International Society of Travel Medicine
Established 1991
Promoting healthy travel worldwide

Rapid Communication

Quantifying the transmission advantage associated with N501Y substitution of SARS-CoV-2 in the UK: an early data-driven analysis

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

[1]JC School of Public Health and Primary Care, Chinese University of Hong Kong, Hong Kong, [2]CUHK Shenzhen Research Institute, Shenzhen, China, [3]Department of Microbiology, Chinese University of Hong Kong, Hong Kong and [4]Department of Paediatrics, Chinese University of Hong Kong, Hong Kong

*To whom correspondence should be addressed. Email: maggiew@cuhk.edu.hk (MHW)

Abstract

The emerging N501Y mutation in severe acute respiratory syndrome coronavirus 2, which becomes prevalent in the UK rapidly, is one of the major challenges of COVID-19 control. To explore the transmission advantage, we estimate that the N501Y substitution increases the infectivity by 52% (95% confidence interval: 46, 58) in terms of the reproduction number.

Key words: COVID-19, mutation, transmission advantage, reproduction number, statistical modelling, B.1.1.7 lineage

Main text

Exploring the relationship between mutation activities and disease transmissibility is important to manage the epidemiological outcomes at population scale. In September 2020, genetic variants carrying the N501Y substitution on the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) spike (S) protein began to spread first in Wales and then otherwhere in the UK[1] and reached dominance among SARS-CoV-2 variants in other lineages by the end of 2020. The rapid spread of 501Y variant indicates a possible transmission advantage over the 501N variant. In this work, we adopted a statistical framework to infer the transmission advantage associated with N501Y substitution using the surveillance and sequencing data in the UK.

We reconstructed the variant-specified instantaneous reproduction number by using the commonly adopted renewable equation.[2] The transmission advantage of 501Y variant against 501N variant is defined as the ratio, η, between variant-specified reproduction numbers, see Supplementary Information S1. In other words, given the reproduction number of cases infected by 501N variant as R_t, the reproduction number of cases infected by the 501Y variant is $\eta \cdot R_t$. Thus, if $\eta > 1$, the mutated variant may be more infectious than the original genetic variant and *vice versa*.

We obtained the SARS-CoV-2 sequencing data and COVID-19 surveillance data (Figure 1A) collected from 1 August to 31 December 2020, via the public databases released by GISAID and World Health Organization, respectively, see Supplementary Information S2. With the likelihood-based inference framework, we reconstruct the variant-specified instantaneous reproduction number in Figure 1B. The reproduction number of 501Y variant appears higher than that of 501N variant. Although R_t of the 501N variant fluctuated around 1, the $\eta \cdot R_t$ of 501Y variant appears generally above 1 during the same period, which implies a higher potential for causing larger epidemics, and more intensive intervention strategies are thus desired. Figure 1C shows that the estimated proportion of 501Y variant fits the observed proportion well. We estimated that the 501Y variant is 52% (95% confidence interval: 46, 58) more transmissible than the 501N variant, which indicates a transmission advantage of N501Y substitution. To check the sensitivity of estimates, we repeat the estimating process of η with alternative shorter or longer mean generation time. We find
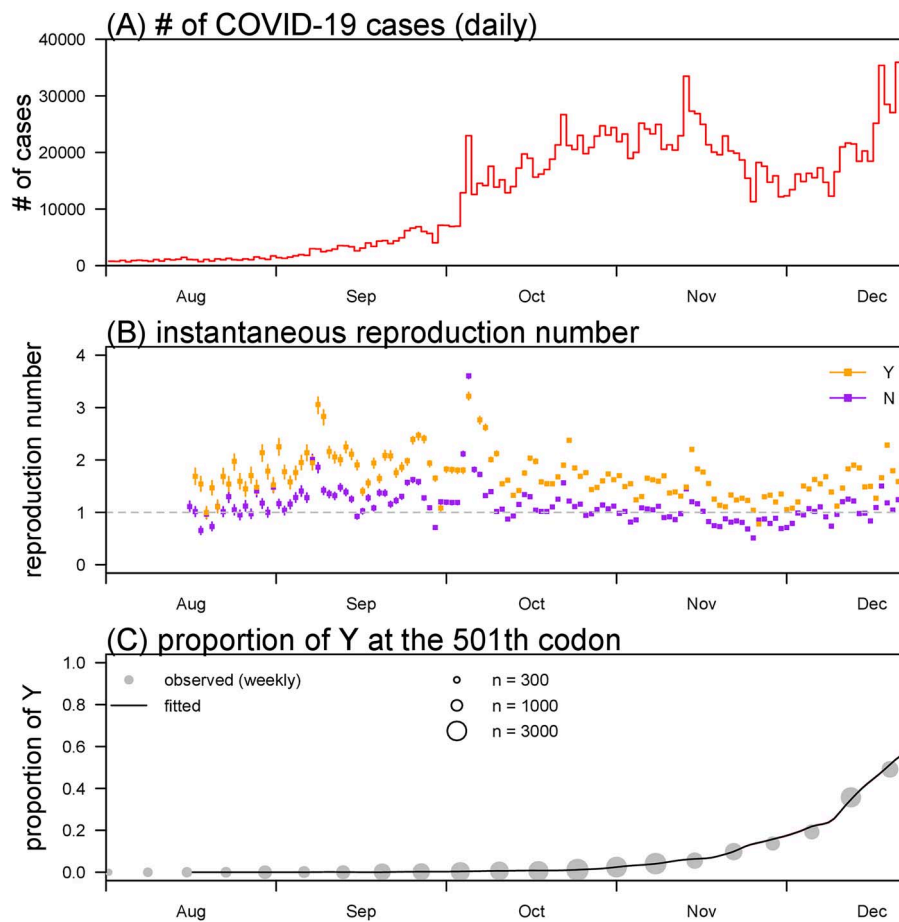
Figure 1. The daily number of COVID-19 cases (panel A), the reconstructed reproduction number (R_t, panel B) and proportion of the Y on the 501th codon of the S protein (panel C). Panel A shows the daily number of COVID-19 time series in the UK. Panel B shows the estimated reproduction numbers of Y (in orange) and N (in purple). The dots are the estimates, and bars are the 95% confidence intervals. Panel C shows the observed (dots) and fitted (curve) proportion of the Y on the 501th codon of the S protein

that the η estimates are consistently and significantly larger than 1 in similar scales (data not shown), which validates our main results.

The increase in transmissibility associated with N501Y substitution appears consistent with the statistical evidence and structural biological insights in literature.[3,4] Similarly, we learn from the recent D614G substitution of SARS-CoV-2, as well as the seasonal influenza virus, that single but key AA substitution may alter the conformation of protein trimer and binding-competent profiles.[5]

Provided the rapidly growing prevalence and spread of 501Y variant, several concerns of public health importance are raised, which need further exploration. The transmission advantage of the mutation implies increase in both herd immunity threshold and intrinsic growth rate of epidemic, and thus, the non-pharmaceutical interventions for COVID-19 control may require enforcement to flatten the epidemic curve. Clinical severity remains largely unassessed for the 501Y variant, which needs relatively a long-term observation, and unexpected clinical outcomes may warrant adjustments in the treatment strategies. Given the challenges in the control of spatial spreading of COVID-19, we highlight the potential risks of SARS-CoV-2 with 501Y when it invades places having groups of population with vulnerability. There is a limited investigation

for the roles of other mutations co-circulating with N501Y, which might be of importance in compensating or enhancing the viral fitness or functionality. In addition, the matching level of vaccine candidates in development or in production is also unknown, which requires further analysis from both biological and epidemiological sides. Some limitations about the analytical procedures, mainly due to lack of data or information, are discussed in Supplementary Information S3.

Supplementary data

Supplementary data are available at *JTM* online.

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.

Availability of materials

All data used in this work are publicly available.

Consent for publication

Not applicable.

Authors' contributions

Acknowledgements

Funding

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.

Conflict of interests

M.H.W. is a shareholder of Beth Bioinformatics Co., Ltd. B.C.Y.Z. is a shareholder of Beth Bioinformatics Co., Ltd and Health View Bioanalytics Ltd. Other authors declared no competing interests.

References

1. Tang JW, Tambyah PA, Hui DSC. Emergence of a new sars-cov-2 variant in the UK. *J Infect*. doi: 10.1016/j.jinf.2020.12.024.
2. 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**:1505–12.
3. Leung K, Shum MH, Leung GM *et al*. Early transmissibility assessment of the n501y mutant strains of sars-cov-2 in the United Kingdom, October to November 2020. *Eurosurveillance* 2021; **26**: 2002106.
4. Teruel N, Maihot O, Najmanovich RJ. Modelling conformational state dynamics and its role on infection for sars-cov-2 spike protein variants. *bioRxiv* 2020; 2020.12.16.423118.
5. Yurkovetskiy L, Wang X, Pascal KE *et al*. Structural and functional analysis of the d614g sars-cov-2 spike protein variant. *Cell* 2020; **183**:739–51.e8.