Inferring SARS-CoV-2 variant within-host kinetics

Baptiste Elie[1,*], Bénédicte Roquebert[2], Mircea T. Sofonea[1],
Sabine Trombert-Paolantoni[2], Vincent Foulongne[3], Jérémie Guedj[4]
Stpéhanie Haim-Boukobza[2], Samuel Alizon[5,*]

[1] MIVEGEC, CNRS, IRD, Université de Montpellier, Montpellier, France
[2] Cerba Laboratory, Saint Ouen L'Aumône, France
[3] Laboratoire de Virologie, CHU de Montpellier, France
[4] Université de Paris, INSERM, IAME, F-75018 Paris, France
[5] Center for Interdisciplinary Research in Biology (CIRB), College de France, CNRS, INSERM, Université PSL, Paris, France
[*] Corresponding authors: baptiste.elie@ird.fr samuel.alizon@cnrs.fr

Abstract

Since early 2021, SARS-CoV-2 variants of concern (VOCs) have been causing epidemic rebounds in many countries. Their properties are well characterized at the epidemiological level but the potential underlying within-host determinants remain poorly understood. Thanks to a large database of qPCR VOC screening tests performed in the general population and hospitals in France between February 6 and August 21, 2021, we analyse a longitudinal cohort of 6,944 individuals with 14,304 cycle threshold (Ct) values. Using a droplet digital PCR assay, we converted Ct values into numbers of virus copies, which we assumed to be correlated with virus load. We show that the number of viral genome copies reaches a higher peak value and has a slower decay rate in infections caused by Alpha variant compared to that caused by historical lineages. Following the evidence that viral genome copies in upper respiratory tract swabs are informative on contagiousness, we show that the kinetics of the Alpha variant translate into significantly higher transmission potentials, especially in older populations. Finally, comparing infections caused by the Alpha and Delta variants, we find no significant difference in the peak viral copy number, but Delta variant infections seem to have a faster decline in viral copy number over time.

Keywords SARS-Cov2, viral dynamics, within-host, variant of concern, ddPCR

Introduction

The end of the year 2020 saw the identification of three SARS-CoV-2 'Variants of Concern' (VOC) causing phenotypically different infections from 'historical' lineages with demonstrated increases in contagiousness [1, 2, 3, 4] and virulence [1, 5]. Later in 2021, these variants were largely replaced by the new Delta VOC, which was itself more contagious [6] and, potentially, more virulent [7] than the Alpha variant. All VOCs were associated with deadly epidemic rebounds in several countries, which has led to their close monitoring through full genome sequencing but also targeted Reverse-Transcription Polymerase Chain Reaction (RT-PCR) screening. The latter is less precise than the former but more affordable, allowing for wider testing [8].

While clear phenotypic differences could be observed between the infections caused by those lineages, differences at the within-host level is less clear. Part of the reason for this is that most of the data analyse relies on a quantitative value, the cycle threshold value (Ct), which is often used as a proxy for the amount of virus genetic material [9, 10] and care must be taken when making biological interpretations of these Ct values, especially because of coronaviruses' life cycle [11]. However, their temporal variations in the same individual can shed light on the infection. Indeed, cross-sectional studies alone are not sufficient to characterize the impact of the variant on individual viral load because they are more subject to biases *e.g.* in function of the epidemic dynamic [12, 13, 14].

The field of within-host kinetics has grown focusing mainly on chronic infections, but some studies consider acute infections [15]. In the case of SARS-CoV-2, some studies used Ct values as a proxy for virus load to report temporal variations within individuals. Some performed statistical analyses on longitudinal data from hospitalized patients [16], health workers [17], and experimental infections in non-human primates [18]. More recently, studies focusing on the effect of the vaccine showed an overall decrease in viral loads among vaccinated infected patients [19], or a faster viral load decrease [20, 21], although this remains unclear for symptomatic infections [22]. These studies illustrate that viral kinetics reflect some properties of the infection and can explain some of the variations in the detection probability. Finally, regarding the variants of VOCs, cross-sectional studies have reported lower Ct values for Alpha than for wild type infections [1, 23] suggesting that VOCs could be causing infections with higher virus loads.

In this study, we analyse SARS-CoV-2 kinetics in a large number of individuals using linear mixed models, which allows us to explore how virus load dynamics may vary depending on the setting (general population or hospitals) or the variant causing the infection.

Material and Methods

Data

This study was approved by the Institutional Review Board (IRB) of the CHU of Montpellier and is registered at ClinicalTrials.gov under the identifier NCT04653844. The data originates from more than 250,000 variant-specific RT-PCR tests performed in France between February 06 and August 21, 2021 on SARS-CoV-2 positive samples [23, 13].

The assays use, in multiplex, probes targeting variant-specific mutations (see Table 1), as well as a region in the N virus gene for control purposes. We use the Ct of the later probe in our analyses.

Given the specificity of each assay used, different formatting steps were used. In particular, the initial IDS1 test designed to distinguish between the historical lineages and the Alpha VOC may increase the proportion of the latter VOC for high Ct values. Further details are shown in the flow diagram in Figure 1 and can be found in earlier studies [8, 23, 13].

Digital droplet PCR

Using Ct measures as a proxy for virus load has several limitations, especially in the case of coronaviruses [11]. Here, we do not attempt to equate the two but rather assume that temporal variations in Ct values are associated in changes in infectiousness.

To further investigate the biological significance of Ct values, we analysed samples from infections by a known virus lineage, i.e. historical or VOC, using both a variant-specific PCR and a digital droplet PCR (ddPCR). More precisely, we used two couples of primers targeting regions N1 and N2 in the virus, and the cellular human RNAse P for internal control.

Table 1: Summary of the assays used to screen for VOCs among positive tests. WT stands for 'wild type', i.e. ancestral lineages.

Assay	Detailed name	Targets	Variants
IDS1	ID^{TM} SARS-CoV-2/UK/SA Variant Triplex (ID Solution)	N501Y, Δ69-70	WT/δ vs. α vs. β/γ
IDS2	ID^{TM} SARS-CoV-2 / N501Y/ E484K Quadruplex (ID Solution)	N501Y, E484K	WT/δ vs. α vs. β/γ
IDS3	ID^{TM} SARS-CoV-2 / VOC evolution Pentaplex (ID Solution)	L452R, E484K, E484Q	WT/α vs. β/γ vs. δ
Perkin	VariantDetect^{TM} SARS-CoV-2 RT-PCR (PerkinElmer)	L452R, E484K, E484Q	WT/α vs. β/γ vs. δ

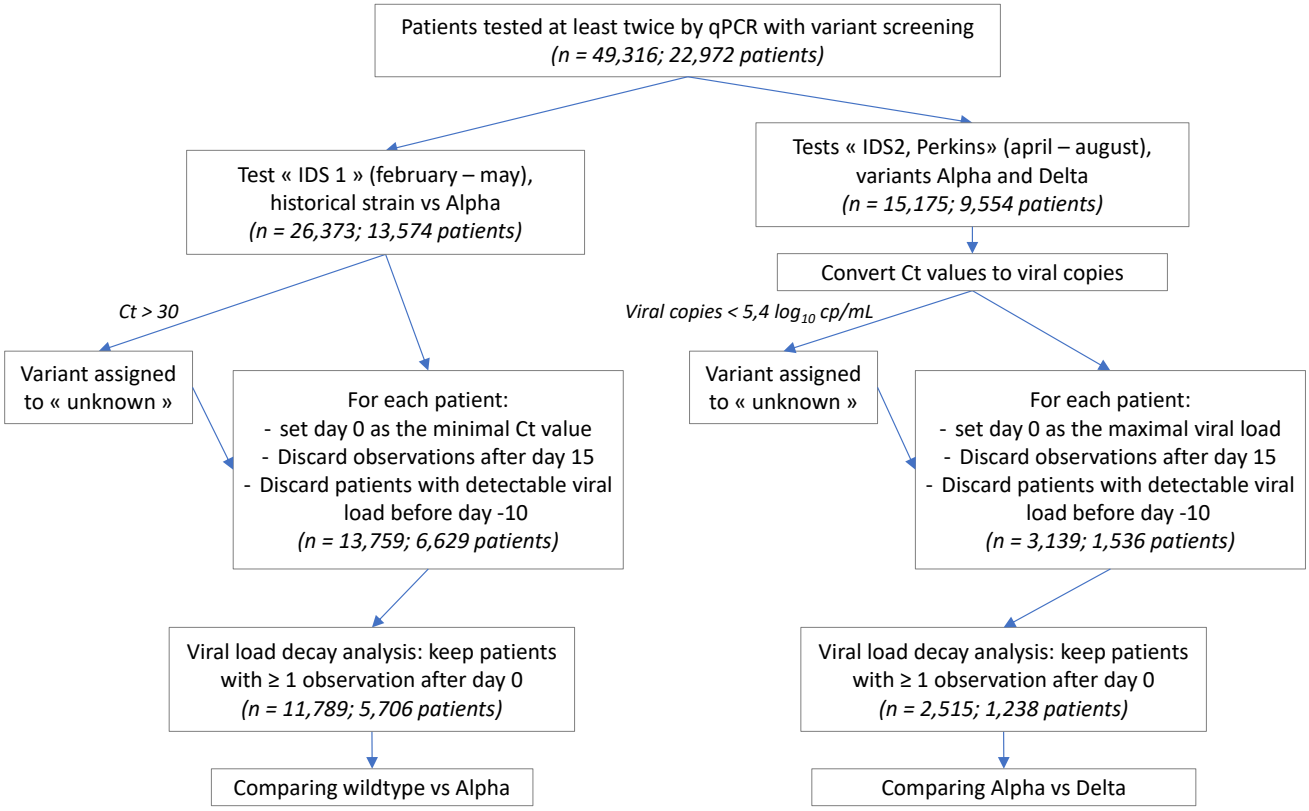

Figure 1: Dataset formatting steps. n indicates the number of tests, i.e. Ct values, analysed. For the IDS1 assay, the analyses were performed directly on the Ct values because the provider did not have any remaining tests to use with the ddPCR calibration. Furthermore, the assessment of the virus lineage (i.e. wild type or VOC) was only performed for tests with Ct values lower or equal than 30. For results obtained using the IDS2 and Perkin assays, Ct values were converted into viral genome copies before subsequent analysis and the assessment of the virus lineage was only performed for tests with more than $5.4 log_{10}$ copies/mL.

Statistical analyses

Linear mixed models

We used a linear model to study the potential link between the Ct value of the screening test and the number of virus copies obtained using ddPCR, using virus lineage and test assay as a co-factor. The results of the statistical model were used to convert Ct values from all assays into a number of virus copies.

We analysed the longitudinal data of number of virus copies (or Ct values for IDS1) with linear mixed models and used the R package lme4 [24] to fit the restricted maximum likelihood parameters to the data. The response variable was the number of virus copies (or the Ct value for IDS1), and we included two random effects on the individual and on the region of sampling, assuming a normal distribution of both effects, which was verified using Shapiro test.

To select which additional effects to include in the linear mixed model, we compared models with all the possible combinations of the following parameters: the virus strain (Alpha, Delta, or wild type), the day (day 0 being that with the lowest Ct value in an individual), the hospitalization status, the age, the variant specific reproduction number stratified by region, the delay between first and second test, and the vaccine coverage at the time of infection for the corresponding age class. We also included interactions between the day and all the previous effects, and between the age and the variant, the hospitalization status and the vaccine coverage. The best model was chosen based on the Akaike Information Criterion (AIC) [24]. For this, the AIC was computed on the corresponding linear mixed models, refitted to their maximal likelihood value.

We verified that models with $\Delta_{AIC} < 2$ were qualitatively identical to the model with the best AIC [25]. We also verified that the censored data points (*i.e.* the Ct values above the limit of detection, which was set at 37, or the viral copies number below 10,000 copies/mL) did not have a significant impact on the results by computing all the linear mixed models selected by a $\Delta_{AIC} < 2$ with a censored effect using the lmec R package.

Differences in Ct values between populations from the linear model outputs were statistically assessed using the Tukey adjustment and the emmeans R package.

Variant specific reproduction number

Regarding the variant specific reproduction number, i.e. the average number of secondary infections caused by an infected person at a given date [26], we first calculated the global reproduction number of the epidemic in each French region using hospital admission data [27]. We then adjusted this number by using the estimated relative proportion of each variant among the daily infections caused and computed in earlier studies [6, 8, 23], and the mean transmission

advantage computed by independent studies on a different dataset, estimating a transmission advantage of the Alpha variant over the historical strain of 29%, and of the Delta variant over the Alpha of 55% [28].

Transmission potential

To investigate the consequences of within-host kinetics at the population level, we had to find a way to link the Ct value (or number of virus copies) in a host at a given time and its contagiousness at the time. Since such data is extremely scarce (but see [29] for a study confirming that this link exists), we decided to assume that the decrease in the Ct value after the infection peak was correlated with the decrease in infectivity after the peak infectivity.

For this, we used the infectiousness profile estimated from the data of [30] by [31], *i.e.* a shifted Gamma distribution, with shape 97.19, rate 3.72, and shift 25.63. This analysis being restricted to one test (IDS1), we studied the correlation between the Ct value and the instantaneous infectiousness, *i.e.* the infectiousness profile density. We then used this mapping to infer a transmission potential, which can be seen as a proxy for the basic reproduction number [32], using the outputs of the linear mixed model capturing within-patient dynamics. This was done by performing an integration from day 0 to day 15 of the infectiousness values obtained from the Ct values.

Data availability

Data and code are available on the gitlab repository: https://gitlab.in2p3.fr/ete/sars-cov2_kinetics

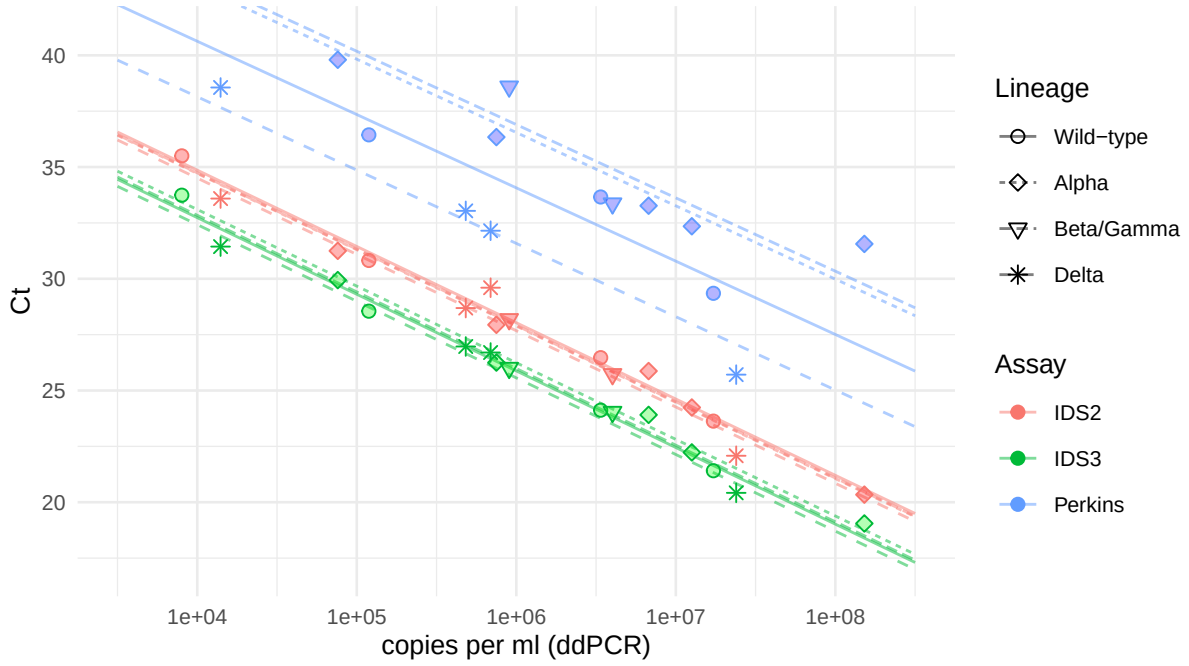

Figure 2: Ct value as a function of SARS-CoV2 copy number. The shape indicates the virus lineage and the color the variant screening assay used to obtain the Ct value for the control probe targeting the N gene. Virus copy number was estimated using a digital droplet PCR targeting the N gene.

Results

From Ct values to number of virus copies

For a given qPCR assay, we found a log-linear relationship between the Ct value and the absolute number of viral copies measured by ddPCR (Figure 2). Ct values were significantly different between assays. Compared to the reference (*i.e.* IDS2), IDS3 yielded a Ct value 1.96 lower (Student t-test, Tukey adjustment: $p < 10^{-4}$), and Perkin a Ct value 6.9 higher ($p < 10^{-4}$). For the Perkin assay, we also detected a significant effect of the variant on the Ct value. For the same number of virus copies, we found a Ct difference of 5.0 between samples originating from an infection by a Delta or an Alpha variant ($p < 10^{-4}$).

Within-host kinetics of the historical strain compared to the Alpha variant

After formatting the database for the IDS1 assay (Figure 1), we identified 12,536 suitable samples from 6,064 individuals. The characteristics of those samples is summarised in Table 2. Due to the limited number of observations, we removed patients infected lineages consistent with the Beta/Eta/Gamma VOCs, and analyzed the remaining 11,789 samples split into historical lineages and Alpha VOC. Most of the individuals were sampled twice with one-week interval.

Hospitalized patients tended to be sampled more regularly (not shown) but represented a minority of the data (358/6,064, 6%). Most samples originated from the Ile-de-France region (3863/6,064, 64%) and were caused by the Alpha variant (4495/6,064, 74%), which is consistent with the state of the epidemic in France at the time where the samples were collected [33].

To analyze the virus load kinetics, we inferred linear mixed models and selected the one with the lowest AIC (see the Methods for details). The estimated values of the model are shown in Table 3.

The linear mixed model revealed significant differences in viral genome copies dynamics between the historical lineages and the Alpha VOC (Figure 3). First, we found an interaction between the individual's age and the lineage causing the infection. Compared to historical lineages infections, the peak virus load appeared to increase with age for infections caused by the Alpha VOC (Table 3). Overall, for the French demographic age structure, the peak viral genome copies in Alpha infections was significantly higher than in historical lineages infections (-0.67 Ct [-0.87,-0.46]). Furthermore, in infections caused by the Alpha VOC, the rate at which the virus load declined over the course of the infection was lower than in historical lineages infections. As a result, 7 days after the viral copy number peak, the Ct difference was higher (-1.08 Ct [-1.32,-0.85]).

We also observed a significant impact of the hospitalization status, with a higher viral genome copies peak size. A shorter sampling delay between first and second test was associated with

Table 2: Properties of the longitudinal dataset based on the ISD1 assay. The data contain 12,536 Ct values from 6,064 individuals.

Variable	Value	Median (IQR) or n %
sampling context	*general population (ref)*	5706 (94%)
	hospital	358 (6%)
age		41 [25-58] years
follow up duration		8 [6-10] days
date of first sampling (in 2021)		Mar 9 [Feb 20 -Mar 22]
lineage	*wildtype (ref)*	1211 (20%)
	Alpha	4495 (74%)
	Beta/Gamma/Eta	358 (6%)
samples per individual	*2*	5686 (93%)
	3	352 (6%)
	>3	26 (1%)
administrative region	*Ile-de-France*	3863 (64%)
	Normandie	659 (11%)
	Hauts-de-France	612 (10%)
	PACA	273 (4%)
	other	657 (11%)

a higher Ct value at intercept. This effect was included in order to correct for the behavioural shift during the sampling period. Finally, the random effect on the region of sampling were significantly associated with the Ct values. These are consistent with the fact that Ct values are known to vary depending on the facility that performed the sampling [9].

These results are robust to the assumptions made regarding the dataset formatting, such as the inclusion or not of individuals above 80, of hospitalized patients, or of patients with a long clinical follow-up (more than 15 days).

Alpha transmission potential advantage

To further investigate the implications of these results at the population level, we performed a mapping between the decline of the daily infectivity after its peak and the daily Ct value for the wild type strain inferred from the linear model. The resulting significant correlation (Figure 4A, $R^2 = 0.95$) supports a linear relationship between Ct and infectiousness, which is consistent with the assumption that both metrics are correlated to the log of the virus load [34].

We used this mapping between Ct values and infectivity to study transmission potential differences between lineages. We found that the advantage of the Alpha VOC over the historical strain was more pronounced in countries with older populations, such as France or Japan, than in countries with younger population, such as Burkina Faso or Brazil (Figure 4B).

Table 3: Linear mixed model parameters estimates. Bold rows correspond to estimates with p < 0.05. The notation a:b indicates an interaction between factors a and b. CI stands for 'confidence interval'. The random effect of the patients on the intercept has a standard deviation of 2.075 [1.95,2.19], the random effect of the region on the intercept has a standard deviation of 1.051 [0.61,1.72] and the residues have a standard deviation of 3.89 [3.82,3.96].

predictor		**Estimate**	**95% CI**
intercept		**24.9**	**(24.1,25.7)**
day		**1.25**	**(1.20,1.30)**
age		-0.0049	(-0.011,0.001)
hospitalisation	*no*	ref	—
	yes	**-1.36**	**(-1.75,-0.97)**
delay between first and second test		**-0.40**	**(-0.43,-0.37)**
day:age		**-0.0029**	**(-0.0036,-0.0022)**
day:strain	***Historical***	ref	—
	Alpha	**-0.059**	**(-0.097,-0.021)**
age:strain	***Historical***	ref	—
	Alpha	**-0.016**	**(-0.020,-0.011)**

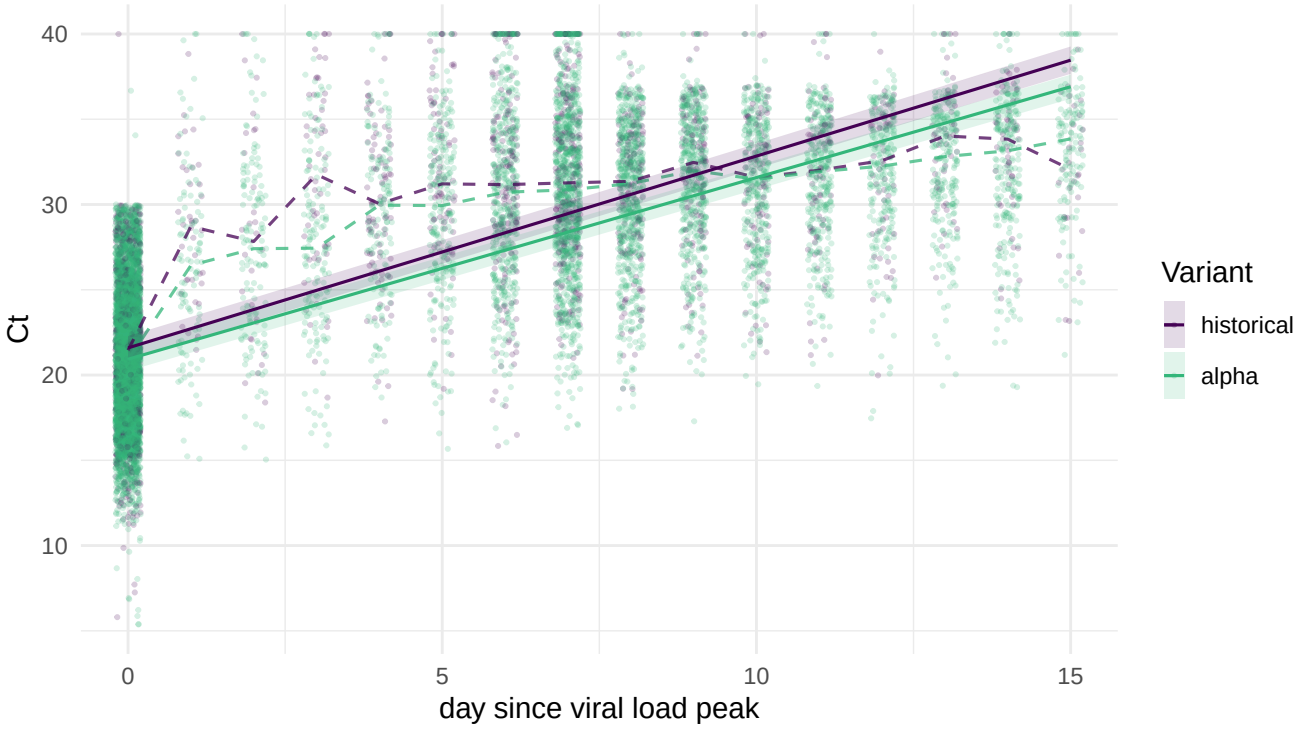

Figure 3: Within-host longitudinal Ct data as a function of the virus lineage. The dots represent the observed values. The dashed lines represent the mean value, for each day and each strain. The lines represent the linear model for an average patient (median age, non hospitalised).

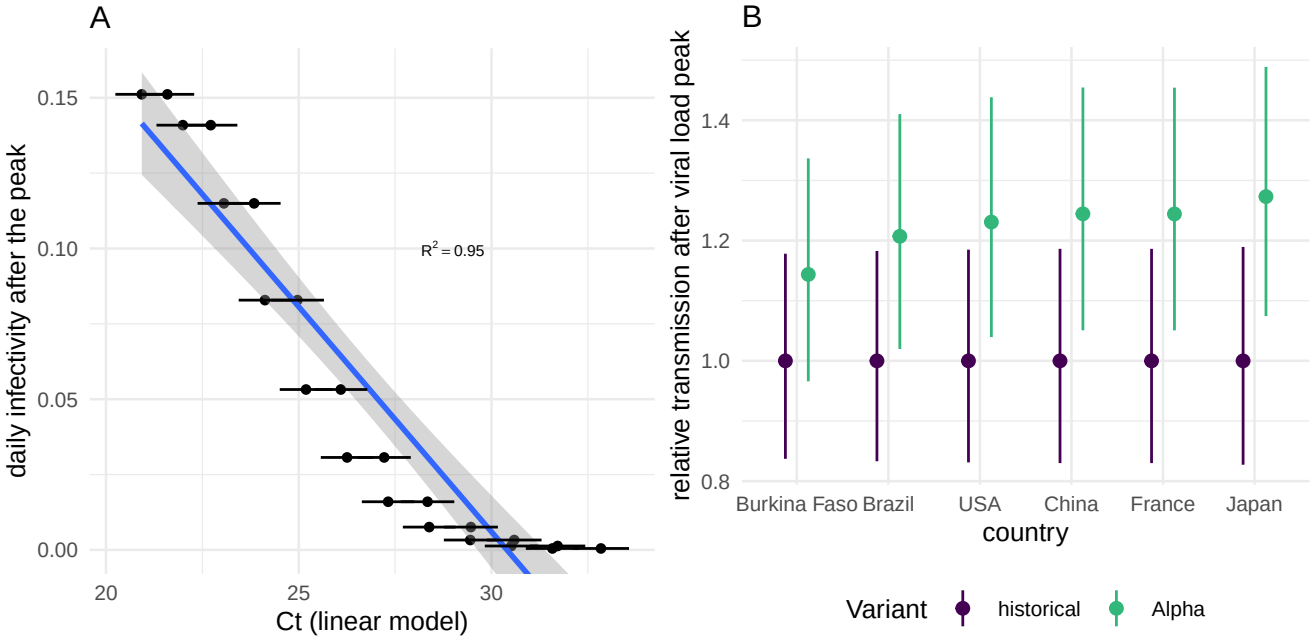

Figure 4: Impact of the differential viral genome copies on the transmissibility. A. Relationship between infectivity and Ct, after the viral genome copies peak, for the historical strain. B. The linear models parameters were used, with the demography of each country, to infer the transmission advantage of the variants. The error bars represent the 95% bootstrap quantile.

Comparing within-host kinetics of the Alpha and Delta variants

We then analysed the data between 13 April 2021 and 21 August 2021, which involved Ct values from 3 other assays that we were able to convert into virus copy numbers (see the Methods and Figure 2). In the following, we use a log base 2 relationship with time since peak of viral genome copies to compare the orders of magnitude with the previous results.

Infections consistent with viruses from Beta/Gamma/Eta lineages were too rare to be analysed and removed them from the dataset. Overall, we were able to compare infections caused by the Alpha or the Delta variant by analyzing the remaining 2,515 samples from 1,238 individuals (see Table 4 for additional details). Compared to the previous dataset, patients were slightly younger (median age 35 vs 41) and the follow-up duration was shorter (median 6 vs 8 days). This latter difference was corrected by the variable taking into account the delay between first and second test.

The best model according to the AIC criterion is detailed in Table 5 and the dynamics of the number of viral genome copies is shown in Figure 5. On 15 out of 17 models with a $\Delta AIC < 2$, we found no significant difference between the kinetics of the infections caused by the Alpha and the Delta VOC on the intercept. However, we found a significant difference on the decay

Table 4: Properties of the longitudinal dataset from April 13 and August 21. The data contain 2,751 Ct values from 1,239 individuals. It involves the IDS2, IDS3, and Perkin assays.

Variable	Value	Median (IQR) or n %
sampling context	*general population (ref)*	1263 (93%)
	hospital	92 (7%)
age		35 [23-54] years
follow up duration		6 [5-7] days
date of first swab (in 2021)		May 5 [Apr 21 - Jul 21]
lineage	*Alpha (ref)*	849 (63%)
	Delta	390 (29%)
	historical	6 (0.4%)
	Beta/Gamma/Eta	110 (8%)
samples per individual	*2*	1313 (97%)
	>2	42 (3%)
sampling region	*Ile-de-France*	676 (50%)
	PACA	204 (15%)
	Normandie	181 (13%)
	Hauts-de-France	142 (10%)
	other	151 (12%)
Screening assay	*IDS2*	1,928 (70%)
	Perkin	824 (30%)
	IDS3	1 (<1%)

rate on 13 models. When taking into account the censoring effect on the copies number estimate on those models, 15 out 17 have a significantly faster decay rate with Delta VOC.

In order to make sure that this result was not due to confounding factors, we included several covariates in our analysis which revealed a strong effect of the sampling date, which could interfere with the variant effect (since the Delta VOC rapidly replaced the Alpha VOC). The vaccine coverage in the population, which was used a proxy for the probability that the individual was vaccinated, as well as the delay between first and second test, measuring behavioural differences, were retained as a cofactor in the best linear model. Furthermore, the variant-specific reproduction number $R(t)$, estimated at the regional level in France was also retained as a significant covariate, which is consistent with previous observations [13, 14], although the effect was not statistically significant.

In the end, the factors that appear to significantly affect within-host virus kinetics are the patient's age, the vaccine coverage at the date of infection, the delay between first and second test, and the effects of the age, the variant and the hospitalization status on the virus copy number decay.

Table 5: Linear model parameters estimates. Bold rows correspond to estimates with p<0.05. CI stands for 'confidence interval'. The notation a:b indicates an interaction between factors a and b. The random effect of the patients on the intercept has a standard deviation of 1.69 [1.42, 1.91], and the residues have a standard deviation of 3.50 [3.37,3.64].

Predictor		**Estimate**	**95% CI**
Intercept		**20.14**	**(18.9,21.4)**
day		**-1.33**	**(-1.42,-1.24)**
	age	**0.024**	**(0.013,0.034)**
R(t) variant specific		0.90	(-0.19,1.98)
Vaccine coverage		**-0.034**	**(-0.045,-0.023)**
Delay between first and second test		**0.67**	**(0.59,0.74)**
day:age		**0.0061**	**(0.0043,0.0080)**
day:hospital		**0.12**	**(0.0018,0.25)**
day:strain	***Alpha***	ref	—
	Delta	**-0.24**	**(-0.33,-0.15)**

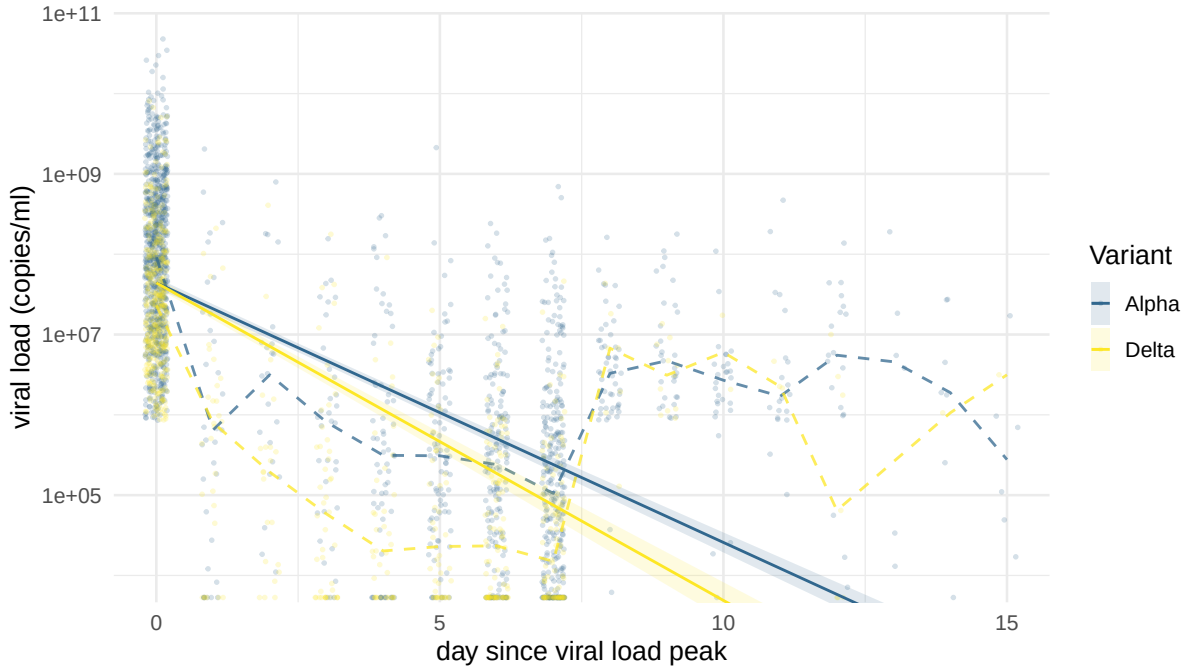

Figure 5: Temporal variation of the number of virus copies as a function of the virus lineage. The dots represent the observed values. The dashed lines represent the mean value, for each day and each strain. The lines represent the linear model for an average patient (median age, non hospitalised).

Discussion

Understanding the within-host kinetics of SARS-CoV-2 infections already yielded original insights on infection virulence [16], or efficiency of screening strategies [17]. Here, we analyzed a large national dataset of longitudinal RT-PCR Ct values to test the hypothesis that epidemic rebounds associated with SARS-CoV-2 variants could be explained by phenotypic differences in the infections they cause.

A linear mixed model indicated that infections caused by SARS-CoV-2 Alpha variant have higher virus loads, with a significant age-dependence. Furthermore, the temporal decrease in virus load was found to be slower when infections were caused by Alpha instead of the historical lineages. The results are consistent with unpublished results from a different cohort, which had fewer individuals with 2 or more samples (2,633 versus 6,064 here) but had self-reported symptom onset dates [35].

To further investigate the consequences of these variations in within-host kinetics at the epidemiological level, we first performed a correlation between the SARS-CoV-2 infectiousness profile and the estimated kinetics. This revealed a linear correlation between Ct value and daily infectiousness. Given that Ct values reflect logarithmic differences in virus load, this trend is consistent with an earlier study [34] and confirmed by a modelling approach [29] showing that a log-linear relationship between the viral load and transmissibility could explain the

observations. Given that we only consider relative transmission, we do not expect the result to depend on a particular regression coefficient, as long as it remains log-linear. We then translated the estimated Ct kinetics into transmission potential profiles. This revealed that variants have a higher transmission potential than wild type strains. Furthermore, this advantage was found to be more pronounced for Alpha in older populations (e.g. with a median advantage of 28% in Japan).

Kinetics of samples collected in hospitals exhibited higher peak viral genome copies (i.e. lower Ct values), and we also found a longer period under a Ct of 30 in hospital settings, which is consistent with earlier studies [16]. Note that the other results were unaffected when we removed the data from the hospitals from the analysis.

Unfortunately, we were unable to convert the Ct values from the first assays into number of virus copies. This would have allowed us to compare the Delta variant to the ancestral lineages and, potentially, to have enough data to also analyse the other variants.

We therefore limited our comparison of the Delta variant to the infections caused by the Alpha variant. Our models tend to show no difference on the peak of viral copy number, but a faster decrease than infections caused by Alpha variants. This can be counter-intuitive given that studies suggest that the latter has a 52 to 110% transmission advantage [6, 28]. This could be explained by the fact that ddPCR itself is limited because it does not count the number of infectious virions [11]. This qualitative difference could explain why we do not find marked differences between virus loads in infections caused by the Alpha and the Delta variant. This is also consistent with existing data showing, for instance, that although infections in vaccinated and non-vaccinated individuals may exhibit similar virus loads, the latter contain a higher proportion of infectious viruses [36].

A limitation of this analysis is that we do not have any indication regarding the date of the infection or of the symptom onset. This uncertainty prevents us from analyzing more mechanistic models with nonlinear mixed-effect models [16]. However, since the nature of the virus causing the infection is unlikely to affect the number of days between infection and screening, we do not expect our assumption that the lowest Ct value corresponds to the peak viral genome copies to introduce biases.

Our comparison between the Alpha and Delta VOCs is potentially subject to additional biases. Indeed, this part of the analysis covers a larger period of sampling, with an important increase in the vaccine coverage, and potential behavioural changes. We attempted to correct for those biases by including covariates such as vaccine coverage in the general population, variant reproduction number, or the delay between the first two tests of an infected individual. The estimated parameters are coherent with the expectations. In particular, the use of antigenic tests as a first line of screening is likely to have delayed the delay between infection and qPCR testing.

These tests increased from 20% of the total Covid-19 screening tests to 50% of the total between May and August in France [37]. We observe that the median delay between the first and the second test (which is usually done 7 days after the symptoms) decreased from 7 to 5 days during the analysis. Therefore, we can interpret the effect of the delay variable as a shift in the day 0: if the delay between the first two tests is shorter, assuming that the second tests was made around the seventh day after symptoms onset, the first test was not done on day 0 but slightly later. It is then coherent that a shorter delay increases the Ct value, or decreases the viral copies at intercept.

Overall, our study illustrates the insights that the combination of large screening data and statistical analyses can bring to the understanding of within-patient kinetics and population spread, as illustrated by our comparison between the Alpha variant and ancestral lineages. It also shows the limitations of Ct values and the added value of infectivity assays.

Acknowledgements

The authors acknowledge the ISO 9001 certified IRD i-Trop HPC (South Green Platform) at IRD montpellier for providing HPC resources that have contributed to the research results reported within this paper.

We thank all the ETE modeling team for thoughtful discussions.

Authors contribution

MTS, SHB, BR, and SA conceived the project, STP, VF, SHB, and BR collected the data, BE, MTS, and SA analysed the data, JG provided statistical expertise, BE and SA wrote the first draft the manuscript, and all authors contributed to the final version of the manuscript.

Conflict of interest

The authors declare that there are no conflicts of interest.

References

[1] Davies NG, et al. (2021) Increased mortality in community-tested cases of SARS-CoV-2 lineage B.1.1.7. *Nature* 593:270–274.

[2] Volz E, et al. (2021) Assessing transmissibility of SARS-CoV-2 lineage B.1.1.7 in England. *Nature* pp. 1–17.

[3] Tegally H, et al. (2021) Detection of a SARS-CoV-2 variant of concern in South Africa. *Nature* 592(7854):438–443.

[4] Faria NR, et al. (2021) Genomics and epidemiology of the P.1 SARS-CoV-2 lineage in Manaus, Brazil. *Science* 372(6544):815–821.

[5] Challen R, et al. (2021) Risk of mortality in patients infected with SARS-CoV-2 variant of concern 202012/1: Matched cohort study. *BMJ* 372:n579.

[6] Alizon S, et al. (2021) Rapid spread of the SARS-CoV-2 Delta variant in some French regions, June 2021. *Eurosurveillance* 26(28):2100573.

[7] Fisman DN, Tuite AR (2021) Evaluation of the relative virulence of novel SARS-CoV-2 variants: A retrospective cohort study in Ontario, Canada. *CMAJ : Canadian Medical Association Journal* 193(42):E1619–E1625.

[8] Haim-Boukobza S, et al. (2021) Detection of Rapid SARS-CoV-2 Variant Spread, France, January 26–February 16, 2021. *Emerging Infectious Diseases* 27(5):1496–1499.

[9] Alizon S, et al. (2021) Epidemiological and clinical insights from SARS-CoV-2 RT-PCR cycle amplification values. *medRxiv*.

[10] Walker AS, et al. (2021) Increased infections, but not viral burden, with a new SARS-CoV-2 variant. *medRxiv*.

[11] Michalakis Y, Sofonea MT, Alizon S, Bravo IG (2021) SARS-CoV-2 viral RNA levels are not 'viral load'. *Trends in Microbiology* 29(11):970–972.

[12] Walker AS, et al. (2021) Ct threshold values, a proxy for viral load in community SARS-CoV-2 cases, demonstrate wide variation across populations and over time. *eLife* 10:e64683.

[13] Alizon S (2021) Superspreading genomes. *Science* 371(6529):574–575.

[14] Hay JA, et al. (2021) Estimating epidemiologic dynamics from cross-sectional viral load distributions. *Science* 373(6552):eabh0635.

[15] Canini L, Perelson AS (2014) Viral kinetic modeling: State of the art. *J Pharmacokinet Pharmacodyn* 41(5):431–443.

[16] Néant N, et al. (2021) Modeling SARS-CoV-2 viral kinetics and association with mortality in hospitalized patients from the French COVID cohort. *Proceedings of the National Academy of Sciences* 118(8).

[17] Hellewell J, et al. (2021) Estimating the effectiveness of routine asymptomatic PCR testing at different frequencies for the detection of SARS-CoV-2 infections. *BMC Medicine* 19(1):106.

[18] Gonçalves A, et al. (2021) SARS-CoV-2 viral dynamics in non-human primates. *PLOS Computational Biology* 17(3):e1008785.

[19] Ke R, et al. (2021) Longitudinal analysis of SARS-CoV-2 vaccine breakthrough infections reveal limited infectious virus shedding and restricted tissue distribution, (Infectious Diseases (except HIV/AIDS)), Preprint.

[20] Singanayagam A, et al. (2021) Community transmission and viral load kinetics of the SARS-CoV-2 delta (B.1.617.2) variant in vaccinated and unvaccinated individuals in the UK: A prospective, longitudinal, cohort study. *The Lancet Infectious Diseases*.

[21] Kissler SM, et al. (2021) Viral Dynamics of SARS-CoV-2 Variants in Vaccinated and Unvaccinated Persons. *New England Journal of Medicine* p. NEJMc2102507.

[22] Blanquart F, et al. (2021) Characterisation of vaccine breakthrough infections of SARS-CoV-2 Delta and Alpha variants and within-host viral load dynamics in the community, France, June to July 2021. *Eurosurveillance* 26(37):2100824.

[23] Roquebert B, et al. (2021) SARS-CoV-2 variants of concern are associated with lower RT-PCR amplification cycles between January and March 2021 in France. *International Journal of Infectious Diseases* 113:12–14.

[24] Bolker BM, et al. (2009) Generalized linear mixed models: A practical guide for ecology and evolution. *Trends in Ecology & Evolution* 24(3):127–135.

[25] Burnham KP, Anderson DR, Burnham KP (2002) *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*. (Springer, New York), 2nd ed edition.

[26] Anderson RM, May RM (1991) *Infectious Diseases of Humans. Dynamics and Control*. (Oxford University Press, Oxford).

[27] Reyné B, Danesh G, Alizon S, Sofonea MT (2020) Rt2: Computing and visualising COVID-19 epidemics temporal reproduction number. *medRxiv* p. 2020.12.05.20244376.

[28] Campbell F, et al. (2021) Increased transmissibility and global spread of SARS-CoV-2 variants of concern as at June 2021. *Eurosurveillance* 26(24):2100509.

[29] Marc A, et al. (2021) Quantifying the relationship between SARS-CoV-2 viral load and infectiousness. *eLife* 10:e69302.

[30] He X, et al. (2020) Temporal dynamics in viral shedding and transmissibility of COVID-19. *Nature Medicine* 26(5):672–675.

[31] Ashcroft P, et al. (2020) COVID-19 infectivity profile correction. *Swiss Medical Weekly* 150(3132).

[32] Fraser C, Hollingsworth TD, Chapman R, de Wolf F, Hanage WP (2007) Variation in HIV-1 set-point viral load: Epidemiological analysis and an evolutionary hypothesis. *Proceedings of the National Academy of Sciences* 104(44):17441–17446.

[33] Sofonea MT, Alizon S (2021) Anticipating COVID-19 intensive care unit capacity strain: A look back at epidemiological projections in France. *Anaesthesia Critical Care & Pain Medicine* 40(4):100943.

[34] Marks M, et al. (2021) Transmission of COVID-19 in 282 clusters in Catalonia, Spain: A cohort study. *The Lancet Infectious Diseases* 0(0).

[35] Cosentino G, et al. (2021) SARS-CoV-2 viral dynamics in infections with variants of concern in the French community.

[36] Despres HW, et al. (2021) Quantitative measurement of infectious virus in SARS-CoV-2 Alpha, Delta and Epsilon variants reveals higher infectivity (viral titer:RNA ratio) in clinical samples containing the Delta and Epsilon variants. *medRxiv*.

[37] (2021) COVID: Exécutions et ventes de tests antigéniques rapides en pharmacie - data.gouv.fr (https://www.data.gouv.fr/fr/datasets/covid-executions-et-ventes-de-tests-antigeniques-rapides-en-pharmacie/).