BMC Infectious Diseases

RESEARCH Open Access

Real-time quantification of the transmission advantage associated with a single mutation in pathogen genomes: a case study on the D614G substitution of SARS-CoV-2

Shi Zhao[1,2], Jingzhi Lou[1], Lirong Cao[1], Hong Zheng[1], Marc K. C. Chong[1,2], Zigui Chen[3], Renee W. Y. Chan[4,5,6,7], Benny C. Y. Zee[1,2], Paul K. S. Chan[3] and Maggie H. Wang[1,2*]

Abstract

Background: The COVID-19 pandemic poses serious threats to global health, and the emerging mutation in SARS-CoV-2 genomes, e.g., the D614G substitution, is one of the major challenges of disease control. Characterizing the role of the mutation activities is of importance to understand how the evolution of pathogen shapes the epidemiological outcomes at population scale.

Methods: We developed a statistical framework to reconstruct variant-specific reproduction numbers and estimate transmission advantage associated with the mutation activities marked by single substitution empirically. Using likelihood-based approach, the model is exemplified with the COVID-19 surveillance data from January 1 to June 30, 2020 in California, USA. We explore the potential of this framework to generate early warning signals for detecting transmission advantage on a real-time basis.

Results: The modelling framework in this study links together the mutation activity at molecular scale and COVID-19 transmissibility at population scale. We find a significant transmission advantage of COVID-19 associated with the D614G substitution, which increases the infectivity by 54% (95%CI: 36, 72). For the early alarming potentials, the analytical framework is demonstrated to detect this transmission advantage, before the mutation reaches dominance, on a real-time basis.

Conclusions: We reported an evidence of transmission advantage associated with D614G substitution, and highlighted the real-time estimating potentials of modelling framework.

Keywords: COVID-19, Mutation, Transmission advantage, Real-time estimation, Statistical modelling

Introduction

The dynamics of the transmission of an infectious disease is largely determined by the pathogen's infectiousness and the course of the transmission [1, 2]. The control of a contagious disease with high infectiousness requires the knowledge of the driven factors that may affect the transmission process [3, 4]. Virus mutation is one of the major challenges for controlling epidemics [5, 6]. The profile of pathogen in terms of viral fitness and functionality may be altered by mutations [7, 8], and in consequence change its transmissibility. Referring to the previous literature on seasonal influenza [9], a few key amino acid (AA) substitutions may lead to remarkable changing dynamics of

*Correspondence: maggiew@cuhk.edu.hk
[1] JC School of Public Health and Primary Care, Chinese University of Hong Kong, Hong Kong, China
Full list of author information is available at the end of the article

antigenic property and epidemiological outcomes at population scale [10, 11]. Similar findings were also reported for other viral pathogens [12, 13].

The coronavirus disease 2019 (COVID-19), whose etiological agent is the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) [14], swept the world in a short period of time [15], and the ongoing COVID-19 pandemic poses serious threat to public health [16]. As of December 31, 2020, over 81 million COVID-19 cases are confirmed in the world with over 1.8 million associated deaths. In February 2020, genetic variants carrying the D614G substitution on the SARS-CoV-2 spike (S) protein began to spread first in Europe [17] and otherwhere globally, reaching fixation in many places rapidly. The D614G is potentially affecting viral transmission [5, 18]. Recent modelling analysis reported statistical evidence that SARS-CoV-2 strains with D614G substitution are likely to have an increased infectivity retrospectively [19]. In 2021, although 614D still can be detected in some places, e.g., Australia with around 25% frequency, the variants carrying 614G is predominant globally.

Some of these variant genomes upon the different selection pressure increase their frequency in the population. Recently, the SARS-CoV-2 Delta variants composed of several novel mutations on Spike protein increased their frequency [20]. This becomes one of the major challenges of COVID-19 control because these variants have more competitive pathological features such higher transmission or resistance to vaccines [21, 22]. Exploring the relationship between the mutation activities and the disease transmissibility is of importance to understand how the evolutionary patterns at molecular scale may shape the epidemiological outcomes at population scale. Quantifying the advantage of mutations that affects the transmission may inform the disease control strategic decision-making process [23].

Given the intensity and the risk scale of the ongoing COVID-19 pandemic, real-time surveillance and inference of the role of key mutations may be crucial for fighting against the pandemic. In this study, we adopted a statistical inference framework to estimate the transmission advantage associated with a single mutation in pathogen genomes empirically, and exemplify by using the COVID-19 data in California, USA. We demonstrate the potentials of this analytical framework to produce an early warning signal for detecting transmission advantage on a real-time basis.

Methods

Reproduction number and transmission advantage: parameterization and likelihood framework

The time-varying reproduction number is commonly adopted to quantify the instantaneous transmissibility of infectious disease in an epidemic. Using the estimation framework in [24], the epidemic growth is modelled as a branching process, thus can be expressed as the ratio of $C(t)$ over $\int_0^\infty w(k)C(t-k)\mathrm{d}k$, which is commonly known as the renewable equation [25]. Here, the $C(t)$ is the observed number of new COVID-19 cases on the t-th day. The function $w(\cdot)$ is the distribution of the generation time (GT) of the disease. The GT is defined as the time interval between the time of exposure, i.e., being infected, of a primary case and that of his associated secondary case in the consecutive transmission generation [26]. The distribution $w(\cdot)$ is predefined in our model, which is commonly estimated from contract tracing surveillance data [27–30].

The transmission advantage of the mutated variant against the original type is defined as the ratio, denoted by η, of the strain-specified reproduction numbers. We denote the reproduction number of cases infected by the original variant as R_t, and thus the reproduction number of cases infected by the mutated variant is $\eta{\cdot}R_t$. If $\eta>1$, the mutated variant may be more infectious than the original genetic variant, and vice versa.

The observed proportion of original genetic variant is denoted by q_t, and the observed proportion of mutated variant is denoted by p_t. Since we consider the binary AA substituting process, we have $p_t+q_t=1$ for all ts. By using the renewable equation backwardly, we model the expected number of cases on the t-th day in Eq. (1).

$$\begin{aligned}\mathbf{E}[C_t] =& R_t \cdot \left[\int_0^\infty w(k)\cdot q(t-k)\cdot C(t-k)\mathrm{d}k \right.\\ &\left. +\eta\int_0^\infty w(k)\cdot p(t-k)\cdot C(t-k)\mathrm{d}k\right].\end{aligned} \tag{1}$$

Here, the $\mathbf{E}[\cdot]$ denotes the expectation function. Therefore, we construct the likelihood function $L_t^{(c)}$ of the daily number of cases using a Poisson-distributed framework with observation at C_t and rate parameter at $\mathbf{E}[C_t]$ as in Eq. (2).

$$L_t^{(c)}(C_t|\mathbf{E}[C_t]) = \frac{\mathbf{E}[C_t]^{C_t}\cdot e^{-\mathbf{E}[C_t]}}{C_t!}. \tag{2}$$

Here, the superscript '(c)' merely indicated the likelihood function is for the number of cases, which does not indicate the power. In addition, the overall reproduction number is $(q_t+\eta{\cdot}p_t){\cdot}R_t$.

For the observed sequencing data, we denote the numbers of original and mutated strains by m_t and n_t, respectively, for the t-th day. The expected chance (or probability) that a randomly selected strain at the t-th day carrying a specific mutation is given in Eq. (3).

$$\mathbf{E}[p_t] = \frac{\eta R_t \int_0^\infty w(k) \cdot p(t-k) \cdot C(t-k)\mathrm{d}k}{\mathbf{E}[C_t]}. \quad (3)$$

Then, we have $\mathbf{E}[q_t] = 1 - \mathbf{E}[p_t]$, which can be modelled with the same fashion. As such, by modelling the sampling of the genetic variants as a Bernoulli process, we construct the likelihood function ($L_t^{(s)}$) of the observed genotype using a Bernoulli-distributed framework with probability at $\mathbf{E}[p_t]$ as in Eq. (4).

$$L_t^{(s)}(m_t, n_t|\mathbf{E}[p_t]) = (1 - \mathbf{E}[p_t])^{m_t} \cdot \mathbf{E}[p_t]^{n_t}. \quad (4)$$

Here, the superscript '(s)' merely indicated the likelihood function is for genetic variants, which does not indicate the power.

With Eqs. (2) and (4), we reconstruct the R_t time series, denoted by $\{R_t\}$, and estimate η using the overall likelihood function defined in Eq. (5).

$$L(\{R_t\}, \eta|\{C_t\}, \{m_t\}, \{n_t\}) = \prod_t \left[L_t^{(c)} \cdot L_t^{(s)}\right]. \quad (5)$$

Similar formulations were adopted in previous studies [19, 31, 32].

COVID-19 surveillance data and SARS-CoV-2 sequencing data

To demonstrate the application of the framework, we adopted the data of COVID-19 in California, USA, and estimated the transmission advantage η of the D614G substitution. The surveillance data of daily number of COVID-19 cases are collected from the **R** package "*nCov2019*" [33], which is extracted from the COVID-19 surveillance platform launched by the New York Times. Figure 1A shows the daily number of COVID-19 cases time series in California.

The SARS-CoV-2 strains are obtained via the Global initiative on sharing all influenza data (GISAID) with collection dates ranging from January 1 to June 30, 2020 in California [34]. A total of 4268 full-length human SARS-CoV-2 strains are retrieved on December 31, 2020. All SARS-CoV-2 strains used for analysis are provided in the appendix (Additional file 1). We consider the study period from January 1 to June 30, 2020 when other mutated lineages, e.g., B.1.1.7, P.1, or B.1.617.2, were not yet detected. Multiple sequence alignment is performed using Clustal Omega [35], and the SARS-CoV-2 strain '*China/Wuhan-Hu-1/2019|EPI_ISL_402125*' is considered as the reference sequence.

Likelihood-based inference and real-time estimation

To setup the model, we considered the w as a Gamma distribution having mean ($\pm$SD) values of 5.3 ($\pm$2.1) days by averaging the GT estimates for COVID-19 from the existing literatures [27–29, 36, 37]. Slight variation in the settings of the GT will not affect our main findings.

Using the likelihood framework defined in Eq. (5), we calculate the maximum likelihood estimation (MLE) of η to determine transmission advantage of D614G substitution. The 95% confidence intervals (95%CI) are calculated using the profile likelihood estimation framework with a Chi-square quantile as cutoff [38, 39], which is also adopted in [40–43].

For the real-time estimation, we repeat the statistical inference process of η using a part of dataset, instead of the full dataset, divided by the observing date. For example, the real-time estimate of η on the τ-th day is calculated by using the dataset with reporting date from the first day (i.e., January 1, 2020) to the τ-th day. We compare the consistency of the η estimates on a real-time basis in terms of their scales and 95%CIs. Moreover, we define the early warning signal as that a real-time estimate of η larger than 1 and of statistical significance can be obtained before the mutated strains (i.e., those SARS-CoV-2 strains with amino acid G) reach the dominance level in the population. For dominance level, it is considered as the proportion of the mutated strains (p_t) over 0.5, i.e., $p_t > 0.5$, which can be observed empirically. An early warning signal indicates the real-time estimating potentials of our analytical framework in detecting the transmission advantage due to mutation.

Results

In California, the epidemic curve grew since February, see Fig. 1A, peaked in July with daily number of COVID-19 case over 10,000, declined in August, and has maintained at a steady level since September (data not shown). We reconstruct the daily instantaneous reproduction numbers of the cases infected by SARS-CoV-2 strains with D614 or G614 type in Fig. 1B. We observe that the overall trends of reproduction numbers are relatively high in the early March, but gradually decreasing thereafter since the local 'stay-at-home' order was issued on March 19, 2020 in California [44]. During the first half of March, which is regarded as the early phase of the outbreak, the average reproduction number is 2.4, which is largely consistent with most of previous estimates [15, 16, 45–47].

We report the estimated proportion of D614G substitution $\mathbf{E}[p_t]$ fits the observed sequencing data well, see Fig. 1C. We infer the transmission advantage η at 1.54 (95%CI: 1.36, 1.72), which means the D614G substitution increases 54% of the transmissibility. Hence, in Fig. 1B, the reproduction number of the SARS-CoV-2 variant with 614G appears higher than that of the original genotype. Although reproduction number R_t of the 614D are below 1 for most of the time after April 2020, the

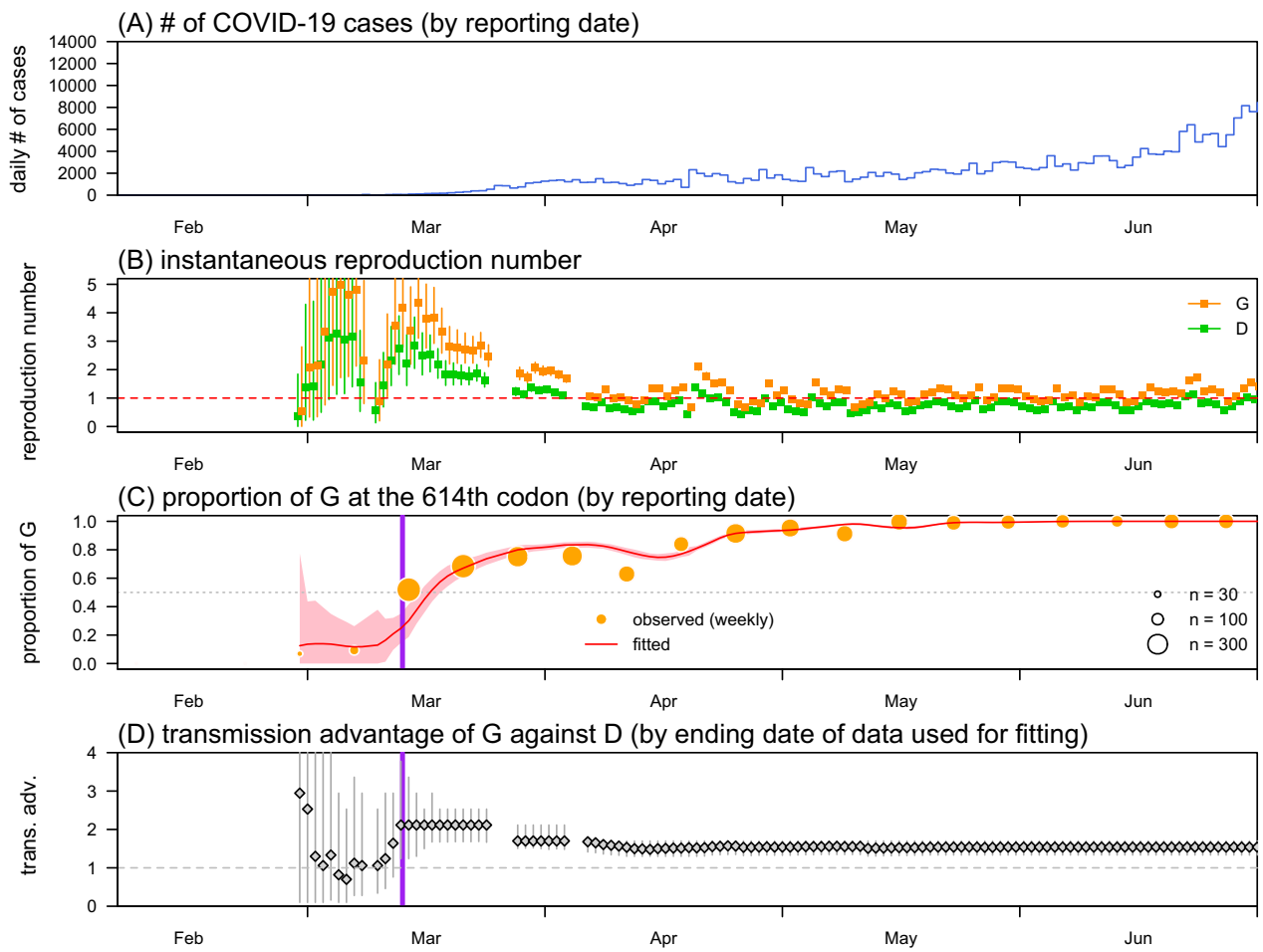

Fig. 1 The daily number of COVID-19 cases (panel **A**), the reconstructed reproduction number (R_t, panel **B**), proportion of the G on the 614-th codon of the S protein (panel **C**), and estimated transmission advantage of G on the real-time basis (η, panel **D**). Panel A shows the daily number of COVID-19 time series in California state, USA. Panel **B** shows the estimated R_ts of G (in orange) and **D** (in green). Panel **C** shows the observed (dots) and fitted (curve) proportion of the Glycine (G) on the 614-th codon of the S protein. Panel **D** shows the real-time estimates of the transmission advantage of G (η). In panels **B** and **D**, the dots are the estimates, and bars are the 95%CIs. In panel **C**, the curve indicates the mean fitting results, the shading area indicates the 95%CIs, and horizontal dashed grey line represents proportion level at 0.5. In panels C and D, the vertical bold purple line represents the date, March 12, 2020 when the η estimates yields an 'early warning signal', i.e., $\eta > 0$ significantly, detecting the transmission advantage of G (against **D**)

reproduction number $\eta \cdot R_t$ of type G fluctuated around 1 during the same period, which led to a large-scale epidemic wave in California during summer in 2020 (see Fig. 1A).

For the real-time estimating potentials, we find that the real-time estimates of η appear unstable in February and early March, when the D614G substitution emerge, and gradually converge and stabilize since March 12, see Fig. 1D. Specially, on March 12 (highlighted in Fig. 1C and Fig. 1D), when the proportion of D614G substitution (p_t) reaches 35% (<0.5), the η estimate is 2.12 (95%CI: 1.24, 3.78), which is significantly larger than 1.

Discussion

Although the variants carrying D614G substitution might be introduced to California from aboard during the first few months of pandemic, the observed changes in SARS-CoV-2 mutations (p_t) were likely due to the spread of virus locally after the implementation of strict travel-ban measurements. The significant increase in transmissibility associated with the D614G substitution is biologically reasonable according to similar findings reported in previous studies. Consistent evidences of the transmission advantage of D614G substitution were also reported in previous literature both statistically [19, 48]

and experimentally [49–53]. The D614G replacement leads to increased infectivity and stability of the virion and is shown to enhance viral replication in human lung epithelial cells [51, 52]. The interaction of the SARS-CoV-2 S protein with multiple epithelium components, e.g., glycocalyx, and proteases, govern the cellular entry [54]. Thus, the mutations on S protein with more effective interaction with these epithelium components enables SARS-CoV-2 variants to infect with relatively lower virus titer. Previous analysis implied that the D614G substitution may alter the conformation of spike protein trimer that shifted toward an ACE2 binding-competent state [50], and thus may functionally improve receptor binding capacity from a theoretical perspective [17, 18, 53]. The D614G substitution increases host cell entry via ACE2 and transmembrane protease serine 2 (TMPRSS2) [54]. Comparing to substitution, we learn from the influenza virus that major antigenic changes can be caused by a single AA substitution related to the receptor binding domain (RBD) [55].

Although a significant transmission advantage of D614G is found, we notice that the proportion of 614G variant generally increased, while the reproduction number series decreased in March and then remained constant. The reasons may include that the increase in transmissibility associated with D614G was counteracted by the effects of local non-pharmaceutical interventions that reduced the overall transmission of COVID-19. For sensitivity checking, we repeat the estimating process of η with alternative mean GT using a shorter estimate of 4.0 days [30] and a longer estimate of 7.5 days [15]. We find that the η estimates are consistently and significantly larger than 1 in similar scales (data not shown), which validates our main results. The statistical inference framework is empirical, and thus can be extended to explore the transmission advantage attributed to single mutations for other infectious diseases.

Our analytical framework can yield an early warning signal in detecting the transmission advantage due to D614G substitution before the mutation reaching dominance on a real-time basis. Although some recent studies indicate that the D614G mutation is unlikely to undermine the neutralization from current SARS-CoV-2 vaccine candidates [53, 56], there are also other studies suggest the concerns should be raised oppositely [57, 58]. Similar concerns of the changes in the protective effect from vaccine or prior infection are frequently raised regarding other recent SARS-CoV-2 varaints [22, 59–62]. Under selection, viral quasispecies including closely related viral genomes might be generated by the accumulation of mutations [63]. As such, the early warning signal provides an opportunity for improving disease control strategies and healthcare planning against the mutated strains, which might have different diagnostic conditions or clinical outcomes [19, 50, 53]. Hence, we highlight the importance of our analytical framework, such that the public health risks related to viral mutations may be controllable with early preparedness.

For the limitations of this study, we have the following remarks. First, the reconstruction of R_t relies on the setting of the generation time (GT). We model the GT distribution, i.e., $w(\cdot)$, of COVID-19 as a fixed Gamma distribution, which follows previous studies [27–30, 36]. In the real-world situation, the time interval between transmission generations might be varying [45], which may affect the reconstruction of reproduction number. However, the overall trends of R_t estimates are unlikely changed due to slight variation in GT [45]. Thus, we consider the impact of this limitation on the inference of transmission advantage may be negligible, and our model can be extended to a more complex context with the time-varying GT data available. Second, theoretically, the GT distribution might be altered by the mutated strains. However, by screening the literature of COVID-19, we find no evidence that GT is varied associated with the D614G substitution in SARS-CoV-2, and thus we presume $w(\cdot)$ to be a fixed distribution. Third, for the R_t estimation parts, $C(t)$ in the 'methods' section should be the numbers of COVID-19 cases onset at time t. However, due to the surveillance data by onset date are unavailable, we adopted the current dataset by reporting data as a proxy for the COVID-19 incidence time series. If one considers a constant reporting lag, the R_t estimates will have exactly same trends but shifted for the reporting lag. Considering the similar reporting delay also occurred for the SARS-CoV-2 sequencing data, the effects of the two reporting lags may be counteracted. We remark that this approximation in analysis is unlikely to affect the main conclusions in this study. Furthermore, with detailed reporting lag information of each individual case, adjustment for reporting delay can surely be carried out based on our current analytical framework. Fourth, this study focuses on exploring the effects on changing the disease transmissibility associated with a single mutation, e.g., D614G, but the real-world biological mechanisms, which are usually more complex, remain uncovered. As an example, on one hand, the R384G substitution in influenza A/H3N2 virus enhances ability of in-host immune-escape [64], which indicates an increase in infectivity [9], but this substitution appears detrimental. On the other hand, the co-mutations of R384G in nucleoprotein (NP) could improve and compensate the viral fitness or functionality of [7, 8], such that the mutated strains reached fixation rapidly in 1993–1994 flu season. Future studies are needed for exploring the mechanisms of how D614G in

SARS-CoV-2 affects the transmissibility of COVID-19. Fifth, the transmission advantage can be contributed by multiple factors such as increase in infectiousness or viral viability, change in the infection risk to different group of hosts [65], change in the escape from antibodies, shortening of generation interval, changes in clinical conditions, population size dynamics, and selective pressures [66]. Our analytical framework cannot disentangle the effects of each factor, which requires more complex methods, and detailed information [67]. Sixth, homogeneous mixing and equal contribution of all cases were assumed in our model. Thus, the reproduction numbers and transmission advantage estimates are interpreted as the average scales for the whole population in California. Seventh, there are multiple mutations in the SARS-CoV-2 variants carrying 614G, and we remark that the independent effects of each mutation cannot be disentangled in this study, where the interactions among these mutations are unassessed. Lastly, as a data-driven study, the estimated association should be interpreted with caution. With ecological setting, though our analysis provides statistical evidence about the likelihood of causality, the findings in this study cannot guarantee the causality, which needs further biomedical experiments in more sophisticated contexts.

Conclusions

The modelling framework in this study links together the mutation activity at molecular scale and COVID-19 transmissibility at population scale. We report statistical evidence of the transmission advantage associated with the D614G substitution in SARS-CoV-2. We highlight that an early warning signal in detecting this transmission advantage can be generated on a real-time basis. Future studies on exploring the mechanism between SARS-CoV-2 mutation and COVID-19 infectivity are needed.

Abbreviations

AA: Amino acid; COVID-19: Coronavirus disease 2019; D614G: The amino acid substitution changing from Aspartic Acid (D) to Glycine (G) on the 614-th codon (of the S protein of SARS-CoV-2); GT: Generation time; LR: Likelihood-ratio; GISAID: Global initiative on sharing all influenza data; MLE: Maximum likelihood estimation; RBD: Receptor binding domain; SARS-CoV-2: Severe acute respiratory syndrome coronavirus 2; SD: Standard deviation; 95%CI: 95% Confidence interval.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12879-021-06729-w.

Additional file 1. The acknowledgement table of SARS-CoV-2 strain sequences used in this study.

Acknowledgements

This study is conducted using the resources of Alibaba Cloud Intelligence High Performance Cluster computing facilities, which is made free for COVID-19 research.

Disclaimer

The funding agencies had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; or decision to submit the manuscript for publication.

Authors' contributions

Conceptualization: SZ. Methodology: SZ, and MKCC. Software: SZ. Validation: SZ. Formal analysis: SZ. Investigation: SZ. Resources: SZ, and JZ. Data Curation: SZ, and JZ. Writing—Original Draft: SZ. Writing—Review and Editing: JZ, LC, HZ, MKCC, ZC, RWYC, BCYZ, PKSC, and MHW. Visualization: SZ. Supervision: MHW. Project Administration: JZ. Funding acquisition: MHW. All authors critically read the manuscript, and gave final approval for publication.

Funding

This work is supported by CUHK grant [PIEF/Ph2/COVID/06, 4054456], the Health and Medical Research Fund (HMRF) Commissioned Research on COVID-19 [INF-CUHK-1] of Hong Kong, China, and partially supported by the National Natural Science Foundation of China (NSFC) [31871340, 71974165].

Availability of data and materials

All data used in this work are publicly available. The processed data and codes are available via https://github.com/plxzpnxZBD/real-time_TransAdv.

Declarations

Ethics approval and consent to participate

The COVID-19 number of cases and sequencing data are collected via public domains, and thus neither ethical approval nor individual consent is applicable.

Consent for publication

Not applicable.

Competing interests

MHW is a shareholder of Beth Bioinformatics Co., Ltd. BCYZ is a shareholder of Beth Bioinformatics Co., Ltd and Health View Bioanalytics Ltd. Other authors declared no competing interests.

Author details

[1]JC School of Public Health and Primary Care, Chinese University of Hong Kong, Hong Kong, China. [2]CUHK Shenzhen Research Institute, Shenzhen, China. [3]Department of Microbiology, Chinese University of Hong Kong, Hong Kong, China. [4]Department of Paediatrics, Chinese University of Hong Kong, Hong Kong, China. [5]Hong Kong Hub of Pediatric Excellence, Chinese University of Hong Kong, Shatin, N.T., Hong Kong, China. [6]CUHK-UMCU Joint Research Laboratory of Respiratory Virus & Immunobiology, Chinese University of Hong Kong, Shatin, N.T., Hong Kong, China. [7]Li Ka Shing Institute of Health Sciences, Faculty of Medicine, Chinese University of Hong Kong, Shatin, N.T., Hong Kong, China.

Received: 7 May 2021 Accepted: 20 September 2021
Published online: 07 October 2021

References

1. Tuite AR, Fisman DN. Reporting, Epidemic Growth, and Reproduction Numbers for the 2019 Novel Coronavirus (2019-nCoV) Epidemic. Ann Intern Med. 2020;172(8):567–8.
2. Riou J, Althaus CL: Pattern of early human-to-human transmission of Wuhan 2019 novel coronavirus (2019-nCoV), December 2019 to January 2020. Euro surveillance : bulletin Europeen sur les maladies transmissibles = European communicable disease bulletin 2020, 25(4):2000058.

3. Kutter JS, Spronken MI, Fraaij PL, Fouchier RA, Herfst S. Transmission routes of respiratory viruses among humans. Curr Opin Virol. 2018;28:142–51.
4. Fraser C, Riley S, Anderson RM, Ferguson NM. Factors that make an infectious disease outbreak controllable. Proc Natl Acad Sci U S A. 2004;101(16):6146–51.
5. Baum A, Fulton BO, Wloga E, Copin R, Pascal KE, Russo V, Giordano S, Lanza K, Negron N, Ni M, et al. Antibody cocktail to SARS-CoV-2 spike protein prevents rapid mutational escape seen with individual antibodies. Science. 2020;369(6506):1014–8.
6. Tsetsarkin KA, Vanlandingham DL, McGee CE, Higgs S: A single mutation in chikungunya virus affects vector specificity and epidemic potential. PLoS Pathog 2007, 3(12):e201.
7. Rimmelzwaan GF, Berkhoff EGM, Nieuwkoop NJ, Fouchier RAM, Osterhaus A. Functional compensation of a detrimental amino acid substitution in a cytotoxic-T-lymphocyte epitope of influenza a viruses by comutations. J Virol. 2004;78(16):8946–9.
8. Rimmelzwaan GF, Berkhoff EGM, Nieuwkoop NJ, Smith DJ, Fouchier RAM, Osterhaus A. Full restoration of viral fitness by multiple compensatory co-mutations in the nucleoprotein of influenza A virus cytotoxic T-lymphocyte escape mutants. J Gen Virol. 2005;86(6):1801–5.
9. Gog JR, Rimmelzwaan GF, Osterhaus ADME, Grenfell BT. Population dynamics of rapid fixation in cytotoxic T lymphocyte escape mutants of influenza A. Proc Natl Acad Sci. 2003;100(19):11143–7.
10. Smith DJ, Lapedes AS, de Jong JC, Bestebroer TM, Rimmelzwaan GF, Osterhaus AD, Fouchier RA. Mapping the antigenic and genetic evolution of influenza virus. Science. 2004;305(5682):371–6.
11. Zhao S, Lou J, Cao L, Chen Z, Chan RW, Chong MK, Zee BC, Chan PK, Wang MH. Quantifying the importance of the key sites on haemagglutinin in determining the selection advantage of influenza virus: Using A/H3N2 as an example. J Infect. 2020;81(3):452–82.
12. Botosso VF, Zanotto PM, Ueda M, Arruda E, Gilio AE, Vieira SE, Stewien KE, Peret TC, Jamal LF, Pardini MI et al: Positive selection results in frequent reversible amino acid replacements in the G protein gene of human respiratory syncytial virus. PLoS Pathog 2009, 5(1):e1000254.
13. Tolle MA. Mosquito-borne diseases. Curr Probl Pediatr Adolesc Health Care. 2009;39(4):97–140.
14. Hu B, Guo H, Zhou P, Shi ZL. Characteristics of SARS-CoV-2 and COVID-19. Nat Rev Microbiol. 2021;19(3):141–54.
15. Li Q, Guan X, Wu P, Wang X, Zhou L, Tong Y, Ren R, Leung KSM, Lau EHY, Wong JY, et al. Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus-Infected Pneumonia. N Engl J Med. 2020;382(13):1199–207.
16. Wu JT, Leung K, Leung GM. Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study. Lancet. 2020;395(10225):689–97.
17. Wan Y, Shang J, Graham R, Baric RS, Li F: Receptor Recognition by the Novel Coronavirus from Wuhan: an Analysis Based on Decade-Long Structural Studies of SARS Coronavirus. J Virol 2020, 94(7).
18. Benvenuto D, Demir AB, Giovanetti M, Bianchi M, Angeletti S, Pascarella S, Cauda R, Ciccozzi M, Cassone A. Evidence for mutations in SARS-CoV-2 Italian isolates potentially affecting virus transmission. J Med Virol. 2020;92(10):2232–7.
19. Volz E, Hill V, McCrone JT, Price A, Jorgensen D, O'Toole A, Southgate J, Johnson R, Jackson B, Nascimento FF et al: Evaluating the Effects of SARS-CoV-2 Spike Mutation D614G on Transmissibility and Pathogenicity. Cell 2021, 184(1):64–75 e11.
20. Ito K, Piantham C, Nishiura H. Predicted domination of variant Delta of SARS-CoV-2 before Tokyo Olympic games, Japan. Eurosurveillance. 2021;26(27):2100570.
21. Yadav PD, Sapkal GN, Abraham P, Ella R, Deshpande G, Patil DY, Nyayanit DA, Gupta N, Sahay RR, Shete AM et al: Neutralization of Variant Under Investigation B.1.617.1 With Sera of BBV152 Vaccinees. Clin Infect Dis 2021.
22. Planas D, Veyer D, Baidaliuk A, Staropoli I, Guivel-Benhassine F, Rajah MM, Planchais C, Porrot F, Robillard N, Puech J et al: Reduced sensitivity of SARS-CoV-2 variant Delta to antibody neutralization. Nature 2021.
23. Ferguson NM, Cummings DA, Cauchemez S, Fraser C, Riley S, Meeyai A, Iamsirithaworn S, Burke DS. Strategies for containing an emerging influenza pandemic in Southeast Asia. Nature. 2005;437(7056):209–14.
24. Cori A, Ferguson NM, Fraser C, Cauchemez S. A new framework and software to estimate time-varying reproduction numbers during epidemics. Am J Epidemiol. 2013;178(9):1505–12.
25. Zhao S, Musa SS, Hebert JT, Cao P, Ran J, Meng J, He D, Qin J: Modelling the effective reproduction number of vector-borne diseases: the yellow fever outbreak in Luanda, Angola 2015–2016 as an example. PeerJ 2020, 8:e8601.
26. Wallinga J, Lipsitch M. How generation intervals shape the relationship between growth rates and reproductive numbers. Proc Biol Sci. 2007;274(1609):599–604.
27. Ferretti L, Wymant C, Kendall M, Zhao L, Nurtay A, Abeler-Dorner L, Parker M, Bonsall D, Fraser C: Quantifying SARS-CoV-2 transmission suggests epidemic control with digital contact tracing. Science 2020, 368(6491):eabb6936.
28. Ganyani T, Kremer C, Chen D, Torneri A, Faes C, Wallinga J, Hens N: Estimating the generation interval for coronavirus disease (COVID-19) based on symptom onset data, March 2020. Euro surveillance : bulletin Europeen sur les maladies transmissibles = European communicable disease bulletin 2020, 25(17):2000257.
29. Tindale LC, Stockdale JE, Coombe M, Garlock ES, Lau WYV, Saraswat M, Zhang L, Chen D, Wallinga J, Colijn C: Evidence for transmission of COVID-19 prior to symptom onset. Elife 2020, 9:e57149.
30. Zhao S. Estimating the time interval between transmission generations when negative values occur in the serial interval data: using COVID-19 as an example. Math Biosci Eng. 2020;17(4):3512–9.
31. Leung K, Lipsitch M, Yuen KY, Wu JT. Monitoring the fitness of antiviral-resistant influenza strains during an epidemic: a mathematical modelling study. Lancet Infect Dis. 2017;17(3):339–47.
32. Zhao S, Lou J, Cao L, Zheng H, Chong MKC, Chen Z, Chan RWY, Zee BCY, Chan PKS, Wang MH: Quantifying the transmission advantage associated with N501Y substitution of SARS-CoV-2 in the UK: an early data-driven analysis. J Travel Med 2021, 28(2):taab011.
33. Wu T, Hu E, Ge X, Yu G. nCov2019: an R package for studying the COVID-19 coronavirus pandemic. PeerJ. 2021;9:e11421.
34. Shu Y, McCauley J. GISAID: Global initiative on sharing all influenza data–from vision to reality. Eurosurveillance. 2017;22(13):30494.
35. Sievers F, Higgins DG: Clustal Omega, accurate alignment of very large numbers of sequences. In: Multiple sequence alignment methods. edn.: Springer; 2014: 105–116.
36. He X, Lau EHY, Wu P, Deng X, Wang J, Hao X, Lau YC, Wong JY, Guan Y, Tan X: Temporal dynamics in viral shedding and transmissibility of COVID-19. Nat Med 2020:1–4.
37. Zhao S, Tang B, Musa SS, Ma S, Zhang J, Zeng M, Yun Q, Guo W, Zheng Y, Yang Z et al: Estimating the generation interval and inferring the latent period of COVID-19 from the contact tracing data. Epidemics 2021, 36:100482.
38. Fan JQ, Huang T. Profile likelihood inferences on semiparametric varying-coefficient partially linear models. Bernoulli. 2005;11(6):1031–57.
39. Bolker BM: Ecological models and data in R: Princeton University Press; 2008.
40. Breto C, He DH, Ionides EL, King AA. Time Series Analysis Via Mechanistic Models. Annals of Applied Statistics. 2009;3(1):319–48.
41. He D, Ionides EL, King AA. Plug-and-play inference for disease dynamics: measles in large and small populations as a case study. J R Soc Interface. 2010;7(43):271–83.
42. Lin Q, Chiu AP, Zhao S, He D. Modeling the spread of Middle East respiratory syndrome coronavirus in Saudi Arabia. Stat Methods Med Res. 2018;27(7):1968–78.
43. Zhao S, Lou J, Chong MKC, Cao L, Zheng H, Chen Z, Chan RWY, Zee BCY, Chan PKS, Wang MH: Inferring the Association between the Risk of COVID-19 Case Fatality and N501Y Substitution in SARS-CoV-2. Viruses 2021, 13(4).
44. California Health Officials Announce a Regional Stay at Home Order [https://www.gov.ca.gov/wp-content/uploads/2020/03/3.19.20-attested-EO-N-33-20-COVID-19-HEALTH-ORDER.pdf]
45. Ali ST, Wang L, Lau EHY, Xu XK, Du Z, Wu Y, Leung GM, Cowling BJ. Serial interval of SARS-CoV-2 was shortened over time by nonpharmaceutical interventions. Science. 2020;369(6507):1106–9.
46. Chinazzi M, Davis JT, Ajelli M, Gioannini C, Litvinova M, Merler S, Pastore YPA, Mu K, Rossi L, Sun K, et al. The effect of travel restrictions on the

spread of the 2019 novel coronavirus (COVID-19) outbreak. Science. 2020;368(6489):395–400.
47. Gatto M, Bertuzzo E, Mari L, Miccoli S, Carraro L, Casagrandi R, Rinaldo A. Spread and dynamics of the COVID-19 epidemic in Italy: Effects of emergency containment measures. Proc Natl Acad Sci U S A. 2020;117(19):10484–91.
48. Leung K, Pei Y, Leung GM, Lam TTY, Wu JT: Empirical transmission advantage of the D614G mutant strain of SARS-CoV-2. medRxiv 2020; 10.1101/2020.09.22.20199810
49. Weissman D, Alameh MG, de Silva T, Collini P, Hornsby H, Brown R, LaBranche CC, Edwards RJ, Sutherland L, Santra S et al: D614G Spike Mutation Increases SARS CoV-2 Susceptibility to Neutralization. Cell Host Microbe 2021, 29(1):23–31 e24.
50. Yurkovetskiy L, Wang X, Pascal KE, Tomkins-Tinch C, Nyalile TP, Wang Y, Baum A, Diehl WE, Dauphin A, Carbone C et al: Structural and Functional Analysis of the D614G SARS-CoV-2 Spike Protein Variant. Cell 2020, 183(3):739–751 e738.
51. Plante JA, Liu Y, Liu J, Xia H, Johnson BA, Lokugamage KG, Zhang X, Muruato AE, Zou J, Fontes-Garfias CR et al: Spike mutation D614G alters SARS-CoV-2 fitness. Nature 2020.
52. Hou YJ, Chiba S, Halfmann P, Ehre C, Kuroda M, Dinnon KH 3rd, Leist SR, Schafer A, Nakajima N, Takahashi K, et al. SARS-CoV-2 D614G variant exhibits efficient replication ex vivo and transmission in vivo. Science. 2020;370(6523):1464–8.
53. Korber B, Fischer WM, Gnanakaran S, Yoon H, Theiler J, Abfalterer W, Hengartner N, Giorgi EE, Bhattacharya T, Foley B et al: Tracking Changes in SARS-CoV-2 Spike: Evidence that D614G Increases Infectivity of the COVID-19 Virus. Cell 2020, 182(4):812–827 e819.
54. Seyran M, Takayama K, Uversky VN, Lundstrom K, Palù G, Sherchan SP, Attrish D, Rezaei N, Aljabali AAA, Ghosh S et al: The structural basis of accelerated host cell entry by SARS-CoV-2. The FEBS Journal 2020, n/a(n/a).
55. Koel BF, Burke DF, Bestebroer TM, van der Vliet S, Zondag GC, Vervaet G, Skepner E, Lewis NS, Spronken MI, Russell CA, et al. Substitutions near the receptor binding site determine major antigenic change during influenza virus evolution. Science. 2013;342(6161):976–9.
56. Dearlove B, Lewitus E, Bai H, Li Y, Reeves DB, Joyce MG, Scott PT, Amare MF, Vasan S, Michael NL, et al. A SARS-CoV-2 vaccine candidate would likely match all currently circulating variants. Proc Natl Acad Sci U S A. 2020;117(38):23652–62.
57. van Dorp L, Acman M, Richard D, Shaw LP, Ford CE, Ormond L, Owen CJ, Pang J, Tan CCS, Boshier FAT et al: Emergence of genomic diversity and recurrent mutations in SARS-CoV-2. Infect Genet Evol 2020, 83:104351.
58. Weisblum Y, Schmidt F, Zhang F, DaSilva J, Poston D, Lorenzi JC, Muecksch F, Rutkowska M, Hoffmann HH, Michailidis E et al: Escape from neutralizing antibodies by SARS-CoV-2 spike protein variants. Elife 2020, 9:e61312.
59. Xie XP, Liu Y, Liu JY, Zhang XW, Zou J, Fontes-Garfias CR, Xia HJ, Swanson KA, Cutler M, Cooper D et al: Neutralization of SARS-CoV-2 spike 69/70 deletion, E484K and N501Y variants by BNT162b2 vaccine-elicited sera. Nat Med 2021:1–2.
60. Moore JP, Offit PA. SARS-CoV-2 Vaccines and the Growing Threat of Viral Variants. JAMA. 2021;325(9):821–2.
61. Muik A, Wallisch A-K, Sänger B, Swanson KA, Mühl J, Chen W, Cai H, Maurus D, Sarkar R, Türeci Ö: Neutralization of SARS-CoV-2 lineage B. 1.1. 7 pseudovirus by BNT162b2 vaccine–elicited human sera. Science 2021.
62. Supasa P, Zhou D, Dejnirattisai W, Liu C, Mentzer AJ, Ginn HM, Zhao Y, Duyvesteyn HME, Nutalai R, Tuekprakhon A: Reduced neutralization of SARS-CoV-2 B. 1.1. 7 variant by convalescent and vaccine sera. Cell 2021.
63. Andino R, Domingo E. Viral quasispecies. Virology. 2015;479–480:46–51.
64. Berkhoff EGM, Boon ACM, Nieuwkoop NJ, Fouchier RAM, Sintnicolaas K, Osterhaus A, Rimmelzwaan GF. A mutation in the HLA-B* 2705-restricted NP383-391 epitope affects the human influenza A virus-specific cytotoxic T-lymphocyte response in vitro. J Virol. 2004;78(10):5216–22.
65. Faria NR, Mellan TA, Whittaker C, Claro IM, Candido DdS, Mishra S, Crispim MAE, Sales FCS, Hawryluk I, McCrone JT: Genomics and epidemiology of the P. 1 SARS-CoV-2 lineage in Manaus, Brazil. Science 2021.
66. Saad-Roy CM, Morris SE, Metcalf CJE, Mina MJ, Baker RE, Farrar J, Holmes EC, Pybus OG, Graham AL, Levin SA. Epidemiological and evolutionary considerations of SARS-CoV-2 vaccine dosing regimes. Science. 2021;372(6540):363–70.
67. Ong SWX, Young BE, Lye DC: Lack of detail in population-level data impedes analysis of SARS-CoV-2 variants of concern and clinical outcomes. The Lancet Infectious Diseases.

Publisher's Note