

Temperature and Latitude Correlate with SARS-CoV-2 Epidemiological Variables but not with Genomic Change Worldwide

Prakruthi Burra¹, Katiria Soto-Díaz², Izan Chalen³,
Rafael Jaime Gonzalez-Ricon³, Dave Istanto⁴
and Gustavo Caetano-Anollés^{4,5}

¹Department of Computer Science, University of Illinois at Urbana-Champaign, USA.

²Neuroscience Program, University of Illinois at Urbana-Champaign, USA. ³Department of Animal Sciences, University of Illinois at Urbana-Champaign, USA. ⁴Department of Crop Sciences, University of Illinois at Urbana-Champaign, USA. ⁵C.R. Woese Institute for Genomic Biology, University of Illinois at Urbana-Champaign, USA.

Evolutionary Bioinformatics
Volume 17: 1–8
© The Author(s) 2021
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/1176934321989695



ABSTRACT: The SARS-CoV-2 virus that causes the COVID-19 disease has spread quickly and massively around the entire globe, causing millions of confirmed cases and deaths worldwide. The disease poses a serious ongoing threat to public health. This study aims to understand the disease potential of the virus in different regions by studying how average spring temperature and its strong predictor, latitude, affect epidemiological variables such as disease incidence, mortality, recovery cases, active cases, testing rate, and hospitalization. We also seek to understand the association of temperature and geographic coordinates with viral genomics. Epidemiological data along with temperature, latitude, longitude, and preparedness index were collected for different countries and US states during the early stages of the pandemic. Our worldwide epidemiological analysis showed a significant correlation between temperature and incidence, mortality, recovery cases and active cases. The same tendency was found with latitude, but not with longitude. In the US, we observed no correlation between temperature or latitude and epidemiological variables. Interestingly, longitude was correlated with incidence, mortality, active cases, and hospitalization. An analysis of mutational change and mutational change per time in 55 453 aligned SARS-CoV-2 genome sequences revealed these parameters were uncorrelated with temperature and geographic coordinates. The epidemiological trends we observed worldwide suggest a seasonal effect for the disease that is not directly controlled by the genomic makeup of the virus. Future studies will need to determine if correlations are more likely the result of effects associated with the environment or the innate immunity of the host.

KEYWORDS: SARS-CoV-2, COVID-19, incidence, mortality, recovery patients, active cases, testing rate, hospitalization, genome analysis, mutation

RECEIVED: September 23, 2020. **ACCEPTED:** December 30, 2020.

TYPE: Original Research

FUNDING: The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: Office of Research and Office of International Programs, College of Agricultural, Consumer and Environmental Sciences at the University of Illinois at Urbana-Champaign.

DECLARATION OF CONFLICTING INTERESTS: The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

CORRESPONDING AUTHOR: Gustavo Caetano-Anollés, Department of Crop Sciences, University of Illinois at Urbana-Champaign, 332 NSRC, 1101 W Peabody Drive, Urbana, IL 61801, USA. Email: gca@illinois.edu

Introduction

The zoonotic SARS-CoV-2 virus that likely originated in the city of Wuhan of Central China has spread more rapidly across the globe than other SARS-like β -coronavirus strains.¹ As of April 15, 2020 (time of data download of our study), the COVID-19 disease was present in at least 211 countries around the globe killing ~143 000 people and infecting ~2 million. Six months later, there were ~30 million confirmed cases and a million deaths worldwide, illustrating the massive worldwide spread of the disease. Compared to other pandemics, including H1N1 in 2009, COVID-19 seems to spread at faster speeds.² Interestingly, the viral nucleic acid shedding pattern of patients infected with COVID-19 resembles that of influenza patients, suggesting that transmission of COVID-19 may occur during the first few days after the onset of symptoms.¹ Therefore, it is assumed that viral transmission occurs at illness onset and even with mild or no symptoms. However, the pattern of transmission observed in COVID-19 is distinctive from that of SARS-CoV and indicates that spread may not be

effectively controlled by isolating the patient after the onset of the disease.¹

Genome analyses help mitigate viral spread and facilitate treatment.¹ The CoV genome has been shown to contain a variable number (6–11) of open reading frames (ORFs)³ where two thirds of the viral RNA, located mainly in the first ORF (ORF1a/b), translates into 2 polyproteins, pp1a and pp1ab, and encodes 16 non-structural proteins (NSP), while the remaining ORFs encode structural and accessory proteins.⁴ These proteins include the essential spike glycoprotein (S), the small envelope protein (E), the matrix protein (M), and the nucleocapsid protein (N)⁵ and various accessory proteins that interfere with the host's innate immune response. A group of researchers⁶ performed deep meta-transcriptomic sequencing on the Wuhan-Hu-1 coronavirus (WHCV), revealing it exhibits some genomic and phylogenetic similarity to SARS-CoV, particularly in the glycoprotein S gene and its receptor binding domain (RBD), and indicating the capacity for direct human transmission. In turn, screening for β -coronavirus



receptors showed that human cells expressing ACE2, but not dipeptidyl peptidase-4 (DPP4) or aminopeptidase N (APN), increased SARS-CoV-2 entry.⁷ At the protein level, an initial comparative analysis of the first 3 SARS-CoV-1 genomes to SARS-like coronaviruses showed significant differences in the length and amino acid makeup of their proteins but showed no substitutions in NSP7, NSP13, envelope, matrix or accessory proteins 6 and 8b.⁸ Other recent research suggested that mutations in NSP2 and NSP3 play a role in the infectious capacity and differentiation mechanism of SARS-CoV-2.⁹ An analysis of the genotypes of COVID-19 in different patients from various provinces found that SARS-CoV-2 had been mutating in different patients in China.¹⁰ While the degree of diversification of SARS-CoV-2 appears less than that of the H7N9 avian influenza¹¹, a population genetic analysis of 103 genomes revealed 2 prevalent evolutionary types of SARS-CoV-2: type L (70%) and type S (30%), where the strains in type L were more aggressive and contagious than those of type S.¹² A recent worldwide entropy study of diversification of 15 342 genomes during the early start of the pandemic that was conducted in parallel with the present study revealed 27 high-entropy mutant genotypes spreading through continents.¹³

The pathophysiology of SARS-CoV-2 infection is very similar to that of SARS-CoV infection, with aggressive inflammatory responses strongly implicated in airway damage.¹⁴ The severity of the disease in patients is not only due to viral infection but is also dependent on the host response.¹⁵ The pattern of increasing severity with age is also broadly consistent with the epidemiology of SARS-CoV and MERS-CoV.^{16,17} The acute respiratory distress syndrome (ARDS) observed in COVID-19 is characterized by shortness of breath and low blood oxygen levels.¹⁸ ARDS can directly lead to respiratory failure, which is the cause of death in 70% of fatal cases of COVID-19,¹⁸ often linked to the large release of cytokines by the immune system in response to a viral infection. A cytokine storm and sepsis symptoms are the cause of death in 28% of fatal cases of COVID-19.¹⁸ In these cases, uncontrolled inflammation is capable of producing damage to multiple organs leading to organ failure, especially of the heart, liver, and kidney systems. It should also be noted that the majority of patients with SARS-CoV infection that progressed to kidney failure eventually died.¹⁹

Since worldwide reports reveal that the level of COVID-19 contagion and infectivity appears higher in certain regions¹⁹ and taking into consideration all the similarities with influenza infection, we sought to investigate if COVID-19 infection is seasonal. We hypothesize that COVID-19 epidemiology and genetic makeup will be affected by this seasonality phenomenon. There are good grounds for the premise. For example, high temperatures and humidity affect the half-life of the SARS-CoV-2 virus in fomite transmission environments.²⁰ Similarly, high temperature and humidity significantly affect COVID-19 daily new cases and deaths using a log-linear generalized additive model.²¹ To test our hypothesis, we explored

how spring temperature, a highly correlated parameter of seasonality,^{22,23} and geographic coordinates are impacting COVID-19 epidemiology, including disease incidence and mortality, patient recovery, and country preparedness, as well as viral genomic makeup.

Materials and Methods

Epidemiological data collection

All the epidemiological data of COVID-19, including incidence, mortality, recovery rate, number of active cases, testing rate and hospitalization rate, were collected from the Worldometer reference website²⁴ on April 15, 2020, a date that represents the middle of the Spring season and brackets maximal temperature variation among regions of the world.²² Incidence was defined as the number of cases reported, mortality was defined as the number of deaths, recovery rate represented the number of patients that recovered from the infection, active cases were patients that were positive for the virus by the time of sampling, testing rates represented the number of tests being performed, and hospitalization rate was defined as the number of hospitalizations reported. Epidemiological data were normalized to account for differences in the populations of World countries and US states by dividing data entries by total populations and expressing values as percentages. Supplemental Table S1 provides separate spreadsheets with data for countries and states. Total populations as of December 2019 were retrieved from several databases listed in Supplemental Table S1. For example, the incidence of the disease in the US was 641 299 total cases and the population 331 002 651 people. Normalized incidence was therefore 0.194 (ie, $[641\,299/331\,002\,651] \times 100$). In order to measure the effect of temperature on incidence, mortality, recovery rate, number of active cases, testing rate and hospitalization rate, average spring temperatures of every country and state of the US were collected from internet resources listed in Supplemental Table S1. Spring temperatures were used because it had been shown that after April 15 change in zonal-mean surface temperature anomaly decreases at all latitudes of Earth.²² We also collected average latitudes and longitudes of countries' capitals and states of the US from resources listed in Supplemental Table S1. Population, temperature and coordinate data were retrieved from different sources because data in databases were not all-encompassing.

Risk index quantification

In order to contextualize the health risk imposed by the pandemic in different countries, we defined a "risk index" of country preparedness and morbidity associated with the viral disease. This index aims to provide key public health information related to the pandemic. Using available data from the World Health Organization (WHO),²⁵ we generated an indicator of risk that merges both health indicators as well as available infrastructure in the analyzed country. We collected

the country preparedness index from the International Health Regulations (IHR) core capacity index; this index provides an estimator about how equipped a country is, in terms of health infrastructure. We then subtracted from IHR, expressed on a scale of 1 to 100, the probability of dying by health factors that are linked to higher probability of death. This other factor is the probability of death from cardiovascular disease, cancer, diabetes, or chronic respiratory disease between ages 30 and 70.

For the 2 indicators, the most updated information was retrieved from the WHO web page.²⁵ After using the available information, we searched for relationships between the risk index and the total number of cases, active cases, deaths, and recoveries for each of the analyzed countries. Finally, we calculated correlation coefficients between the variables.

We used the risk index to classify countries into 6 different categories. Category 1 countries have risk indices below zero; they hold the highest epidemiological risk due to limitations in health and infrastructure. Category 2, 3, 4, and 5 countries have risk indices ranging from 0 to 50, 50 to 70, 70 to 80, and 80 to 90, respectively. Finally, Category 6 countries have risk indices higher than 90 and hold high preparedness indices and low death probabilities.

Genomic data collection and processing

GISAID,²⁶ a repository of viral data was used for this analysis. A multiple sequence alignment (MSA) of 70 832 viral sequences was downloaded on August 5, 2020. A metadata file containing information pertinent to sequences such as their strain, country, length, etc. was downloaded on September 5, 2020. The “gisaid_epi_isl” field was chosen as the key for each sequence in the metadata.

We processed our genomic data as follows. Sequences from non-human host samples were eliminated as were sequences from countries that were not analyzed in our epidemiological work. We only looked at sequences in the MSA which had a corresponding gisaid_epi_isl field in the metadata file. Those sequences for which the “date” field consisted only of a year were removed from consideration. For sequences whose “date” field only consisted of a month and a year, we assumed that the sequence had a day value of 15. We only analyzed sequences whose date field was at most until the end of April 2020. In all, we selected 55 455 sequences. Our chosen reference sequence EPI_ISL_402125, and the sequence EPI_ISL_406798, with the same value for the “date” field as the reference, were removed from the analysis to take care of division by zero errors. We compiled the genomic coordinates of the SARS-CoV-2 genes encoding the orf1a polyprotein (pp1a), NSP2, RBD, and spike protein S1 and S2 domains (spike_s1 and spike_s2) from the UCSC genome browser.

The genomic change of every sequence with respect to the reference was quantified using the multiple alignment. Specifically, genomic change of a sequence was defined as the

number of base positions where it differed from the reference. The availability of date field values for each sequence enabled us to quantify its genomic change per time. We also computed the genomic change and genomic change per unit time for the genes and domains listed above. Our end goal was to check if there was an association between genomic change or genomic change per unit time and temperature and geographic coordinates of latitude and longitude of the country to which the sequences belonged.

Statistical analysis

Epidemiological data were normalized based on total population as reported by December 2019. The ROUT test from the GraphPad Prism software (<https://www.graphpad.com>) was used to identify epidemiological outliers. In the worldwide analysis of epidemiological data, 2 outliers (UAE and Morocco) were identified and removed. No significant outliers were identified in the US nationwide analysis. In order to determine the interaction between epidemiological data and factors such as temperature, geographic coordinates and preparedness variables, Pearson's correlation coefficients (r) and P -values were computed using an online calculator (<https://www.answerminer.com/calculators/correlation-test/>). To explore if there was a relationship between genomic change and temperature and geographic coordinates, we performed a Pearson correlation test using the Python's scipy package²⁷ and checked significance using 2-tailed P -values. In all cases, $P < .05$ was considered statistically significant. We chose to perform a Pearson correlation analysis because epidemiological data represented continuous random variables that were not normally distributed. Normality was tested using the D'Agostino's K-squared test, Cramer-von Mises criterion, the Anderson-Darling test, and the Shapiro-Francia test.

Results

It has been informally suggested that high environmental temperature decreases the impact of COVID-19 without enough evidence to support the “seasonal” hypothesis. In order to establish a direct relationship between environmental temperatures and epidemiological variables of the early stages of the pandemic (epidemiological data downloaded April 15, 2020) we used bivariate Pearson correlation analyses to test if average temperatures during the spring season and geographic coordinates of latitude and longitude were associated with population-normalized data of incidence, mortality, recovery of patients, active cases and testing rate in every country that showed at least one case of COVID-19, a total of 211 countries (Figure 1; Supplemental Figure S1, Supplemental Table S1). We chose the Pearson correlation because the method is independent of making assumptions of normality either in the marginal distributions or in the bivariate surface with reasonable sample sizes of $N > 20$, especially because a number of statistical tests rejected the null hypothesis that the data were

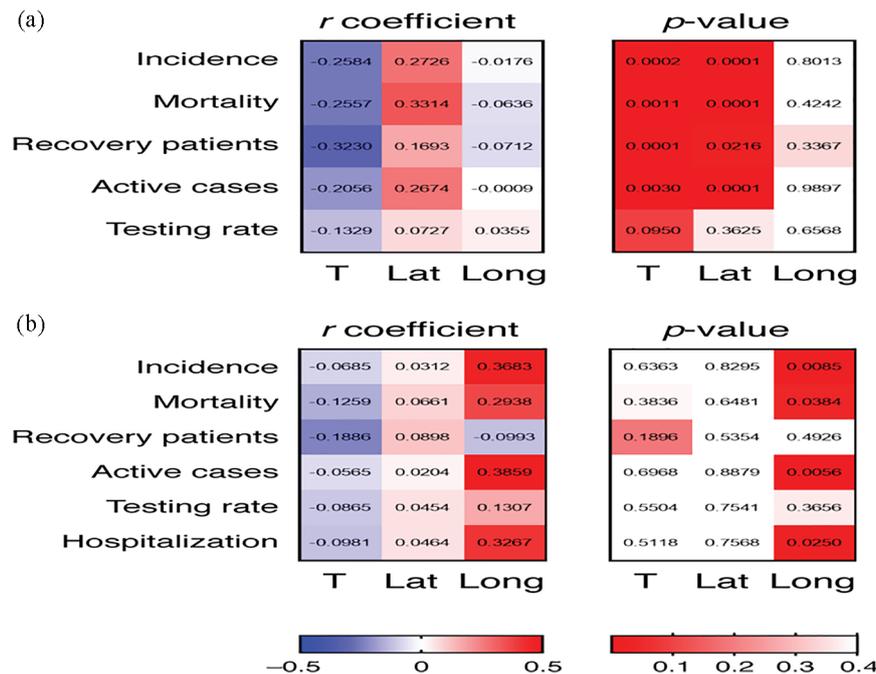


Figure 1. Effect of temperature and geographic coordinates on worldwide and nationwide epidemiological data. Heatmaps of the Pearson correlation coefficients (r) and associated P -values describing the relationship of spring average temperatures (T) or geographic coordinates of latitude (Lat) and longitude (Long) with epidemiological variables for world countries (a) or US states (b). Correlations were considered significant when P -values were less than .05 and association strengths had coefficients r higher than 0.1.

normally distributed (Supplemental Table S2). The method does assume finite variances and co-variances. Since the correlation method is extremely sensitive to outliers, we deleted extreme values to avoid skewed distributions. We note that traveling speed and spatiotemporal patterns of COVID-19 spread have confounding effects on correlation analyses. Diminishing these effects by modeling spatiotemporal parameters would likely increase association strengths and significant detection of borderline correlations in our analyses but would add uncertainty to data normalization. Their effects however are mitigated by the fact that worldwide spread of the disease was already substantial at the beginning of April of 2020. Analysis of data from the COVID-19 CSSE data repository of John Hopkins University (<https://github.com/CSSEGISandData/COVID-19>) revealed that on April 1, 2020, the pandemic had just abandoned the lag phase of accumulation with 50 344 cumulative deaths (following 952 172 cases). At that time most countries had total border closures and a few had bans on high-risk regions. On April 15, 2020, time of data download, worldwide spread of the disease as a first wave was already substantial. Numbers reached 143 023 deaths (2.081 million cases) and had already established a global linear accumulation pattern. Most deaths and cases occurred in Europe (63.6% and 46.7%, respectively), the US (25.3% and 33.2%) and Asia (8.4% and 16.0%).

Worldwide correlation analyses revealed that temperatures were negatively correlated to incidence, mortality, recovery and active cases with statistical significance ($P = .0001-.0030$ for a 2-tailed test) and association strengths ranging from weak to moderate (r ranging from -0.205 to -0.332).

In contrast, no significant correlation was observed between temperature and testing rate. Since distance from the Equator measured as latitude is a strong predictor of temperature, we validated temperature effects by studying correlations of geographic coordinates with epidemiological data. As expected, we found latitudes were positively correlated to incidence, mortality, recovery and active cases with statistical significance ($P = .0001-.0216$) and association strengths ranging from weak to moderate (r ranging from 0.169 to 0.331) (Figure 1; Supplemental Figure S1). Again, no significant correlation was observed between latitude and testing rate. As negative control, we used the geographic coordinate of longitude to test the expectation of no correlation. Indeed, longitude did not correlate with any of the epidemiological variables studied (r ranging from 0.0009 to 0.0712 with $P = .337-.990$) (Figure 1; Supplemental Figure S1).

The correlation patterns we observed resulted in a dichotomy of temperatures and latitudes of the 5 top countries exhibiting either the highest or lowest levels of epidemiological descriptors. For example, the temperatures of countries with the highest and lowest incidence, mortality, recovery of patients, active cases and testing rate had temperatures that ranged 48.5 to 73.4 °F and 69 to 87 °F, 48.5 to 73.4 °F and 46 to 90 °F, 30 to 62 °F and 66.3 to 98.5 °F, 48.5 to 73.4 °F and 32 to 87 °F, and 30 to 66.6 °F and 67.1 to 90 °F, respectively (Supplemental Table S1). In all cases temperature values were higher for countries with lowest descriptors.

The US is currently the country with the most reported COVID-19 cases. We therefore tested if temperatures and geographic coordinates for US states were correlated with

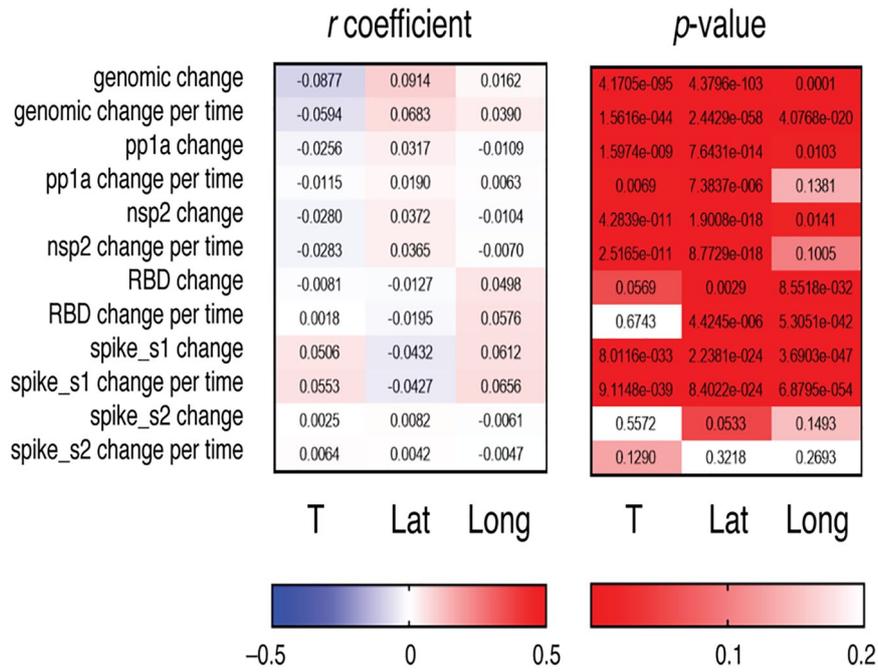


Figure 2. Effect of temperature and geographic coordinates on SARS-CoV-2 genomic change worldwide. Heatmaps of the Pearson correlation coefficients (r) and associated 2-tailed P -values describing the relationship of spring average temperatures (T) or geographic coordinates of latitude (Lat) and longitude (Long) with genomic change or genomic change per unit time are shown for the entire SARS-Co-V-2 genome or selected genomic segments coding for the orf1a polyprotein (pp1a), nonstructural protein 2 (nsp2), receptor-binding domain (RBD), and the spike protein domains S1 and S2 (spike_s1 and spike_s2). A coefficients r value higher than 0.1 indicates that 2 variables are correlated. Our threshold for significance was a P -value of .05.

normalized epidemiological data, which in this case also included hospitalization rates. We found no significant correlation between temperature and latitude with epidemiological variables (Supplemental Figure S2; Supplemental Table S1). On the other hand, we identified a significant correlation between longitude and normalized incidence, mortality, active cases, and hospitalization rate with weak to moderate association strengths (r ranging from 0.294 to 0.386 with $P=.0085-.038$) but no significant correlations with recovery and testing rate. These results are in sharp contrast with those observed worldwide. The correlation patterns did not result in a dichotomy of temperatures of the 5 top states exhibiting either the highest or lowest levels of epidemiological descriptors (data not shown).

We also studied the relationships between a risk index that measures country preparedness and morbidity and epidemiological data worldwide (Supplemental Figure S3). The distribution of the index ranged from -21 (Somalia) to 90.8 (Norway). Countries were then split into categories according to their indices (Figure S4). Six countries had indices with negative values, 52 countries had indices from 1 to 50, 37 countries had indices with values between 50 and 70, 20 countries had indices with values between 70 and 80, 29 countries had indices with values between 80 and 90, and only 9 countries had indices higher than 90 (Figure S3). No significant correlations were found between the risk index and total number of cases, active cases, deaths, and recoveries (Supplemental Figure S3).

Finally, we sought to establish a link between temperature-latitude effects with genomic change to determine if temperature-related epidemiological effects were controlled by the virus in its interaction with the host. Genomic change and genomic change per unit time were computed from an alignment of 55 453 SARS-CoV-2 genome sequences to determine if there were significant statistical correlations with temperatures and geographic coordinates of the countries from where genomes were collected. Mutation accumulation and rates were calculated for the entire genome and for specific regions known for significant pathways of mutational change. Statistically significant r values computed in a Pearson correlation test were smaller than 0.1. Thus, we were unable to find any significant correlation that would indicate a positive or negative association strength (Figure 2).

Discussion

Effect of temperature on COVID-19 epidemiology

Solar energy received by any region of the world varies with time of day, seasons, and latitude, a phenomenon that creates temperature variations. Temperature is also affected by differences in topographical surface and altitude. For example, continents are generally warmer than oceanic regions in the Northern hemisphere, while this reverses in the Southern Hemisphere tempered by the scarcity of land masses. We chose to collect epidemiological data during a time in which temperature differences between countries continue to be maximal according to a latitude-calendar month profile of surface air

temperature across the globe.²² Our worldwide analysis of epidemiological data collected for 211 countries (April 15, 2020) suggests that there is a significant negative correlation between environmental temperature and normalized COVID-19 epidemiological data, including disease incidence, mortality, recovery rate and the number of active cases (Figure 1). Two recent studies are in line with the evidence we here present. A recent study showed that weather affected disease incidence in Indonesia and might be an important factor in decreasing the number of COVID-19 cases in that country.²⁸ Indeed, Indonesia was one of the countries with the lowest incidence rates in our study (Supplemental Table S1). Similarly, Demongeot et al.²⁹ showed that higher temperatures decreased infection rates in both French administrative regions and in 21 countries spread throughout continents. In contrast, a correlation study at the onset of the pandemic when only 5768 deaths had been reported (January 22–March 16, 2020) mostly in China, Italy and Iran revealed some strong correlations between epidemiological variables but no significant correlation with temperature.³⁰ The inability to find a correlation with temperature may simply stem from the study being conducted during late winter. Instead, our study appropriately used spring temperatures, acquiring the bulk of epidemiological data during late March and early April. We find significant correlations during the initial phases of the pandemic despite the limited temperature ranges of the spring; top countries exhibiting either highest or lowest levels of epidemiological descriptors had temperatures ranging from 30 to 98.5 °F (Supplemental Table S1).

Viral disease slows down during the summer for certain viruses, including influenza.³¹ Pearson's correlation was previously used to study global and local patterns controlling influenza-like virus seasonality.³² Low temperatures in weather patterns showed a high negative correlation with the incidence of cases, a result that is relevant for COVID-19 research. This phenomenon is often observed with several respiratory viruses and is also implicated in virus survival and transmission.³² The spread of SARS-CoV during the 2003 epidemic has been shown to be temperature-dependent.³³ Moreover, Wu et al.²¹ saw that temperature was significantly correlated with the spread of SARS after adjustment for several factors, including the number of patients in intensive care units. However, explaining the cause of a temperature effect on epidemiology can be difficult. For example, high environmental temperatures may decrease both survival and infectivity of the virus. Unfortunately, at this time there is not enough evidence to conclusively support the hypothesis. Real-world viral infections can be hard to recreate in a laboratory, so studying temperature effects on viral survival or infection is challenging. Some virus strains are more environmentally susceptible than others and their survival can be affected by regions, climate and weather. The massive global spread of COVID-19 suggests that SARS-CoV-2 can spread very quickly despite warm and

humid weather. However, the effects of the environment on COVID-19 are still being explored. While 56°C temperatures kill the virus at ~10 000 viral units per 15 minutes, these environmental conditions are not reached during this spring season and very few countries will reach that temperature during the summer.³⁴

The WHO suggests disease incidence, mortality, recovery rate and number of active cases are properties related to onset of the infection. However, both mortality and recovery rate are more dependent on the organism's immune system,³⁵ which suggests that patients from regions with high temperatures are also more resistant to the virus. An alternative explanation of our data is that countries with higher temperatures have less environmental airflow and high humidity which are important factors that decrease the success of viral infections.³³ In the US, temperature does not vary as much as in other countries of the World, even though the average temperature of the states was 50 °F and normalized epidemiological variables are close to those of countries with the same temperatures (Table S1). This temperature invariance could explain why US data does not show correlations with epidemiological data.

Effect of geographic coordinates

Latitude is a strong predictor of temperature while longitude is not. Thus, geographic coordinates can be used to confirm the validity of temperature effects worldwide. Indeed, we find a significant correlation between latitude and epidemiological variables. Increases in latitude away from the equator were positively correlated with incidence, mortality, recovery rate and number of active cases, but showed no correlation with testing rate, which is probably influenced by a number of complicating factors (Figure 1). Remarkably, the latitude effect was not present in the US, where the latitude range of 19.74° and 66.16° was smaller than that of the worldwide analysis. As expected, a worldwide correlation analysis of longitude and epidemiological data revealed no significant effects. In sharp contrast, a strong correlation between longitude and incidence, mortality, number of active cases and hospitalization rate was found in the US. Here, incidence is more related to the onset of the disease and mortality, the spread of the disease in both highly populated coasts of the US, and the differential response to the disease across US states. Number of active cases and hospitalization rate are more related to the host immunity. A different perspective would suggest that recovery rate and testing rate are more related to how individual states manage the disease, which is limited by statewide economical and medical support, and their strategy to lift restrictions.

Effect of preparedness and morbidity

Several factors contribute to coronavirus-elicited ARDS, including obesity and cardiovascular and respiratory disease.³⁶ The role of such co-morbidities in hospitalization rates or

deaths remains understudied.³⁷ Reports suggest obesity, diabetes, and old age are highly related to the probability of hospitalization due to coronavirus infection.³⁷ There is evidence that patients with diabetes may be more vulnerable to infection.³⁷ We introduced a risk index of preparedness that incorporates morbidities. This risk index however did not show significant correlations with either number of cases or deaths. The absence of this relationship may be explained by specific factors affecting the spread of the virus and the pandemic in every country such as demographics and geography.³⁸ Those factors may not only include morbidity variables but also response decisions to address the pandemic.

The case of Nepal suggests that government preparedness and response had a crucial role in the development of the pandemic.³⁹ The role of infrastructure in preparedness is not only limited to health infrastructure but, also, to communications and transportation, among others.³⁸ We note that for future reference, it is recommended to validate the relationships proposed in this study with the risk index using post-pandemic data. This kind of longitudinal study would allow us to test the strength of the proposed index and would help understand how the factors considered in the risk index affect mortality rates and the spread of the virus. This tool can also be used in the future for the prioritization of health resources.

Effect of genomic change

A recent parallel study of 15342 indexed virus genome sequences revealed novel pathways of mutational change during the early stages of the COVID-19 pandemic.³⁸ The analysis predicted an ongoing mutational shift from the spike and replication proteins to other regions of the proteome, especially those known to represent major β -interferon antagonists that subvert the immune response, including the nucleocapsid protein and the viroporin 3a protein. To test if these and other mutational pathways were responsible for the temperature and latitude effects we here report, we explored if there were significant statistical correlations of genomic change or genomic change per time with temperatures and geographic coordinates of the countries from where genomes were collected (Figure 2). However, r coefficients computed in a Pearson correlation test falsified the hypothesis. Apportioning genomic change to selected regions of the genome that encode for a protein of significance, such as the replication complex encoded by ORF1a/b, the domains of the spike protein, and the NSP2 protease regulator, did not show significant correlations either. Thus, mutational changes in the virus genomic makeup appear unrelated to the temperature modulation of the COVID-19 disease, prompting to consider that the effects seen in the epidemiological data are directly, but not exclusively, dependent on the environment and the host immune system.

Previous studies with influenza have shown that ambient temperature and nutritional status can control the virus-specific adaptive immune response.⁴⁰ This mechanism might be

modulated by type I IFN, which under warm temperatures, can restrict viral replication.⁴¹ Research on mice suggests that exposure to high levels of heat severely impaired adaptive immune responses following respiratory influenza virus infection, which also affects virus clearance.⁴⁰ Even though the relationship between environmental temperatures and immune system regulation is still unclear, some studies have planted the idea of high environmental temperatures and host nutritional status regulating the generation of virus-specific CD4+ and CD8+ T cells and antibody responses following respiratory viral infection. These specific CD8+ T cells require IL-1R signaling in lung dendritic cells (DCs)⁴² and under high environmental temperatures, the levels of autophagy in the lung increase. Since autophagy and mitophagy restrict inflammasome-dependent cytokine release by regulating the amounts of pro-IL-1 β and damaged mitochondria, respectively,⁴³ it is possible that elevated levels of autophagy in high heat-exposed patients suppresses IL-1 β secretion in the lung following viral infection.⁴³ This provides a better understanding of how environmental temperatures affect epidemiological variables by suggesting the importance of IFNs and virus-specific T cells.

Other biological factors can also be relevant. Low-temperature seasons are often associated with vitamin D deficiency because of seasonal reductions in exposure to ultra-violet (UV) radiation.⁴⁴ Low levels of vitamin D have been shown to impair the body's antimicrobial peptide system, which is responsible for regulating the immune response.^{45,46} Seasonal changes in length of day can interfere with an individual's circadian rhythm, which is regulated by the release of the hormone melatonin. This interference can weaken the immune system and increase the risk of infection.⁴⁴ Changes in photoperiod and sunlight exposure have been used to explain the observed latitudinal migration of influenza activity during the winter season.⁴⁵ As observed with influenza infection latitudinal belts, lower temperatures are associated with increases in morbidity and mortality. Thus, the interaction of the environment with COVID-19 epidemiological data is important and merits further study.

Acknowledgements

This study began as a class research project in CPSC 567, a course in bioinformatics and systems biology taught by G.C.-A. at the University of Illinois in the spring of 2020. We thank Tre Tomaszewski for his input on the choice of genomic data and other members of the class who provided extensive feedback on the project. We also thank Herve Seligmann for sharing results prior to publication. Finally, we acknowledge with gratitude, the work of all researchers who obtained specimens, and generated and shared genetic sequence data on GISAID. COVID-19 research in the laboratory of G.C.-A is supported by the Office of Research and Office of International Programs in the College of Agricultural, Consumer and Environmental Sciences at the University of Illinois at Urbana-Champaign.

Author Contributions

K.S.-D. and P.B designed the project and contributed as first authors equally to this work. K.S.-D. found the most relevant genes for the genomic analysis. P.B collected and analyzed the genomic component of the project, R.J.G.R. collected all the epidemiological data. K.S.-D. collected and analyze the temperature and coordinates data. I.C. and R.J.G.R. collected and analyzed the preparedness and morbidity data. All co-authors contributed to data acquisition, data analysis, and data interpretation and wrote, edited and approved the manuscript.

ORCID iDs

Katria Soto-Díaz  <https://orcid.org/0000-0002-7425-6995>

Izan Chalen  <https://orcid.org/0000-0002-4607-8140>

Gustavo Caetano-Anollés  <https://orcid.org/0000-0001-5854-4121>

Supplemental material

Supplemental material for this article is available online.

REFERENCES

- Park SE. Epidemiology, virology, and clinical features of severe acute respiratory syndrome-coronavirus-2 (SARS-CoV-2; Coronavirus Disease-19). *Clin Exp Pediatr*. 2020;63:119-124.
- World Health Organization. World now at the start of 2009 influenza pandemic. Statement to the press by WHO-Director General Dr. Margaret Chan. 2009. https://www.who.int/mediacentre/news/statements/2009/h1n1_pandemic_phase6_20090611/en/
- Song Z, Xu Y, Bao L, et al. From SARS to MERS, thrusting coronaviruses into the spotlight. *Viruses*. 2019;11:59.
- Guo YR, Cao QD, Hong ZS, et al. The origin, transmission and clinical therapies on coronavirus disease 2019 (COVID-19) outbreak: an update on the status. *Mil Med Res*. 2020;7:11.
- Cui J, Li F, Shi ZL. Origin and evolution of pathogenic coronaviruses. *Nat Rev Microbiol*. 2019;17:181-192.
- Wu F, Zhao S, Yu B, et al. A new coronavirus associated with human respiratory disease in China [published correction appears in *Nature*. 2020 Apr;580:E7]. *Nature*. 2020;579:265-269.
- Letko M, Marzi A, Munster V. Functional assessment of cell entry and receptor usage for SARS-CoV-2 and other lineage B betacoronaviruses. *Nat Microbiol*. 2020;5:562-569.
- Wu A, Peng Y, Huang B, et al. Genome composition and divergence of the novel coronavirus (2019-nCoV) originating in China. *Cell Host Microbe*. 2020;27:325-328.
- Angeletti S, Benvenuto D, Bianchi M, Giovanetti M, Pascarella S, Ciccozzi M. COVID-2019: the role of the nsp2 and nsp3 in its pathogenesis. *J Med Virol*. 2020;92:584-588.
- Zhang L, Shen FM, Chen F, Lin Z. Origin and evolution of the 2019 novel coronavirus. *Clin Infect Dis*. 2020;71:882-883.
- Wu D, Zou S, Bai T, et al. Poultry farms as a source of avian influenza A (H7N9) virus reassortment and human infection. *Sci Rep*. 2015;5:7630.
- Tan X, Wu C, Li X, et al. On the origin and continuing evolution of SARS-CoV-2. *Nat Sci Rev*. 2020;7:1012-1023.
- Tomaszewski T, DeVries R, Dong M, et al. New pathways of mutational change in SARS-CoV-2 proteomes involve regions of intrinsic disorder important for virus replication and release. *Evol Bioinf*. 2020;16:1-18.
- Wong CK, Lam CW, Wu AK, et al. Plasma inflammatory cytokines and chemokines in severe acute respiratory syndrome. *Clin Exp Immunol*. 2004;136:95-103.
- Tay MZ, Poh CM, Rénia L, MacAry PA, Ng LFP. The trinity of COVID-19: immunity, inflammation and intervention. *Nat Rev Immunol*. 2020;20:363-374.
- Huang C, Wang Y, Li X, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China [published correction appears in *Lancet*. 2020 Jan 30]. *Lancet*. 2020;395:497-506.
- Chen N, Zhou M, Dong X, et al. Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study. *Lancet*. 2020;395:507-513.
- Zhang B, Zhou X, Qiu Y, et al. Clinical characteristics of 82 cases of death from COVID-19. *PLoS One*. 2020;15:e0235458.
- Chu KH, Tsang WK, Tang CS, et al. Acute renal impairment in coronavirus-associated severe acute respiratory syndrome. *Kidney Int*. 2005;67:698-705.
- Biryukov J, Boydston JA, Dunning RA, et al. Increasing temperature and relative humidity accelerates inactivation of SARS-CoV-2 on surfaces. *MSphere*. 2020;5:e00441-20.
- Wu Y, Jing W, Liu J, et al. Effects of temperature and humidity on the daily new cases and new deaths of COVID-19 in 166 countries. *Sci Total Environ*. 2020;729:139051.
- Manabe S, Ploshay J, Lau NC. Seasonal variation of surface temperature change during the last several decades. *J Clim*. 2011;24:3817-3821.
- Uddin MB, Hasan M, Harun-Al-Rashid A, Ahsan ML, Imran MAS, Ahmed SSU. Ancestral origin, antigenic resemblance and epidemiological insights of novel coronavirus (SARS-CoV-2): global burden and Bangladesh perspective. *Infect Genet Evol*. 2020;84:104440.
- Worldometer. 2020 coronavirus update (live) from COVID-19 virus pandemic – worldometer. 2020. Accessed April 15, 2020. https://www.worldometers.info/coronavirus/?utm_campaign=homeAdvegas1220
- World Health Organization. (2020). Global health observatory: explore a world of health data. <https://www.who.int/data/gho>
- Shu Y, McCauley J. GISAID: global initiative on sharing all influenza data – from vision to reality. *Euro Surveill*. 2017;22:30494.
- Virtanen P, Gommers R, Oliphant TE, et al. SciPy 1.0: fundamental algorithms for scientific computing in Python [published correction appears in *Nat Methods*. 2020 Feb 24]. *Nat Methods*. 2020;17:261-272.
- Tosepu R, Gunawan J, Effendy DS, et al. Correlation between weather and Covid-19 pandemic in Jakarta, Indonesia. *Sci Total Environ*. 2020;725:138436.
- Demongeot J, Flet-Berliac Y, Seligmann H. Temperature decreases spread parameters of the new Covid-19 case dynamics. *Biology (Basel)*. 2020;9:94.
- Pawari S, Stanam A, Chaudhari M, Rayudu D. Effects of temperature on COVID-19 transmission. Preprint. Posted online March 30, 2020. medRxiv preprint. doi:10.1101/2020.03.29.20044461.
- Moriyama M, Hugentobler WJ, Iwasaki A. Seasonality of respiratory viral infections. *Annu Rev Virol*. 2020;7:83-101.
- Nimbalkar PM, Tripathi NK. Space-time epidemiology and effect of meteorological parameters on influenza-like illness in Phitsanulok, a northern province in Thailand. *Geospat Health*. 2016;11:447.
- Lin K, Fong DYT, Zhu B, Karlberg J. Environmental factors on the SARS epidemic: air temperature, passage of time and multiplicative effect of hospital infection. *Epidemiol Infect*. 2006;134:223-230.
- Yao Y, Pan J, Liu Z, et al. No association of COVID-19 transmission with temperature or UV radiation in Chinese cities. *Eur Respir J*. 2020;55:2000517.
- Delano MJ, Ward PA. The immune system's role in sepsis progression, resolution, and long-term outcome. *Immunol Rev*. 2016;274:330-353.
- Simonnet A, Chetboun M, Poissy J, et al. High prevalence of obesity in severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2) requiring invasive mechanical ventilation. *Obesity (Silver Spring)*. 2020;28:1195-1199.
- Muniyappa R, Gubbi S. COVID-19 pandemic, coronaviruses, and diabetes mellitus. *Am J Physiol Endocrinol Metab*. 2020;318:E736-E741.
- Phelan AL, Katz R, Gostin LO. The novel coronavirus originating in Wuhan, China: challenges for global health governance. *JAMA*. 2020;323:709-710.
- Koirala J, Acharya S, Neupane M, Phuyal M, Rijal N, Khanal U. Government preparedness and response for 2020 pandemic disaster in Nepal: a case study of COVID-19. *Management*. Published online March 30, 2020. doi: 10.2139/ssrn.3564214
- Moriyama M, Ichinohe T. High ambient temperature dampens adaptive immune responses to influenza A virus infection. *Proc Natl Acad Sci USA*. 2019;116:3118-3125.
- Boonarkart C, Suptawiwat O, Sakorn K, Puthavathana P, Auewarakul P. Exposure to cold impairs interferon-induced antiviral defense. *Arch Virol*. 2017;162:2231-2237.
- Ichinohe T, Lee HK, Ogura Y, Flavell R, Iwasaki A. Inflammasome recognition of influenza virus is essential for adaptive immune responses. *J Exp Med*. 2009;206:79-87.
- Harris J, Hartman M, Roche C, et al. Autophagy controls IL-1beta secretion by targeting pro-IL-1beta for degradation. *J Biol Chem*. 2011;286:9587-9597.
- Fuhrmann C. The effects of weather and climate on the seasonality of influenza: what we know and what we need to know. *Geography Compass*. 2010;4:718-730.
- Cannell JJ, Zaslouff M, Garland CF, Scragg R, Giovannucci E. On the epidemiology of influenza. *Viral J*. 2008;5:29.
- Cannell JJ, Vieth R, Umhau JC, et al. Epidemic influenza and vitamin D. *Epidemiol Infect*. 2006;134:1129-1140.