

1 SARS-Cov2 seasonality and adaptation are driven by solar
2 activity.

3 Octavio Gonzalez-Lugo^{*}

4 ^{*}Corresponding author *No affiliation*. Email: ogonzalezl@outlook.com

5 **Abstract**

6 Since its isolation in Wuhan SARS-CoV2 showed a high mutation rate hindering the ability
7 to properly characterize. Also as a consequence of its size, traditional sequence analysis meth-
8 ods were computationally constrained. However, applying variational autoencoders (VAEs) to
9 custom sequence representations results in a series of clusters sorted by the sunshine duration
10 (SD) rate of change (SDRC) and other solar-derived features. The transition between clusters
11 is characterized by changes in viral genome size, apparent deletions can be found through-
12 out the SARS-CoV2 genome. This series of deletions might behave as an internal clock inside
13 the genome. SDRC-derived features synchronize COVID-19 cases into a single period. Both
14 SDRC-derived features and solar features correlate with COVID-19 cases pointing towards a
15 solar-dependent seasonality. Atmospheric changes that affect solar radiation also showed a cor-
16 relation with COVID-19 cases. Analyzing viral genome composition as time series displays
17 an attractor-like behavior under different solar-derived time scales. While clustering them by
18 environmental conditions shows a similar pattern as the one found by the VAE models. Fur-
19 ther development of analysis techniques will help us to better understand the seasonality and
20 adaptation of pathogenic organisms.

21 **Introduction**

22 The ongoing COVID-19 pandemic generated a large quantity of data as a result of tracking both the
23 virus and the disease cases. Those data resources were published regularly and openly throughout the
24 pandemic. However, the size of the datasets restricted the ability to analyze the data to a few centers
25 with high computational resources. Such a barrier is particularly true for genome sequence data.
26 As the genome size hindered the ability to use large samples size to understand the evolution and
27 adaptation of the virus [1]. The development of new sequence representation schemes as well as the
28 application of deep learning methods to the new datasets could fast-track the characterization of new
29 emerging or neglected pathogens. The following describes the use of different custom SARS-CoV2
30 sequence representation schemes to train VAE models. And how to combine different information
31 with the insights obtained by the model. Continuous development and improvement of sequence

32 representation schemes will lead to a faster and more accurate characterization of emerging or
33 neglected viral pathogens.

34 **Generative models for sequence representation.**

35 Generative models are a special kind of machine-learning models that can generate new samples
36 from a simple input. A VAE is one of the many different generative machine learning architectures.
37 A VAE is a special neural network architecture, constituted by an encoder and decoder network,
38 trained with the same data for input and output. As a consequence of the training, the encoder
39 network finds a low-dimensional representation of the data that can be used for other applications
40 [2]. When there are no labels available for the data the training is also known as unsupervised
41 learning.

42 The low dimensional representation is also known as learned representation and each axis encodes
43 an attribute inside the data. Specifics of the attribute will depend on the patterns inside the data,
44 the number of samples, possible biases in those samples, and the model architecture. This allows
45 for controlling the behavior or attributes that the newly generated sample will have. (Figure 1)

46 One of the main bottlenecks for sequence modeling is the size of the sequence, as the sequence
47 grows in size long-range dependencies start to be lost [3]. Thus designing different data structures
48 that better capture those dependencies or find other kinds of patterns inside the data is of great
49 interest. Furthermore, the application of generative modeling to genomic data, especially to fast-
50 adapting viruses, could fast-track the development of new treatments and the understanding of
51 emerging diseases.

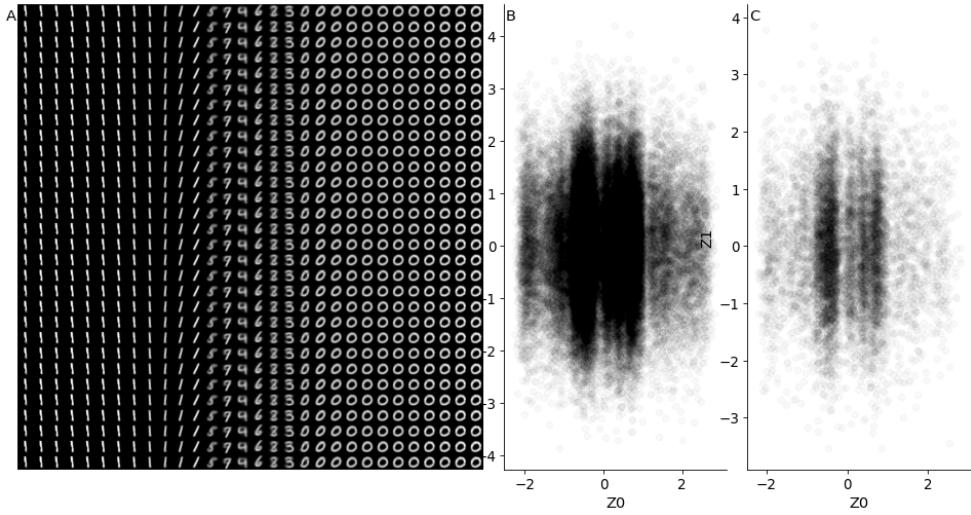


Figure 1: **VAE learned representation from the MNIST dataset.** A Visualization of the learned representation across Z_0 and Z_1 , the model learns the "roundness" of each number and aligns it across the Z_0 axis. B, C learned representation, B shows the train data, and C the test data. Each dot represents a single sample from the MNIST dataset, encoding a 28x28 image into a 2D space.

52 Stacked K-mers

53 The first sequence representation consists of the normalized stacked frequencies of sliding sampled
 54 K-mers. The sliding sampling scheme enables the analysis of different reading frames inside the
 55 SARS-CoV2 genome. Trained VAE model results in a low dimensional representation that clusters
 56 the sequences into a series of clusters ordered on one axis (Figure 2 A,B,C). Correlation analysis
 57 shows that the latent dimension Z0 highly correlates with the SDRC the day of the year(DOY) as
 58 well as different components in the solar radiation spectrum (Figure 2 D).

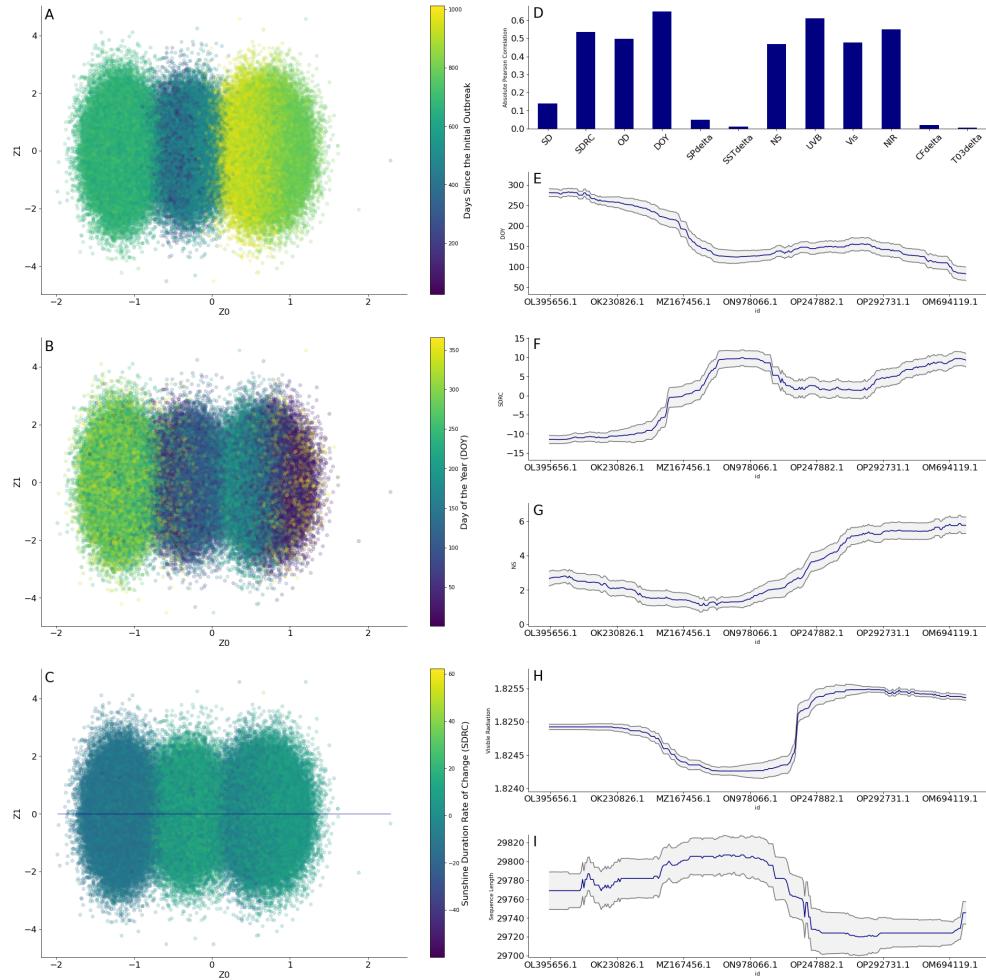


Figure 2: **SARS-CoV2 encodes temporal information inside its genome.** A, B, C Latent space visualization, each dot represents a unique genome and colors the corresponding time scale. The line on C shows the range of genomes selected for further analysis. D Correlation between the Z0 latent dimension from the selected sequences and the different time-related features. E, F, G, and H Changes followed by the selected genomes under different time scales. I Changes in genome size through the found order. E, F, G, H, I 95% confidence interval in gray.

59 To further characterize the information inside the latent dimension Z0 a series of sequences are
 60 selected close to zero in the latent dimension Z1. Changes in the SDRC and DY show an upward

and downward trend similar to the pseudo color in Figure 2 E,F. While solar radiation features such as the normalized number of sunspots (NS) and the irradiation of visible radiation follow a profile similar to the SARS-CoV2 genome size Figure 2 G,H,I.

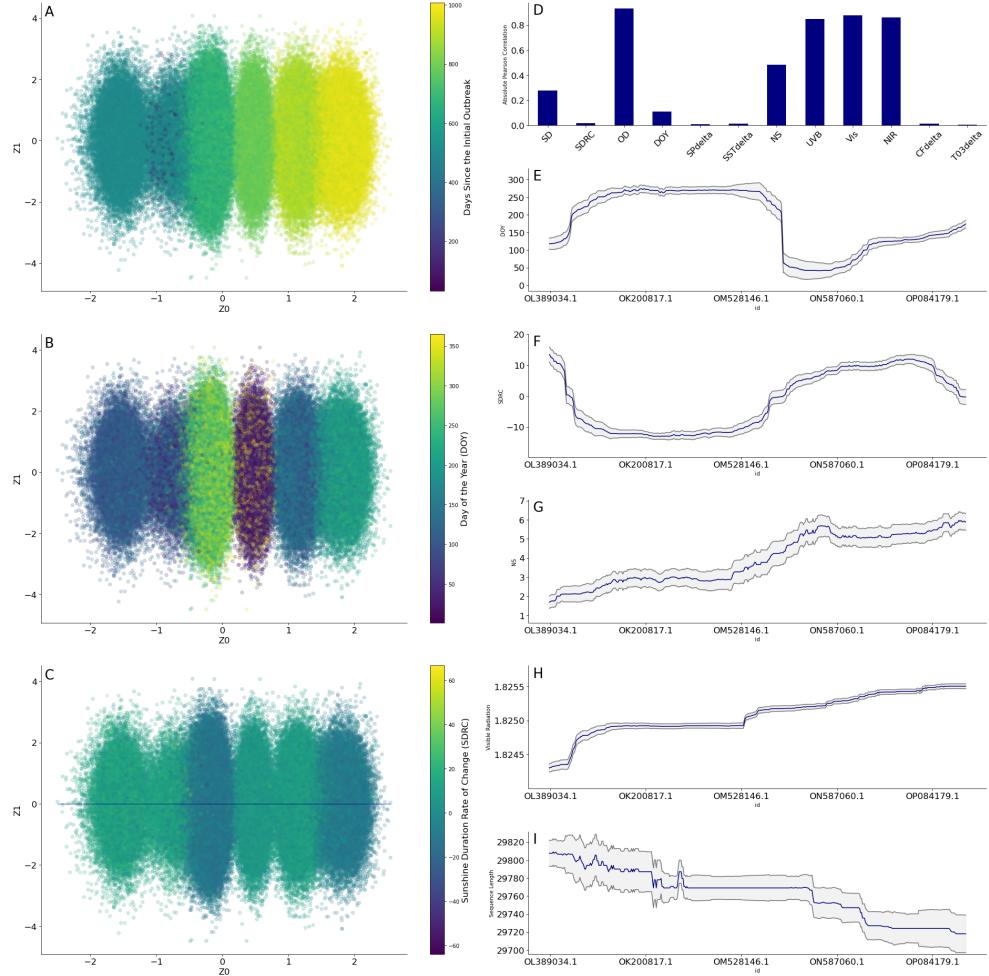


Figure 3: Subsampling the genomic data biased the learned representation. A, B, C Latent space visualization, each dot represents a unique genome and colors the corresponding time scale. The line on C shows the range of genomes selected for further analysis. D Correlation between the Z0 latent dimension from the selected sequences and the different time-related features. E, F, G, and H Changes followed by the selected genomes under different time scales. I Changes in genome size through the found order. E, F, G, H, I 95% confidence interval in gray.

However, genomic sampling is biased towards the second year of the pandemic and forward. This bias could influence the order found in the learned representation. To remove sampling bias as much as possible genomes are randomly selected to contain approximately the same number of genomes per day. Max number of genomes was adjusted around the peaks of the first and second COVID-19 waves. The subsampled dataset also showed a high correlation towards solar-derived features. However, the high correlation towards SDRC is diminished several folds (Figure 3 D).

70 Changes in the correlation patterns point towards a minimum set of samples needed to shift
 71 the learned representation towards a specific space. The full dataset contains an excess of samples
 72 at the year level, thus the learned representation is biased towards patterns that better represent
 73 that characteristic. While the subsampled data contains fewer samples at a year level biasing the
 74 representation towards the long-term representation. Although specific changes related to a yearly
 75 or long-term adaptation have not been found yet. Changes in the learned representation hint at its
 76 existence.

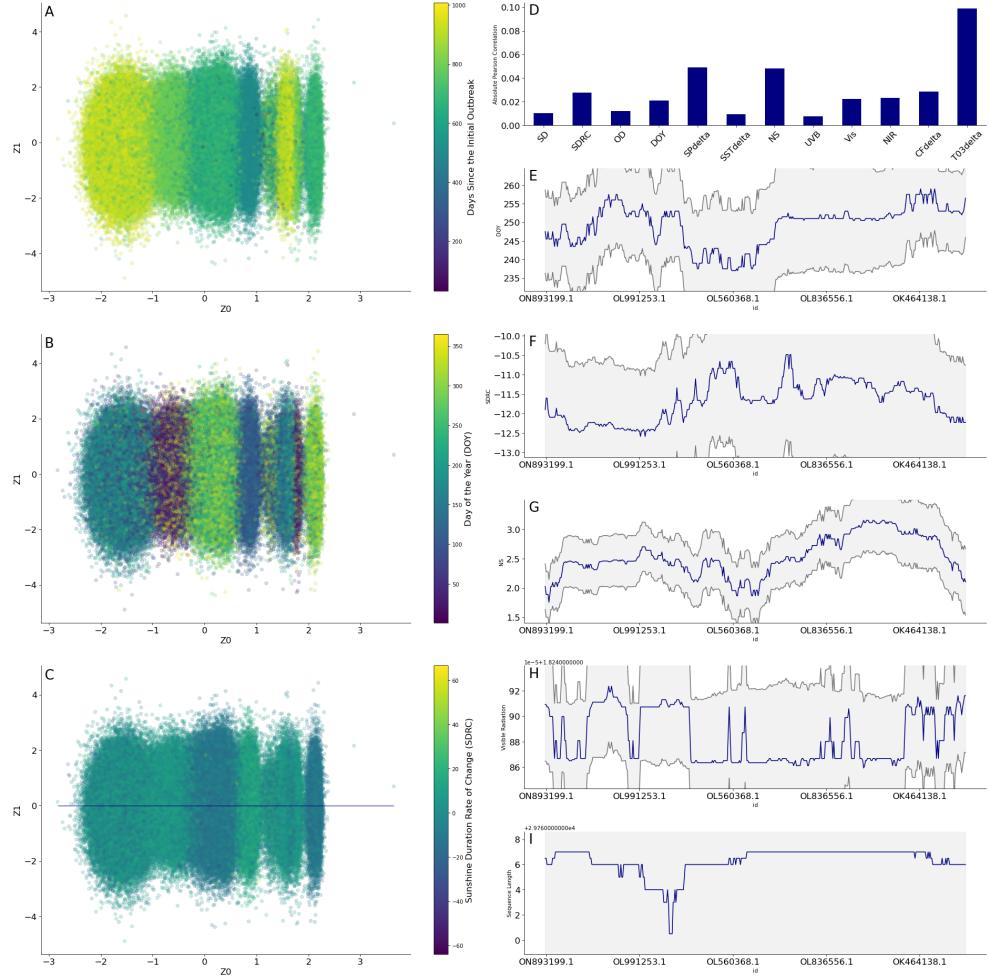


Figure 4: **Fragment-based representation better encode environmental features.** A, B, C Latent space visualization, each dot represents a unique genome and colors the corresponding time scale. The line on C shows the range of genomes selected for further analysis. D Correlation between the Z0 latent dimension from the selected sequences and the different time-related features. E, F, G, and H Changes followed by the selected genomes under different time scales. I Changes in genome size through the found order. E, F, G, H, I 95% confidence interval in gray.

77 **Adjacency Matrices**

78 The previous dataset is capable to retrieve a specific correlation between the SARS Cov 2 genome
 79 and temporal scales, as well as other environmental features. However, it can't retrieve information
 80 regarding specific locations in the SARS-Cov2 genome. Therefore the design of different sequence
 81 representation schemes could ease the detection of such patterns.

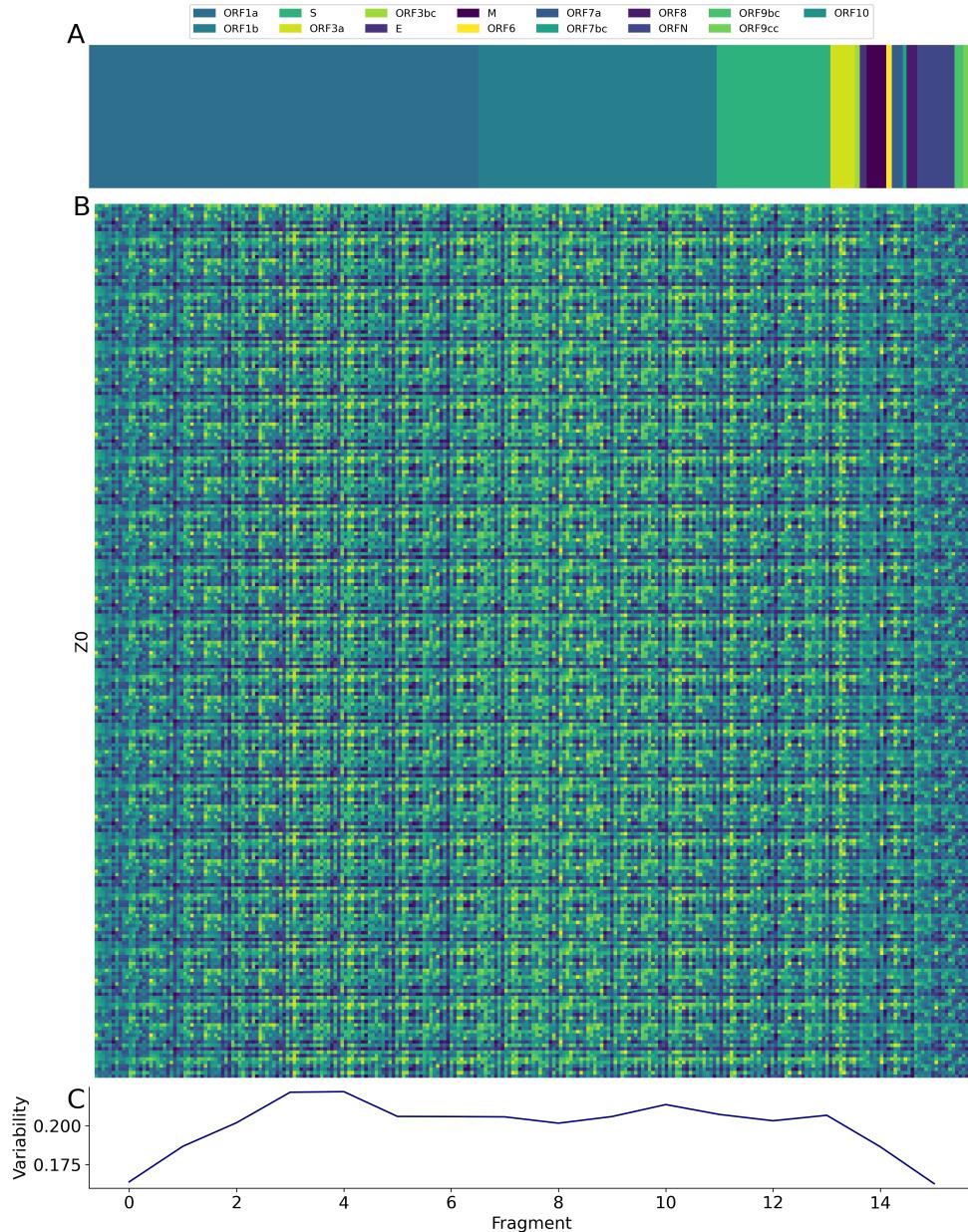


Figure 5: **Extremes of the SARS Cov2 viral genome are less variable.** A SARS Cov2 genome B Random walk around Z0, C Genome variability measured as the standard deviation of each fragment.

82 The second dataset encodes the sequence by dividing it into 16 fragments and calculating the
83 adjacency matrix of sliding consecutive 2-mers. The same encoding process is repeated for the
84 reversed fragment. This will encode the sequences by the connectivity between the different compo-
85 nents rather than the frequency of each component. This will result in a hypercube of (16,16,16,2)
86 dimensions.

87 Learned representation also retrieves a temporal pattern, however, the correlation between the
88 solar features is diminished several folds, except for the daily change in the total column of ozone
89 (TO3delta). Although several features were tested no other significant correlation was found (Figure
90 4 D).

91 Yet the generative nature of the VAE model can also be useful to address how the genome changes
92 through time and region. The decoder part of the model can be used to predict the changes along
93 the learned representation. This kind of analysis is also known as a random walk, as the only input
94 needed for the decoder part of the model is a series of randomly generated numbers. To ease the
95 interpretation the input values for the random walk only differ along Z0 leaving Z1 equal to zero.

96 The standard deviation of the extreme regions of the SARS-CoV2 genome showed lower variability
97 compared to the remaining regions (Figure 5 D). this characteristic of the genome could be used to
98 target viral components in those regions as they will be less likely to be mutations.

99 Particularly at the 5' end of the genome Nsp1 and Nsp2 proteins could be potential targets due
100 to the low variability in that region. Particularly Montelukast has been shown to inhibit Nsp1 by
101 itself [4] or in combination with Ponatinib and Rilpivirine [5].

102 On the 3' of the viral genome both the nucleocapsid protein as well as the membrane protein
103 align to the region with less variability. The nucleocapsid protein has been suggested to be added
104 to new vaccine formulations as is less prone to mutations [6]. Spike and nucleocapsid fusion protein
105 using an adenoviral vector showed a specific response in animal models [7]. Using only the nucleo-
106 capsid protein for vaccination showed that inoculations of Lewis and Wistar rats with recombinant
107 nucleoprotein resulted in no worrying side effects as well as the production of specific antibodies [8].
108 Regarding the membrane protein, inoculations with the membrane and envelope proteins in mice
109 showed partial protection [9]. Both combination or single nucleocapsid proteins could offer a better
110 long-term inoculation strategy due to low mutation in the nucleocapsid protein.

111 Dimensional expansion.

112 Previous models further confirmed the presence of a temporal pattern inside the SARS-CoV2 genome
113 and this pattern is highly correlated to solar activity. Either by the calculation of solar-derived
114 features such as the SD or SDRC as well as direct measurements of solar radiation at different wave-
115 lengths, or solar activity by the number of sunspots. They also showed the effect of data sampling
116 in the learned representation, as biasing the sampling towards a specific temporal characteristic will
117 result in a model that better represents such patterns. And that expanding the dimensionality of
118 the sequence representation could be a better strategy for large sequences. As the rearrangement
119 could bring closer different regulatory or similar regions mimicking the 3D structure of the genome.
120 To further test that hypothesis a third dataset was designed to predict the full-length SARS-CoV2

sequence. The dataset consists of the full length of one-hot-encoded sequences, and the resulting encoding is reshaped to a hypercube of dimensions (32,32,32,4). Also, the number of samples is lowered to reduce the yearly pattern influence and skew the results toward the long-range pattern.

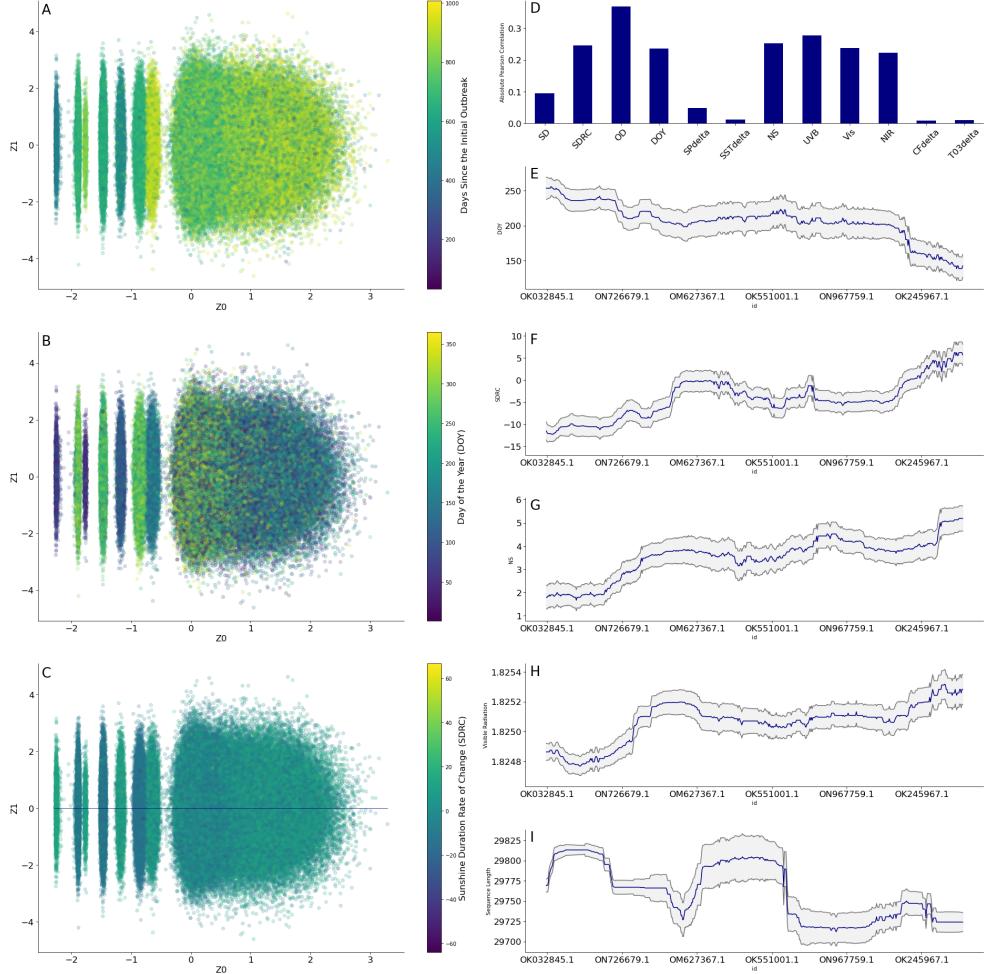


Figure 6: SARS Cov2 follows patterns of solar activity A, B, C Latent space visualization, each dot represents a unique genome and colors the corresponding time scale. The line on C shows the range of genomes selected for further analysis. D Correlation between the Z0 latent dimension from the selected sequences and the different time-related features. E, F, G, and H Changes followed by the selected genomes under different time scales. I Changes in genome size through the found order. E, F, G, H, I 95% confidence interval in gray.

As expected the model can retrieve a representation that better fits the long-range pattern, however, is also capable to retrieve the yearly pattern encoded inside the SD and SDRC (Figure 6 D). This multiplicity suggests a mixture of different time scales playing a role in SARS Cov2 adaptation and seasonality. A long-range dynamic, a yearly dynamic, and perhaps many others.

To test that hypothesis the genome size is used as a measurement of viral adaptation as genome size also follows the time scale inside the virus. Time since the initial outbreak shows a continuous

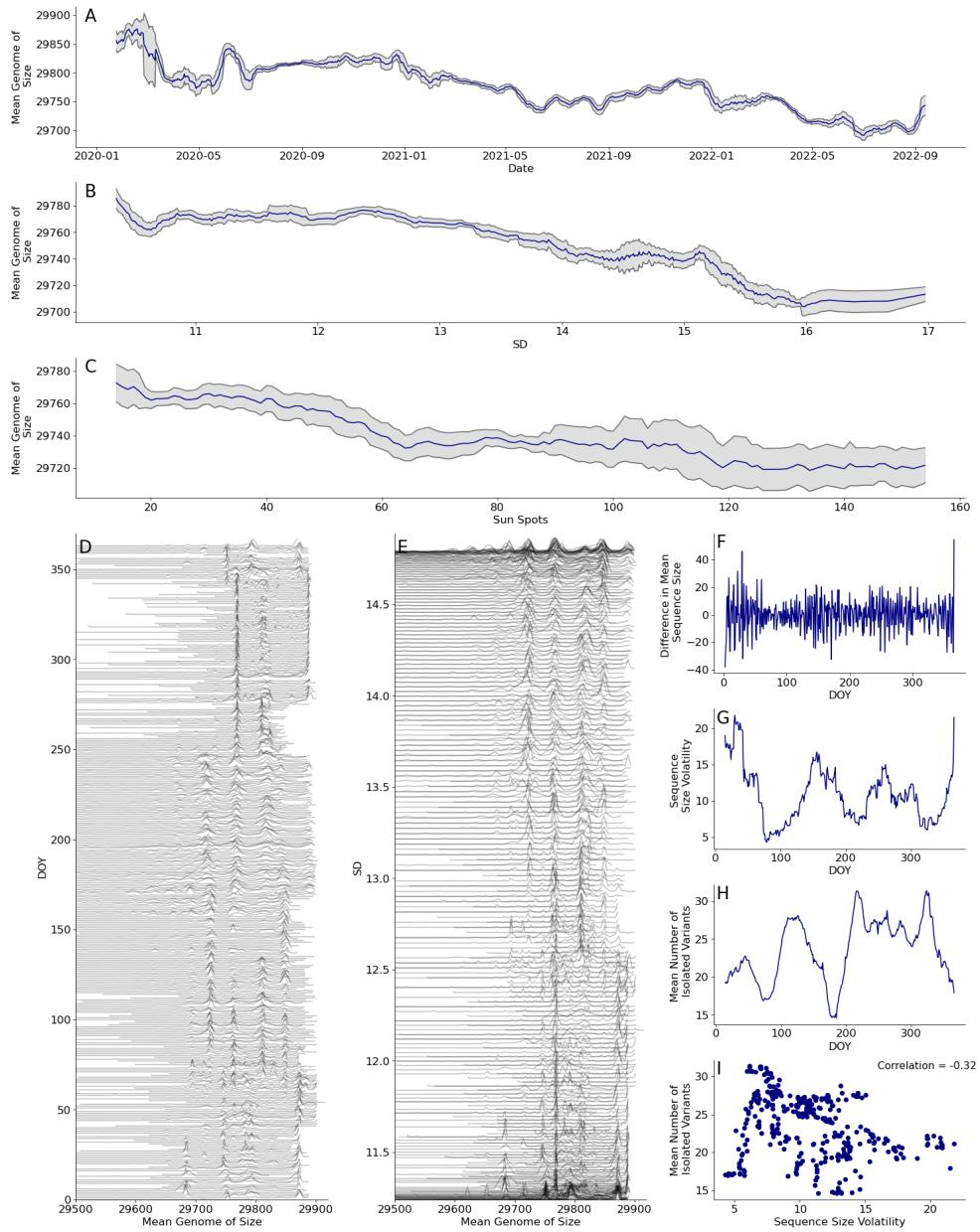


Figure 7: Genome size follows different scales of solar activity. A Mean SARS-CoV2 genome size from Dec-2012 to Sep-2022, B Mean SARS-CoV2 genome size grouped by SD. C Mean SARS-CoV2 genome size grouped by the number of sunspots. D Changes in genome size by day at a single location. E Changes in genome size by SD at a single location. F Changes in mean genome size by day of the year. G Volatility or standard deviation of the mean change in genome size by day of the year. H Mean number of isolated variants at the location by day of the year. I Correlation between Genome size volatility and the number of isolated variants.

decline in average genome size. The pattern is again repeated when the genomes are sorted by the SD or by the number of sunspots (Figure 7 A,B,C). This points to specific frequency components of

132 solar activity, the yearly component encoded by SD, and the solar or Schwabe cycle [10]. A 12-year
133 cycle that is measured by the number of sunspots. A third component will be the daily changes in
134 solar radiation that vary throughout the day. SD measures the duration of such a period but not
135 how solar radiation changes through the day. Genome size histograms at one specific location show
136 a series of peaks that skew toward smaller genome sizes at periods with high SD and vice-versa.
137 Furthermore, each day contains a series of distributions that might point to isolations at different
138 times of the day (Figure 7 D,E).

139 Continuous rearrangement of the genome could be one of the many drivers behind the generation
140 of new variants. Rapid changes in solar activity could drive the rearrangement of the viral genome
141 leading to the generation of new variants. Measuring the genome rearrangement as the volatility of
142 the mean genome size deviation and the genomic diversity as the number of variants shows a small
143 negative correlation between both metrics (Figure 7 F,G,H,I). Although the correlation is small
144 it could point toward specific mutational hotspots where the generation of variants is more likely.
145 Those mutational hotspot could be repeated through the different frequency components of solar
146 activity.

147 The combination of different frequency components could lead also to recombinant-like genomes.
148 Identification of recombinant variants might point towards a genomic rearrangement derived from
149 a low-frequency component within the solar cycle. The first isolation of omicron was on November
150 2021 when the Schwabe cycle was entering a linear range. As the solar cycle was entering a new phase
151 the transition could lead to a recombinant-like virus. First isolations of delta-cron a recombinant
152 variant between omicron and delta were reported in January 2022 [11]. The linear phase of the
153 Schwabe cycle is forecasted to end by 2024 NOAA [12]. Thus it's likely that more recombinant-like
154 variants either with similarity to previously extinct clades or the generation of a novel clade will be
155 isolated in the following months.

156 Genome rearrangements on the daily component have also been reported, although there were
157 scarce at first. A Case report where the main infecting variant changed between samples taken
158 at different times was published by [13]. Also, co-infections and genome variability started to be
159 reported [14] [15]. Variability within the host might point towards a high-frequency component or
160 the combination of different frequency components within the solar cycle rather than a co-infection
161 scenario. Easement in the detection of the different genomic rearrangements could be the result of
162 a higher adaptation to the codon usage bias(CUB) of the host. As the virus better adapts to the
163 host's CUB fewer viral particles will be needed to infect the host, effectively increasing the R₀ of the
164 virus. CUB adaptations could remain silent as a synonym codon will be favored over another [16].

165 Dependence on solar activity could explain the benefit of different pharmacological and non-
166 pharmacological interventions that mimic the action of solar radiation. Supplementation with vi-
167 tamin D has been found to confer protection against COVID-19 [17] [18]. Melatonin treatment
168 also increases the clinical recovery rate [19]. IR radiation increases the recovery time of COVID-19
169 patients [20]. These different interventions by themselves are not able to prevent SARS-CoV2 infec-
170 tions. However, in conjunction, they show that correct circadian signaling is key for fast COVID-19
171 recovery.

172 Although the different models are also capable to classify the viral genomes into several clusters,

they do not match any given variant. The different clusters might represent viral pseudo species or a specific branch in the SARS-CoV2 phylogenetic tree. Different clusters also point to shared patterns among the different genomes in the same cluster. Although the specific nature of the pattern is not investigated its existence and classification could facilitate the design of new treatments. Recurrent mutation patterns could constrain the number of possible immune evasive variants or point towards a conserved mutation path. The lack of recognition of those patterns could be the result of the nonlinear nature of the viral adaptation or heavily relying on the Wuhan isolate or ancestral strain. The fast mutation rate of SARS-CoV2 also implies that the ancestral strain could be another variant and not the initial virus that originated the outbreak.

Despite the fact that there are many examples of the application of machine learning models for SARS-CoV2 genomic data with even more sophisticated architectures [21] [22], the solar correlation has not been presented before. Perhaps the main reason behind that is how the model was used, rather than a tool for prediction, the model was used as a high dimensional sorting algorithm. Then different features were used to find the meaning behind the latent dimensions.

Sorting the data by autoencoders has been previously used to obtain meaningful coordinate systems from complex data sources [23]. This allows us to better understand the dynamics of a system by applying known dynamical models. Data obtained by genomic surveillance might not represent entirely the continuous adaptation of the virus as a series of recurrent ancestors. But rather a series of snapshots at different times of the interaction between a tissue within the host and SARS-CoV2.

Epidemic curves under a generalized coordinate system

Although there was a difference between folds and models the SDRC showed a high correlation, suggesting that it could encode the seasonality inside the COVID-19 waves. Displaying the number of cases with SD as a time scale for COVID-19 cases shows cases clustered at both ends of the time scale. While SDRC swaps the order of the peaks, suggesting that there could be an intermediate step between SD and SDRC that synchronizes both peaks into a single period. As SDRC is the first derivative of SD a smooth transition between SD and SDRC could be obtained by fractionally differentiating SD. Fractional differentiation is the generalization of the derivation and integration operation over fractional orders [24].

Fractionally differentiated SD (fSD) slowly synchronized the different peaks into a single one (Figure 8 A). Synchronizing the different COVID-19 waves by adapting the fractional order shows a relatively similar value except near the Equator and latitudes higher than 40 deg (Figure 8 B).

The fSD provides a general time scale to analyze the cases data regardless of the location. The synchronization also allows to evaluate the correlation between cases and other environmental variables. Analyzing the mean number of cases over the generalized time scale shows high correlations between different wavelengths of solar radiation. Particularly a linear correlation is found between UVA and UVB and the number of cases of COVID-19 (Figure 9).

Environmental variables obtained from the Aqua/AIRS L3 Daily Standard Physical Retrieval dataset shows also linear correlations, particularly over environmental variables that influence solar

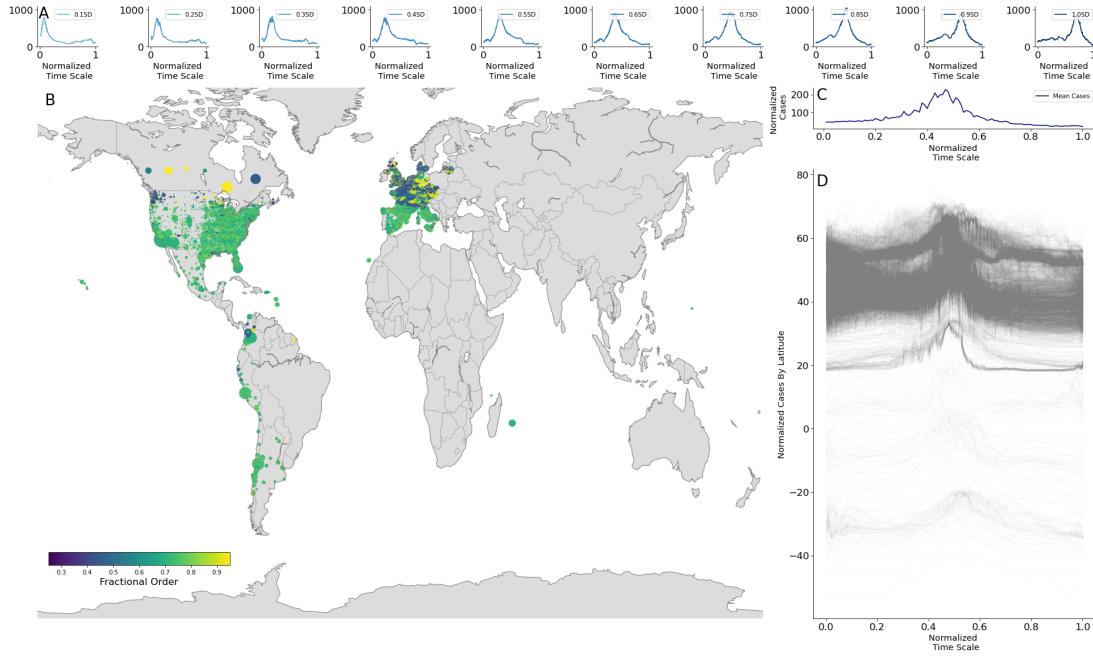


Figure 8: COVID-19 pandemic dynamics follows the fractional SDRC. A. Synchronization of epidemic curves by fractionally differentiating SD as a time scale. From left to right increasing order of fractional differentiation. B. Fractional order that synchronizes the epidemic curve at a specific geographic location. The marker shape is equal to the mean number of reported cases and the color is equal to the fractional order that synchronizes the epidemic curve. C. Mean synchronized curve. D. Waterfall plot of the synchronized curves by latitude.

radiation. Changes in the total column of ozone as well as changes in the cloud fraction. Ozone is an atmospheric component that absorbs the majority of UV radiation at the different wavelengths. While clouds, whose main component is water, absorb infrared radiation and UV radiation [25]. Correlation with other environmental variables are also found, yet are less in magnitude compared to radiation wavelengths and atmospheric variables (Figure 9).

Coordinate system discovery from the different models allowed the synchronization of the different epidemic curves at a city level. This specific data treatment will remove most of the socio-economic and behavioral effects of combining the data by country, age, or other proposed segmentation. It also removed the effect of outliers either due to high or low testing frequency, yet advantages of other kinds of segmentation were lost.

The relationship between solar activity and susceptibility to pathogenic diseases is a hypothesis previously proposed,[10] showed the correlation between different outbreaks through time and stationary points in the Schwabe cycle. Dependence between the susceptibility to SARS-Cov2 and the yearly change in solar radiation can be interpreted as a higher frequency component of solar activity. An even higher frequency component will be the daily changes in solar radiation that encompass the circadian rhythm. Dependence of COVID-19 on daily changes in solar radiation has been presented previously. Detection efficiency peaks at around the middle of the day [26] [27] [28]. Also, hospital admission has been found to have a particular diurnal dependence as reported by [29].

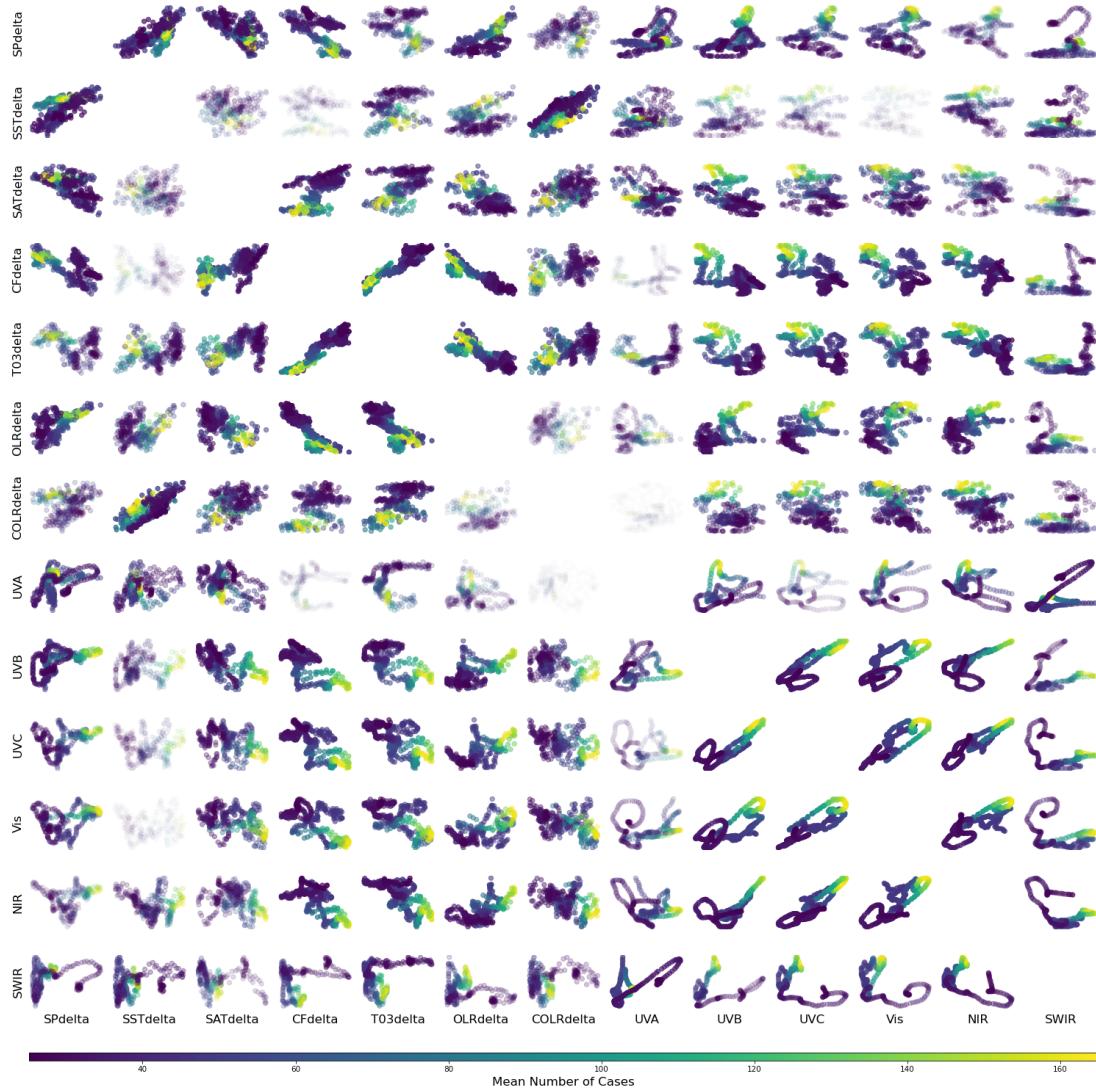


Figure 9: Correlation between COVID-19 cases and environmental features under a generalized coordinate system. Color represents the mean number of cases while brightness correlation between each feature. environmental features are calculated as the difference between the ascending and descending scans from the reconstructed dataset. SP (Surface pressure from forecast. (hPa)), SST (Surface skin temperature. (Kelvin)), SAT (Temperature of the atmosphere at the Earth's surface. (Kelvin)), CF (Combined layer cloud fraction. (0- 1). (Unitless)), T03 (Total integrated column ozone burden. (Dobson units)), OLR (Outgoing long-wave radiation flux. (watts/m²)), COLR (Clear-sky outgoing long-wave radiation flux. (watts/m²)).

As stated before correct circadian signaling is of extreme importance for covid-19 recovery and protection. Stationary points of solar radiation such as those found in the middle of the year and winter, or at sunrise, sunset, or the middle of the day might impact the ability to signal those events if parts of the pathway are hindered either by environmental factors or host factors.

Another environmental factor that might impair circadian signaling, especially during the start

235 of the outbreak, was the low solar activity. Initial cases of COVID-19 were reported at the end of
236 2019 where sunspots were almost absent as we were entering solar cycle 25. If another frequency
237 component of solar activity is affecting circadian signaling most likely will be of lower frequency.
238 The Gleissberg cycle is a cycle with a duration of around 100 years [30], low levels of solar radiation
239 at the Gleissberg frequency also correlate to the start of the COVID-19 pandemic as well as to other
240 major outbreaks, particularly the Spanish Flu outbreak

241 Conditions found at stationary points at the different frequency components of solar activity
242 could impair circadian signaling, especially if the host is compromised over those pathways.

243 Reframing viral genome reconstruction.

244 Previous models suggest a deterministic mechanism behind viral replication and the role of random
245 mutations most likely is less than previously suggested. How the virus replicates its genome and
246 reconstructs the viral particle is an area of continuous research and very challenging. As the virus
247 adapts to the host new strategies might evolve as a result. An example of such adaptation is the
248 discovery of new entry pathways of SARS-Cov2 into the cell. The first entry mechanism was by
249 the interaction between ACE2 receptors, while the second involved TMPRSS2 [31]. Viral entry
250 and other phenotypes could be the result of the adaptation inside the viral genome and might be
251 preferred for a specific cluster of SARS-Cov2 variants. How those different strategies emerge could
252 be addressed by reframing the viral reconstruction process as different problems. They might share
253 some properties among them but the methodology used to analyze them will differ between them.

254 As a dynamical system.

255 The learned representation obtained from the different models showed a cyclical pattern in one
256 dimension, while the meaning of the second dimension remains to be addressed. One possibility
257 of interpretation behind the different clusters and the learned dimensions is that they encode the
258 interaction between the host and the virus. Each dimension might encode the resource availability
259 in the host and the resource expenditure from the virus. This interaction could be modeled as a
260 predator-prey model, where the available resources inside the host are the prey and the predator is
261 the virus. One simple way to address such behavior is by modifying the VAE architecture to also
262 retrieve a dynamical system. The learned representation is derived and the different constants of
263 the predator-pray model are added to the optimization algorithm to be found. The use of autoen-
264 coders for the detection of parsimonious dynamical systems has already been proposed as a feasible
265 technique for complex data sources [23].

266 Retrieved dynamics (assuming the time scale is in days) show a peak around the third day and
267 a duration of the transitory phase of around 5 days. This duration roughly matches the incubation
268 period of SARS-Cov2 [32] pointing to the importance of applying such models, specifically at the
269 beginning of an outbreak where information is scarce. Furthermore, adding the dynamic part to the
270 VAE stabilizes training. As VAEs trained using similar training conditions and the same samples
271 can't sort the viral genomes in comparison with their dynamical counterparts (Figure 10).

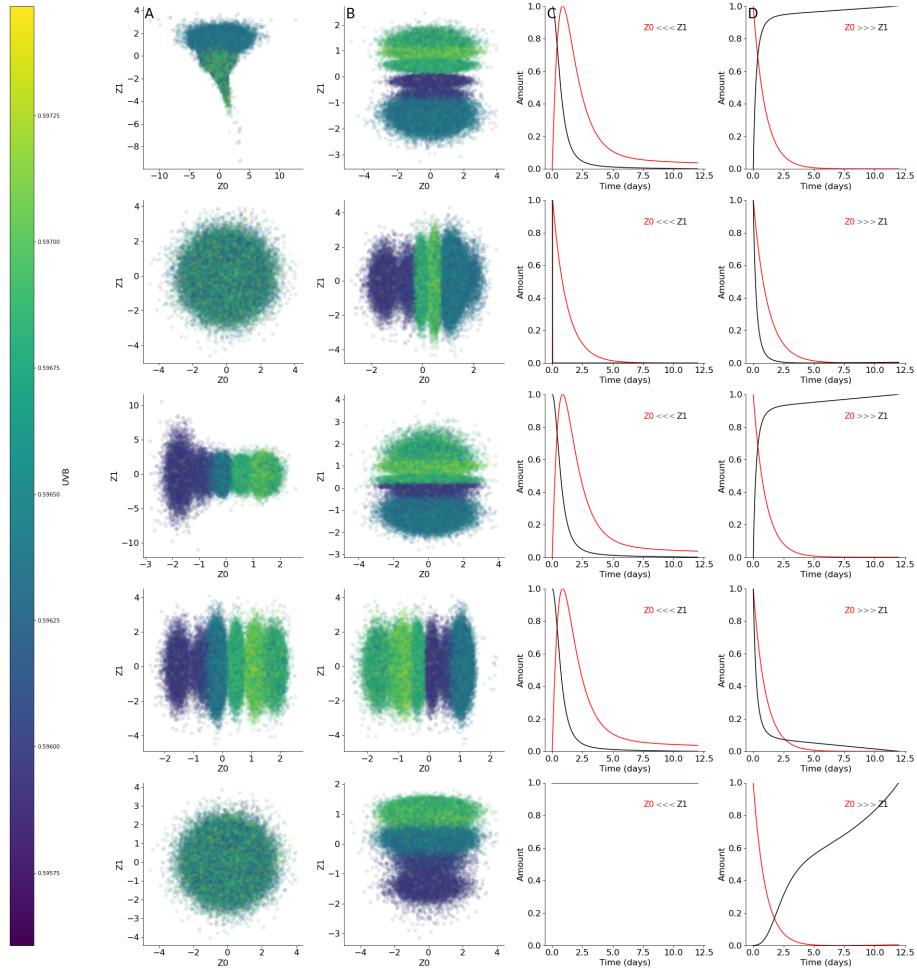


Figure 10: Dynamical VAE stabilizes training and allows the clustering of SARS Cov2 genomes. A Learned representation was obtained without dynamical loss, B Learned representation was obtained with a dynamical loss, same samples were used to train each model row-wise. D, E Retrieved dynamical system, initial condition displayed in the plot

Being able to approximate the incubation time from genomic surveillance data shows the importance of genomic data as well as one of the many applications of such data other than mutational patterns. However, at the beginning of an outbreak, the amount of information is scarce making it difficult to obtain such an approximation. Dividing the complete dataset into 6 different periods using SD as a time scale shows that each model can retrieve a dynamical model with the same duration of the transitory phase.

Each period has a duration of about 2 months and if the samples are equally distributed around large geographical extensions the number of samples needed per location will decrease dramatically. Sharing data at the early stages of an outbreak of a novel pathogen could be the only way such approximations could be obtained (Figure 11).

Nevertheless, caution is needed with the interpretation of these results. As most likely only

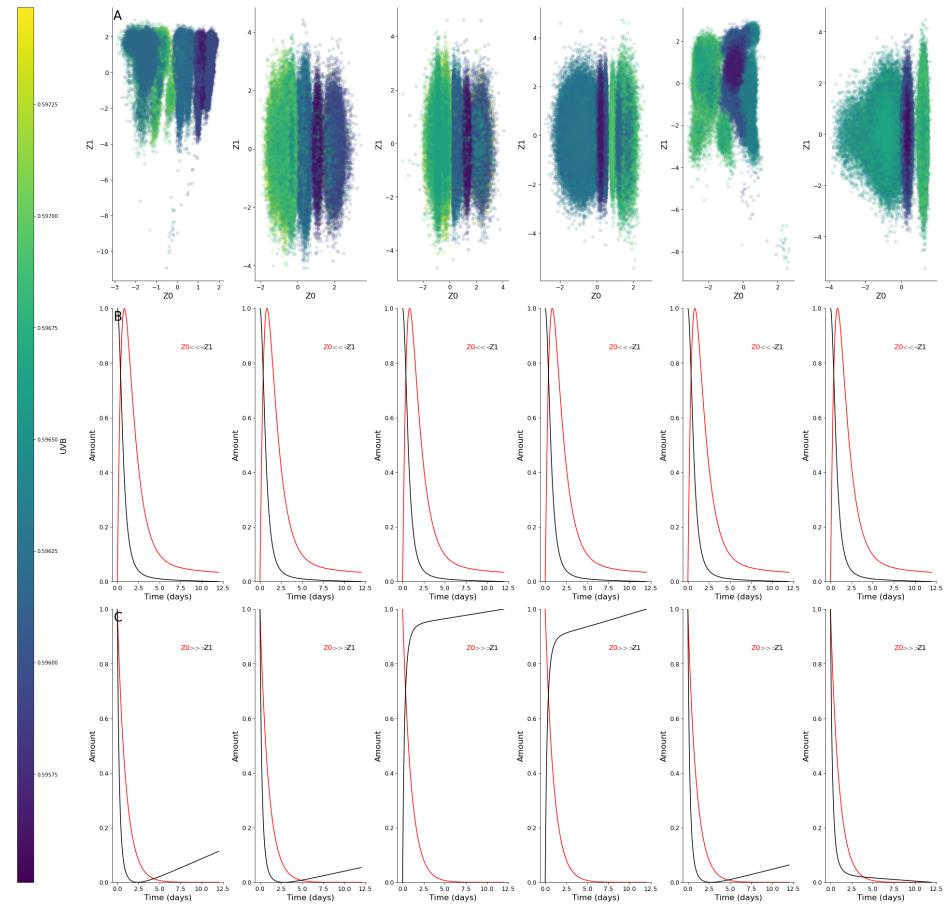


Figure 11: Dynamics can be retrieved by a single fragment of the yearly cycle. A learned representation was obtained by segmentation of the data into 6 consecutive periods by SD. B, C Retrieved dynamical system under different initial conditions.

represent the interaction with a specific tissue within the host, particularly the upper respiratory tract as is the tissue origin for most of the samples. Recurrent COVID-19 symptoms with a duration greater than the incubation time could be suggestive of intrahost infections or a persistent infection.

Dynamic behavior inside the SARS Cov2 genome suggests that the genome can be analyzed using a time series approach. Smoothed single nucleotide time series by rolling mean showed a series of oscillations in the content of the different nucleotides (Figure 12 A,C,E,G). Oscillations are nearly absent when the nucleotide frequency is grouped by the number of sunspots (Figure 12 E). Reconstructing the path followed by the four different nucleotides under different time scales resulted in a series of attractor-like behavior when the data for the time scale contains one full cycle (Figure 12 B,D,F,H). In the case of the SN time scale, only a single arc is observed.

Oscillatory behavior inside the SARS-Cov2 nucleotide composition can be used to schedule nucleotide-analogs antiviral treatment. As changes in the nucleotide content will increase or decrease the efficiency of those analogs. Particularly for SARS-Cov2 the different oscillations could explain the mixed reports for Remdesivir an adenine analog at the beginning of the pandemic. The

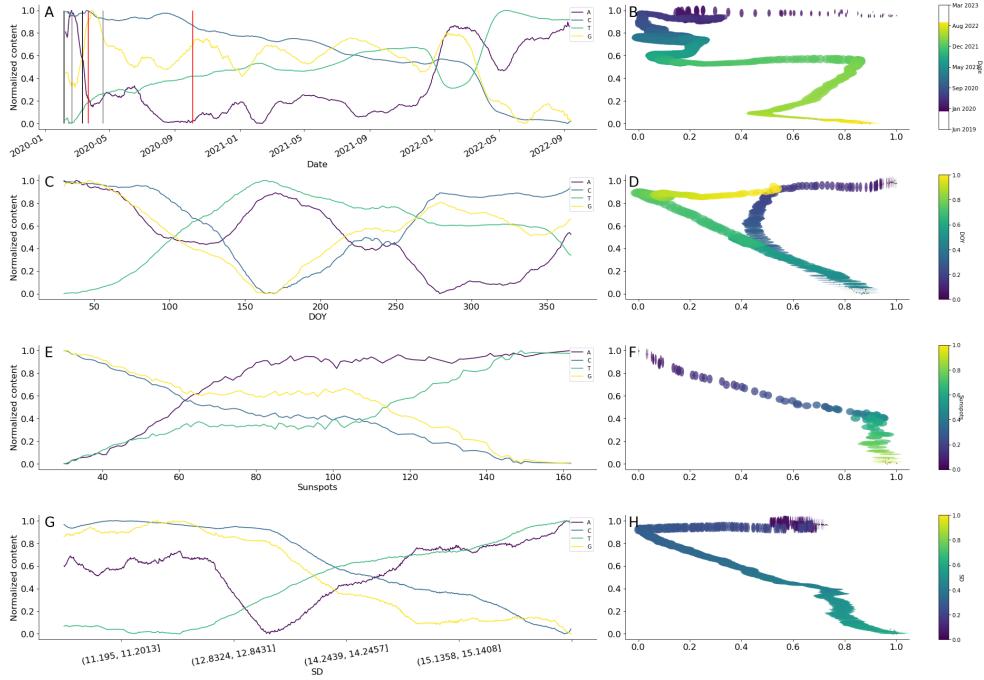


Figure 12: Attractor-like behavior of SARS-CoV2 genome composition under different time scales. A. Single nucleotide dynamics by date, vertical lines show the time on which different Remdesivir trials took place. Black and grey Wang et al and Beigel et al trials and red WHO consortium trial. B, D, F, and H Composition dynamics, color equal to the selected time scale, X and Y equal to [A] and [C] content, width and height of the marker are equal to [T] and [G] content. C Yearly single nucleotide dynamics. E Single nucleotide dynamics under a solar time scale, the number of sunspots are used as a measure of solar activity. G Yearly single nucleotide dynamics under the SD timescale.

earliest trial by Wang et al [33] was conducted from February to March the next one was conducted by Beigel et al [34] from late February to middle April 2020 both trials found a benefit from Remdesivir treatment. While the WHO solidarity consortium trial [35], which found little to no effect of Remdesivir treatment, took place from late March to early October 2020 when the adenine content was reduced between seven and ten fold (Figure 12 A).

The high dimensionality of the data hinders the ability to properly visualize how the genome adapts. One simple method to represent such information is by encoding the k-mer composition as the shape of a closed curve. This simple representation allows to analyze changes in composition under different time scales and their correlation to environmental variables previously selected.

K-mer composition under the environmental variables TO3delta and CFdelta grouped by SD showed a similar pattern as the one found in the stacked Kmer models. Similar shapes are near together and as the environmental variables change a transition between different shapes is also observed. While UVA and UVB showed a linear correlation between the environmental scale and viral genome composition and a smooth transition between the different shapes (Figure 13).

This new representation further confirms the dependence between SARS-CoV2 adaptation and

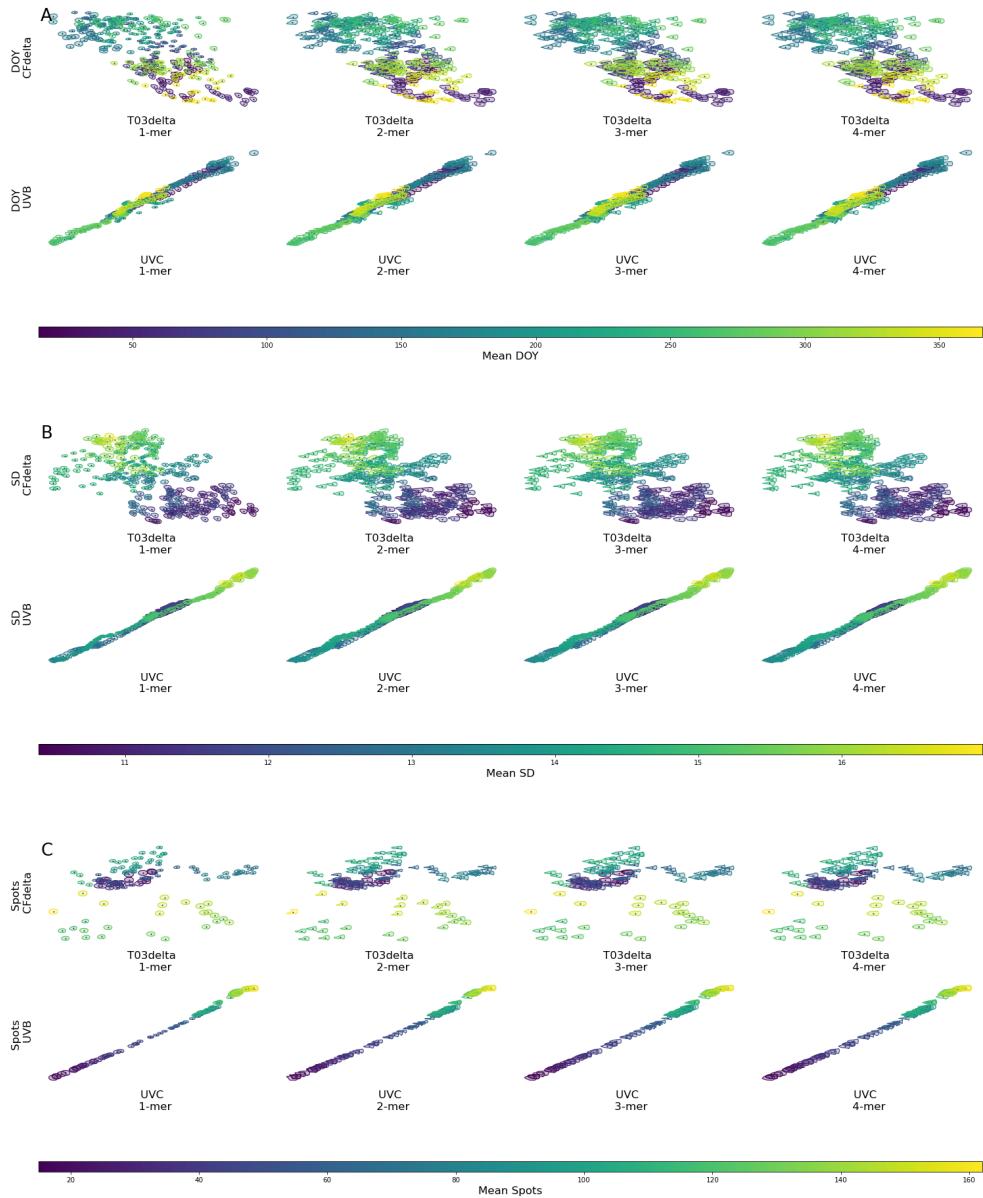


Figure 13: SARS-CoV2 genome adaptation under different time scales and environmental changes. Genome composition is encoded as the shape of a closed curve, each column represents the k-mer composition and rows the time scale. Only highly correlated features were analyzed.

312 solar activity. Also starts to provide a possible interpretation of the learned representation of the
 313 stacked K-mers models. The frequency of specific K-mers could encode an environmental advantage
 314 to the viral genome. As UV radiation is able to inactivate viruses by generating new bonds within
 315 paired bases it is possible that by reducing the frequency of specific K-mers such bonds are less likely
 316 to be formed [36]. Furthermore changes in genome size could also hint at an adaptation response to
 317 radiation. A smaller genome will be less likely to be inactivated by high levels of UV radiation that

318 happen in the summer or at the Shwabe maximum. While a larger genome will be favored at low
319 levels of radiation such as winter and the Shwabe minimum.

320 Changes in genome size could also affect the perceived severity of the infection by adding or
321 removing parts of the genome. A larger genome will contain the full range of information to generate
322 a wider range of symptoms. While a smaller one will be more constrained in that regard allowing
323 the virus to infect with less perceived severity.

324 For example, Orf8 is a viral protein that plasma levels correlated with disease severity and
325 mortality [37]. Infecting viruses lacking Orf8 was associated with better outcomes. Orf8 also inhibits
326 the presentation of antigens by the MHC of class 1 [38]. The shrinking SARS-CoV2 genome could
327 delete or partially delete this orf as an adaptation response to changing levels of radiation. Other
328 Orfs inside the genome viruses could also be subjected to a similar adaptation mechanism.

329 As resource optimization.

330 Seasonal adaptation (genome rearrangement) and susceptibility(high spread rate) of the host to
331 SARS-CoV2 point towards a specific dependence of SARS-CoV2 to cellular conditions. Specific
332 seasonal expression patterns might drive the susceptibility to SARS-CoV2 as the main components
333 of the virus are RNA and proteins. One possible mechanism used by the virus to adapt to those
334 conditions will be to adapt its genome to the CUB of the host. CUB optimization effectively
335 optimizes the nucleotide and amino acid availability inside the cell. Changes inside the genomic
336 composition of the cell due to seasonal changes will alter the available resources to synthesize those
337 genes, transcripts, and proteins. Thus even a virus highly adapted to the CUB of a particular cell
338 type will face periods with low resources to replicate itself lowering its R0.

339 CUB is coded by the availability of specific codons inside the genes expressed inside a specific
340 cell [39]. Seasonal gene expression will change a subset of genes leading to a dynamical CUB thus if
341 a virus optimizing CUB will likely also follow seasonal gene expression. By mimicking the genetic
342 composition of the cell a virus will be able to follow a particular seasonal pattern. Mimicry will
343 also increase the chance of copying large enough pieces of a gene/transcript and synthesizing a viral
344 protein with a fragment highly similar to the host leading to autoimmunity.

345 Seasonal genes with a similar composition to SARS-CoV2 will match both nucleotide and amino
346 acid pools needed for the virus to replicate itself. To test that hypothesis a new dataset is generated
347 with the GRCh 38 reference transcripts sequences and compared with the mean SARS-CoV2 com-
348 position using the Mahalanobis distance. Selected transcripts are then predicted with the trained
349 model and the transcripts with the lowest reconstruction error are selected. Information about the
350 retrieved transcripts is then obtained with mygene service.

351 The previous filter results on a series of transcripts but information was obtained for only a
352 subset of them. Interpro classification showed among the main coded domains were the armadillo-
353 like domain, the Protein Kinase domain, and the zinc finger domain (Figure 14 B). Pathway analysis
354 showed the involvement of the immune system, gene expression, and signal transduction among
355 others (Figure 14 C). However, the involvement of GPCR signaling was observed under different
356 categories. While the involvement of the different transcripts in diseases showed many hereditary

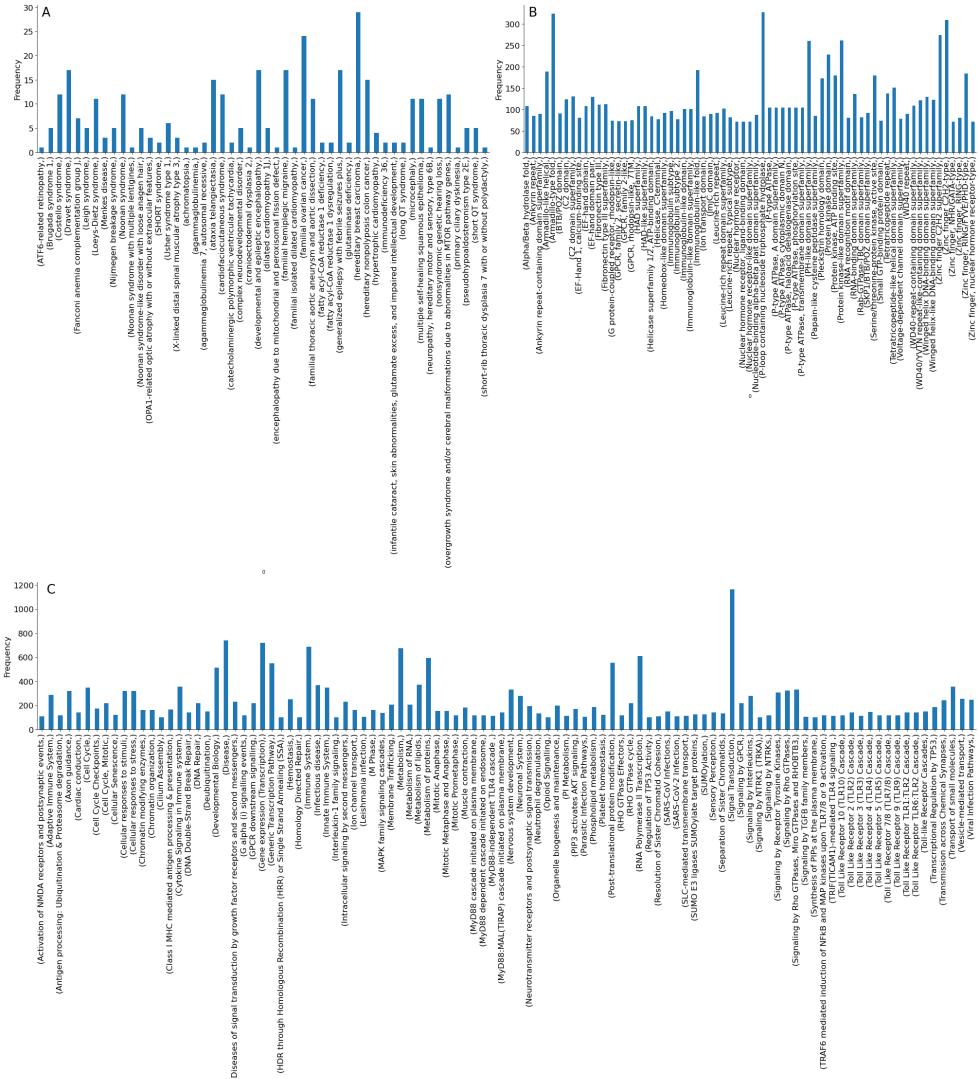


Figure 14: Data mining of compositionally similar transcripts. A. Disease involvement of the selected transcripts. B. Most common domains coded by the selected transcripts. C. Pathway involvement of the proteins coded by the selected transcripts.

disorders that affect the brain and the heart (Figure 14 A).

Although in the case of SARS-CoV2, the information regarding following seasonal expression patterns is scarce, mainly due to two reasons. The low amount of information regarding seasonal expression patterns. And that most of the functional information is skewed towards a few subsets of proteins/genes [40], however valuable information still can be retrieved.

For example, two seasonal genes *CLOCK* and *RORA* were selected by the previous filter. *CLOCK* is mainly expressed over the winter season, while *RORA* is expressed over the summer season [41]. Direct involvement in COVID-19 has also been found for both genes. *BMAL1* is a protein that interacts with *CLOCK* and controls the expression of the ACE2 receptor also regulating the entry

366 of SARS-CoV2 [42]. While RORA is differentially expressed in COVID-19 as well as in heart failure
367 [43]. These two examples show two mechanisms that can be concurrent. Differentially regulated
368 genes could increase the susceptibility to COVID-19 like in the case of RORA. Genomic mimicry
369 could copy a fragment of a host protein to facilitate the manipulation of the molecular machinery,
370 as suggested by CLOCK. A third possibility is when a host protein presents self-interaction, in this
371 case, the similarity is on a specific region. One example of this kind of mimicry is the TMPRSS15
372 protein member of the TMPRSS and some members of this family undergo auto activation [44].
373 TMPRSS15 is also found by the previous analysis and is another entry factor of SARS-CoV2 [45].

374 A manual search of the different genes results in several genes with little information regarding
375 their direct involvement with SARS-CoV2 infection. Most of the genes didn't show any kind of
376 involvement, however addressing them as a whole also gave a little insight into the possible subversion
377 mechanism used by SARS-CoV2.

378 For example, the selection of different olfactory receptors, other GPCRs, channels, and trans-
379 porters point towards the selection of a series of proteins that target the membrane. This specific
380 targeting is important for the release of viral particles to the extracellular space [46]. Similarity with
381 other proteins involved in endosomal trafficking such as Rab27B involved in the delivery of secretory
382 granules [47]. Or Rab3C is involved in the last steps of exocytosis and the release of neurotransmit-
383 ters [48]. HOOK1 another selected protein, plays a role in trafficking via the microtubule network
384 [49]. Members of the SNX (SNX2, SNX16, SNX10, SNX14, SNX13) family regulate GPCR signaling
385 by the trafficking of those receptors [50]. Those particular proteins contain a BAR domain that
386 interacts with different membrane lipids in the membrane, this allows them to modify the shape of
387 the membrane to create an endocytic vesicles [51]. Membrane composition can also be modified to
388 create those anchoring points for BAR domains. Particularly PIP2 is regulated by proteins such as
389 PI3K and INPP4 [52] and different isoforms of those proteins were found in the previous analysis.

390 Addressing the pathway involved by manually searching the function of the different selected
391 genes continues to be challenging. Also as the virus continues to adapt to the environment, selected
392 genes will likely change. This will skew the selection of different proteins within the same family and
393 perform the same function. Therefore a broader selection could be more useful than a set of specific
394 ones. However this only further confirms the need to develop an automated method to analyze the
395 selected genes.

396 Post Acute Sequelae

397 The circadian nature of SARS Cov2 points to circadian dysregulations in the post-acute phase.
398 Following the same logic, treatments that restore circadian rhythmicity could also be useful to
399 manage post-acute sequelae. The circadian nature of SARS Cov2 has also been proposed before
400 [53] and some of the symptoms also show some circadian rhythmicity. Light intolerance is perhaps
401 one of the most easily recognizable symptom with circadian rhythmicity [54]. Specifics on how
402 such dysregulations drive post-acute sequelae are presented as epigenetic manipulation and two
403 mechanisms are proposed.

404 One mechanism used to lower the replication of a virus is the degradation of nucleotides. Re-

405 sulting in an embargo/blockade of the resources needed to replicate the virus [55]. This strategy by
406 itself will not be able to prevent the spread of the virus, but it will lower the burden on the immune
407 system. If the infection became persistent and affects other organs a broader blockade could be
408 obtained by the downregulation of genes with similar composition to the infecting virus. Or shifting
409 the kind of isoform being expressed. This will enable the cell to produce the protein, and perform
410 the same function but with different genomic composition. Downregulation will also reduce the
411 probability for the virus to copy a protein fragment that further refines the ability of the virus to
412 manipulate cellular molecular machinery. Copying specific protein fragments of the host could also
413 derive in autoimmunity if a specific fragment is found in the virus by the immune system.

414 The previous mechanism proposes a one-sided response and frames the post-acute sequelae as
415 a starving strategy. Yet the virus could also generate such dysregulation as a strategy to freeze
416 in time those conditions that enable it to generate a large number of copies of itself. Thus viral
417 infections could also generate epigenetic changes to increase the expression of compositional similar
418 genes at a cellular level. This could create miscommunication between neighboring cells and prevent
419 the function of the tissue as a whole.

420 Both mechanisms propose epigenetic changes that modulate resource availability and it could
421 be complicated to assess the correctness of either of those. However, epigenetic changes occur post
422 SARS-Cov2 infection [56], and manipulation of epigenetic machinery such as HDAC2 has already
423 been reported [57]. Therefore epigenetic modifications are a result of SASR-Cov2 infection but its
424 meaning remains to be addressed. Moreover, those repercussions might not be noticeable by the
425 host until they reach a certain threshold.

426 Epigenetic modification could also increase the resources for another kind of virus, activating dor-
427 mant infections. Viral activation after SARS-Cov2 infection has been previously reported, however,
428 has not been linked to any kind of epigenetic modifications [58].

429 Suggested mechanisms rely on the dysregulation of genes with high compositional similarity
430 to SARS-Cov2, thus identifying the phenotype that results from such dysregulation could provide
431 insight into post-acute sequelae. Mentioning some of the previously selected genes CLOCK dysreg-
432 ulation by gene mutations could result in delayed sleep phase disorder [59]. A disease that presents
433 difficulties in sleep and misregulation of body temperature, symptoms already reported as post-acute
434 sequelae of COVID-19 [60]. While RORA dysregulation by gene mutation could result in Intellectual
435 Developmental Disorder with or Without Epilepsy or Cerebellar Ataxia, a condition with cognitive
436 impairment [61], a symptom also reported for post-acute sequelae of COVID-19 [62].

437 Circadian dysregulation could also be the result of impairing other circadian signaling pathways.
438 Particularly protein glycosylation prevents the degradation of proteins modifying their signaling
439 time [63]. And calcium oscillatory signaling synchronizes the circadian rhythm in the central nervous
440 system [64]. Lack of synchronization between the different oscillators, local and global could result in
441 miscommunication between different systems inside the host leading to dysautonomia. Glycosylation
442 and calcium signaling could be influenced by genes selected from their high compositional similarity
443 to SARS-Cov2 as well as to generate exercise intolerance by two different mechanisms.

444 AGL is a gene that encodes for the glycogen debranching enzyme, mutations in AGL can generate
445 a condition known as Glycogen Storage Disease III. This particular disease is characterized by the

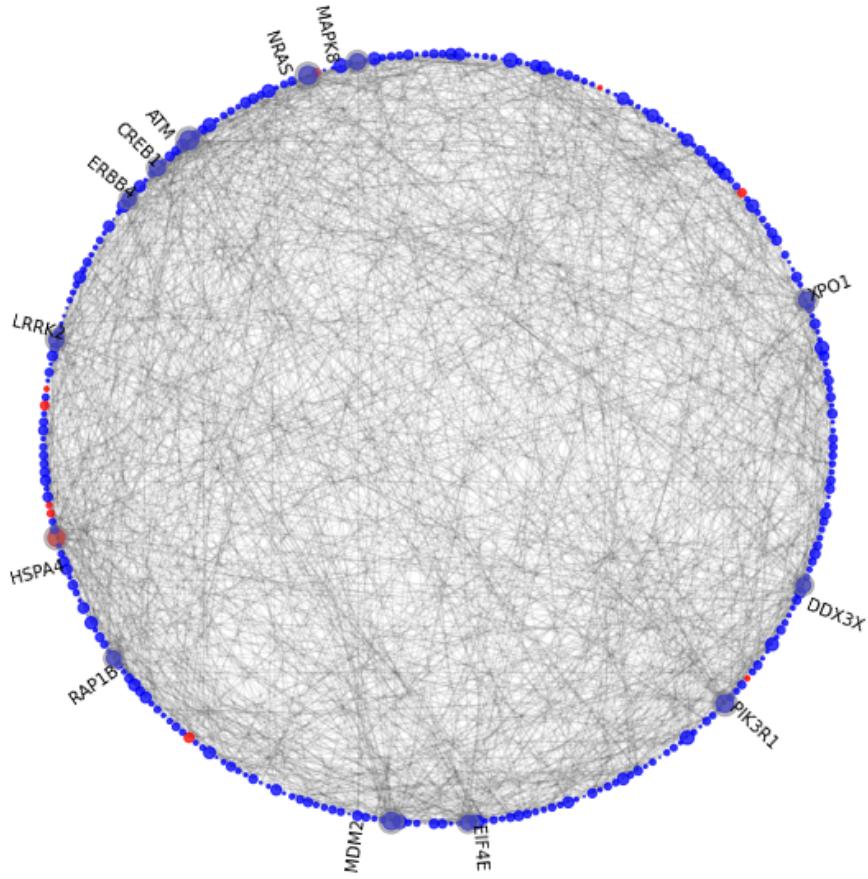


Figure 15: **Protein interactions between the selected compositionally similar genes.** Blue selected genes, red not selected.

446 accumulation of glycogen and affects the liver and the skeletal muscle [65]. This condition can
 447 also generate exercise intolerance and the proposed mechanism behind is the result of low energy
 448 availability to the skeletal muscle [66].

449 Calcium mishandling could also derive in exercise intolerance, particularly mutations in calmod-
 450 ulin 2 lead to a condition known as catecholaminergic polymorphic ventricular tachycardia (CPVT).
 451 Among the different symptoms of CPVT are abnormal heart rate and exercise intolerance. Mutations
 452 inside calmodulin-2 that generate CPVT lower its affinity to calcium [67]. Calmodulin-2 is
 453 also a compositionally similar gene to SARS cov2.

454 Framing post-acute sequela as the miss regulation of compositionally similar genes could help
 455 to increase the number of possible treatments and will have the advantage to provide in some cases
 456 already approved treatments. For example, lithium cation that targets CLOCK has been used to
 457 increase the recovery rate in COVID-19 patients [68] as well as being evaluated as a treatment
 458 for long Covid (ClinicalTrials.gov Identifier NCT05618587). Drugs targeting AGL also showed to

459 inhibit viral replication [69]. While treatments recommended for CPVT such as Flecainide showed
460 a reduction in the odds of being hospitalized for COVID-19 [70]. Beta-blockers such as metropolol
461 also showed a protective effect in COVID-19 patients [71], and it's being evaluated for long covid
462 treatment (ClinicalTrials.gov Identifier NCT05096884).

463 Furthermore, if compositionally similar genes are dysregulated they could disrupt the function
464 of a protein on which they converge due to their different interactions. Analyzing the interaction
465 network of the compositionally similar genes showed a series of targets with a high number of
466 interactions. Of all those different proteins only one presented the condition of interacting with a
467 high number of compositionally similar genes and not being compositionally similar to SASRS-Cov2
468 (Figure 15).

469 HSPA4 encodes for the HSP70 protein, a protein that has a neuroprotective role in many cerebral
470 insults [72]. Its also involved in transient cerebral ischemia, a stroke that lasts for a few minutes, and
471 the symptoms disappear between one and 24 hours. Symptoms include, numbness or weakness in
472 the face, arm, or leg, especially on one side of the body, trouble seeing in one or both eyes, difficulty
473 with walking, dizziness, confusion or difficulty in talking or understanding speech, loss of balance
474 and coordination. HSP70 is also manipulated by SARS-CoV2 infection making it another target to
475 develop treatments [73].

476 Viral persistence is needed for those two previous mechanisms to be activated. Viral persistence
477 could drive either the blockade/starvation or the establishment mechanism. The specific variant,
478 infection date, and previous infections most likely will determine the specific kind of dysregulation.
479 Viral persistence will also increase the risk of copying fragments of host proteins and developing
480 autoimmunity as well as immune activation. Thus addressing viral persistence will need to be
481 concomitant with other treatments aiming to restore or manage some of the symptoms of post-acute
482 sequelae.

483 As more than one intervention will likely be needed to restore post-acute sequelae there's a higher
484 risk of interaction between treatments. Thus careful considerations need to be taken to design an
485 adjust possible treatments.

486 Conclusions

- 487 • The existence of different molecular clocks encoded inside the SARS-CoV2 genomic sequence
488 shows a deterministic mechanism used by the virus to replicate its genome. Other clock
489 components remain to be found if they exist. Nevertheless, the suggested random nature of
490 viral mutation is less than previously suggested or non-existing.
- 491 • Detection of other frequency components inside solar activity and radiation could provide other
492 periods of high SARS-CoV2 susceptibility.
- 493 • Non-linearity inside the environmental driver for SARS-CoV2 adaptation and susceptibility
494 adds a confounding variable to the different risk/benefit assessments. Interpretation of the
495 efficacy and safety of pharmacological and non-pharmacological interventions might need a
496 reanalysis.

- 497 • SD is not a perfect time scale for SARS-Cov2 seasonality but it offers an easy metric to follow
498 for the implementation of different risk reduction measures as well as vaccination schedules.
- 499 • Addressing nutritional deficiencies that impact circadian signaling could be used as a prophylactic
500 treatment at periods of low solar activity.
- 501 • Low-frequency components of low solar activity provide an easy metric for pandemic preparedness.
- 503 • Although SARS-Cov2 follows UVB radiation other consequences of low solar activity could
504 also ease its transmission.
- 505 • Targeting low-variable components inside the SARS Cov2 genome might provide a better
506 long-term treatment strategy.
- 507 • Further improvements in the different sequence representation schemes and machine learning
508 models will fast-track the characterization of emerging and neglected pathogens.

509 Acknowledgments

510 Access to data sources from different government organizations was key for the development of the
511 previous analysis. I would like to thank everyone involved in the acquisition, curation, and other
512 support staff as well as the policymakers who made the data available to anyone. Publicly available
513 data offers a model of silent collaboration from which everyone benefits from it.

514 Declarations

515 The author declares no competing interests. Although the author is raising funds to continue research
516 under different platforms listed here. <https://linktr.ee/tavoglc>

517 Methods

518 Genomic sequences were obtained from the NCBI SARS-Cov2 resource site up to September 2022
519 [74]. Only complete genomes without any ambiguous characters were selected for the analysis.
520 Available metadata was used to reverse geolocate the corresponding geographical location. The
521 isolation date was used as a time descriptor for each sequence as well as for the calculation of the
522 SD [75].

523 Three sequence representations schemes are proposed to analyze SARS Cov 2 genomes. The first
524 one consists of the stacked frequency of k-size fragments or k-mers. SARS-Cov2 genome sequences
525 are spitted in a sliding manner leading to fragments with an overlap of (k-1) nucleotides and (n-k)
526 fragments. Frequency is normalized for each k and the different vectors are combined in a single
527 one. The same method is applied to the reference transcripts dataset GRCh 38.

528 For the second one, each sequence is divided into 16 fragments of equal size. Then for each
529 fragment, a graph is constructed by setting as nodes the 2-mer and adding a link between consecutive
530 2-mers derived from a sliding fragmentation. The same process is applied to the reversed fragment
531 and the normalized adjacency matrix is used as sequence representation.

532 The third dataset consists of the one-hot-encoded full genomic sequence, yet the sequence is
533 transformed into a 4D array. Where the first three dimensions represent the location of each base
534 in the genome and the last one is the kind of nucleotide. This specific rearrangement could bring
535 closer different genomic sequences, mimicking the 3D arrangement of the viral genome inside the
536 viral particle.

537 For each dataset, a specific machine learning model is used in a representation learning task. For
538 the stacked k-mers dataset the VAE architecture consists of a multilayer perceptron, and a modified
539 version to retrieve a dynamical system. For the remaining datasets, the VAE architecture consists of
540 a deep convolutional network of different architectures for each dataset. A five-fold cross-validation
541 was applied and only the best fold is shown.

542 Case data were obtained by the different government agencies involved, as testing and update
543 frequency changed between countries data was collected up to May 2022 [76] [77] [78] [79] [80] [81]
544 [82] [83] [84]. Reverse geolocation is used to retrieve the latitude and longitude at a city level for
545 the available epidemic curves.

546 Solar activity data consists of sunspots data from the WDC-SILSO Royal Observatory of Bel-
547 gium Brussels [85] and the TSIS SIM level 3 solar spectral irradiance 24hr means [86]. While
548 environmental features are obtained from the Aqua/AIRS L3 Daily Standard Physical Retrieval
549 dataset. Missing values are predicted with a random forest regressor trained on a sliding win-
550 dows of 70 days [87]. A complete summary of the different scripts and datasets can be found at
551 <https://github.com/TavoGLC/SARSCov2Solar>

552 References

- 553 [1] Á. O'Toole *et al.*, "Assignment of epidemiological lineages in an emerging pandemic using
554 the pangolin tool," *Virus Evolution*, vol. 7, no. 2, Jul. 2021, veab064, ISSN: 2057-1577. DOI:
555 [10.1093/ve/veab064](https://doi.org/10.1093/ve/veab064). eprint: <https://academic.oup.com/ve/article-pdf/7/2/veab064/49179570/veab064.pdf>. [Online]. Available: <https://doi.org/10.1093/ve/veab064>.
- 557 [2] D. Bank, N. Koenigstein, and R. Giryes, *Autoencoders*, 2021. arXiv: 2003.05991 [cs.LG].
- 558 [3] T. Trinh, A. Dai, T. Luong, and Q. Le, "Learning longer-term dependencies in RNNs with
559 auxiliary losses," in *Proceedings of the 35th International Conference on Machine Learning*, J.
560 Dy and A. Krause, Eds., ser. Proceedings of Machine Learning Research, vol. 80, PMLR, Oct.
561 2018, pp. 4965–4974. [Online]. Available: <https://proceedings.mlr.press/v80/trinh18a.html>.
- 563 [4] M. Afsar *et al.*, "Drug targeting nsp1-ribosomal complex shows antiviral activity against sars-
564 cov-2," *eLife*, vol. 11, S. Haider, M. Zaidi, and S. Haider, Eds., e74877, Mar. 2022, ISSN:

- 565 2050-084X. DOI: 10.7554/eLife.74877. [Online]. Available: <https://doi.org/10.7554/eLife.74877>.
- 566
- 567 [5] H.-T. Kao, A. Orry, M. G. Palfreyman, and B. Porton, "Synergistic interactions of repurposed
568 drugs that inhibit nsp1, a major virulence factor for covid-19," *Scientific Reports*, vol. 12,
569 no. 1, p. 10174, Jun. 2022, ISSN: 2045-2322. DOI: 10.1038/s41598-022-14194-x. [Online].
570 Available: <https://doi.org/10.1038/s41598-022-14194-x>.
- 571 [6] B. Oronsky *et al.*, "Nucleocapsid as a next-generation covid-19 vaccine candidate," *International
572 Journal of Infectious Diseases*, vol. 122, pp. 529–530, Sep. 2022, ISSN: 1201-9712. DOI:
573 10.1016/j.ijid.2022.06.046. [Online]. Available: <https://doi.org/10.1016/j.ijid.2022.06.046>.
- 575 [7] M. S. Khan *et al.*, "Adenovirus-vectored SARS-CoV-2 vaccine expressing S1-N fusion protein,"
576 *Antibody Therapeutics*, vol. 5, no. 3, pp. 177–191, Jul. 2022, ISSN: 2516-4236. DOI: 10.1093/
577 abt/tbac015. eprint: <https://academic.oup.com/abt/article-pdf/5/3/177/45348595/tbac015.pdf>. [Online]. Available: <https://doi.org/10.1093/abt/tbac015>.
- 579 [8] E. K. V. B. Silva *et al.*, "Immunization with sars-cov-2 nucleocapsid protein triggers a pul-
580 monary immune response in rats," *PLOS ONE*, vol. 17, no. 5, pp. 1–17, May 2022. DOI:
581 10.1371/journal.pone.0268434. [Online]. Available: <https://doi.org/10.1371/journal.pone.0268434>.
- 583 [9] K. L. O'Donnell, T. Gourdine, P. Fletcher, C. S. Clancy, and A. Marzi, "Protection from
584 covid-19 with a vsv-based vaccine expressing the spike and nucleocapsid proteins," *Frontiers
585 in Immunology*, vol. 13, 2022, ISSN: 1664-3224. DOI: 10.3389/fimmu.2022.1025500. [Online].
586 Available: <https://www.frontiersin.org/articles/10.3389/fimmu.2022.1025500>.
- 587 [10] J. Martel, S.-H. Chang, G. Chevalier, D. M. Ojcius, and J. D. Young, "Influence of electromag-
588 netic fields on the circadian rhythm: Implications for human health and disease," *Biomedical
589 Journal*, vol. 46, no. 1, pp. 48–59, 2023, ISSN: 2319-4170. DOI: <https://doi.org/10.1016/j.bj.2023.01.003>. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S2319417023000033>.
- 592 [11] T. A. G. on SARS-CoV-2 Virus Evolution, "Historical working definitions and primary ac-
593 tions for sars-cov-2 variants," Mar. 2023. [Online]. Available: <https://www.who.int/publications/m/item/historical-working-definitions-and-primary-actions-for-sars-cov-2-variants>.
- 596 [12] S. W. P. C. (U.S. Dept. of Commerce NOAA, "Solar cycle progression," Apr. 2023. [Online].
597 Available: <https://www.swpc.noaa.gov/products/solar-cycle-progression>.
- 598 [13] A. E. Samoilov *et al.*, "Case report: Change of dominant strain during dual sars-cov-2 in-
599 fection," *BMC Infectious Diseases*, vol. 21, no. 1, p. 959, Sep. 2021, ISSN: 1471-2334. DOI:
600 10.1186/s12879-021-06664-w. [Online]. Available: <https://doi.org/10.1186/s12879-021-06664-w>.

- 602 [14] H.-Y. Zhou *et al.*, “Genomic evidence for divergent co-infections of sars-cov-2 lineages,” *bioRxiv*,
603 2021. DOI: 10.1101/2021.09.03.458951. eprint: <https://www.biorxiv.org/content/early/2021/09/04/2021.09.03.458951.full.pdf>. [Online]. Available: <https://www.biorxiv.org/content/early/2021/09/04/2021.09.03.458951>.
- 606 [15] T. Karamitros, G. Papadopoulou, M. Bousali, A. Mexias, S. Tsiodras, and A. Mentis, “Sars-
607 cov-2 exhibits intra-host genomic plasticity and low-frequency polymorphic quasispecies,”
608 *Journal of Clinical Virology*, vol. 131, p. 104585, 2020, ISSN: 1386-6532. DOI: <https://doi.org/10.1016/j.jcv.2020.104585>. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S1386653220303279>.
- 611 [16] F. Chen *et al.*, “Dissimilation of synonymous codon usage bias in virus–host coevolution due
612 to translational selection,” *Nature Ecology & Evolution*, vol. 4, no. 4, pp. 589–600, Apr. 2020,
613 ISSN: 2397-334X. DOI: 10.1038/s41559-020-1124-7. [Online]. Available: <https://doi.org/10.1038/s41559-020-1124-7>.
- 615 [17] J. B. Gibbons *et al.*, “Association between vitamin d supplementation and covid-19 infection
616 and mortality,” *Scientific Reports*, vol. 12, no. 1, p. 19397, Nov. 2022, ISSN: 2045-2322. DOI:
617 10.1038/s41598-022-24053-4. [Online]. Available: <https://doi.org/10.1038/s41598-022-24053-4>.
- 619 [18] T. I. Hariyanto, D. Intan, J. E. Hananto, H. Harapan, and A. Kurniawan, “Vitamin d supple-
620 mentation and covid-19 outcomes: A systematic review, meta-analysis and meta-regression,”
621 *Reviews in Medical Virology*, vol. 32, no. 2, e2269, 2022. DOI: <https://doi.org/10.1002/rmv.2269>. eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/rmv.2269>. [Online]. Available: <https://onlinelibrary.wiley.com/doi/abs/10.1002/rmv.2269>.
- 624 [19] S.-H. Lan, H.-Z. Lee, C.-M. Chao, S.-P. Chang, L.-C. Lu, and C.-C. Lai, “Efficacy of melato-
625 tonin in the treatment of patients with covid-19: A systematic review and meta-analysis of
626 randomized controlled trials,” *Journal of Medical Virology*, vol. 94, no. 5, pp. 2102–2107, 2022.
627 DOI: <https://doi.org/10.1002/jmv.27595>. eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/jmv.27595>. [Online]. Available: <https://onlinelibrary.wiley.com/doi/abs/10.1002/jmv.27595>.
- 630 [20] P. C. Pereira, C. J. de Lima, A. B. Fernandes, R. A. Zângaro, and A. B. Villaverde, “Car-
631 diopulmonary and hematological effects of infrared led photobiomodulation in the treatment of
632 sars-cov2,” *Journal of Photochemistry and Photobiology B: Biology*, vol. 238, p. 112619, 2023,
633 ISSN: 1011-1344. DOI: <https://doi.org/10.1016/j.jphotobiol.2022.112619>. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S1011134422002342>.
- 635 [21] M. Zvyagin *et al.*, “Genslms: Genome-scale language models reveal sars-cov-2 evolutionary
636 dynamics,” *bioRxiv*, 2022. DOI: 10.1101/2022.10.10.511571. eprint: <https://www.biorxiv.org/content/early/2022/11/23/2022.10.10.511571.full.pdf>. [Online]. Available:
638 <https://www.biorxiv.org/content/early/2022/11/23/2022.10.10.511571>.

- 639 [22] D. Lebatteux, H. Soudeyns, I. Boucoiran, S. Gantt, and A. B. Diallo, “Machine learning-
640 based approach kevolve efficiently identifies sars-cov-2 variant-specific genomic signatures,”
641 *bioRxiv*, 2022. DOI: 10.1101/2022.02.07.479343. eprint: [https://www.biorxiv.org/content/early/2022/02/07/2022.02.07.479343](https://www.biorxiv.org/content/early/2022/02/07/2022.02.07.479343.full.pdf). [Online]. Available: <https://www.biorxiv.org/content/early/2022/02/07/2022.02.07.479343>.
- 644 [23] K. Champion, B. Lusch, J. N. Kutz, and S. L. Brunton, “Data-driven discovery of coordinates
645 and governing equations,” *Proceedings of the National Academy of Sciences*, vol. 116, no. 45,
646 pp. 22 445–22 451, 2019. DOI: 10.1073/pnas.1906995116. eprint: <https://www.pnas.org/doi/pdf/10.1073/pnas.1906995116>. [Online]. Available: <https://www.pnas.org/doi/abs/10.1073/pnas.1906995116>.
- 649 [24] B. Ross, “The development of fractional calculus 1695–1900,” *Historia Mathematica*, vol. 4,
650 no. 1, pp. 75–89, 1977, ISSN: 0315-0860. DOI: [https://doi.org/10.1016/0315-0860\(77\)90039-8](https://doi.org/10.1016/0315-0860(77)90039-8). [Online]. Available: <https://www.sciencedirect.com/science/article/pii/0315086077900398>.
- 653 [25] J. Calbó, D. Pagès, and J.-A. González, “Empirical studies of cloud effects on uv radiation:
654 A review,” *Reviews of Geophysics*, vol. 43, no. 2, 2005. DOI: <https://doi.org/10.1029/2004RG000155>. eprint: <https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2004RG000155>. [Online]. Available: <https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2004RG000155>.
- 658 [26] C. D. McNaughton, N. M. Adams, C. H. Johnson, M. J. Ward, J. E. Schmitz, and T. A.
659 Lasko, “Diurnal variation in sars-cov-2 pcr test results: Test accuracy may vary by time of
660 day,” *Journal of Biological Rhythms*, vol. 36, no. 6, pp. 595–601, 2021, PMID: 34696614. DOI:
661 [10.1177/07487304211051841](https://doi.org/10.1177/07487304211051841). eprint: <https://doi.org/10.1177/07487304211051841>.
662 [Online]. Available: <https://doi.org/10.1177/07487304211051841>.
- 663 [27] A. V. Winnett *et al.*, “Morning sars-cov-2 testing yields better detection of infection due to
664 higher viral loads in saliva and nasal swabs upon waking,” *Microbiology Spectrum*, vol. 10,
665 no. 6, e03873–22, 2022. DOI: 10.1128/spectrum.03873-22. eprint: <https://journals.asm.org/doi/pdf/10.1128/spectrum.03873-22>. [Online]. Available: <https://journals.asm.org/doi/abs/10.1128/spectrum.03873-22>.
- 668 [28] X. Zhuang *et al.*, “Time-of-day variation in sars-cov-2 rna levels during the second wave of
669 covid-19,” *Viruses*, vol. 14, no. 8, 2022, ISSN: 1999-4915. DOI: 10.3390/v14081728. [Online].
670 Available: <https://www.mdpi.com/1999-4915/14/8/1728>.
- 671 [29] M. Romdhani *et al.*, “Is there a diurnal variation of covid-19 patients warranting presenta-
672 tion to the health centre? a chronobiological observational cross-sectional study,” *Annals of
673 Medicine*, vol. 54, no. 1, pp. 3059–3067, 2022, PMID: 36308396. DOI: 10.1080/07853890.
674 2022.2136399. eprint: <https://doi.org/10.1080/07853890.2022.2136399>. [Online].
675 Available: <https://doi.org/10.1080/07853890.2022.2136399>.

- 676 [30] J. Feynman and A. Ruzmaikin, “The centennial gleissberg cycle and its association with ex-
677 tended minima,” *Journal of Geophysical Research: Space Physics*, vol. 119, no. 8, pp. 6027–
678 6041, 2014. DOI: <https://doi.org/10.1002/2013JA019478>. eprint: <https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1002/2013JA019478>. [Online]. Available: <https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2013JA019478>.
- 681 [31] C. B. Jackson, M. Farzan, B. Chen, and H. Choe, “Mechanisms of sars-cov-2 entry into cells,”
682 *Nature Reviews Molecular Cell Biology*, vol. 23, no. 1, pp. 3–20, Jan. 2022, ISSN: 1471-0080. DOI:
683 [10.1038/s41580-021-00418-x](https://doi.org/10.1038/s41580-021-00418-x). [Online]. Available: <https://doi.org/10.1038/s41580-021-00418-x>.
- 685 [32] C. McAloon *et al.*, “Incubation period of covid-19: A rapid systematic review and meta-
686 analysis of observational research,” *BMJ Open*, vol. 10, no. 8, 2020, ISSN: 2044-6055. DOI:
687 [10.1136/bmjopen-2020-039652](https://doi.org/10.1136/bmjopen-2020-039652). eprint: <https://bmjopen.bmjjournals.com/content/10/8/e039652.full.pdf>. [Online]. Available: <https://bmjopen.bmjjournals.com/content/10/8/e039652>.
- 689 [33] Y. Wang *et al.*, “Remdesivir in adults with severe covid-19: A randomised, double-blind,
690 placebo-controlled, multicentre trial,” *The Lancet*, vol. 395, no. 10236, pp. 1569–1578, 2020,
691 ISSN: 0140-6736. DOI: [https://doi.org/10.1016/S0140-6736\(20\)31022-9](https://doi.org/10.1016/S0140-6736(20)31022-9). [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0140673620310229>.
- 693 [34] J. H. Beigel *et al.*, “Remdesivir for the treatment of covid-19 — final report,” *New England
694 Journal of Medicine*, vol. 383, no. 19, pp. 1813–1826, 2020, PMID: 32445440. DOI: [10.1056/NEJMoa2007764](https://doi.org/10.1056/NEJMoa2007764). eprint: <https://doi.org/10.1056/NEJMoa2007764>. [Online]. Available: <https://doi.org/10.1056/NEJMoa2007764>.
- 697 [35] “Repurposed antiviral drugs for covid-19 — interim who solidarity trial results,” *New England
698 Journal of Medicine*, vol. 384, no. 6, pp. 497–511, 2021, PMID: 33264556. DOI: [10.1056/NEJMoa2023184](https://doi.org/10.1056/NEJMoa2023184). eprint: <https://doi.org/10.1056/NEJMoa2023184>. [Online]. Available: <https://doi.org/10.1056/NEJMoa2023184>.
- 701 [36] R. Saha and I. A. Chen, “Effect of uv radiation on fluorescent rna aptamers’ functional and
702 templating ability,” *ChemBioChem*, vol. 20, no. 20, pp. 2609–2617, 2019. DOI: <https://doi.org/10.1002/cbic.201900261>. eprint: <https://chemistry-europe.onlinelibrary.wiley.com/doi/pdf/10.1002/cbic.201900261>. [Online]. Available: <https://chemistry-europe.onlinelibrary.wiley.com/doi/abs/10.1002/cbic.201900261>.
- 706 [37] X. Wu *et al.*, “Secreted orf8 is a pathogenic cause of severe covid-19 and is potentially targetable
707 with select nlrp3 inhibitors,” *bioRxiv*, 2022. DOI: [10.1101/2021.12.02.470978](https://doi.org/10.1101/2021.12.02.470978). eprint: <https://www.biorxiv.org/content/early/2022/08/22/2021.12.02.470978.full.pdf>. [Online]. Available: <https://www.biorxiv.org/content/early/2022/08/22/2021.12.02.470978>.
- 710 [38] L. Zinzula, “Lost in deletion: The enigmatic orf8 protein of sars-cov-2,” *Biochemical and Bio-
711 physical Research Communications*, vol. 538, pp. 116–124, 2021, COVID-19, ISSN: 0006-291X.
712 DOI: <https://doi.org/10.1016/j.bbrc.2020.10.045>. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0006291X20319628>.

- 714 [39] S. T. Parvathy, V. Udayasuriyan, and V. Bhadana, "Codon usage bias," *Molecular Biology*
715 *Reports*, vol. 49, no. 1, pp. 539–565, Jan. 2022, ISSN: 1573-4978. DOI: 10.1007/s11033-021-
716 06749-4. [Online]. Available: <https://doi.org/10.1007/s11033-021-06749-4>.
- 717 [40] E. Dolgin, "The most popular genes in the human genome," *Nature*, vol. 551, no. 7681, pp. 427–
718 431, Nov. 2017, ISSN: 1476-4687. DOI: 10.1038/d41586-017-07291-9. [Online]. Available:
719 <https://doi.org/10.1038/d41586-017-07291-9>.
- 720 [41] X. C. Dopico *et al.*, "Widespread seasonal gene expression reveals annual differences in human
721 immunity and physiology," *Nature Communications*, vol. 6, no. 1, p. 7000, May 2015, ISSN:
722 2041-1723. DOI: 10.1038/ncomms8000. [Online]. Available: <https://doi.org/10.1038/ncomms8000>.
- 723 [42] X. Zhuang *et al.*, "The circadian clock component bmal1 regulates sars-cov-2 entry and repli-
724 cation in lung epithelial cells," *iScience*, vol. 24, no. 10, p. 103144, 2021, ISSN: 2589-0042.
725 DOI: <https://doi.org/10.1016/j.isci.2021.103144>. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S2589004221011123>.
- 726 [43] G. Zhang *et al.*, "Uncovering the genetic links of sars-cov-2 infections on heart failure co-
727 morbidity by a systems biology approach," *ESC Heart Failure*, vol. 9, no. 5, pp. 2937–2954,
728 2022. DOI: <https://doi.org/10.1002/ehf2.14003>. eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/ehf2.14003>. [Online]. Available: <https://onlinelibrary.wiley.com/doi/abs/10.1002/ehf2.14003>.
- 729 [44] T. H. Bugge, T. M. Antalis, and Q. Wu, "Type ii transmembrane serine proteases * β sup*i*/ γ sup*i*,"
730 *Journal of Biological Chemistry*, vol. 284, no. 35, pp. 23177–23181, Aug. 2009, ISSN: 0021-9258.
731 DOI: 10.1074/jbc.R109.021006. [Online]. Available: <https://doi.org/10.1074/jbc.R109.021006>.
- 732 [45] Q. Zhang *et al.*, "Environmentally-induced mdig is a major contributor to the severity of
733 covid-19 through fostering expression of sars-cov-2 receptor nrps and glycan metabolism,"
734 *bioRxiv*, 2021. DOI: 10.1101/2021.01.31.429010. eprint: <https://www.biorxiv.org/content/early/2021/02/01/2021.01.31.429010.full.pdf>. [Online]. Available: <https://www.biorxiv.org/content/early/2021/02/01/2021.01.31.429010>.
- 735 [46] P. V'kovski, A. Kratzel, S. Steiner, H. Stalder, and V. Thiel, "Coronavirus biology and repli-
736 cation: Implications for sars-cov-2," *Nature Reviews Microbiology*, vol. 19, no. 3, pp. 155–
737 170, Mar. 2021, ISSN: 1740-1534. DOI: 10.1038/s41579-020-00468-6. [Online]. Available:
738 <https://doi.org/10.1038/s41579-020-00468-6>.
- 739 [47] H. Gomi, K. Mori, S. Itohara, and T. Izumi, "Rab27b is expressed in a wide range of exocytic
740 cells and involved in the delivery of secretory granules near the plasma membrane," *Molecular*
741 *Biology of the Cell*, vol. 18, no. 11, pp. 4377–4386, 2007, PMID: 17761531. DOI: 10.1091/mbc.e07-05-0409.
742 eprint: <https://doi.org/10.1091/mbc.e07-05-0409>. [Online]. Available:
743 <https://doi.org/10.1091/mbc.e07-05-040>.

- 751 [48] G. Fischer von Mollard, B. Stahl, A. Khokhlatchev, T. Südhof, and R. Jahn, “Rab3c is a
752 synaptic vesicle protein that dissociates from synaptic vesicles after stimulation of exocytosis,”
753 *Journal of Biological Chemistry*, vol. 269, no. 15, pp. 10971–10974, 1994, ISSN: 0021-9258.
754 DOI: [https://doi.org/10.1016/S0021-9258\(19\)78076-4](https://doi.org/10.1016/S0021-9258(19)78076-4). [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0021925819780764>.
- 755 [49] M. A. Olenick, R. Dominguez, and E. L. Holzbaur, “Dynein activator Hook1 is required for
756 trafficking of BDNF-signaling endosomes in neurons,” *Journal of Cell Biology*, vol. 218, no. 1,
757 pp. 220–233, Oct. 2018, ISSN: 0021-9525. DOI: 10.1083/jcb.201805016. eprint: https://rupress.org/jcb/article-pdf/218/1/220/1379611/jcb_201805016.pdf. [Online].
758 Available: <https://doi.org/10.1083/jcb.201805016>.
- 759 [50] B. Amatya *et al.*, “Snx-pxa-rgs-pxc subfamily of snxs in the regulation of receptor-mediated
760 signaling and membrane trafficking,” *International Journal of Molecular Sciences*, vol. 22,
761 no. 5, 2021, ISSN: 1422-0067. DOI: 10.3390/ijms22052319. [Online]. Available: <https://www.mdpi.com/1422-0067/22/5/2319>.
- 762 [51] S. Ahmed, W. Bu, R. T. C. Lee, S. Maurer-Stroh, and W. I. Goh, “F-bar domain proteins,”
763 *Communicative & Integrative Biology*, vol. 3, no. 2, pp. 116–121, 2010, PMID: 20585502. DOI:
764 10.4161/cib.3.2.10808. eprint: <https://doi.org/10.4161/cib.3.2.10808>. [Online].
765 Available: <https://doi.org/10.4161/cib.3.2.10808>.
- 766 [52] H. Wang *et al.*, “Phosphatidylinositol 3,4-bisphosphate synthesis and turnover are spatially seg-
767 regated in the endocytic pathway,” *Journal of Biological Chemistry*, vol. 295, no. 4, pp. 1091–
768 1104, 2020, ISSN: 0021-9258. DOI: [https://doi.org/10.1016/S0021-9258\(17\)49918-2](https://doi.org/10.1016/S0021-9258(17)49918-2). [On-
769 line]. Available: <https://www.sciencedirect.com/science/article/pii/S0021925817499182>.
- 770 [53] C. A. Goldstein, D. Kagan, M. Rizvydeen, S. Warshaw, J. P. Troost, and H. J. Burgess, “The
771 possibility of circadian rhythm disruption in long covid,” *Brain, Behavior, Immunity - Health*,
772 vol. 23, p. 100476, 2022, ISSN: 2666-3546. DOI: <https://doi.org/10.1016/j.bbih.2022.100476>. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S2666354622000667>.
- 773 [54] M. Trott, R. Driscoll, and S. Pardhan, “The prevalence of sensory changes in post-covid syn-
774 drome: A systematic review and meta-analysis,” *Frontiers in Medicine*, vol. 9, 2022, ISSN: 2296-
775 858X. DOI: 10.3389/fmed.2022.980253. [Online]. Available: <https://www.frontiersin.org/articles/10.3389/fmed.2022.980253>.
- 776 [55] D. Ayinde, N. Casartelli, and O. Schwartz, “Restricting hiv the samhd1 way: Through nu-
777 cleotide starvation,” *Nature Reviews Microbiology*, vol. 10, no. 10, pp. 675–680, Oct. 2012,
778 ISSN: 1740-1534. DOI: 10.1038/nrmicro2862. [Online]. Available: <https://doi.org/10.1038/nrmicro2862>.
- 779 [56] R. Wang *et al.*, “Sars-cov-2 restructures host chromatin architecture,” *Nature Microbiology*,
780 vol. 8, no. 4, pp. 679–694, Apr. 2023, ISSN: 2058-5276. DOI: 10.1038/s41564-023-01344-8.
781 [Online]. Available: <https://doi.org/10.1038/s41564-023-01344-8>.
- 782
- 783
- 784
- 785
- 786
- 787
- 788

- 789 [57] N. Taefehshokr *et al.*, “Sars-cov-2 nsp5 antagonizes mhc ii expression by subverting histone
790 deacetylase 2,” *bioRxiv*, 2023. DOI: 10.1101/2023.02.10.528032. eprint: <https://www.biorxiv.org/content/early/2023/04/14/2023.02.10.528032.full.pdf>. [Online].
791 Available: <https://www.biorxiv.org/content/early/2023/04/14/2023.02.10.528032>.
- 793 [58] H. D. Desai, K. Sharma, J. V. Patoliya, E. Ahadov, and N. N. Patel, “A rare case of varicella-
794 zoster virus reactivation following recovery from covid-19,” *Cureus*, vol. 13, no. 1, e12423, Jan.
795 2021, ISSN: 2168-8184. DOI: 10.7759/cureus.12423. [Online]. Available: <https://doi.org/10.7759/cureus.12423>.
- 797 [59] T. Iwase *et al.*, “Mutation screening of the human clock gene in circadian rhythm sleep
798 disorders,” *Psychiatry Research*, vol. 109, no. 2, pp. 121–128, 2002, ISSN: 0165-1781. DOI:
799 [https://doi.org/10.1016/S0165-1781\(02\)00006-9](https://doi.org/10.1016/S0165-1781(02)00006-9). [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0165178102000069>.
- 801 [60] R. Tedjasukmana, A. Budikayanti, W. R. Islamiyah, A. M. A. L. Witjaksono, and M. Hakim,
802 “Sleep disturbance in post covid-19 conditions: Prevalence and quality of life,” *Frontiers in
803 Neurology*, vol. 13, 2023, ISSN: 1664-2295. DOI: 10.3389/fneur.2022.1095606. [Online].
804 Available: <https://www.frontiersin.org/articles/10.3389/fneur.2022.1095606>.
- 805 [61] C. Guissart *et al.*, “Dual molecular effects of dominant $\text{jem}_1\text{roraj}/\text{em}_1$ mutations cause two
806 variants of syndromic intellectual disability with either autism or cerebellar ataxia,” *The American
807 Journal of Human Genetics*, vol. 102, no. 5, pp. 744–759, May 2018, ISSN: 0002-9297. DOI:
808 [10.1016/j.ajhg.2018.02.021](https://doi.org/10.1016/j.ajhg.2018.02.021). [Online]. Available: <https://doi.org/10.1016/j.ajhg.2018.02.021>.
- 810 [62] V. Venkataramani and F. Winkler, “Cognitive deficits in long covid-19,” *New England Journal
811 of Medicine*, vol. 387, no. 19, pp. 1813–1815, 2022. DOI: 10.1056/NEJMcibr2210069. eprint:
812 <https://doi.org/10.1056/NEJMcibr2210069>. [Online]. Available: <https://doi.org/10.1056/NEJMcibr2210069>.
- 814 [63] M.-D. Li *et al.*, “ $\text{jem}_1\text{oi}/\text{em}_1\text{-glcnac}$ signaling entrains the circadian clock by inhibiting bmal1/clock
815 ubiquitination,” *Cell Metabolism*, vol. 17, no. 2, pp. 303–310, Feb. 2013, ISSN: 1550-4131. DOI:
816 [10.1016/j.cmet.2012.12.015](https://doi.org/10.1016/j.cmet.2012.12.015). [Online]. Available: <https://doi.org/10.1016/j.cmet.2012.12.015>.
- 818 [64] J. Cavieres-Lepe and J. Ewer, “Reciprocal relationship between calcium signaling and circadian
819 clocks: Implications for calcium homeostasis, clock function, and therapeutics,” *Frontiers in
820 Molecular Neuroscience*, vol. 14, 2021, ISSN: 1662-5099. DOI: 10.3389/fnmol.2021.666673.
821 [Online]. Available: <https://www.frontiersin.org/articles/10.3389/fnmol.2021.666673>.
- 823 [65] M. Okubo, Y. Aoyama, and T. Murase, “A novel donor splice site mutation in the glycogen
824 debranching enzyme gene is associated with glycogen storage disease type iii,” *Biochemical and Biophysical Research Communications*, vol. 224, no. 2, pp. 493–499, 1996, ISSN: 0006-
825 291X. DOI: <https://doi.org/10.1006/bbrc.1996.1055>. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0006291X96910554>.

- 828 [66] N. Preisler *et al.*, “Exercise intolerance in glycogen storage disease type iii: Weakness or energy
829 deficiency?” *Molecular Genetics and Metabolism*, vol. 109, no. 1, pp. 14–20, 2013, ISSN: 1096-
830 7192. DOI: <https://doi.org/10.1016/j.ymgme.2013.02.008>. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S1096719213000607>.
- 832 [67] A. Leenhardt, I. Denjoy, and P. Guicheney, “Catecholaminergic polymorphic ventricular tachycardia,” *Circulation: Arrhythmia and Electrophysiology*, vol. 5, no. 5, pp. 1044–1052, 2012. DOI:
833 10.1161/CIRCEP.111.962027. eprint: <https://www.ahajournals.org/doi/pdf/10.1161/CIRCEP.111.962027>. [Online]. Available: <https://www.ahajournals.org/doi/abs/10.1161/CIRCEP.111.962027>.
- 837 [68] C. Spuch *et al.*, “Efficacy and safety of lithium treatment in sars-cov-2 infected patients,” *Frontiers in Pharmacology*, vol. 13, 2022, ISSN: 1663-9812. DOI: 10.3389/fphar.2022.850583.
838 [Online]. Available: <https://www.frontiersin.org/articles/10.3389/fphar.2022.850583>.
- 841 [69] E. C. Clarke, R. A. Nofchissey, C. Ye, and S. B. Bradfute, “The iminosugars celgosivir, castanospermine and UV-4 inhibit SARS-CoV-2 replication,” *Glycobiology*, vol. 31, no. 4, pp. 378–
842 384, Sep. 2020, ISSN: 1460-2423. DOI: 10.1093/glycob/cwaa091. eprint: <https://academic.oup.com/glycob/article-pdf/31/4/378/37761638/cwaa091.pdf>. [Online]. Available:
844 <https://doi.org/10.1093/glycob/cwaa091>.
- 846 [70] A. Israel *et al.*, “Identification of drugs associated with reduced severity of covid-19: A case-control study in a large population,” *medRxiv*, 2021. DOI: 10.1101/2020.10.13.20211953.
847 eprint: <https://www.medrxiv.org/content/early/2021/03/24/2020.10.13.20211953.full.pdf>. [Online]. Available: <https://www.medrxiv.org/content/early/2021/03/24/2020.10.13.20211953>.
- 851 [71] A. Clemente-Moragón *et al.*, “Metoprolol in critically ill patients with covid-19,” *Journal of the American College of Cardiology*, vol. 78, no. 10, pp. 1001–1011, 2021, ISSN: 0735-1097.
852 DOI: <https://doi.org/10.1016/j.jacc.2021.07.003>. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S073510972105590X>.
- 855 [72] H. Tomimoto, O. Takemoto, I. Akiguchi, and T. Yanagihara, “Immunoelectron microscopic study of c-fos, c-jun and heat shock protein after transient cerebral ischemia in gerbils,” *Acta Neuropathologica*, vol. 97, no. 1, pp. 22–30, Jan. 1999, ISSN: 1432-0533. DOI: 10.1007/s004010050951. [Online]. Available: <https://doi.org/10.1007/s004010050951>.
- 859 [73] P. Joshi *et al.*, “Host inducible-hsp70a1a is an irresistible drug target to combat sars-cov2 infection and pathogenesis,” *bioRxiv*, 2023. DOI: 10.1101/2023.05.05.539661. eprint: <https://www.biorxiv.org/content/early/2023/05/08/2023.05.05.539661.full.pdf>. [Online]. Available: <https://www.biorxiv.org/content/early/2023/05/08/2023.05.05.539661>.
- 863 [74] E. L. Hatcher *et al.*, “Virus Variation Resource – improved response to emergent viral outbreaks,” *Nucleic Acids Research*, vol. 45, no. D1, pp. D482–D490, Nov. 2016, ISSN: 0305-1048.
864 DOI: 10.1093/nar/gkw1065. eprint: <https://academic.oup.com/nar/article-pdf/45/D1/D482/8846723/gkw1065.pdf>. [Online]. Available: <https://doi.org/10.1093/nar/gkw1065>.

- 867 [75] W. C. Forsythe, E. J. Rykiel, R. S. Stahl, H.-i. Wu, and R. M. Schoolfield, “A model com-
868 parison for daylength as a function of latitude and day of year,” *Ecological Modelling*, vol. 80,
869 no. 1, pp. 87–95, 1995, ISSN: 0304-3800. DOI: [https://doi.org/10.1016/0304-3800\(94\)
870 00034-F](https://doi.org/10.1016/0304-3800(94)00034-F). [Online]. Available: [https://www.sciencedirect.com/science/article/pii/
871 030438009400034F](https://www.sciencedirect.com/science/article/pii/030438009400034F).
- 872 [76] E. Guidotti and D. Ardia, “Covid-19 data hub,” *Journal of Open Source Software*, vol. 5,
873 no. 51, p. 2376, 2020. DOI: [10.21105/joss.02376](https://doi.org/10.21105/joss.02376).
- 874 [77] E. Guidotti, “A worldwide epidemiological database for covid-19 at fine-grained spatial reso-
875 lution,” *Scientific Data*, vol. 9, no. 1, p. 112, 2022. DOI: [10.1038/s41597-022-01245-1](https://doi.org/10.1038/s41597-022-01245-1).
- 876 [78] I. Berry *et al.*, “A sub-national real-time epidemiological and vaccination database for the
877 covid-19 pandemic in canada,” *Scientific Data*, vol. 8, no. 1, Jul. 2021. DOI: [10.1038/s41597-021-00955-2](https://doi.org/10.1038/s41597-021-00955-2). [Online]. Available: <https://doi.org/10.1038/s41597-021-00955-2>.
- 878 [79] E. Dong, H. Du, and L. Gardner, “An interactive web-based dashboard to track covid-19 in
879 real time,” *The Lancet Infectious Diseases*, vol. 20, no. 5, pp. 533–534, May 2020, ISSN: 1473-
880 3099. DOI: [10.1016/S1473-3099\(20\)30120-1](https://doi.org/10.1016/S1473-3099(20)30120-1). [Online]. Available: [https://doi.org/10.1016/S1473-3099\(20\)30120-1](https://doi.org/10.1016/S1473-3099(20)30120-1).
- 881 [80] S. de Salud Mexico, “Información referente a casos covid-19 en méxico,” May 2022. [Online].
882 Available: <https://datos.gob.mx/busca/dataset/informacion-referente-a-casos-covid-19-en-mexico>.
- 883 [81] M. de Salud Argentina, “Covid-19. casos registrados en la república argentina,” May 2022.
884 [Online]. Available: <http://datos.salud.gob.ar/dataset/covid-19-casos-registrados-en-la-republica-argentina>.
- 885 [82] M. de Ciencia and M. de Salud, “Dp3 - casos totales por región,” May 2022. [Online]. Available:
886 <https://github.com/MinCiencia/Datos-COVID19/tree/master/output/producto3>.
- 887 [83] M. de Salud - MINSA, “Casos positivos por covid-19,” May 2022. [Online]. Available: <https://www.datosabiertos.gob.pe/dataset/casos-positivos-por-covid-19-ministerio-de-salud-minsa>.
- 888 [84] P. de Datos Abiertos Gobierno de Colombia, “Casos positivos de covid-19 en colombia,” May
889 2022. [Online]. Available: <https://www.datos.gov.co/Salud-y-Proteccion-Social/Casos-positivos-de-COVID-19-en-Colombia/gt2j-8ykr/data>.
- 890 [85] S. W. D. Center, “The international sunspot number,” *International Sunspot Number Monthly
891 Bulletin and online catalogue*, [Online]. Available: <http://www.sidc.be/silso/>.
- 892 [86] E. Richard, “Tsis sim level 3 solar spectral irradiance 24-hour means v09,” *Goddard Earth
893 Sciences Data and Information Services Center (GES DISC)*, 2022. DOI: [DOI:10.5067/TSIS-SIM/DATA318](https://doi.org/10.5067/TSIS-SIM/DATA318). [Online]. Available: https://disc.gsfc.nasa.gov/datacollection/TSIS_SSI_L3_24HR_09.html.

- 903 [87] G. E. S. Data and I. S. C. (DISC), “Aqua/airs l3 daily standard physical retrieval (airs-only)
904 1 degree x 1 degree v7.0,” *AIRS project (2019)*, 2023. [Online]. Available: <https://doi.org/10.5067/U03Q64CTTS1U>, .
905