Research Paper

Association of SARS-CoV-2 clades with clinical, inflammatory and virologic outcomes: An observational study

Barnaby E Young[a,b,c], Wycliffe E Wei[d,e], Siew-Wai Fong[f,g,h], Tze-Minn Mak[a], Danielle E Anderson[i], Yi-Hao Chan[f,g], Rachael Pung[e,j], Cheryl SY Heng[e], Li Wei Ang[d], Adrian Kang Eng Zheng[i], Bernett Lee[g], Shirin Kalimuddin[i,k], Surinder Pada[l], Paul A Tambyah[m,n], Purnima Parthasarathy[o], Seow Yen Tan[p], Louisa Sun[q], Gavin JD Smith[i], Raymond Tzer Pin Lin[a], Yee-Sin Leo[a,b,c,n,r], Laurent Renia[f,g], Lin-Fa Wang[i], Lisa FP Ng[f,g], Sebastian Maurer-Stroh[s,*], David Chien Lye[a,b,c,n], Vernon J Lee[e,r,*]

[a] National Centre for Infectious Diseases, Singapore
[b] Deaprtment of Infectious Diseases, Tan Tock Seng Hospital, Singapore
[c] Lee Kong Chian School of Medicine, Nanyang Technological University, Singapore
[d] National Public Health and Epidemiology Unit, National Centre for Infectious Diseases, Singapore
[e] Singapore Ministry of Health, Singapore
[f] A*STAR ID Labs, Agency for Science, Technology and Research (A*STAR), Singapore
[g] Singapore Immunology Network, A*STAR, Singapore
[h] Department of Biological Science, National University of Singapore, Singapore
[i] Duke NUS Medical School, Singapore
[j] London School of Hygiene and Tropical Medicine, London, UK
[k] Singapore General Hospital of Singapore, Singapore
[l] Ng Teng Fong General Hospital, Singapore
[m] National University Health System, Singapore
[n] Yong Loo Lin School of Medicine, National University of Singapore, Singapore
[o] Khoo Teck Puat Hospital, Singapore
[p] Changi General Hospital, Singapore
[q] Alexandra Hospital, Singapore
[r] Saw Swee Hock School of Public Health, National University of Singapore, Singapore
[s] Bioinformatics Institute, Agency for Science, Technology and Research, Singapore

ARTICLE INFO

ABSTRACT

Background: Host determinants of severe coronavirus disease 2019 include advanced age, comorbidities and male sex. Virologic factors may also be important in determining clinical outcome and transmission rates, but limited patient-level data is available.

Methods: We conducted an observational cohort study at seven public hospitals in Singapore. Clinical and laboratory data were collected and compared between individuals infected with different SARS-CoV-2 clades. Firth's logistic regression was used to examine the association between SARS-CoV-2 clade and development of hypoxia, and quasi-Poisson regression to compare transmission rates. Plasma samples were tested for immune mediator levels and the kinetics of viral replication in cell culture were compared.

Findings: 319 patients with PCR-confirmed SARS-CoV-2 infection had clinical and virologic data available for analysis. 29 (9%) were infected with clade S, 90 (28%) with clade L/V, 96 (30%) with clade G (containing D614G variant), and 104 (33%) with other clades 'O' were assigned to lineage B.6. After adjusting for age and other covariates, infections with clade S (adjusted odds ratio (aOR) 0·030 (95% confidence intervals (CI): 0·0002−0·29)) or clade O (B·6) (aOR 0·26 (95% CI 0·064−0·93)) were associated with lower odds of developing hypoxia requiring supplemental oxygen compared with clade L/V. Patients infected with clade L/V had more pronounced systemic inflammation with higher concentrations of pro-inflammatory cytokines, chemokines and growth factors. No significant difference in the severity of clade G infections was observed (aOR 0·95 (95% CI: 0·35−2·52). Though viral loads were significantly higher, there was no evidence of increased transmissibility of clade G, and replicative fitness in cell culture was similar for all clades.

Interpretation: Infection with clades L/V was associated with increased severity and more systemic release of pro-inflammatory cytokines. Infection with clade G was not associated with changes in severity, and despite higher viral loads there was no evidence of increased transmissibility.

Research in context

Evidence before this study

The global initiative on sharing all influenza data (GISAID) nomenclature describes the emergence of four major SARS-CoV-2 clades: clade *L*, clade *V* (variant of the ORF3a coding protein NS3-G251), clade *G* (variant of the spike protein *S*-D614G with further subclades GR, GH and GV collectively called clade *G* here), and clade *S* (variant ORF8-L84S). *S* and *L* are the earliest known strains co-occurring in December 2019/January 2020 with Clade *L* being characteristic for the Wuhan outbreak. *G* and *V* clades are descendants of *L*, though *V* and *L* remain genetically closely related. "*O*" for "Other clades" is assigned for viruses which do not meet these clade definitions, in which case the more detailed lineage information (e.g. B.6) can be used. Similar to other classification schemes this nomenclature describes genetic differences, but does not attempt to reflect antigenic or other phenotypic properties of the virus. A literature search was done through Medline up to October 23, 2020 with the following keywords: Coronavirus disease 2019; COVID-19; SARS-CoV-2; clade; lineage; D614G. Viruses containing the genetic variant D614G (clade *G*) have become the most common reported worldwide since March 2020, and *in vitro* and animal models suggest this virus may be more infectious. Country-level data has also correlated the prevalence of clade G with mortality, however, little patient level data was available, and no study compared the outcomes of all four major clades.

Added value of this study

In this cohort study of 319 patients with PCR-confirmed SARS-CoV-2 infection, we found that compared with clade L/V, the odds of developing hypoxia requiring supplemental oxygen in patients infected with clade S or O (Lineage B.6) were significantly lower after adjusting for age group, sex, presence of comorbidities and other co-variates. Patients infected with clade-L/V also had more pronounced systemic inflammation with higher concentrations of pro-inflammatory cytokines, chemokines and growth factors. While there was no evidence of increased severity of clade-G infections, viral loads (as indicated by PCR cycle threshold values) were significantly higher with this clade. However, there was no evidence of increased transmissibility of this virus, and replicative fitness in cell culture was similar for all clades.

Implications of all the available evidence

Infection with clades *L/V* was associated with a more severe infection and more systemic release of pro-inflammatory cytokines compared with Clades *G, S* and *O* (B.6). No increased severity from infection with Clade-*G* viruses was observed, and while viral loads with this virus were significantly higher than other clades there was no evidence of increased infectivity. Further study of these variants may have implications for development of treatments and vaccines.

1. Introduction

The severity of coronavirus disease 2019 (COVID-19) ranges from asymptomatic infection to pneumonia, critical illness and death [1–3]. Major determinants of disease severity include host age, pre-existing comorbidities and sex, but virologic factors may also be important [4,5]. The characterisation of a severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2) variant with a 382-nucletide deletion in open reading frame 8 (ORF8) and its association with milder clinical illness is illustrative [6]. However, currently clinical comparison of the transmissibility and virulence of common SARS-CoV-2 variants is lacking. The recent emergence of several variants of concerns (VOC) in South Africa and the UK, and suspected implications for the effectiveness of vaccines and monoclonal antibody therapeutics highlights the importance of this type of analysis [7].

Several classification schemes have been introduced to describe the diversity of SARS-CoV-2 viruses [8]. The global initiative on sharing all influenza data (GISAID) nomenclature describes four major SARS-CoV-2 clades in the early outbreak: clade *L*, clade *V* (variant of the ORF3a coding protein NS3-G251), clade *G* (variant of the spike protein *S*-D614G with further subclades GR, GH and GV collectively called clade *G* here), and clade *S* (variant ORF8-L84S). *S* and *L* are the earliest known clades, co-occurring from December 2019 to January 2020, with Clade *L* being characteristic of the Wuhan outbreak. *G* and *V* clades are descendants of *L*, though *V* and *L* remain genetically closely related [9]. Nextstrain nomenclature offers a similar high level annotation, while SARS-CoV-2 lineages (cov-lineages.org) provide a more granular view with hundreds of named lineages incorporating epidemiological and sequence information corresponding to outbreaks in different regions [10]. The high-level GISAID clades described above cover 97·5% of all strains while the rest are labelled as "*O*" for "Other clades" in which case the more detailed lineage information (e.g. B.6) can be used.

The various SARS-CoV-2 nomenclatures are generally in agreement in their categorisation of genetic diversity but importantly, they do not attempt to reflect antigenic or other phenotypic properties of the virus. Viruses containing the genetic variant G614 (*G*-clade) have become the most common reported worldwide since March 2020, leading to the hypothesis that this variant is more infectious [11]. However, the current dominance of G614 viruses may reflect a founder effect, while country-level data correlating the prevalence of G614 with mortality requires a more detailed study [12–14]. Whether other clades are associated with phenotypic differences is unknown.

Singapore experienced a wide diversity of SARS-CoV-2 clade infections over the first three months of the pandemic, with cases imported from China, Europe, the United Sates and other major travel routes [15]. We aim to compare clinical outcomes and transmissibility following infection with different SARS-CoV-2 clades using patient-level data generated through comprehensive contact tracing, case finding and isolation measures in Singapore since the beginning of the local outbreak.

2. Methods

2.1. Study design and participants

Details of the first 10,000 COVID-19 cases in Singapore (from January 22 to April 22, 2020), by date of notification were extracted from the Ministry of Health database and matched with SARS-CoV-2

sequences shared on GISAID via laboratory accession numbers (see supplementary file for acknowledgment table) [16]. Detailed clinical information was obtained from all patients who were participants of a multi-centre prospective observational cohort study conducted at seven major public hospitals in Singapore (Appendix p1). Over the first three months of the COVID-19 pandemic this study attempted to recruit all individuals admitted with virologically-confirmed SARS-CoV-2 infection. Serial blood samples were collected from consenting participants during hospitalisation and post-discharge. Clinical data was obtained from a review of medical records. In addition, we obtained clinical data from a retrospective study of all patients with COVID-19 who were admitted to the National Centre for Infectious Diseases, the main infectious diseases management centre in Singapore.

2.2. Epidemiological investigation and testing

All COVID-19 cases were interviewed by the Singapore Ministry of Health to determine activity histories from 14 days preceding symptom onset until isolation in hospital, including recent travel history and possible contact with confirmed cases. Contact tracing was also initiated to identify contacts who had significant interactions with each case. Active case finding through testing of symptomatic contacts was performed to detect additional COVID-19 cases amongst these contacts and amongst travellers quarantined on arrival in Singapore. In addition, COVID-19 testing was mandated for persons meeting suspect case definitions, including persons who presented with pneumonia and influenza-like illnesses [17].

2.3. Estimating transmissibility by clade

The number of secondary cases that resulted from each case was derived from the epidemiological findings of all 10,000 cases. It was calculated through repeated probabilistic sampling (1000 repetitions) based on the assumption that in a cluster with 3 or more cases, cases with symptoms onset after an index case could have acquired the infection from the index case or from other secondary case(s). We determine the possible parent infectors for an offspring infectee after factoring for (a) the serial intervals between the infectee and all other possible infectors and (b) the overall serial interval distribution that we have observed amongst cases in Singapore [15]. Only symptomatic cases were considered.

For estimating transmissibility clades for cases that were not genotyped were imputed from those of epidemiologically linked cases that were gentoyped. Clusters with more than one clade implicated were excluded. The number of secondary cases amongst cases with assigned viral clade were analysed for associations between viral clade and disease transmissibility. Due to significantly different transmission patterns, cases in the dormitory setting where there are large outbreaks were excluded in the analysis of secondary transmission.

2.4. Clinical management

Over the course of this study, all patients with COVID-19 were isolated with airborne transmission precautions regardless of disease severity. Supplemental oxygen was administered for pulse oximeter saturations $\leq 94\%$. Patients with moderate to severe hypoxia (defined as requiring fraction of inspired oxygen [FiO_2] $\geq 40\%$) were transferred to the intensive care unit (ICU) for high-flow oxygen via nasal cannula and invasive mechanical ventilation as required. De-isolation was contingent on two consecutive nasopharyngeal swabs > 24 h apart that were negative for SARS-CoV-2 by polymerase chain reaction (PCR).

2.5. Correlation with clinical outcomes

Primary endpoint was the proportion of patients who developed severe COVID-19, defined as hypoxia requiring supplemental oxygen. This endpoint was selected due to the low incidence of ICU admission or death in Singapore. Secondary outcomes were the levels of immune mediators in plasma samples and the SARS-CoV-2 viral load as defined by the PCR cycle threshold (Ct) value from the first available nasopharyngeal swab. Other clinical outcomes are exploratory.

2.6. Viral RNA sequencing and clade determination

SARS-CoV-2 PCR was performed with a range of commercially available assays used at the participating hospitals. Sequencing of SARS-CoV-2 was performed in one of four laboratories in Singapore together with GISAID submission and case matching. Pangolin COVID-19 Lineage Assigner and CoVsurver were used to assign lineage and clade to each sequence, respectively [10,18]. For analysis, L and V clades were merged given their close genetic relatedness, similarly G, GH and GR were merged (Appendix p2; GV had not been described at this point in the pandemic). Sequences labelled as 'O' (other clades) were analysed based on assigned Pangolin lineage.

2.7. SARS-CoV-2 culture

Nasopharyngeal swabs were used to inoculate Vero-E6 cells for virus isolation in an Animal Biological Safety Level 3 (ABSL3) laboratory. Cells were cultured at 37 °C until cytopathic effect (CPE) was observed. Positive isolation was confirmed by the observation of CPE and virus-specific PCR. Following virus isolation, virus stocks were prepared in Vero-E6 cells.

For virus growth kinetics, Vero-E6 cells were seeded into 12-well plates at a density of 2×10^5 cells per well. Cells were infected with the respective viruses at a multiplicity of infection (MOI) of 0·01. Cell-associated virus was harvested at 24, 48, 72 and 96 h post-infection and titrated by limiting dilution on Vero-E6 cells and are expressed as the 50% tissue culture infectious dose ($TCID_{50}$) per millilitre.

2.8. Multiplex microbead-based immunoassay

Levels of specific immune mediators in the first plasma samples collected during hospitalisation were quantified by multiplex microbead-based immunoassays. Plasma samples were treated with 1% Triton™ X-100 for virus inactivation [19]. Immune mediator levels were measured with the Luminex™ assay using the Cytokine/Chemokine/Growth Factor 45-plex Human ProcartaPlex™ Panel 1 (ThermoFisher Scientific, Waltham, MA) (Appendix p3). Patient samples with concentrations out of measurement range were assigned the value of logarithmic transformation of Limit of Quantification (LOQ). Data analysis was conducted with Bio-Plex Manager™ 6.1.1 software (Bio-Rad, Hercules, CA).

2.9. Data processing and statistical analysis

Kruskal-Wallis test and Conover's post-hoc pairwise multiple comparison test (for continuous variables) and chi-square test (for categorical variables) were used to compare characteristics and outcomes between groups. Firth's logistic regression was used to examine the association between the SARS-CoV-2 clade and development of hypoxia due to the presence of separation with zero cell frequency of hypoxia for a particular clade. The following covariates were selected for inclusion in the multivariable logistic regression model: age group, sex, comorbidities, diagnosis through active case finding, month of COVID-19 diagnosis and clade/lineage. An multivariable ordinal regression model was also developed using the same

covariates and outcomes ordered as: 0 = mild (no pneumonia), 1 = pneumonia, no hypoxia, 2 = hypoxia requiring supplemental oxygen, 3 = critical illness and/or death. A quasi-Poisson regression was used to examine the association between SARS-CoV-2 clades and transmission rate, using the number of secondary cases resulting from a case as the dependant variable and the days between disease onset to isolation as an offset to the linear predictor. The latter is used as a surrogate for the period of transmissibility, the true period of which is generally longer due to pre-symptomatic transmission. Subsequent adjustment was performed for this model, for age (represented by three bands: < 45 years (referent), $45-64$ years and ≥ 65 years) and sex (male as referent).

One-way analysis of variance (ANOVA) followed by post-hoc *t*-test with Bonferroni correction was performed on the logarithmically transformed concentrations in order to discern the differences in immune mediator levels amongst patients with different SARS-CoV-2 clade infections. Heatmap and plots were generated using GraphPad Prism version 8 (San Diego, CA). In the heat map presentation, the concentrations of immune mediators were scaled between 0 and 1 for visualisation.

Data processing and analysis were done in the *R* statistical language (version 3.6.1) and Stata version 15 (StataCorp, College Station, TX). All statistical tests were two-sided, and *p*-values < 0.05 considered statistically significant. Adjustment for multiple testing was not performed.

2.10. Ethics statement

The epidemiological investigation was conducted under the Infectious Diseases Act (Singapore). Study protocols were approved by ethics committees of the National Healthcare Group and SingHealth. Written informed consent was obtained from participants for clinical data and biological sample collection as part of the PROTECT study (2012/00917; 2018/3045). A waiver of informed consent for retrospective data collection only was granted for individuals admitted to the National Centre of Infectious Diseases (2020/01122). Healthy donor samples were collected under study numbers 2017/2806 and NUS IRB 04-140.

2.11. Role of the funding source

3. Results

3.1. Descriptive epidemiology

The first confirmed COVID-19 case in Singapore was an imported infection notified on January 22, 2020, and the first locally transmitted cases were reported on February 3, 2020 (Fig. 1). In response to the outbreak travel restrictions were progressively introduced from February 1, 2020 and a heighted disease alert level was announced on February 7, 2020. A surge in imported cases occurred in March 2020 when residents returned from overseas, with entry-restrictions and 14-day quarantine for arrivals imposed on affected countries. The first foreign worker dormitory cluster was announced at the end of March 2020. This spread to involve numerous foreign worker dormitories, resulting in a steep increase in cases that peaked at more than 1000 cases per day by late April 2020. The increase in cases triggered social distancing restrictions on March 26, 2020, with broad restrictions imposed from April 7, 2020 similar to the lockdowns in other countries.

Of the first 10,000 COVID-19 cases reported in Singapore, 8132 (81·3%) were male foreign workers residing in dormitories. Overall, the proportion of elderly cases (age 65 and above) was 2·4% (240/10,000); decreasing from 11·0% (98/865) in January to March, to 1·5% (134/9033) in April 2020. 0·2% (23/10,000) died with COVID-19 listed as a direct cause. 540 of the first 10,000 cases had SARS-CoV-2
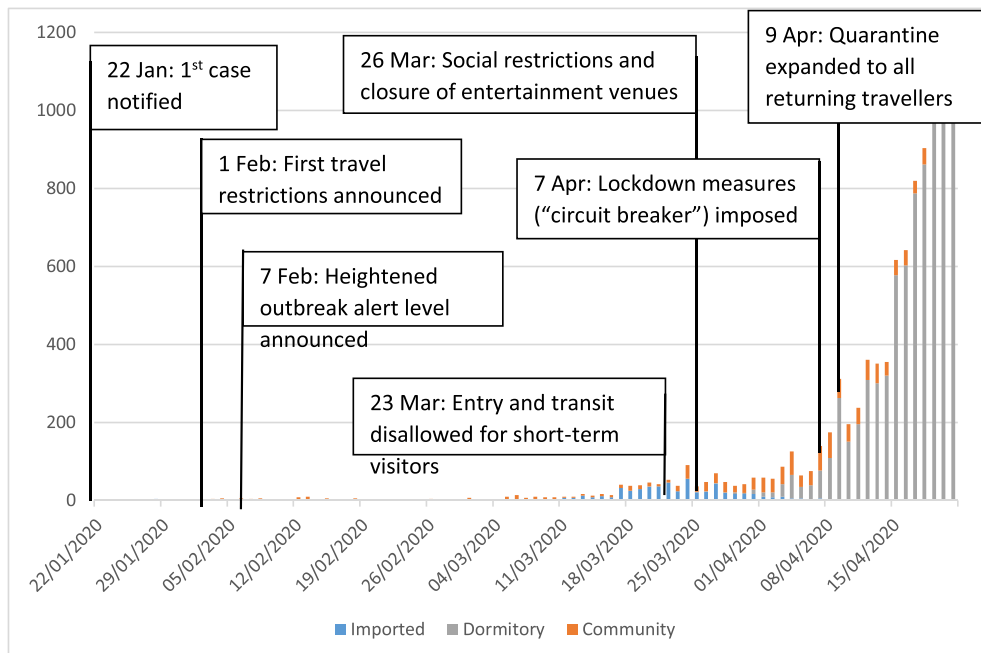

Fig. 1. Epidemiological curve of COVID-19 cases, Singapore, January 22 to April 21, 2020 (*n* = 9956).

genotyping information available through whole genome sequencing. The four major clades, *S, L, V* and *G* were all present in Singapore in January 2020 with repeated importation and local clusters of infection until April 2020. The large outbreak amongst foreign workers was almost exclusively due to *O* (B.6) [20].

3.2. Cohort study

320/540 (59%) patients with genotyped virus had detailed clinical data available for analysis (Appendix p1). 29 (9%) were clade *S*, of which 18 (62%) were known to be infected with the attenuated Δ382 variant (Table 1). 67 (21%) were infected with clade *L*, 23 (7%) with clade *V*, and 96 (30%) with clade *G* (of which 33 were GH and 41 GR). 104 (33%) infections with *O* label (other clades) were assigned to lineage B.6, while one infection with lineage B.4 was excluded from further analysis.

Overall, the cohort for analysis of 319 individuals included 47 (15%) with severe infection (hypoxia with requirement for supplemental oxygen) and 23 (7%) with critical illness (Table 1). 89% of the cohort was aged less than 65 years, and 82% had a Charlson's comorbidity index of 0. 47% of the cohort was diagnosed through active case finding, a proportion that did not significantly change from January to April (January/February, 26/49 (53%); March 82/189 (43%); April 41/82 (50%), $p = 0.95$, Chi square for trend). There was no significant change in the crude proportion of individuals with severe infection from January to April (January/February 11/49 (22%); March 25/189 (13%); April 11/82 (13%), $p = 0.25$, Chi square for trend).

Significant baseline differences between the groups infected with different clades were observed – including in age, ethnicity and comorbidities. After adjusting for co-variates in the multivariable logistic regression model, compared with *L/V* there was no significant difference in the severity of *G*-clade infection, while infections with clade S (adjusted odds ratio (aOR) 0.030 (95% confidence intervals (CI): 0.0002−0.29)) or clade *O* (B.6) (aOR 0.26 (95% CI 0.064−0.93)) were associated with lower odds of developing hypoxia requiring supplemental oxygen compared with infection (Table 2). There were 30 dormitory residents amongst the 104 individuals infected with *O* (B.6) in the cohort, none of whom had severe infection (dormitory: 0/30 (0%), non-dormitory: 12/74 (16%)). After excluding dormitories residents from the analysis, infections with *O* (B.6) remained significantly less severe compared with *L/V* (aOR 0.29 (95% CI 0.071−1.00), $p = 0.049$, Appendix p4).

Table 1.
Characteristics of 319 genotyped COVID-19 cases.

Clade	All	S	L/V	G	O (B.6)	p-value
N*	319	29	90	96	104	–
Month of diagnosis						< 0.0001
January/February[†]	49 (15%)	23 (80%)	24 (27%)	2 (2%)	0 (0%)	
March	188 (59%)	5 (17%)	65 (72%)	89 (93%)	29 (28%)	
April	82 (26%)	1 (3%)	1 (1%)	5 (5%)	75 (72%)	
Symptom onset to admission (days)	4 (2−7)	4 (2−5)	3.5 (2−7)	3 (2−5)	4 (3−8)	0.031
Diagnosis through active case finding[‡]	149 (47%)	15 (52%)	49 (54%)	34 (35%)	51 (49%)	0.055
Demographics						
Age (years)	42 (30−55.5)	39 (31−53)	48 (32−62)	32 (25−46)	43 (34−57)	< 0.0001
	185 (58%)	18 (62%)				
< 45	100 (31%)	10 (34%)	39 (43%)	69 (72%)	59 (57%)	0.0005
45−64	34 (11%)	1 (3%)	37 (41%)	25 (26%)	28 (27%)	
≥65			14 (16%)	2 (2%)	17 (16%)	
Male sex	198 (62%)	19 (66%)	50 (56%)	60 (63%)	69 (66%)	0.46
Ethnicity						
- Chinese	183 (57%)	19 (66%)	65 (72%)	51 (53%)	48 (46%)	< 0.0001
- Malay	19 (6%)	0 (0%)	8 (9%)	3 (3%)	8 (8%)	
- South Asian[§]	37 (12%)	1 (3%)	4 (4%)	9 (9%)	23 (22%)	
- Other	80 (25%)	9 (31%)	13 (14%)	33 (34%)	25 (24%)	
Charlson score (median, IQR)	0 (0−0)	0 (0−0)	0 (0−0)	0 (0−0)	0 (0−0.5)	0.060
- ≥ 1						
- Diabetes	57 (18%)	3 (10%)	17 (19%)	11 (11%)	26 (25%)	0.058
- Hypertension	22 (7%)	0 (0%)	10 (11%)	3 (3%)	9 (9%)	0.15
	53 (17%)	0 (0%)	22 (24%)	7 (7%)	24 (23%)	0.0003
Co-morbid conditions[‖]	78 (24%)	3 (10%)	27 (30%)	13 (14%)	35 (34%)	0.0014
Laboratory values						
LDH	383 (328−461)	369 (321−418)	405 (339−532)	370 (312−428)	383 (332−457)	0.032
CRP	5.3 (1.7−15.1)	5.0 (1.3−11.7)	7.6 (2.6−23.7)	5.0 (1.5−12.5)	4.4 (1.7−12.8)	0.083
Lymphocyte count	1.3 (0.9−1.8)	1.3 (1.0−1.6)	1.1 (0.9−1.7)	1.1 (0.9−1.4)	1.5 (1.1−2.1)	<0.0001
Neutrophil count	2.9 (2.1−4.0)	2.3 (2.0−3.3)	2.9 (2.1−3.7)	3.0 (2.1−3.9)	3.1 (2.2−4.3)	0.052
Ct Value	24.6 (21.1−29.6)	26.8 (23.4−31.5)	25.7 (22.0−30.4)	21.9 (19.1−26.4)	25.8 (23.7−30.7)	< 0.0001
Outcomes						
Pneumonia	122 (38%)	10 (34%)	45 (50%)	33 (34%)	34 (33%)	0.060
Supplemental oxygen	47 (15%)	0 (0%)	24 (27%)	11 (11%)	12 (12%)	0.0007
ICU/Death	23 (7%)	0 (0%)	14 (16%)	0 (0%)	9 (9%)	0.0002

Data is presented as median (inter-quartile range) or *n* (%) unless indicated.
CRP: *C*-reactive protein; ICU: intensive care unit; LDH: lactate dehydrogenase.
  * One case (lineage B.4) excluded.
  [†] 23 Jan to 29 Feb 2020.
  [‡] Including contact tracing, quarantine orders and surveillance.
  [§] India, Bangladesh, Sri Lanka.
  [‖] Categorical variable defined as Charlson Comorbidity Index ≥1 or hypertension. See Appendix p4 for breakdown of comorbid conditions by clade.

Table 2
Odds ratios of candidate predictors for (a) hypoxia requiring supplemental oxygen from Firth's logistic regression analysis, and (b) clinical outcome ordered from mild infection to critical illness and/or death from multivariable ordinal logistic regression†.

	Univariable LR model		Multivariable LR model		Proportional odds model†	
	OR (95% CI)	p-value	aOR (95% CI)	p-value	aOR (95% CI)	p-value
Age Group (Years)						
< 45	1·00 (reference)	-	1·00 (reference)	-	1·00 (reference)	-
45−64	6·25 (2·83−15·1)	< 0·0001	4·55 (1·93−11·57)	0·0004	4·25 (2·44−7·50)	< 0·0001
≥65	18·6 (7·36−50·6)	< 0·0001	7·80 (2·45−26·6)	0·0005	9·82 (3·95−25·00)	< 0·0001
Female sex	0·84 (0·43−1·57)	0·58	0·70 (0·33−1·47)	0·35	0·85 (0·50−1·44)	0·55
Co-morbid conditions‡	7·06 (3·69−13·8)	< 0·0001	4·57 (1·94−11·1)	0·00053	3·48 (1·84−6·59)	0·00013
Clade						
L/V	1·00 (reference)	-	1·00 (reference)	-	1·00 (reference)	-
S	0·046 (0·004−0·35)	0·00048	0·030 (0·0002−0·29)	0·00059	0·30 (0·10−0·81)	0·021
G	0·37 (0·16−0·77)	0·0082	0·95 (0·35−2·52)	0·91	1·01 (0·51−2·04)	0·97
O (B.6)	0·37 (0·17−0·76)	0·0072	0·26 (0·064−0·93)	0·037	0·36 (0·13−0·92)	0·037
Diagnosis through active case finding	0·68 (0·36−1·26)	0·22	0·89 (0·42−1·90)	0·77	0·83 (0·49−1·40)	0·48
Month of diagnosis						
January/February	1·00 (reference)	-	1·00 (reference)	-	1·00 (reference)	-
March	0·66 (0·29−1·63)	0·36	0·21 (0·061−0·70)	0·012	0·21 (0·088−0·52)	0·00070
April	0·63 (0·24−1·73)	0·36	0·42 (0·07−2·43)	0·32	0·34 (0·10−1·40)	0·086

aOR: adjusted odds ratio; CI: confidence interval; LR: Logistic regression; OR: odds ratio.

† Outcomes ordered as: 0=mild (no pneumonia), 1=pneumonia, no hypoxia, 2=hypoxia requiring supplemental oxygen, 3=critical illness and/or death. Proportional odds assumption examined with Brant test: parallel regression assumptions holds for individual variables and overall model ($p = 0.92$).

‡ Categorical variable defined as Charlson Comorbidity Index ≥ 1 or hypertension.

3.3. Comparison of early immune response

Acute plasma samples at the first collection timepoint upon hospital admission were available from 99 patients (Fig. 2a). No significant difference in the sample collection time point was observed between the different clades (Appendix p6). Collectively, higher concentrations of chemokines MCP-1, RANTES, SDF-1α, and IP-10, pro-inflammatory IFN-γ, IL-6 and IL-1β and the anti-inflammatory IL-1RA were detected in patients infected with clades L and V compared with patients infected with either clade S or G (Fig. 2b) at a median of 5 days post-illness onset (interquartile range 3−8). In addition, these patients had higher levels of growth factors associated with lung injury and regeneration, including SCF, VEGF-A, PlGF-1 and PDGF-BB. Notably, patients infected with clade G SARS-CoV-2 had lower concentrations of growth factors including VEGF-A, HGF, PDGF-BB and PlGF-1, pro-inflammatory IL-1β, IFN-γ and chemokines such as RANTES, SDF-1α and IP-10, at the early acute phase of infection (Fig. 2B).
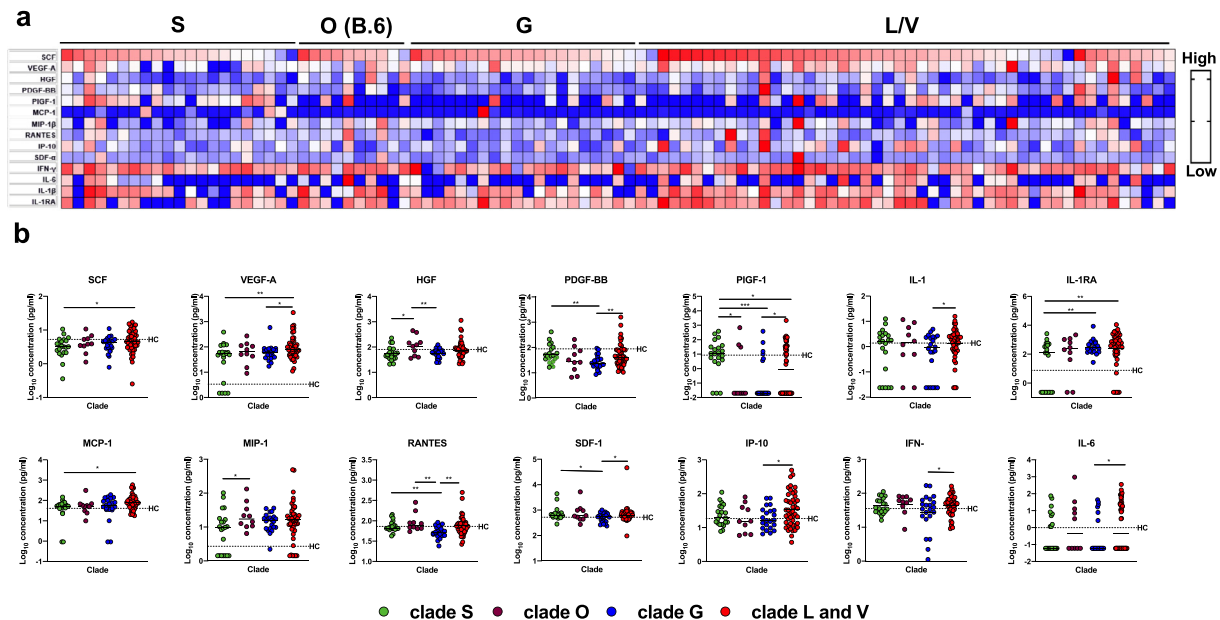
Fig. 2. *Plasma immune mediator levels of COVID-19 patients infected with different SARS-CoV-2 clades.* Concentrations of 45 immune mediators were quantified using a 45-plex microbead-based immunoassay. (a) Heatmap of immune mediator levels in plasma samples of patients infected with different SARS-CoV-2 clades (clade S, n = 21; clade O (B.6), n = 10; clade G, n = 22; clades L and V, n = 46) during first collection timepoint upon hospital admission (median 5 days from symptom onset). Each colour represents the relative concentration of a particular analyte. Blue and red indicates low and high concentration, respectively. (b) Profiles of significant immune mediators of COVID-19 patients infected with different SARS-CoV-2 clades at the first collection timepoint upon hospital admission. Immune mediator levels in plasma fraction samples from first collection timepoint during hospital admission (median 5 days from symptom onset) were compared amongst the patients infected with different SARS-CoV-2 clades. One-way ANOVA followed by post-hoc *t*-test with Bonferroni correction was performed on the logarithmically transformed concentration (*$p < 0\bullet05$; **$p < 0\bullet01$; **$p < 0\bullet0001$). Immune mediator levels for healthy controls (n = 23) are indicated by the black dotted line. Patient samples with concentration out of measurement range are presented as the value of logarithm transformation of Limit of Quantification (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.).

Table 3.

Relative risk of onward transmission (i.e. any secondary transmission) by clade amongst non-dormitory cases with genotyped and inferred clades ($n = 587$).

	Crude RR (95%CI)	p-value	Adjusted RR (95%CI)	p-value
Clade				
L/V	1·00 (reference)	-	1·00 (reference)	-
S	1·28 (0·94−1·73)	0·12	1·28 (0·94−1·75)	0·12
G	0·35 (0·23−0·56)	< 0·001	0·36 (0·23−0·58)	< 0·001
O (B.6)	0·78 (0·61−1)	0·048	0·80 (0·62−1·02)	0·07
Age Group (Years)				
< 45	1·00 (reference)	-	1·00 (reference)	-
45−64	1·17 (0·93−1·47)	0·19	1·11 (0·89−1·39)	0·34
≥ 65	1·07 (0·76−1·5)	0·70	1·02 (0·73−1·44)	0·91
Female sex	1·24 (1−1·53)	0·048	1·19 (0·97−1·46)	0·10

Note: Modified Poisson regression with binary outcome of transmission occurrence as regressand, amongst those who were exposed to the community after symptom onset. CI: confidence interval; RR: relative risk.

3.4. Transmissibility

In addition to the 540 cases with SARS-CoV-2 genotyping information available, a further 5626 cases had virus genotype inferred by epidemiologically linkages. The majority (5460/6166, 88·6%) were dormitory residents and were excluded from the analysis. amongst 706 cases included in the analysis (Appendix p7−8), the average number of secondary transmissions was 0·34, with lowest in clade G (0·11) (Appendix p9). After adjusting for age and gender, the transmission rate was also lowest in *G*, with an adjusted transmission rate ratio of 0·40 (95%CI: 0·23−0·71, p = 0·002) compared to clades *L/V*. The difference in the risk of transmission between clades *L/V* and the other clades were not statistically significant (Table 3).

3.5. Viral culture

There was no significant difference between the viral loads of Clade *L/V, S* and *O* (B.6), whereas Clade *G* was associated with significantly higher viral loads compared with the other three clades (Conover post hoc pairwise comparison, p < 0.05). Day of sample collection was not significantly different (Kruskal−Walis, p = 0.64). The kinetics of viral replication in cell culture were similar between the different clades as measured up to 96 h after inoculation (Fig. 3).

4. Discussion

In this large cohort study we found no evidence for attenuated or increased severity of clade *G* when comparing infections with the *L/V* clades. However, clade *G* was associated with significantly higher viral loads (as measured by PCR Ct value). Clade *S* was associated with less severe infection, likely reflecting infection by the ORF7/8 deletion variant that we previously reported to be associated with attenuated severity [6]. *O* (B.6) was associated with attenuated severity of infection for reasons that are unclear and merits further investigation.

G-clade viruses emerged in late January 2020 and became the most common clade worldwide (based on samples submitted to GISAID) in March 2020 [21]. To date there is little patient level data comparing outcomes from *G*-clade viruses with other variants. A study of 999 patients in Sheffield, UK, reported that infection with G614 was not associated with an increased risk of hospital admission, however, it was associated with higher viral load [22]. A study of 113 infections in Hong Kong reported no significant increase in the severity of infection [23]. We found that clade *G* is not associated with different disease severity compared with *L/V*.

The significantly lower PCR Ct value with G-clade infection observed in our cohort supports other data that this virus may be more infectious [24]. *In vitro* experiments indicate that the G614 variant opens the structure of the spike protein, facilitating binding to the ACE-2 receptor and that this virus is more infectious in tissue culture [21,25]. While previous studies have used a pseudovirus, we observed no difference in viral kinetics between *G* and other clades. In a hamster model, the SARS-CoV-2 virus with the G614 virus
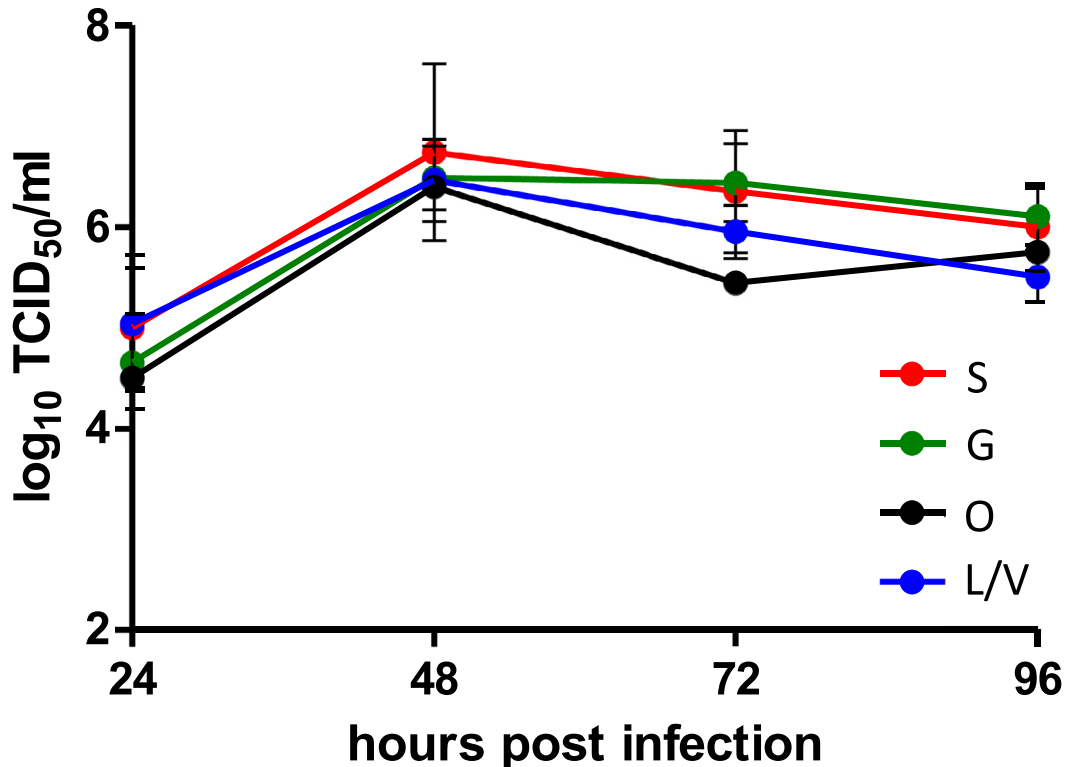
Fig. 3. Kinetics of viral replication of 16 SARS-CoV-2 virus isolates up to 96 h post-infection. *S* 3 isolates; *L/V* 7; *G*, 5; *O* (B.6). TCID: Tissue culture infectious dose.

produced higher infectious titres in nasal washes and trachea, but not lungs, than the D614 virus [26].

At the same time, we observed significantly lower transmission from individuals infected with *G*-clade viruses compared with other clades. This is likely to be due to significantly earlier isolation of this group, and may reflect either improvements in community diagnosis or behavioural differences in then affected population. For all clades the reproduction number, *R*, was significantly below 1, and overall 0·34 reflecting the success of Singapore's COVID-19 public health measures at reducing transmission in the community and indicating that these interventions are effective regardless of clade [15]. While an underestimate because only symptomatic infections are considered, it is still low compared to other *R* values estimated from symptomatic infections [27]. Higher SARS-CoV-2 viral load has been correlated with COVID-19 mortality, suggesting host immune response to G-clade infections may differ and be less pathogenic – either as a direct result of D614G or other genetic variation associated with this clade [28,29].

We found patients infected with clade *L/V* had more pronounced systemic inflammation with higher concentrations of pro-inflammatory cytokines, chemokines and growth factors that are strongly associated with severe COVID-19 during the early acute phase of infection [30,31]. Excessive production of pro-inflammatory cytokines in SARS-CoV-2 infection damages endothelial cells resulting in capillary leakage and vascular injury [32]. The more robust production of pro-inflammatory IL-1β, IL-6, MCP-1, RANTES and MIP-1 β at the early phase of infection observed in patients infected with clade *L/V* could promote lung damage, which may in turn result in severe clinical outcomes, as shown by the higher proportion of patients in this group requiring oxygen. Intriguingly, those with clade *G* infections had a less overactive inflammatory response [22]. In addition, patients with clade *S* infections had lower levels of inflammation compared with those infected with *L/V*, and this may reflect the known association between the ORF8 deletion variant and inflammatory responses [6].

This observational study has a number of limitations. We only included data up to April 2020, a period which saw the greatest diversity of SARS-CoV-2 clades in Singapore. Viral adaptation to infection in humans may have occurred since then which affects COVID-19 severity. We also merged several closely genetically related clades (*L/V* and *G*/GH/GR) to improve study power which prevents us from analysing the effects of these individual clades. SARS-CoV-2 testing during the study period was largely symptom driven and it is possible that asymptomatic or pauci-symptomatic infection rates vary with different clades. As the pandemic progressed beyond March 2020 the cohort study was unable to enrol all individuals with COVID-19 due to the large numbers. Recruitment was biased towards more severe infections, suggesting that for *G* and *O* (B.6) clades in particular severity is likely to have been overestimated. Clinical management evolved over the first few months of the pandemic which may have improved outcomes. However, it remained largely supportive: remdesivir was yet to be licensed, corticosteroids were avoided, while convalescent plasma and monoclonal antibodies were not available [33,34]. SARS-CoV-2 PCR was performed on different commercial platforms, with varied gene targets, limiting the comparability of Ct values. Also, clade distribution in the population is associated with certain subpopulations and clusters. As such, associations of severity with clades may reflect underlying human behaviour or characteristics in the subpopulations, rather than the true virulence or infectiousness of the virus. These differences are difficult to measure and challenging to control for in an observational study.

SARS-CoV-2 genotypes imputed through epidemiological linkage data risks misclassification. However, the community disease incidence in Singapore was low (e.g. 0.7 per 100,000 persons per day in April 2020) in a setting of aggressive contact tracing. As a result it was unlikely that infection within a cluster arose from more than one source of infection providing high confidence in the relationships between cases. Furthermore, clusters with more than one genotyped clade were excluded. While there may be uncertainties in determining the exact direction of transmission within a cluster, a probabilistic approach to transmission was undertaken. A limitation is that only symptomatic cases were considered in assessing transmissibility, as this allows greater confidence in the direction of transmission. The results may not necessarily reflect the transmissibility of asymptomatic disease unless the clade distribution of asymptomatic cases is similar to that of symptomatic disease.

In conclusion, as a newly emerged human pathogen, SARS-CoV-2 is likely to be under strong selection pressure to adapt to infection in humans. The recent emergence of new variants which may also be able to evade host immunity is concerning, though claims of increased transmissibility or changes in virulence require careful evaluation [35]. In our large cohort study, with detailed epidemiological, clinical and immunological data we find patient-level evidence that infections by *G*-clade viruses are associated with higher viral titres, without necessarily affecting disease severity or increasing transmissibility. The association of *O* (B.6) infections with attenuated severity and stronger immune response as compared to early *L/V* vs later *G*-clades is intriguing but requires further study.

Declaration of Competing Interest

BY has received honoraria from Sanofi and Roche, outside the submitted work. All other authors no conflicts of interest declared.

Contributors

Acknowledgments

Data sharing

Qualified researchers may request access to anonymized individual patient level data by sending a request to the corresponding author. Data will be shared after approval of a proposal, with a signed data access agreement.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.ebiom.2021.103319.

References

[1] Guan WJ, Ni ZY, Hu Y, Liang WH, Ou CQ, He JX, et al. Clinical characteristics of coronavirus disease 2019 in China. N Engl J Med. 2020;382(18):1708–20 30.

[2] Long QX, Tang XJ, Shi QL, Li Q, Deng HJ, Yuan J, et al. Clinical and immunological assessment of asymptomatic SARS-CoV-2 infections. Nat Med 2020;26(8):1200–4.

[3] Young BE, Ong SWX, Ng LFP, Anderson DE, Chia WN, Chia PY, et al. Viral dynamics and immune correlates of COVID-19 disease severity. Clin Infect Dis Off Publ Infect Dis Soc Am 2020 Aug 28ciaa1280Online ahead of print.

[4] Zhang JJY, Lee KS, Ang LW, Leo YS, Young BE. Risk factors of severe disease and efficacy of treatment in patients infected with COVID-19: a systematic review, meta-analysis and meta-regression analysis. Clin Infect Dis Off Publ Infect Dis Soc Am 2020;71(16):2199–206.

[5] Zhang X, Tan Y, Ling Y, Lu G, Liu F, Yi Z, et al. Viral and host factors related to the clinical outcome of COVID-19. Nature 2020;583(7816):437–40.

[6] Young BE, Fong SW, Chan YH, Mak TM, Ang LW, Anderson DE, et al. Effects of a major deletion in the SARS-CoV-2 genome on the severity of infection and the inflammatory response: an observational cohort study. Lancet 2020;396(10251):603–11 Lond Engl29.

[7] Lauring AS, Hodcroft EB. Genetic variants of SARS-CoV-2-what do they mean? JAMA 2021 Jan 6Online ahead of print.

[8] Alm E, Broberg EK, Connor T, Hodcroft EB, Komissarov AB, Maurer-Stroh S, et al. Geographical and temporal distribution of SARS-CoV-2 clades in the WHO European Region, January to June 2020. Euro Surveill Bull Eur Sur Mal Transm Eur Commun Dis Bull 2020;25(32):2001410. Aug.

[9] Mercatelli D, Giorgi FM. Geographic and genomic distribution of SARS-CoV-2 mutations. Front Microbiol 2020;11:1180. [Internet]Jul 22.

[10] Rambaut A, Holmes EC, Á O'Toole, Hill V, McCrone JT, Ruis C, et al. A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. Nat Microbiol 2020;5(11):1403–7.

[11] Koyama T, Platt D, Parida L. Variant analysis of SARS-CoV-2 genomes. Bull World Health Organ 2020;98(7):495–504 Jul 1.

[12] Toyoshima Y, Nemoto K, Matsumoto S, Nakamura Y, Kiyotani K. SARS-CoV-2 genomic variations associated with mortality rate of COVID-19. J Hum Genet 2020;65(12):1075–82.

[13] Becerra-Flores M, Cardozo T. SARS-CoV-2 viral spike G614 mutation exhibits higher case fatality rate. Int J Clin Pract 2020;74(8):e13525.

[14] Eaaswarkhanth M, Al Madhoun A, Al-Mulla F. Could the D614G substitution in the SARS-CoV-2 spike (S) protein be associated with higher COVID-19 mortality? Int J Infect Dis IJID Off Publ Int Soc Infect Dis 2020;96:459–60 Jul.

[15] Pung R, Cook AR, Chiew CJ, Clapham HE, Sun Y, Li Z, et al. Effectiveness of containment measures against COVID-19 in Singapore: implications for other national containment efforts. Epidemiol Camb Mass 2020;32(1):79–86.

[16] Shu Y, McCauley JGISAID. Global initiative on sharing all influenza data - from vision to reality. Euro Surveill Bull Eur Sur Mal Transm Eur Commun Dis Bull 2017;22(13) 30.

[17] Lee VJ, Chiew CJ, Khong WX. Interrupting transmission of COVID-19: lessons from containment efforts in Singapore. J Travel Med 2020;27(3):taaa039. May 18.

[18] GISAID - CoV surver mutations app [Internet]. [cited 2021 Feb 8]. Available from: https://www.gisaid.org/epiflu-applications/covsurver-mutations-app/

[19] Darnell MER, Taylor DR. Evaluation of inactivation methods for severe acute respiratory syndrome coronavirus in noncellular blood products. Transfusion 2006;46(10):1770–7 (Paris)Oct.

[20] COVID-19 Situation Report [Internet]. [cited 2021 Feb 8]. Available from: https://covidsitrep.moh.gov.sg/

[21] L Y, X W, Ke P, C TT, Tp N, Y W, et al. Structural and functional analysis of the D614G SARS-CoV-2 spike protein variant [Internet] Cell 2020;183(3):739–51 e8.

[22] Korber B, Fischer WM, Gnanakaran S, Yoon H, Theiler J, Abfalterer W, et al. Tracking changes in SARS-CoV-2 Spike: evidence that D614G increases infectivity of the COVID-19 virus. Cell 2020;182(4):812–27 20e19.

[23] Mak GCK, Lau AWL, Chan AMY, Chan DYW, Tsang DNC. The D614G substitution in the S gene and clinical information for patients with COVID-19 detected in Hong Kong. J Clin Virol 2020;130:104550 Sep.

[24] Bullard J, Dust K, Funk D, Strong JE, Alexander D, Garnett L, et al. Predicting infectious SARS-CoV-2 from diagnostic samples. Clin Infect Dis Off Publ Infect Dis Soc Am 2020 May 22ciaa638Online ahead of print.

[25] Li Q, Wu J, Nie J, Zhang L, Hao H, Liu S, et al. The Impact of mutations in SARS-CoV-2 spike on viral infectivity and antigenicity. Cell 2020;182(5):1284–94 03e9.

[26] Plante JA, Liu Y, Liu J, Xia H, Johnson BA, Lokugamage KG, et al. Spike mutation D614G alters SARS-CoV-2 fitness. Nature 2020 Oct 26Online ahead of print.

[27] Tindale LC, Stockdale JE, Coombe M, Garlock ES, Lau WYV, Saraswat M, et al. Evidence for transmission of COVID-19 prior to symptom onset. Elife 2020;9:e57149. Jun 22.

[28] Pujadas E, Chaudhry F, McBride R, Richter F, Zhao S, Wajnberg A, et al. SARS-CoV-2 viral load predicts COVID-19 mortality. Lancet Respir Med 2020;8(9):e70.

[29] Westblade LF, Brar G, Pinheiro LC, Paidoussis D, Rajan M, Martin P, et al. SARS-CoV-2 viral load predicts mortality in patients with and without cancer who are hospitalized with COVID-19. Cancer Cell 2020;38(5):661–71 Nove2.

[30] Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. Lancet 2020;395(10223):497–506 Lond EnglFeb 15.

[31] Liu J, Li S, Liu J, Liang B, Wang X, Wang H, et al. Longitudinal characteristics of lymphocyte responses and cytokine profiles in the peripheral blood of SARS-CoV-2 infected patients. EBioMedicine 2020;55:102763 May.

[32] He L, Ding Y, Zhang Q, Che X, He Y, Shen H, et al. Expression of elevated levels of pro-inflammatory cytokines in SARS-CoV-infected ACE2+ cells in SARS patients: relation to the acute lung injury and pathogenesis of SARS. J Pathol 2006;210(3):288–97 Nov.

[33] Beigel JH, Tomashek KM, Dodd LE, Mehta AK, Zingman BS, Kalil AC, et al. Remdesivir for the treatment of Covid-19 - final report. N Engl J Med 2020;383(19):1813–26.

[34] RECOVERY Collaborative Group, Horby P, Lim WS, Emberson JR, Mafham M, Bell JL, et al. Dexamethasone in hospitalized patients with Covid-19 - preliminary report. N Engl J Med 2020 Jul 17Online ahead of print.

[35] Galloway SE, Paul P, MacCannell DR, Johansson MA, Brooks JT, MacNeil A, et al. Emergence of SARS-CoV-2 B.1.1.7 lineage - United States, December 29, 2020-January 12, 2021. MMWR. Morb Mortal Wkly Rep 2021;70(3):95–9 Jan 22.