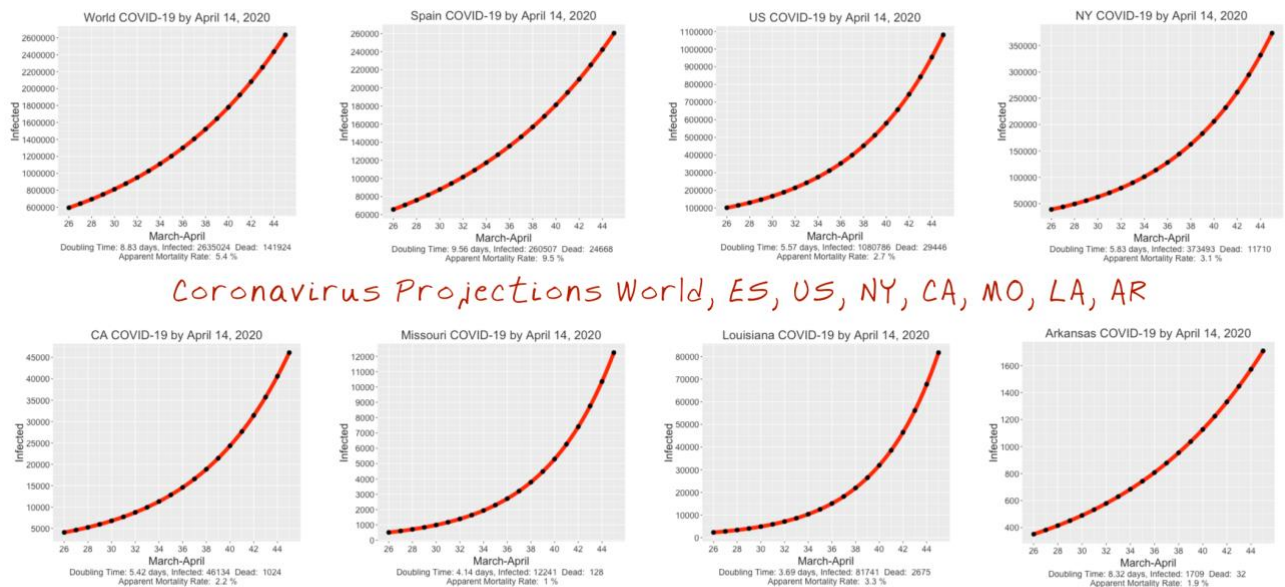
The two-point exponential estimation algorithm described above was applied to the World, United States, New York State, California, Louisiana and Arkansas for the March 30 estimate. Data sources are enumerated in the Appendix 1

By comparing growth, the growth curves of March 30, 2020 with the growth curves of April 4, one can estimate that about 100,000 lives in the US have been saved as a result of stay-at-home measures and sanitary precautions. That is an encouraging number, but we know from cyclicity arguments (provided below) that a resurgence will occur if people are careless, or travel too much, too soon. One can think of this pandemic as a socially-modulated wildfire, in wildfire season where some capable firefighters have gotten the fire under control. It only takes one mistake to reignite it. In the US, 29,000 people are predicted to have died by late April, a mournful statistic. That's half a Viet Nam War of casualties in a two-month span, considering that New York logged its first case on March 1, 2020. Spain and Missouri were added for the April 4 estimate.



Coronavirus Projections World, US, NY, CA, LA, AR
March 30, 12:00am CDT L. Van Warren

Figure 22 - CV Infection and Mortality Forecasts for April 14 Made March 30, 2020



April 4, 2020 12:00 am CDT L. Van Warren

Figure 23 - CV Infection and Mortality Forecasts for April 14 Made April 4, 2020

Kinsa Atypical Illness Fever Map of April 2, 2020

The company Kinsa distributes a digital thermometer to customers across the United States and collects patient temperature data using a companion phone application. This data has been processed to distinguish typical from atypical fevers. The atypical fevers are indicators of the degree

of penetrance of the CV into a region. This data, when contrast-enhanced, gives a clearer picture of that incursion.

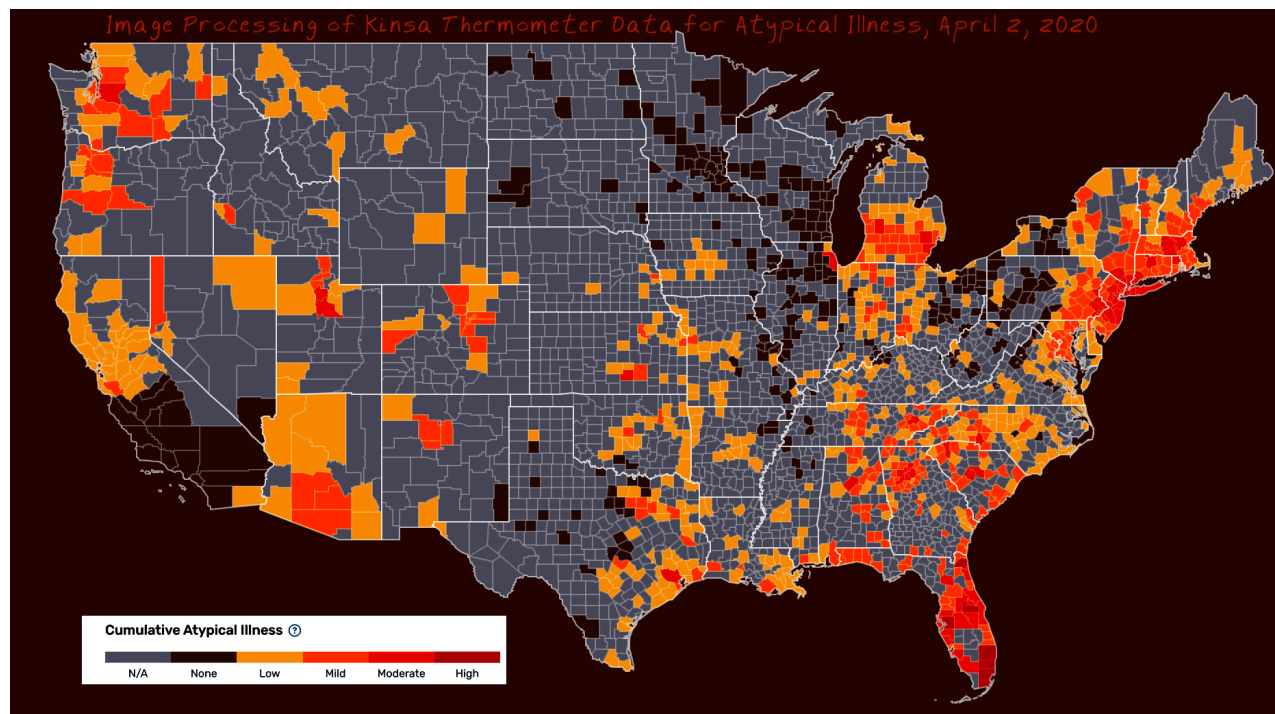


Figure 24 - Contrast Enhanced Kinsa Fever Map of April 2, 2020

Cyclicality

As the initial wave of pandemic spread subsides, there will be a tendency, discussed below in 'Pleasing the Mean' to loosen restrictions on mobility, travel, assembly, and social distancing in the hope of 'getting back to normal'. Until herd immunity figures are reached of between [29% and 74%](#), additional waves of infection are very likely to occur. This phenomenon was seen in the 1918 influenza pandemic:

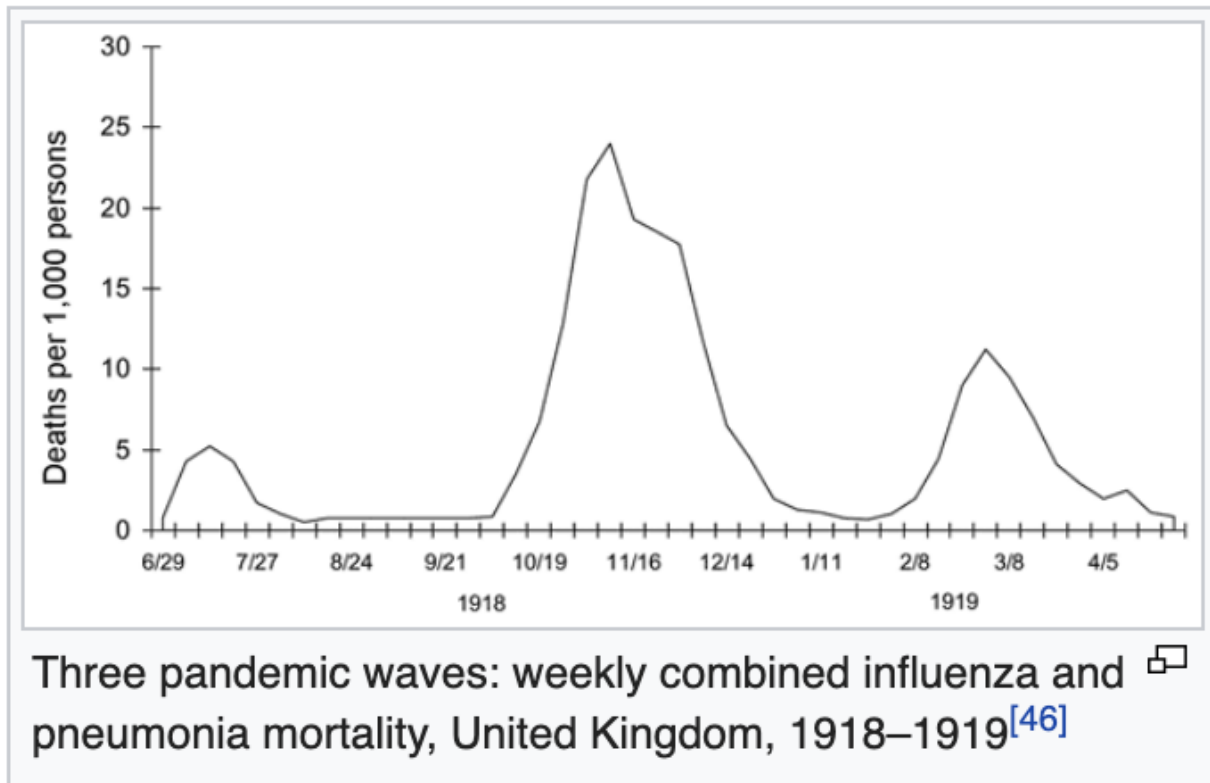


Figure 25 - CDC.gov - [1918 Pandemic Influenza: Three Waves](#)

A more detailed forecast of COVID-19 infections in India has been performed by Rajesh Singh, and R. Adhikari, in their paper, "[Age-structured impact of social distancing on the COVID-19 epidemic in India](#)". It shows the waves resulting from several possible policy permutations:

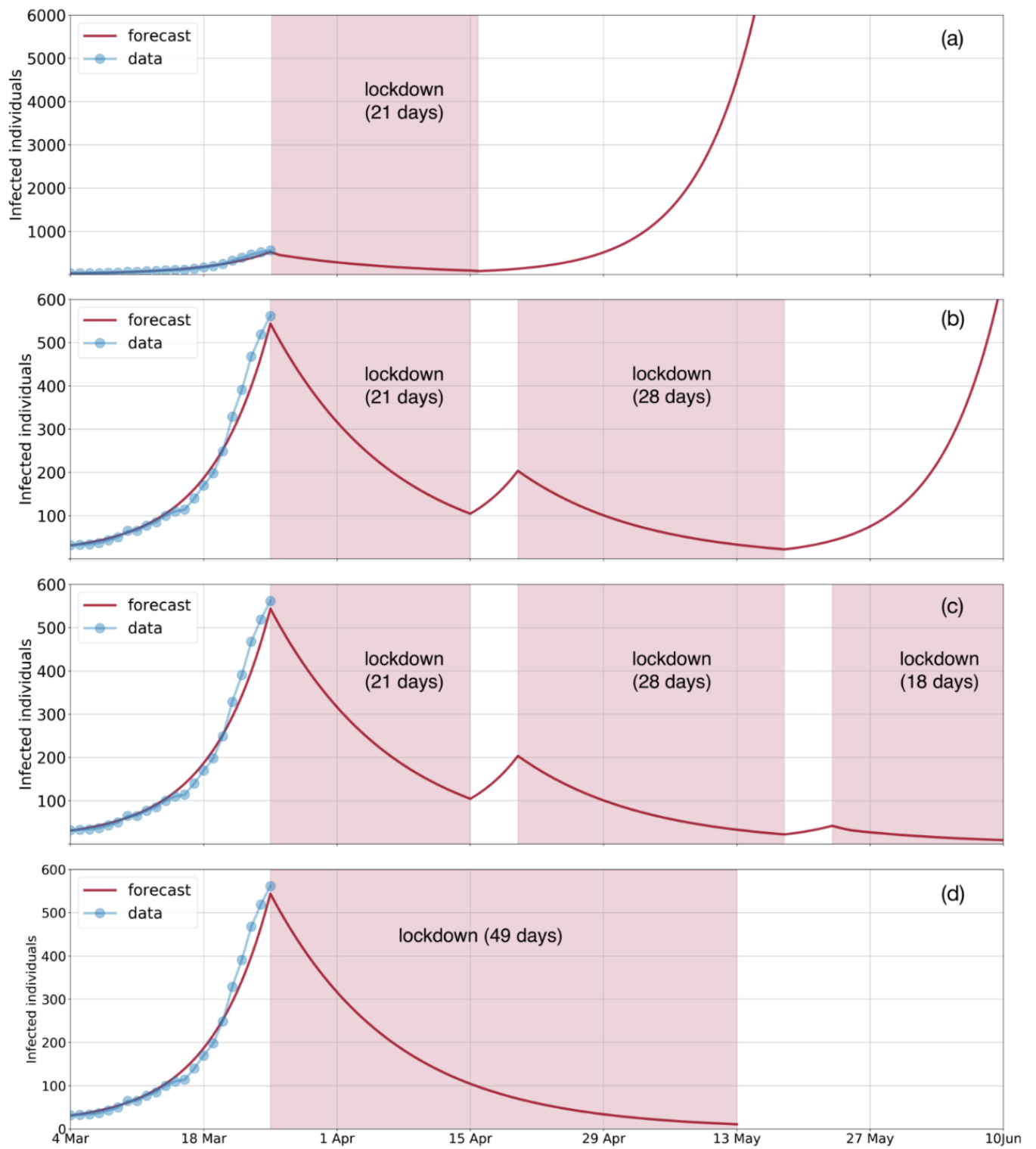


Figure 26 - Singh, and Adhikari – Policy Dependent Forecast of India's Covid-19 Epidemic

Decision-Making in a Pandemic is Not Like Traditional Political Decision Making

Pleasing the Mean

Politicians who are in a position of trying to please the most people at any given time, can, when faced with sensitive issues, produce decisions which 'please the mean'. An example of this is economic well-being vs. health well-being. In the current situation reducing lockdown measures enhances (temporarily) economic well-being at the expense of hurting (temporarily and permanently) health well-being. The trouble is, that with viral contagion, any exposure enables exponential growth of infection. Politicians and administrators who are accustomed to striking what seems like a fair and equitable middle-balance, actually enable the virus to flourish. This habit of 'Pleasing the Mean', while understandable, can result in widespread fatal catastrophe. A better metaphor for decision making is stopping the propagation of the virus by any means necessary.

Technical Communication in Crisis

In rapidly advancing pandemics, whose doubling times are on the order of a few days, (as demonstrated above), traditional publication cycles are too long for circulating up to date clinical information, lessons learned, resource allocations, experimentation proposals and so forth. Accelerating communication life cycles is facilitated by the world-wide web, but requires adjustment of workflows, expectations and optimized use of the communication modalities. These modalities include phone, text, email, blog post, academic paper retrieval, etc. The frequently used modality of face-to-face meeting has been hindered by the need for social distancing to reduce transmission of the virus. Each of the many flavors of workers advancing this effort has some preferred and measurable communication modality that determines their effectiveness as a contributor. This requires rapid shifting from communication modes in one's comfort zone, to new ones by the acquisition of new habits, behavior patterns and skillsets.

Latency in Traditional Communication Pathways

Even in a specific investigation like this it is important to consider the two prime figures of merit in communication, *throughput*, and *latency*. Throughput is a measure of bandwidth; how much data transmission or work occurs per unit time. The second measure of communication effectiveness is latency, which is how long, after the delivery of a work article or dataset, is that item picked up and used in a beneficial way? Common parlance uses the term, 'going viral' for contributions of high interest. In that regard we have much to learn from the strategy this virus uses to get things done. We should insist on low latency and high throughput in our work contributions.

Two additional communication parameters are *interrupted* versus *polled* communication. If a given worker has to continue to poll or check for the arrival of any physical or data resource of interest, they will be expending many useless cycles in that pursuit that could be dedicated to other things. On the other hand, if that same worker, is *interrupted* only at that moment when some key resource arrives, their wasted time is minimized.

Optimizing throughput, latency, interruption and polling are essential goals in streamlining communication operations, whether data intensive or *materiale* in nature.

Future Work

Application of Machine Learning

The explosion in machine learning and data science software, hardware and knowhow presents a unique opportunity for solving difficult problems, from proteomics to delivery optimization. It is important that we widen our toolset and thinking to consider how such capabilities can be used to defeat the virus.

Command and Control for Viral Analysis

Creating operational environments for tracking the spread and the mechanisms of viral infection is similar to a command and control situation where multiple streams and multiple kinds of data must be collected, processed and fused into a meaningful action plan. Whereas in former times such command and control environments existed at centralized locations, they must now, by the very nature of the pandemic be distributed, while enabling the effective communication of a centralized facility.

References and Bibliographic Entries

A complete list of Images, Fasta files, PDB files, bibliographic entries and supplementary documents can be found [here](#), many of these are directly linked into this document.

Acknowledgements

Molecular graphics and analyses performed with UCSF Chimera, developed by the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco, with support from NIH P41-GM103311. See UCSF Chimera--a visualization system for exploratory research and analysis. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE. J Comput Chem. 2004 Oct;25(13):1605-12.

BLASTn analysis done using the web-based tools at National Center for Biotechnology Information (NCBI)[Internet]. Bethesda (MD): National Library of Medicine (US), National Center for Biotechnology Information; [1988] – [cited 2017 Apr 06]. Available from: <https://www.ncbi.nlm.nih.gov/>

Alignment visualization performed using Jalview by Waterhouse, A.M., Procter, J.B., Martin, D.M.A, Clamp, M. and Barton, G. J. (2009) "Jalview Version 2 - a multiple sequence alignment editor and analysis workbench"
Bioinformatics25 (9) 1189-1191 doi: 10.1093/bioinformatics/btp033

Appendix 1: Data Sources for Epidemiological Data

1. [Imperial College London Seminar on Covid-19](#)
2. [Euro Data in R](#)
3. [Prof. Steven Riley Tracking Infection Data Using Bank Movement Data](#)
4. [Johns-Hopkins Coronavirus Dashboard](#)
5. [World & US Data Source](#)
6. [NY, CA, LA, MO & AR Data Source](#)
7. [Bing Covid-19 Dashboard for Day of State Statistics](#)
8. [Worldometer](#)
9. [Prof. Neil Ferguson Feb 4 talks about sampling inaccuracies.](#)
10. [Asserts a 5 Day doubling time on Feb 4, 2020](#)
11. [The Reproduction Number - Undetected:Detected Ratio typically varies from 2.3 to 3.1](#)
12. [SARS outbreak of 2003 and Influenza outbreak of 2008 are companion events](#)
13. [Prof. Neil Ferguson Feb 12 Asserts that only severe cases in Chinese are being tested.](#)
14. ['Infection Prevalence' Where a Closed Group is complete tested and applied to open groups](#)
15. [1957 and 1968 Influenza Pandemics](#)
16. [Mathematical Definitions](#)

Abbreviations / Glossary / Tools

BINF:	Bioinformatics
BLAST:	Basic Local Alignment Search Tool
BP:	Blood Pressure
CHIMERA:	UCSF Chimera – An Extensible Molecular Modeling System
Covid-19:	The illness caused by SARS-COV-2
CV:	Severe Acute Respiratory Syndrome Coronavirus 2: SARS-CoV-2
GO:	The Gene Ontology (GO) Knowledgebase
KEGG:	KEGG: Kyoto Encyclopedia of Genes and Genomes
MERS-COV:	Middle East Respiratory Syndrome Coronavirus
NCBI-BLAST:	Basic Local Alignment Search Tool from NCBI/NIH/NLM
NCBI-VIRUS:	Taxonomy of Viruses
JALVIEW:	Java Multiple Sequence Alignment Editor/Viewer/Workbench
PHYRE2:	Protein Homology/analogY Recognition Engine V 2.0
MAFFT:	Multiple Alignment Using Fast Fourier Transform
NCBI:	National Center for Integrative Biomedical Informatics
AMR:	Apparent Mortality Rate : deaths divided by known positive cases.