SARS-CoV-2 transmissibility compared between variants of concern and vaccination status

Liang Wang[1,*], Xavier Didelot[2], Yuhai Bi[1,3,*], George F Gao[1,3,*]

[1]CAS Key Laboratory of Pathogenic Microbiology and Immunology, Institute of Microbiology, Center for Influenza Research and Early-warning (CASCIRE), CAS-TWAS Center of Excellence for Emerging Infectious Diseases (CEEID), Chinese Academy of Sciences, Beijing 100101, China

[2]School of Life Sciences and Department of Statistics, University of Warwick, Coventry CV47AL, United Kingdom

[3]University of Chinese Academy of Sciences, Beijing 101408, China

* Correspondence: wangliang@im.ac.cn (L.W); beeyh@im.ac.cn (Y.B.); gaof@im.ac.cn (G.F.G)

Abstract

Since the start of the SARS-CoV-2 pandemic in late 2019, several variants of concern (VOC) have been reported to have increased transmissibility. In addition, despite the progress of vaccination against SARS-CoV-2 worldwide, all vaccines currently in used are known to protect only partially from infection and onward transmission. We combined phylogenetic analysis with Bayesian inference under an epidemiological model to infer the reproduction number (R_t) and also trace person-to-person transmission. We ~~also~~ examined the impact of phylogenetic uncertainty and sampling bias on the estimation. Our result indicated that ~~the~~ lineage B had a significantly higher transmissibility than lineage A, and contributed to the global pandemic to a large extent. In addition, although the transmissibility of VOCs ~~has been increased compared with~~is larger than other exponentially growing lineages ~~with exponential growth rate~~, this difference is not very high. The probability of detecting onward transmission from patients infected with SARS-CoV-2 VOCs who had received at least one dose of vaccine was approximate 1.06% (3/284), which was slightly lower but not statistically ~~not~~ significantly different from a probability of 1.21% (10 /828) for unvaccinated individuals. In addition to VOCs, exponentially growing lineages ~~with exponential growth rate~~ in each country should also be ~~paid attention~~account for when tailoring prevention and control strategies. One dose of vaccination could not efficiently prevent the onward transmission of SARS-CoV-2 VOCs. ~~In order to prevent this~~Consequently, non-pharmaceutical interventions (such as ~~low-cost and efficient strategies, like~~ wearing masks and social distancing ~~etc~~) should still be implemented in each country

during the vaccination period.

Keywords

SARS-CoV-2, variants of concern, vaccine, transmissibility, onward transmission

Introduction

Coronavirus diseases 2019 (COVID-19), the biggest pandemic so far in the 21st century, is caused by a novel type of coronaviruses named SARS-CoV-2 (also known as 2019-nCoV, or HCoV-19)[1]. As of 10th October 2021, there are more than 238 million confirmed cases with more than four million deaths[2], posing a global threat to public health. During the SARS-CoV-2 pandemic, several types of SARS-CoV-2 variants of concern (VOC) with increased transmissibility emerged, such as B.1.1.7 (WHO label: Alpha), B.1.351 (WHO label: Beta), P.1 (WHO label: Gamma), and B.1.617.2 (WHO label: Delta)[2-5], the global spread of these VOCs has also further thoroughly taxed the medical systems and global economies.

Although VOCs deserves worldwide attention, those lineages with exponential growth in each country cannot be ignored. Since the advantages of transmissibility for VOCs were mainly concluded by comparing them to all other lineages as a whole[2, 5], it will cause the advantage of transmissibility for some lineages to be overwhelmed. In addition, VOCs have also been reported to be harder to neutralize by convalescent and vaccine sera than others[6-11], indicating they could still infect vaccinated individuals, which therefore could increase the probability of transmission to others. Together with the increased breakthrough infection rates[12], more efforts are needed to identify the transmissibility of lineages with exponential growth other than VOCs in each country and survey the extent of onward transmission caused by vaccinated persons being infected by SARS-CoV-2 VOCs, which is also an indicator for policy makers to tailor

further prevention and control measures during the vaccination and post-vaccination process.

Materials and methods

Data collection and selection

SARS-COV-2 genomic sequences were download from GISAID several times (data for estimating lineage A and B was downloaded at 9th April 2020, data for UK was downloaded at 21st December 2020, data for South Africa and Brazil was downloaded at 16th March 2021, data for India was downloaded at 13th May 2021). For estimating the extent of onward transmission caused by vaccinated persons being infected by SARS-CoV-2 VOCs, genomic sequences and corresponding patients' vaccination status were download from GISAID at 18th June 2021. Totally, we got 408 SARS-CoV-2 genomic sequences, all of which came from patients who had received at least one dose of vaccine before being infected with SARS-CoV-2 VOCs.

Only viral genomes collected before the implementation of national non-pharmaceutical interventions would be included in the analysis of R_t estimation for lineage A and B. In addition, countries that include lineage A and B, and the number of completely viral genomes within each lineage ≥80 would be included in the subsequent analysis. Since only the United States and Australia met the above criteria, the estimation of the transmissibility of lineage A and B was only based on the data of these two countries. The cut-off dates for the collection time in the USA and Australia are

20th and 25th January 2020, respectively, as there were no nationwide epidemic prevention measures were implemented before the date. Due to the high volume of genomic data from sub-lineages in the UK, South Africa, Brazil, and India, the amount of calculation would be too large, especially for reconstruction of dated phylogeny. ~~In this case, we~~ We therefore filtered and ~~also~~ sub-sampled the data for datasets from each sub-lineage. First, the viral genomes of patients who had not had a history of international travel are retained, according to their epidemiological data. Second, the viral genomes should also meet the criteria as follow: length ≥29 KB, and the ratio of N in the genome ≤1%. Third, based on the collection date, if more than 10 genomes were available in a specific date, we randomly select 10 of them, otherwise all genomes would be included. For identifying onward transmission caused by patients being infected with VOCs after receiving at least one dose of vaccine, we first filtered the data based on several following criteria. Only complete SARS-CoV-2 genomes from patients receiving at least one dose of vaccine were retained for further analysis. We then discarded genomic data with no exact collection date (accurate to days). Due to the aim of our study is to identify direct transmission events, we then also collected viral genomic sequences that were highly similar to those SARS-CoV-2 genomes from patients receiving at least one dose of vaccine, as we assumed that SARS-CoV-2 genome sequences from two patients that directly transmitted SARS-CoV-2 to each other were with high sequence similarity. For each SARS-CoV-2 genome from patients receiving at least one dose of vaccine (query), we also used BLAST to find 10 most similar complete genomes (target) and then retained those with exact collection date

(accurate to days) which were also from the same country as each query and their collection times were within 22 days (maximum infectious period)[13] after the collection time of the query. The query and target sequences were then put together and removed redundancy for further analysis. For SARS-CoV-2 Alpha VOC, genomic sequences were split into different datasets based on the country, and only dataset contained more than 70 SARS-CoV-2 genomes was used for further analysis, as the computational cost was extremely large if we combined data from all countries. Since there are still several countries with limited genomic sequences, we then merged them into a dataset. Other VOCs were considered as independent dataset and were not further split anymore. Finally, only 284 genomic sequences of SARS-CoV-2 VOCs, all of which came from patients who had received at least one dose of vaccine before infection, and 828 genomic sequences of SARS-CoV-2 VOCs that close related to the above sequences but all of which came from patients who did not receive vaccine at all were retained for further analysis. Before further analysis, genomic sequences were aligned using Mafft v7.310[14]. Then, we trimmed the uncertain regions in 3′ and 5′ terminals and also masked 30 sites (Supplementary Table 1) that are highly homoplastic and have no phylogenetic signal as previous noted (https://virological.org/t/issues-with-sars-cov-2-sequencing-data/473).

Reconstruction of dated phylogeny

Since recombination could affect the evolutionary signal, we searched for recombination events in these SARS-CoV-2 genomes using RDP4[15]. No evidence

for recombination has been found in our dataset. We used jModelTest v2.1.6[16] to find the best substitution model for each dataset according to the Bayesian information criterion. The best substitution model for each dataset was listed in Supplementary Table 2. The list of genomic sequences used in this study were provided in Supplementary Table 3 &4. The list of genomic sequences used in this study were openly shared via the GISAID initiative[17]. We then used the Bayesian Markov Chain Monte Carlo (MCMC) approach implemented in BEAST v1.10.4[18] to derive a dated phylogeny for each dataset. At least three replicate runs for each 100 million MCMC steps were performed for each dataset, among which sampled parameters and trees every 10,000 steps. For data from lineage A and B in USA and Australia during the early phase of COVID-19, the estimation of the most appropriate combination of molecular clock and coalescent models for Bayesian phylogenetic analysis was determined using both path-sampling and stepping-stone models[19]. In order to reduce the amount of calculation, we assumed that data from sub-lineages followed a strict molecular clock and with an exponential population growth tree prior, as genomic sequences used in each dataset were all from the same sub-lineage and they all had an exponential growth. For dataset of identifying onward transmission caused by patients being infected with VOCs after receiving at least one dose of vaccine, as genomic sequences used in each dataset were all from the same lineage, we assumed that they followed a strict molecular clock. The estimation of the most appropriate coalescent models for Bayesian phylogenetic analysis was determined using both path-sampling and stepping-stone models[19]. The model comparison result for datasets from lineage

A and B in USA and Australia were listed in Supplementary Table 5. Tracer 1.7.1[20] was then used to check the convergence of MCMC chain (effective sample size >200) and to compute marginal posterior distributions of parameters, after discarding 10% of the MCMC chain as burn-in. We determined whether there was sufficient temporal signal in each dataset, as it was the prerequisite for getting a reliable inference when performed phylodynamic analysis. Bayesian evaluation of temporal signal (BETS)[21] was used to evaluate the temporal signal in each dataset. BETS relies on the comparison of marginal likelihoods of two models: the heterochronous model (with tip date) and isochronous (without tip date) model. Analyses were performed with at least three independent replicates of 100 million MCMC steps each, sampling parameters and trees every 10,000 steps with the best substitution model and most appropriate combination of molecular clock and coalescent models determined above for each dataset. The marginal likelihoods were estimated by PS. The Bayes factor (BF) was then calculated based on the likelihoods of two models (heterochronous and isochronous). If the log BF >5 (heterochronous model against isochronous model), it indicated there was sufficient temporal signal in this dataset. The log BF for each dataset was listed in Supplementary Table 6, the result suggested that the temporal signal was sufficiently strong.

Transmission Analysis

As viral genomes were incompletely sampled and the pandemic is currently ongoing, TransPhylo v1.4.4[22] was used to infer the transmission tree using the dated

phylogeny generated above as input. For B.1.617.2 (Delta) dataset of identifying onward transmission caused by patients being infected with VOCs after receiving at least one dose of vaccine, we split them into four subtrees (Supplementary Figure 1) to reduce the amount of computation. The process of split tree into several subtrees did not affect the result, as direct transmission always occurred in patients within close-related branches. The generation time (*i.e.* the time gap from infection to onward transmission, denoted as G) of COVID-19 was previously estimated as 4.8 ± 1.7 days[23], and we used these values to compute the shape and scale parameter of a gamma distribution of G using the R package epitrix[24]. The distribution of sampling time (*i.e.* the time gap from infection to detection and sampling) was set equal to the distribution of generation time. For each dataset, we performed the TransPhylo analysis several replicated runs for each 500,000 iterations simultaneously estimating the transmission tree, the proportion of sampling, the within-host coalescent time Neg, and the two parameters of the negative binomial offspring distribution (which represents the number of secondary cases caused by each infection), and then merge them together. Therefore, R_t could be inferred as the median of the offspring distribution. All results were generated after discarding the first part of the MCMC chains as burn-in. The MCMC mixing and convergence was assessed based on the effective sample size of each parameter (>200) and by visual examination of the MCMC traces (-Supplementary Tables 7 & 8). The probabilities of direct transmission from one host to another were estimated as the proportion of MCMC samples in which this direct transmission event occurred. The expected numbers of intermediates from one host to another were

estimated as the average across the MCMC samples of the number of intermediates between the two hosts. The probability of onward transmission for VOCs caused by unvaccinated persons is calculated by taking the number of direct transmission event caused by unvaccinated persons and dividing by the total number of unvaccinated persons. The probability of onward transmission for VOCs caused by people receiving at least one dose of vaccine is calculated by taking the number of direct transmission event caused by people receiving at least one dose of vaccine and dividing by the total number of people receiving at least one dose of vaccine.

Evaluating the robustness of the estimation

Since dated phylogeny was used to estimate the transmissibility for each lineage, we should test whether and how the phylogenetic uncertainty and sampling bias affect the estimation of R_t. We first tested how the phylogenetic uncertainty affect the result, because only the maximum clade credibility (MCC) tree was used to estimate the transmissibility. We used data from our previous study[25]. Ten dated phylogenetic trees were randomly selected from the MCMC chains. The parameter setting was the same as previous study description. The estimation of R_t from random selected tree from MCMC chain were always lower than for the MCC tree (Supplementary Figure 2). As the MCC tree is more accurate than to trees sampled in MCMC chains, this result suggested that the uncertainty of the phylogeny would cause an underestimation of the R_t. ~~In this case~~Consequently, the use of the MCC tree for estimation of R_t would reduce the impact of phylogenetic uncertainty on the results as much as possible. In addition,

the sampling bias was also a key factor affecting the phylogenetic uncertainty. In order to test if the sampling bias affect the estimation of R_t, we also repeatedly randomly sub-sampled the data five times for each dataset using same criteria (if more than 10 genomes were available in a specific date, we randomly select 10 of them, otherwise all genomes would be included) and then performed the same analysis.

Results

Lineage B has a higher transmissibility than lineage A

The mean R_t for lineage A from Australia and USA were estimated as 1.75 (95% credible intervals (CI) 1.43-2.11) and 1.74 (95% CI 1.61-1.89), respectively (Figure 1A). However, the mean R_t for lineage B from Australia and USA were estimated as 2.33 (95% CI 2.05-2.64) and 3.18 (95% CI 2.76-3.63), respectively (Figure 1A). Firstly, the R_t of lineage B is significantly greater than that of lineage A, indicating higher transmissibility of lineage B compared to lineage A. This might be the reason why strains from lineage B rapidly became dominantly all over the world (Figure 1B). Secondly, the R_t of lineage A from the two countries are very close, however, the R_t of lineage B varied greatly between Australia and USA. We then found that the composition of lineage was significantly different between the datasets from these two countries (Figure 1C and D, $p<0.01$, Fisher's exact test, two-sided). We speculated that different sub-lineages within lineage B might have different transmissibility and then tested the hypothesis by conducting further analysis. Since the data from lineage A was limited, the evaluation of transmissibility for each sub-lineage was mainly focused on

those from lineage B and other emerging lineages in the same country during the same periods.

Some dominant lineages in the UK have similar transmissibility to B.1.1.7

The composition of lineages in the UK is shown in Figure 2A. B.1.177 was the dominant strain before 2021. We also found that the number of viral genomes from England far exceeds that from other parts of the UK (Figure 2B). Besides, according to the accumulation of number of viral genomes from each lineage in England, we could find that only three lineages (B.1.177, B.1.1.37, B.1.1.7) grew exponentially after October 2020 (Figure 2B). The R_t for B.1.177, B.1.1.37, B.1.1.7 were estimated as 1.08 (95% CI 1.072-1.09), 1.068 (95% CI 1.05-1.086), and 1.186 (95% CI 1.158-1.213) (Figure 2C). The B.1.177, B.1.1.37 had similar R_t which were both close to 1. However, B.1.1.7 had a significantly higher transmissibility than these two lineages. We next tested if the significantly high R_t could be affected by sampling bias. After five independently repeated sampling and subsequent analysis, we found that all these R_t for B.1.1.7 were close to each other, ranging from 1.178 to 1.194. ~~Besides~~Furthemore, all the 95% credible intervals from repeated sampling ~~also~~ did not ~~have any intersection~~intersect with those from lineage B.1.177 and B.1.1.37. Thus, the sampling bias had limited effect on the estimation of R_t for each lineage. We also found that B.1.177 had a similar transmissibility than B.1.1.37 (Student's t test, two-sided with Holm–Bonferroni adjusted p =0.1) (Figure 2D).

Slightly lower transmissibility for B.1.1.54 than B.1.351 in South Africa

The composition of lineages in South Africa is shown in Figure 3A. Lineage B.1.1.54 was the dominant strain before October 2020. Since then, the dominant strain in South Africa was switched to lineage B.1.351 gradually. According to the accumulation of number of viral genomes from each lineage in South Africa, we could find that only lineage B.1.1.54 and B.1.351 grew exponentially after July 2020 (Figure 3B). We could find the R_t for B.1.351 and B.1.54 during July 2020 and February 2021 were estimated as 1.05 (95% CI 1.044-1.065) and 1.02 (95% CI 1.011-1.034), respectively (Figure 3C). The difference of transmissibility between B.1.351 and B.1.54 was also significant (Student's t test, two-sided p<0.001) (Figure 3D). ~~In this case~~Consequently, isolates from B.1.351 had a slightly higher transmissibility than those from B.1.154.

P.2 had a slightly lower transmissibility than P.1 in Brazil

The composition of lineages in Brazil is shown in Figure 4A. Lineage B.1.1.33 and B.1.1.28 were the dominated before January 2021. Since October 2020, two novel lineages (P.1 and P.2) had gradually appeared and had shown exponential growth (Figure 4B). We could find the R_t for P.1 and P.2 during December 2020 to February 2021 were estimated as 1.07 (95% credible intervals 1.054-1.084) and 1.06 (95% credible intervals 1.049-1.070) (Figure 4C), respectively. The difference of transmissibility between P.1 and P.2 was also significant (Student's t test, two-sided p=0.016) (Figure 4D). ~~In this case~~Consequently, isolates from P.1 had a slightly higher transmissibility than those from P.2.

B.1.617.2 has a higher transmissibility than other dominant lineages in India

The top five dominant lineages and their corresponding proportion in India are shown in Figure 5A. Since July 2020, several other lineages, like B.1, B.1.36, B.1.36.29, emerged and grew exponentially in India (Figure 5B). ~~In this case~~Consequently, only these five lineages were used to estimate their R_t. The R_t was estimated as 1.013 (95% CI 1.006-1.021), 1.018 (95% CI 1.009 1.027), 1.019 (95% CI 1.010-1.027), 1.033 (95% CI 1.026-1.040), 1.123 (95% CI 1.106-1.140) for B.1, B.1.36, B.1.36.29, B.1.617.1, B.1.617.2, respectively (Figure 5C). After 5 independently repeated sampling and followed analysis for each lineage, we found that both B.1.617.1 and B.1.617.2 had significantly higher transmissibility than B.1, B.1.36, and B.1.36.29 (all Student's t test, two-sided with Holm–Bonferroni adjusted $p<0.001$) (Figure 5D). Furthermore, B.1.617.2 also had a significantly higher transmissibility than B.1.617.1 (Student's t test, two-sided with Holm–Bonferroni adjusted $p<0.001$). In addition, the transmissibility of both B.1.36, and B.1.36.29 is significantly higher than that of B.1 (both Student's t test, two-sided with Holm–Bonferroni adjusted $p<0.001$) (Figure 5D). However, similar transmissibility was found between B.1.36 and B.1.36.29 (Student's t test, two-sided with Holm–Bonferroni adjusted $p=0.057$) (Figure 5D).

Assessment of extent of onward transmission caused by partially vaccinated individuals infected with SARS-CoV-2 VOCs

We found a total of 14 direct transmission events. Four of them, concerning three types

of VOCs, were transmitted by vaccinated patients among three countries (Table 1). For convenience, we labelled patients involved in these four direct transmission pairs identified in this study. V1/V2 and V3/V4 from Belgium and Spain are considered to be transmitted by each other with a bidirectional probability for direct transmission of 0.99 and 0.85, respectively. However, we could not determine the direction of the transmission, as the probabilities of direct transmission from both directions were similar. We also found that these four patients had not been infected by others, as the bidirectional probability for direct transmission between them to others (except the patients who are considered to be their corresponding direct transmission pair) are all extremely low (Supplementary Figure 3). In the dataset of P.1, we also found two patient pairs with bidirectional probability for direct transmission as 0.76 and 0.65, respectively. Furthermore, the direction of transmission was more likely from patients receiving vaccines to those without receiving vaccines, as the probability of direct transmission from one direction (from patients receiving vaccines to those without receiving vaccines) were both >0.5 and significantly higher than that from the opposite direction. Next, we tested if the phylogenetic uncertainty affected the estimation of direct transmission events. We could find that the posterior probability of the branches containing V1/V2, V5/N1, and V6/N2 were 1, 0.99 and 0.99, indicating the extremely low phylogenetic uncertainty on these branches, further suggesting that the direct transmission events estimated based on these branches are highly reliable. However, the posterior probability of the branch containing V3 and V4 was only 0.33, suggesting V3 and V4 did not always clustered together. We could therefore only conclude that

we found definite evidence for three direct transmission events, being transmitted by patients receiving at least one dose of vaccines, with high probability. The probability of detecting onward transmission caused by patients being infected by SARS-CoV-2 VOCs after receiving at least one dose of vaccine was estimated to be 1.06% (3/284). We also calculated the probability of onward transmission caused by patients being infected by SARS-CoV-2 VOCs who had not received any vaccine in the same dataset. Ten direct transmission events were identified in the same datasets (Table 2). After checking the phylogenetic robustness of branch containing these patients, we found that the posterior probability of these branches all >0.9, indicating high phylogenetic robustness (Supplementary Figure 3). The direct transmission events identified on these branches were therefore robust. The probability of detecting transmission from patients infected by SARS-CoV-2 VOCs who had not received any vaccine was 1.21% (10/828). The probability after vaccination was therefore slightly lower, but not significantly different (Fisher exact test, $p>0.5$). This result suggested the vaccine has no obvious effect on suppressing the continued spread of VOC, and so it needs to be implemented in parallel with existing NPIs.

Discussion

Assessing the transmissibility of pathogens is essential to tailor prevention and control strategies. As the COVID-19 pandemic spread, several VOCs have been found. The emergence of these VOCs has caused a significant threat to public health. A previous study had documented that B.1.1.7, B.1.351, P.1, and B.1.617.2 have an increased

transmissibility of 29% (95%CI: 24-33), 25% (95%CI: 20-30), 38% (95%CI: 29-48), and 97% (95%CI: 76-117) compared to other lineages[5]. However, this conclusion was based on comparing non-VOC as a whole with VOC. For some dominant lineages, the number of cases added per day may be much higher than that of other lineages, but due to its large base, the number of cases from these dominant lineages will not increase exponentially. However, if these dominant lineages are grouped together with those lineages in which number of cases have increased exponentially, but the number of cases is not high, the advantages of transmissibility for those exponentially growing lineages will be overwhelmed. ~~In this case~~Consequently, in order to account for ~~not to ignore~~ th~~e~~ose exponentially growing lineages, it will be very important to list them separately as an assessment of their transmissibility.

Our results show that lineage B has a significantly higher transmissibility than lineage A (Figure 1A). Together with the fact that lineage B was the dominant types of SARS-CoV-2 all over the world, it seems that the high transmissibility of lineage B contributed to the global pandemic to a large extent. However, we also found that the transmissibility for lineage B from Australia and USA differed significantly. Considering the significantly different composition of sub-lineages among these two countries, we speculated that different sub-lineage within lineage B would have different transmissibility. We estimated the transmissibility of VOCs and the dominant lineages with exponential growth during same period in each country, so that the impact of non-pharmaceutical interventions on the estimation of R_t will be consistent among

different lineages. Our results also indicated although VOCs had advantage of transmissibility, there are still some lineages in each country with not much lower transmissibility. Theses lineages should also need to be taken seriously in the formulation of prevention and control policies.

Although vaccine manufacturers have been continuously producing vaccines, unequal distribution of vaccines will still cause many people to be unable to get vaccinated in the short term. In addition, even if there is an adequate supply of vaccines and vaccination is being gradually progressed, it takes a relatively long period to achieve complete vaccination in each country. It means that every country will have a certain period of time, during which many people received only one dose of the vaccine, leading to insufficient antibodies produced in their bodies. However, it was still unknown whether and to what extent people receiving at least one dose of vaccines could also transmit VOCs to others. We ~~found~~ estimated the probability of onward transmission caused by patients being infected by SARS-CoV-2 VOCs after receiving at least one dose of vaccine ~~would~~ to be ~~1.23~~1.06%. The similar probability of onward transmission caused by patients being infected by SARS-CoV-2 VOCs without receiving any vaccine indicated that only one dose of vaccine could not prevent individuals from infections of SARS-CoV-2 VOCs. However, the overall extent of onward transmission caused by patients being infected by SARS-CoV-2 VOCs after receiving at least one dose of vaccine could be underestimated. First, not all the viral genomic sequences and clinical information of patients are available. Second, the

criteria used in this study was very strict to reduce the false positive rate. Previous study using household contact data demonstrated that vaccination (most of individuals receiving one dose of vaccine) can reduce the probability of onward transmission by 50% (from 10% to 5%)[26]. However, they did not distinguish between VOCs and non-VOCs. Our results indicated that partially vaccination could not efficiently prevent the onward transmission of SARS-CoV-2 VOCs.

Although the extent of onward transmission caused by patients being infected by SARS-CoV-2 VOCs after receiving at least one dose of vaccine was low, the prevent and control measures should not be loosed intemperately for following reasons. First, the low extent of onward transmission was partially contributed to non-pharmacological interventions implemented in each country. If the prevent and control measures were abolished, the human contact frequency would be increased and then also increase the probability of SARS-CoV-2 infection and further onward transmission. Second, breakthrough infections have been identified in several countries[12, 27, 28], indicating the vaccines against SARS-CoV-2 could not be totally neutralized. The coexistence of SARS-CoV-2 and its antibodies in the human body and the continued spread of the virus among incompletely immunized individuals will make it easier to generate vaccine-escaped variants, which would thoroughly threaten the public health. Therefore, non-pharmaceutical interventions (such as ~~some low-cost and efficient strategies, like~~ wearing masks and social distancing ~~*etc*~~) should be implemented in each country before the vaccination is completed.

Key Points

- ~~Except~~ In addition to VOCs, lineages with exponential growth ~~rate~~ should also be paid attention in each country.
- One dose of vaccination could not efficiently prevent the onward transmission of SARS-CoV-2 VOCs
- Non-pharmaceutical interventions ~~(such as low-cost and efficient strategies, like wearing masks and social distancing *etc*)~~ should continue to ~~still~~ be implemented in each country during the vaccination period.

Biographical note

Liang Wang is an assistant professor at Institute of Microbiology, Chinese Academy of Sciences

Xavier Didelot is a professor at School of Life Sciences and Department of Statistics, University of Warwick

Yuhai Bi is a professor at Institute of Microbiology, Chinese Academy of Sciences

George F. Gao is a professor at Institute of Microbiology, Chinese Academy of Sciences

References

1. Jiang S, Shi Z, Shu Y et al. A distinct name is needed for the new coronavirus, Lancet 2020;395:949.

2. Davies NG, Abbott S, Barnard RC et al. Estimated transmissibility and impact of SARS-CoV-2 lineage B.1.1.7 in England, Science 2021;372:149-+.

3. Volz E, Mishra S, Chand M et al. Assessing transmissibility of SARS-CoV-2 lineage B.1.1.7 in England, Nature 2021;593:266-269.

4. Faria NR, Mellan TA, Whittaker C et al. Genomics and epidemiology of the P.1 SARS-CoV-2 lineage in Manaus, Brazil, Science 2021;372:815-821.

5. Campbell F, Archer B, Laurenson-Schafer H et al. Increased transmissibility and global spread of SARS-CoV-2 variants of concern as at June 2021, Euro Surveill 2021;26.

6. Supasa P, Zhou D, Dejnirattisai W et al. Reduced neutralization of SARS-CoV-2 B.1.1.7 variant by convalescent and vaccine sera, Cell 2021;184:2201-2211 e2207.

7. Abu-Raddad LJ, Chemaitelly H, Butt AA et al. Effectiveness of the BNT162b2 Covid-19 Vaccine against the B.1.1.7 and B.1.351 Variants, N Engl J Med 2021;385:187-189.

8. Hoffmann M, Arora P, Gross R et al. SARS-CoV-2 variants B.1.351 and P.1 escape from neutralizing antibodies, Cell 2021;184:2384-2393 e2312.

9. Zhou D, Dejnirattisai W, Supasa P et al. Evidence of escape of SARS-CoV-2 variant B.1.351 from natural and vaccine-induced sera, Cell 2021;184:2348-2361 e2346.

10. Liu C, Ginn HM, Dejnirattisai W et al. Reduced neutralization of SARS-CoV-2 B.1.617 by vaccine and convalescent serum, Cell 2021;184:4220-4236 e4213.

11. Wall EC, Wu M, Harvey R et al. Neutralising antibody activity against SARS-CoV-2 VOCs B.1.617.2 and B.1.351 by BNT162b2 vaccination, Lancet 2021;397:2331-2333.

12. Kustin T, Harel N, Finkel U et al. Evidence for increased breakthrough rates of SARS-CoV-2 variants of concern in BNT162b2-mRNA-vaccinated individuals, Nat Med 2021;27:1379-1384.

13. Li F, Li YY, Liu MJ et al. Household transmission of SARS-CoV-2 and risk factors for susceptibility and infectivity in Wuhan: a retrospective observational study, Lancet Infect Dis

2021;21:617-628.

14. Katoh K, Misawa K, Kuma K et al. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform, Nucleic Acids Res 2002;30:3059-3066.

15. Martin DP, Murrell B, Golden M et al. RDP4: Detection and analysis of recombination patterns in virus genomes, Virus Evol 2015;1:vev003.

16. Darriba D, Taboada GL, Doallo R et al. jModelTest 2: more models, new heuristics and parallel computing, Nat Methods 2012;9:772.

17. Elbe S, Buckland-Merrett G. Data, disease and diplomacy: GISAID's innovative contribution to global health, Glob Chall 2017;1:33-46.

18. Suchard MA, Lemey P, Baele G et al. Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10, Virus Evol 2018;4:vey016.

19. Baele G, Li WL, Drummond AJ et al. Accurate model selection of relaxed molecular clocks in bayesian phylogenetics, Mol Biol Evol 2013;30:239-243.

20. Rambaut A, Drummond AJ, Xie D et al. Posterior Summarization in Bayesian Phylogenetics Using Tracer 1.7, Syst Biol 2018;67:901-904.

21. Duchene S, Lemey P, Stadler T et al. Bayesian Evaluation of Temporal Signal in Measurably Evolving Populations, Mol Biol Evol 2020;37:3363-3379.

22. Didelot X, Fraser C, Gardy J et al. Genomic Infectious Disease Epidemiology in Partially Sampled and Ongoing Outbreaks, Mol Biol Evol 2017;34:997-1007.

23. Challen R, Brooks-Pollock E, Tsaneva-Atanasova K et al. Meta-analysis of the SARS-CoV-2 serial interval and the impact of parameter uncertainty on the COVID-19 reproduction number, medRxiv 2020:2020.2011.2017.20231548.

24. Jombart T, Cori A, Kamvar ZN et al. epitrix: small helpers and tricks for epidemics analysis. https://CRAN.R-project.org/package=epitrix.

25. Wang L, Didelot X, Yang J et al. Inference of person-to-person transmission of COVID-19 reveals hidden super-spreading events during the early outbreak phase, Nat Commun 2020;11:5006.

26. Harris RJ, Hall JA, Zaidi A et al. Effect of Vaccination on Household Transmission of SARS-CoV-2 in England, N Engl J Med 2021;385:759-760.

27. Brinkley-Rubinstein L, Peterson M, Martin R et al. Breakthrough SARS-CoV-2 Infections in Prison after Vaccination, N Engl J Med 2021;385:1051-1052.

28. Hacisuleyman E, Hale C, Saito Y et al. Vaccine Breakthrough Infections with SARS-CoV-2 Variants, N Engl J Med 2021;384:2212-2218.

Figure Legend

Figure 1. Difference in transmissibility between lineages A and B.

A. The distribution of R_t for each lineage. The black line in each distribution indicated the 95% CI.

B. The cumulative number of SARS-CoV-2 genomes for each lineage all over the world.

C. The heatmap of number of viral genomes for each sub-lineage in lineage A.

D. The heatmap of number of viral genomes for each sub-lineage in lineage B.

Figure 2. Difference in transmissibility for lineages in the UK. ~~Lineage B of SARS-CoV-2 has a higher transmissibility than lineage A~~.

A. The pie chart of SARS-CoV-2 lineage composition in the UK. The circle size was proportion to the number of SARS-CoV-2 genomes.

B. The cumulative number of SARS-CoV-2 genomes for each lineage in different region in the UK. The dash line indicated the earliest collection date of the data used for estimating the transmissibility for each lineage.

C. The distribution of R_t for each lineage. The black line in each distribution indicated the 95% CI.

D. The boxplot of repeated estimation of transmissibility by using 5 independent re-sampling data for each lineage. Upper bound, center, and lower bound of box represent the 75th percentile, the 50th percentile (median), and the 25th percentile, respectively.

Figure 3. Difference in transmissibility for lineages in South Africa.

A. The donut chart of SARS-CoV-2 lineage composition in South Africa.

B. The cumulative number of SARS-CoV-2 genomes for each lineage in South Africa. The dash line indicated the earliest collection date of the data used for estimating the transmissibility for each lineage.

C. The distribution of R_t for each lineage. The black line in each distribution indicated the 95% CI.

D. The boxplot of repeated estimation of transmissibility by using 5 independent re-sampling data for each lineage. Upper bound, center, and lower bound of box represent the 75th percentile, the 50th percentile (median), and the 25th percentile, respectively.

Figure 4. Difference in transmissibility for lineages in Brazil.

A. The donut chart of SARS-CoV-2 lineage composition in Brazil.

B. The cumulative number of SARS-CoV-2 genomes for each lineage in Brazil. The dash line indicated the earliest collection date of the data used for estimating the transmissibility for each lineage.

C. The distribution of R_t for each lineage. The black line in each distribution indicated the 95% CI.

D. The boxplot of repeated estimation of transmissibility by using 5 independent re-sampling data for each lineage. Upper bound, center, and lower bound of box represent the 75th percentile, the 50th percentile (median), and the 25th percentile, respectively.

Figure 5. Difference in transmissibility for lineages in India.

A. The donut chart of SARS-CoV-2 lineage composition in India.

B. The cumulative number of SARS-CoV-2 genomes for each lineage in India. The dash line indicated the earliest collection date of the data used for estimating the transmissibility for each lineage.

C. The distribution of R_t for each lineage. The black line in each distribution indicated the 95% CI.

D. The boxplot of repeated estimation of transmissibility by using 5 independent re-sampling data for each lineage. Upper bound, center, and lower bound of box represent the 75th percentile, the 50th percentile (median), and the 25th percentile, respectively.

E. **Figure 6**. Validation of direct transmission pairs. A. The bidirectional direct transmission probability of patients involved in direct transmission pairs and others (excluding their corresponding direct transmission patient). Upper bound, center, and lower bound of box represent the 75th percentile, the 50th percentile (median), and the 25th percentile, respectively. Whiskers represent 1.5× interquartile range and points are outliers. B. The number of intermediates between patients involved in direct transmission pairs and others (excluding their corresponding direct transmission patient).

Table 1. The statistics of direct transmission pairs (transmission from patients receiving at least one dose of vaccines to others) identified in our study.

VOCs	Country	Patient_1 ID	Patient_2 ID	Probability of Patient_1 transmit to Patient_2	Probability of Patient_2 transmit to Patient_1	Bidirectional probability for direct transmission
B.1.1.7	Belgium	V1	V2	0.42551111	0.56824444	0.99375555
B.1.1.7	Spain	V3	V4	0.47526851	0.37742592	0.85269444
P.1	Brazil	V5	N1	0.58384444	0.17171111	0.75555555
P.1	French Guiana	V6	N2	0.64104444	0.01077777	0.65182222

Table 2. The statistics of direct transmission pairs (transmission between patients who both did not receive vaccine) identified in our study.

VOCs	Country	Patient_1 ID	Patient_2 ID	Probability of Patient_1 transmit to	Probability of Patient_2 transmit to	Bidirectional probability for direct

				Patient_2	Patient_1	transmission
B.1.1.7	Belgium	N3	N4	0.587111	0.156022	0.743133
B.1.1.7	Estonia	N5	N6	0.306932	0.34694	0.653872
B.1.1.7	Italy	N7	N8	0.253991	0.28041	0.534402
B.1.1.7	Italy	N9	N10	0.219726	0.28953	0.509256
B.1.1.7	Spain	N11	N12	0.510713	0.423852	0.934565
B.1.1.7	Spain	N13	N14	0.470843	0.392519	0.863361
B.1.1.7	Spain	N15	N16	0.281417	0.254824	0.536241
B.1.1.7	USA	N17	N18	0.475644	0.039222	0.514867
B.1.351	Belgium	N19	N20	0.382056	0.137167	0.519222
P.1	Brazil	N21	N22	0.2872	0.255111	0.542311

566

567

Supplementary Information

Supplementary Figure 1. The division of subtrees for Delta dataset.

Supplementary Figure 2. The 95% CI distribution of R_t using MCC tree and ten randomly selected trees from the MCMC chains.

Supplementary Figure 3. Overview of the direct transmission events identified in our datasets. The MCC tree is showed for each dataset. Branches with a posterior probability >0.9 are shown by a purple circle. The size of the circle is proportional to the posterior probability. Branches of patients involved in direct transmission identified in this study were marked in red. Patients receiving at least one dose of vaccine were highlighted in green. A. Analysis of dataset of SARS-CoV-2 B.1.1.7 (Alpha) in Belgium; B. Analysis of dataset of SARS-CoV-2 B.1.1.7 (Alpha) in Spain; C. Analysis of dataset of SARS-CoV-2 B.1.1.7 (Alpha) in USA; D. Analysis of dataset of SARS-CoV-2 B.1.1.7 (Alpha) in other countries; E. Analysis of dataset of SARS-CoV-2 B.1.351 (Beta); F. Analysis of dataset of SARS-CoV-2 P.1 (Gamma); G. Analysis of dataset of SARS-CoV-2 B.1.617.2 (Delta).

Supplementary Table 1. List of 30 masked sites in SARS-CoV-2genome.

Supplementary Table 2. The best substitution model for dataset from each dataset.

Supplementary Table 3. The acknowledgement table of viral genomes used for estimating R_t.

Supplementary Table 4. The acknowledgement table of viral genomes used for evaluating the onward transmission caused by patients being infected with SARS-CoV-2 VOCs after receiving at least one dose of vaccine.

Supplementary Table 5. Log-marginal likelihood estimates from model selection by using the path-sampling (PS) and stepping-stone (SS) approaches for lineage A and B.

Supplementary Table 6. Bayesian evaluation for the temporal signal of dataset from each dataset.

Supplementary Table 7. The estimation of R_t and corresponding effective size of each dataset.

Supplementary Table 8. The parameters of offspring distribution estimated for different dataset.

Acknowledgements

We gratefully acknowledge the authors from the originating laboratories and the submitting laboratories where genetic sequence data were generated and shared via GISAID, enabling this research.

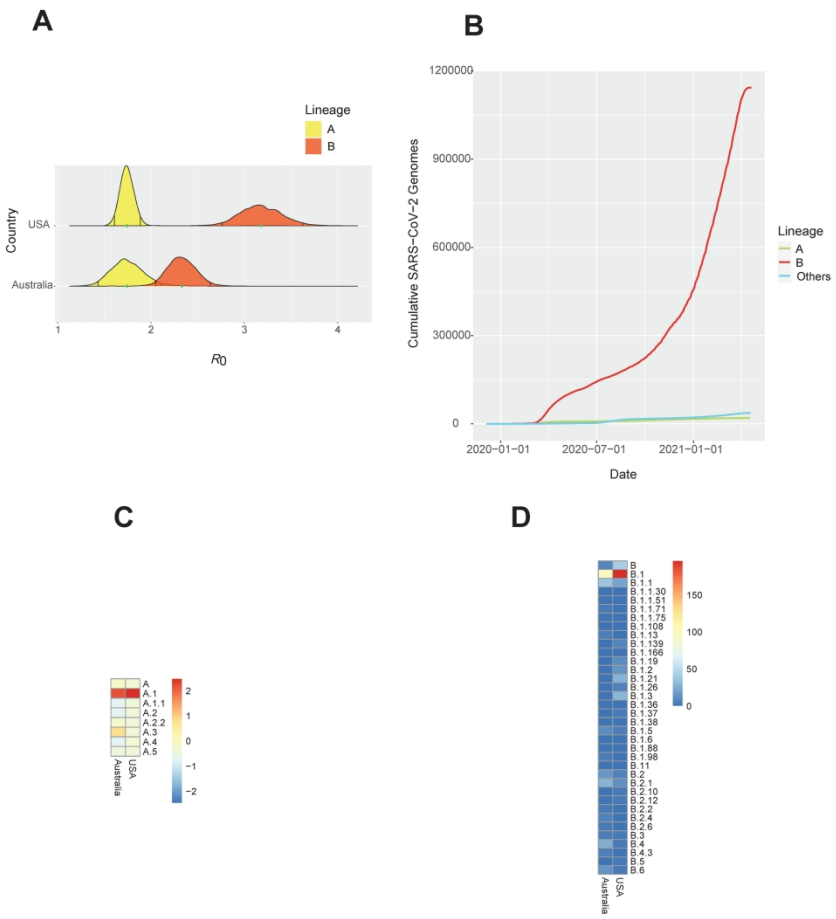
Funding

This work was supported by the Strategic Priority Research Programme of the Chinese Academy of Sciences (Grant numbers XDB29010102 and XDA19090118), National Natural Science Foundation of China (NSFC) (Grant numbers 32041010 and 31900155). Y.B. is supported by the NSFC Outstanding Young Scholars (Grant number 31822055) and Youth Innovation Promotion Association of CAS (Grant number 2017122). X.D. is supported by the NIHR Health Protection Research Unit in Genomics and Enabling Data.

Conflict of Interest

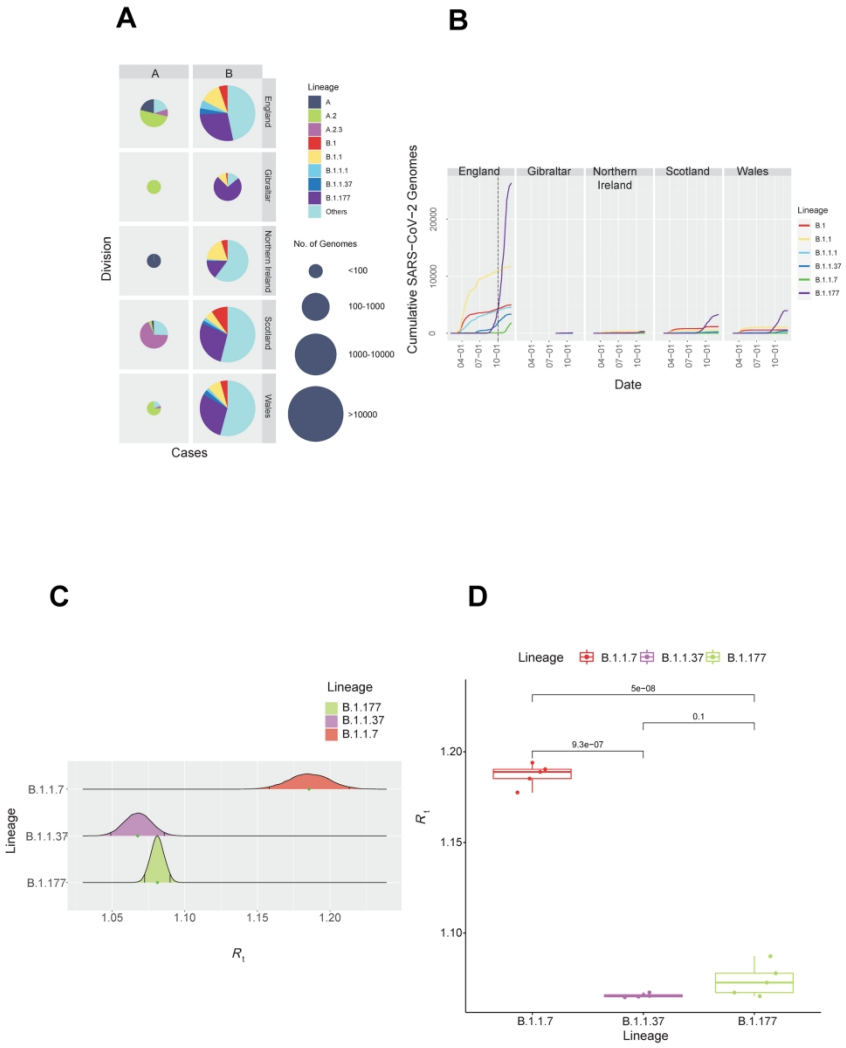
All the authors declared no conf lict of interests

Figure1



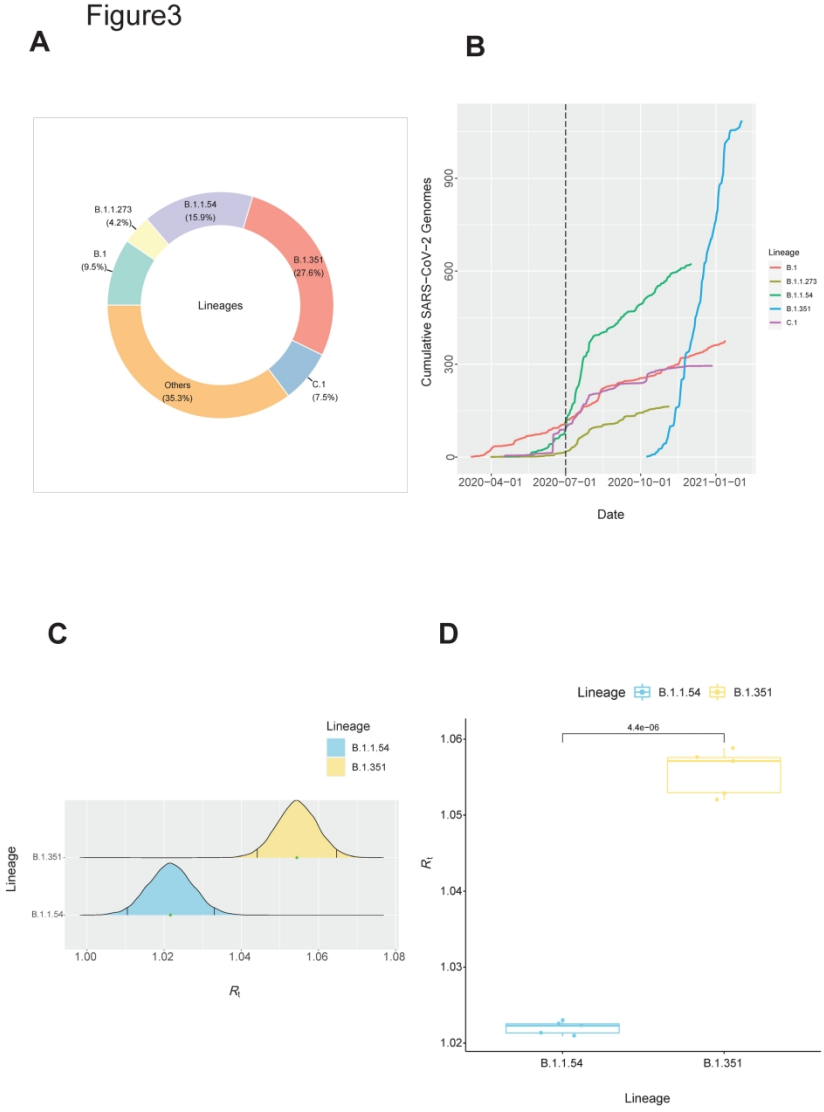
209x297mm (300 x 300 DPI)

Figure2



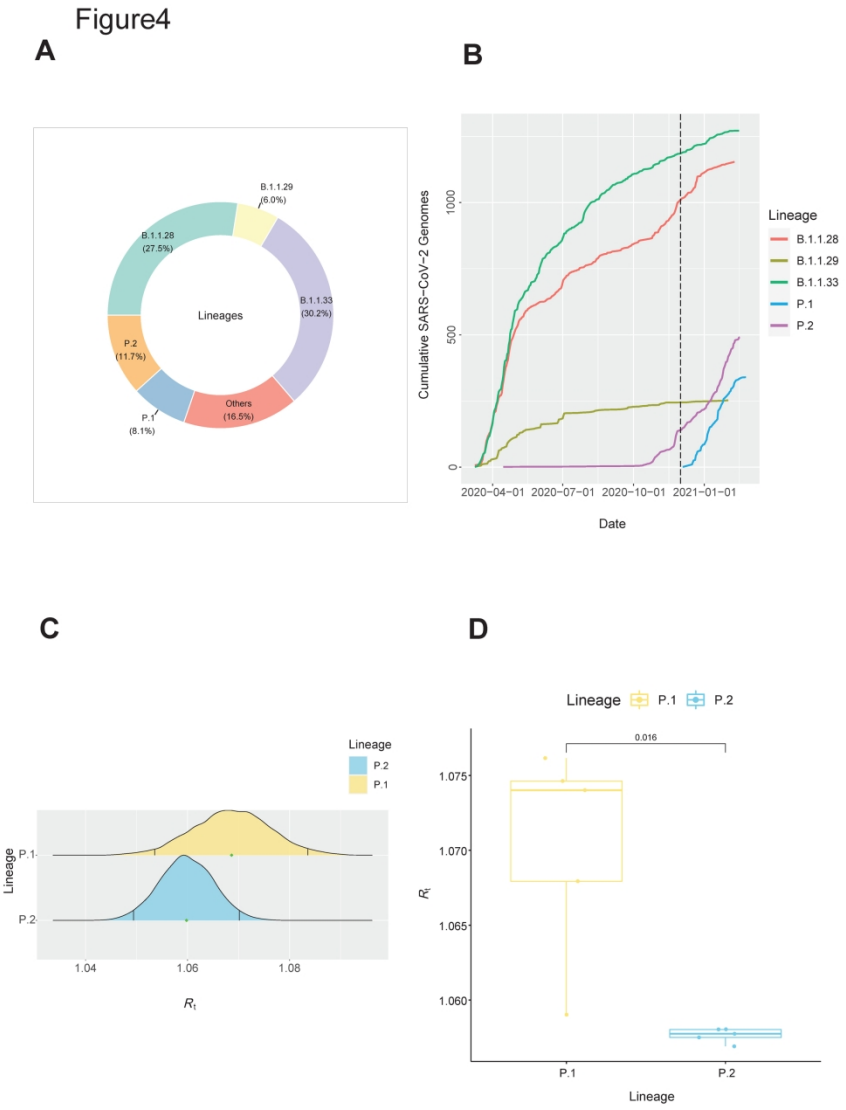
209x297mm (300 x 300 DPI)

Figure3



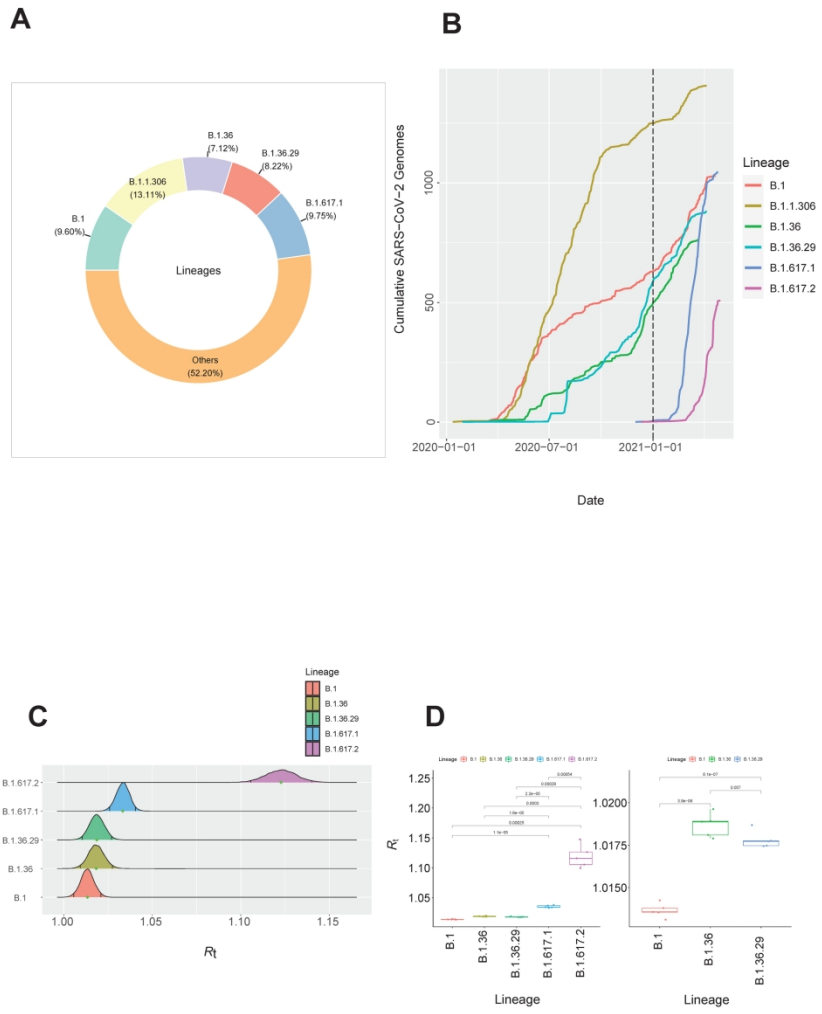
209x297mm (300 x 300 DPI)

Figure4



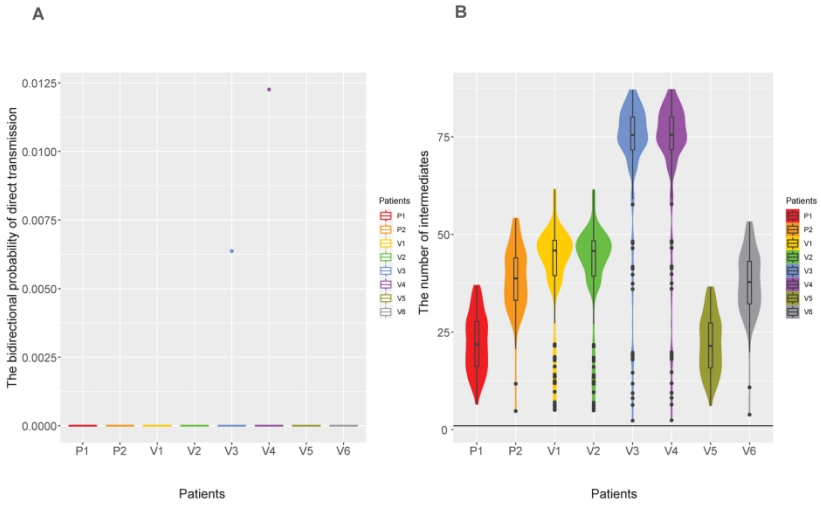
209x297mm (300 x 300 DPI)

Figure5

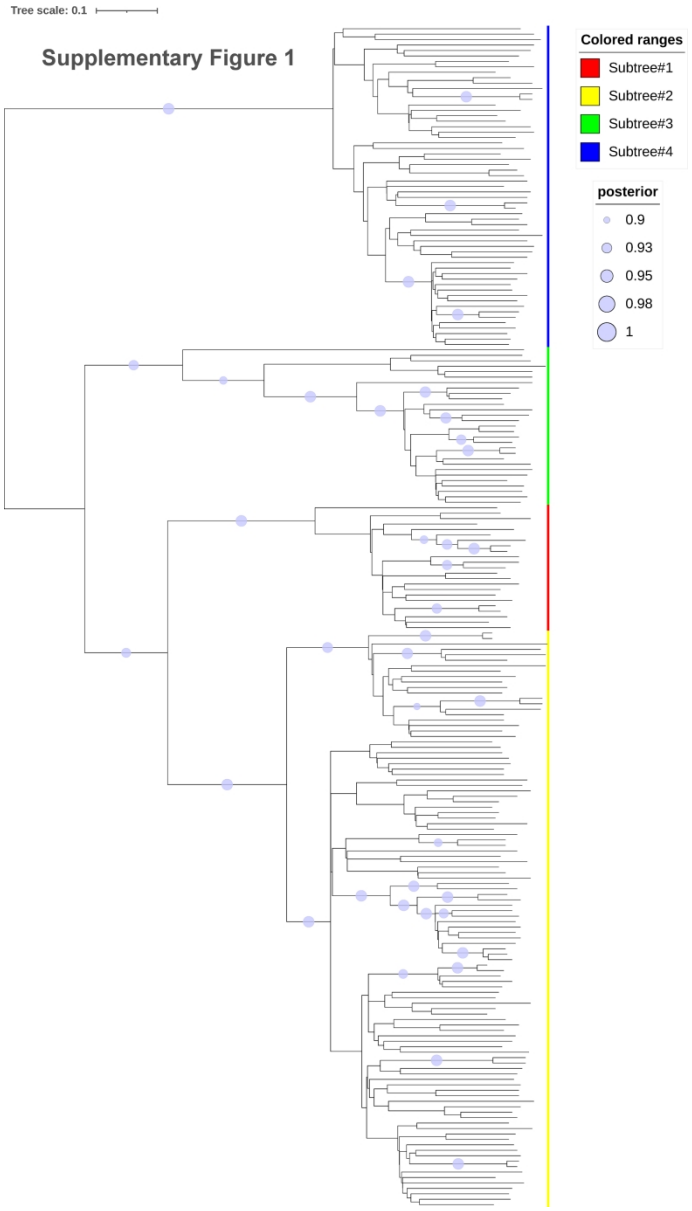


209x297mm (300 x 300 DPI)

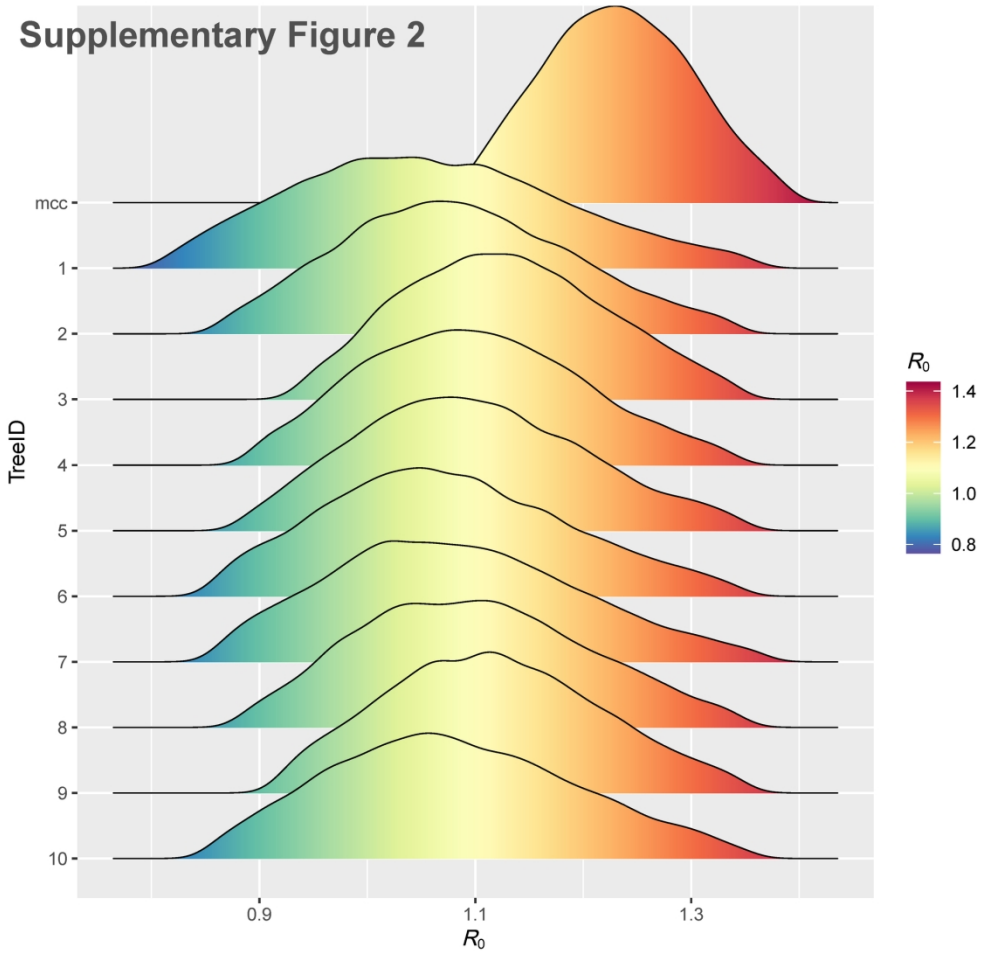
Figure 6



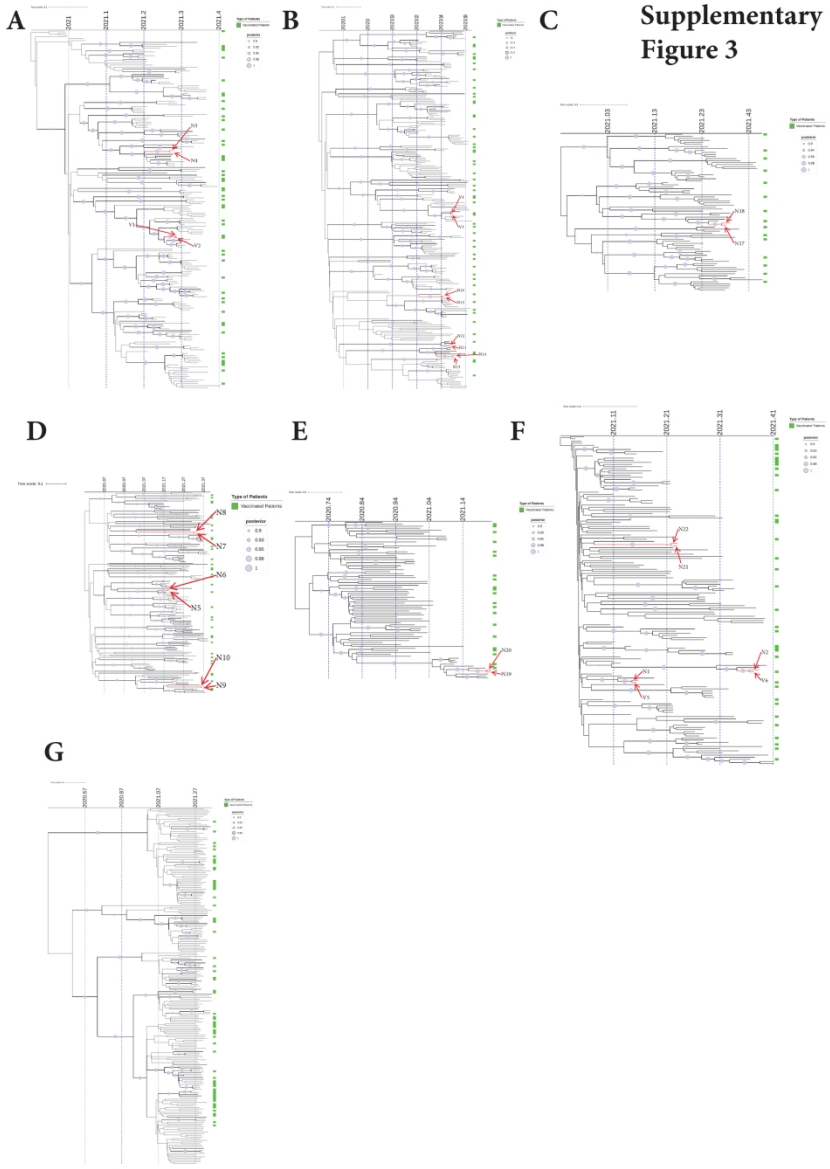
209x297mm (300 x 300 DPI)



194x340mm (300 x 300 DPI)



177x177mm (300 x 300 DPI)



209x297mm (300 x 300 DPI)