SARS-CoV-2 variants of concern are associated with lower RT-PCR amplification cycles between January and March 2021 in France

Benedicte Roquebert, Stéphanie Haim-Boukobza, Sabine Trombert-Paolantoni, Emmanuel Lecorche, Laura Verdurme, Vincent Foulongne, Sonia Burrel, Samuel Alizon, Mircea Sofonea

To cite this version:

HAL Id: hal-03366559
https://hal.archives-ouvertes.fr/hal-03366559
Submitted on 6 Oct 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
SARS-CoV-2 variants of concern are associated with lower RT-PCR amplification cycles between January and March 2021 in France

Benedicte Roquebert\textsuperscript{1}, Stéphanie Haim-Boukobza\textsuperscript{1}, Sabine Trombert-Paolantoni\textsuperscript{1}, Emmanuel Lecorche\textsuperscript{1}, Laura Verdurme\textsuperscript{1}, Vincent Foulongne\textsuperscript{2}, Sonia Burrel\textsuperscript{3}, Samuel Alizon\textsuperscript{4,+}, Mircea T. Sofonea\textsuperscript{4,+,*}

\textsuperscript{1} Laboratoire Cerba - 7-11 r de l’Équerre, 95310 Saint-Ouen-l’Aumône, France

\textsuperscript{2} Laboratoire de Virologie, CHU de Montpellier - 80 av Augustin Fliche, 34295 Montpellier, France

\textsuperscript{3} Sorbonne Université, INSERM U1136, Institut Pierre Louis d’Épidémiologie et de Santé Publique (IPLESP), AP-HP, Hôpital Pitié-Salpêtrière, Service de Virologie - 47-83 bd de l’Hôpital, 75013 Paris, France

\textsuperscript{4} MIVEGEC, Univ. Montpellier, CNRS,IRD – 911 av Agropolis, 34394 Montpellier Cedex 5, France

\textsuperscript{+} equal contribution

\textsuperscript{*} Corresponding author: mircea.sofonea@umontpellier.fr (MIVEGEC (ETE), Centre IRD - 911, av Agropolis 34394 Montpellier Cedex 5, France)

\textbf{word count:} 983
SARS-CoV-2 variants of concern are associated with lower RT-PCR amplification cycles between January and March 2021 in France

**Abstract**

SARS-CoV-2 variants raise concern regarding the mortality caused by COVID-19 epidemics. We analyse 88,375 cycle amplification (Ct) values from variant-specific RT-PCR tests performed between January 26 and March 13, 2021. We estimate that on March 12, nearly 85% of the infections were caused by the Alpha variant and that its transmission advantage over wild type strains was between 38 and 44%. We also find that tests positive for Alpha and Beta/Gamma variants exhibit significantly lower cycle threshold (Ct) values.

**Context**

At least three SARS-CoV-2 variants are currently major sources of concern: Alpha from lineage B.1.1.7 (Davies et al., 2021; Volz et al., 2021), Beta from lineage B.1.351 (Tegally et al., 2021), and Gamma from lineage P.1 (Faria et al., 2021). Alpha and Gamma variants have been shown to be more contagious (Davies et al., 2021; Volz et al., 2021; Faria et al., 2021), while Beta and Gamma seem to evade immune responses (Tegally et al., 2021; Faria et al., 2021). Although the mechanistic bases are still being investigated, the increased transmissibility could be driven by the N501Y mutation and the Δ69-70 deletion in the Spike protein (Davies et al., 2021), where also lies the E484K mutation related to immune escape (Tegally et al., 2021; Faria et al., 2021). Current evidence regarding potential differences in cycle threshold (Ct) values are still unclear (Walker et al., 2020; Faria et al., 2021) as how well they reflect viral loads (Néant et al., 2021; Michalakis et al., 2021).

In France, using data from 11,916 tests performed on Jan 6 and 7, 2021, a study estimated that 3.3% of the infections were caused by the Alpha variant at that time (Gaymard et al., 2021). Another study used 40,777 tests performed between Jan 26 and Feb 16, 2021, and estimated that on Feb 16, 55% of the infections were caused by Alpha, Beta, or Gamma variants (Haim-Boukobza et al., 2021).

Here, we estimate the spread of SARS-CoV-2 variants of concern using 123,867 variant-specific RT-PCRs performed in nasopharyngeal (NP) swabs between Jan 26 and Mar 19, 2021 in France and compare Ct values between variants (the dataset and technical specifications are detailed in the Appendix).
**Alpha variant is dominant**

Using the same methodology described in Haim-Boukobza et al. (2021) and in the Appendix, we calculated the transmission advantage of each variant compared to the wild type strain after correction for several biases (region, sampling date, assay, and patient age). This inference was performed for individuals from 5 to 80 years old and without the data from hospitals (to avoid sampling delay bias). For Alpha, the transmission advantage was 40% (95% confidence interval, CI: [38,42]%). For Beta-Gamma, the estimate was 28% (95%CI: [27,30]%).

We then estimated the proportion of new infections caused by each type of strain on Mar 19, 2021. At a national level, the estimate was 87% for Alpha and 5,4% for Beta/Gamma, but with strong regional heterogeneity (Figure S1).

**Variants have lower Ct values**

We analysed tests and for which the strain has been determined from 1 to 89 year old patients with Ct values lower than 30, i.e. 76,745 tests (91% of all the tests with Ct values).

We used a multiple linear regression to study variations in Ct values between strains. The covariates were the age, the sampling facility (hospital or city), the sampling date, and the region (including the interaction between the last two). We then performed a F-test of a model including the variant identity as a covariate, in interaction with age, against a model without it. The variant effect was found to be statistically significant while the model explained a small fraction of the Ct variability (adj. $R^2=3.8%$), which is consistent with these values being highly variable (Alizon et al., 2021).

The virus strain effect was highly significant in the ANOVA (Figure 1). Samples from Alpha variants had a significantly smaller Ct than that from Beta/Gamma (21.8 vs. 22.1). Both had significantly smaller Ct values than wild type strains (22.8) and other variants.

The model also indicated a significant decrease of Ct with age, in line with existing data (Alizon et al., 2021). Interestingly, this decrease was significantly (more than half) stronger for Alpha, while the inference cannot exclude equivalent slopes for the other variants (note that this trend holds even on the non-Alpha variant subset of Ct values). The sampling date, the sampling region and their interaction were also found to be significant. Samples from hospitals has a slightly higher Ct, likely due to the fact that testing in the general population occurs approximately one week after infection and one week before potential hospitalization. Therefore, hospitalised patients data is likely to reflect an older state of the epidemic.
Figure 1: Cycles threshold (Ct) value for SARS-CoV-2 strains. Median estimates based on the linear model are shown in the box plot, and number of tests in each class are shown in the bottom of the graph. ‘other’ indicate tests with the Δ69-70 deletion and without the N501Y mutation. Start indicate the significance level (**** for a p-values strictly lower than $10^{-4}$ and *** for a p-value of $10^{-4}$, ns : not significant).

Discussion

We show that variant of concern Alpha is now vastly dominant in France compared to wild type strains (87% vs. less than 10%). Beta or Gamma remain limited (approximately 5.4% of the new infections). These results are consistent with earlier reports of a marked transmission-advantage of the Alpha variant (Davies et al., 2021; Volz et al., 2021; Haim-Boukobza et al., 2021; Gaymard et al., 2021).

By investigating the RT-PCR Ct values, which can inform us on clinical features of the infection (Néant et al., 2021; Alizon et al., 2021), we show that infections caused by variants significantly differ from that caused by wild type strains. That variants are associated with lower Ct values could be an indication of higher viral load, although care must be taken because of the biology of SARS-CoV-2 (Michalakis et al., 2021) and of the variability inherent to such values (Alizon et al., 2021).

This result contrasts with earlier findings. One study did not find a significant result when comparing Ct values for tests with or without the S-gene target failure (Walker et al., 2020). However, our results are based on a variant-specific PCR. Another study on the Gamma variant (Faria et al., 2021) did not find a significant difference after accounting for the symptom onset to sampling delay. However, their
study was performed on a limited number of samples \((n = 147)\). Finally, our estimates, based on a large number of interpretable samples with Ct lower than 30, suggest that differences in within-host replication and transcription between Alpha and non-Alpha variants might be lower than previous quantified (Jones et al., 2021).

**Data availability**

The data and scripts used for the analysis will be shared upon publication.

**Conflict of Interest**

None declared.

**Funding Source**

This work received no specific funding.

**Ethical Approval**

This study was approved by the Internal Review Board of the CHU of Montpellier (ClinicalTrial.gov identifier NCT04738331).

**Acknowledgements**

We thank the ETE modelling team for discussion, as well as the CNRS, the IRD, the ANR, and the Région Occitanie for funding (PHYEPI grant).

**References**


Alexandre Gaymard, Paolo Bosetti, Adeline Feri, Gregory Destras, Vincent Enouf, Alessio Andronico,


Houriiyah Tegally, Eduan Wilkinson, Richard J Lessells, Jennifer Gianthari, Sureshnee Pillay, Nokukhanya


Appendix

Supplementary methods

Dataset

We analyse 123,867 variant-specific RT-PCR tests performed in France on the same number of individuals between Jan 26 and Mar 19, 2021. The main assay used was ID\textsuperscript{TM} SARS-CoV-2/UK/SA Variant Triplex (ID SOLUTION) but for 4159 tests (3.4\%) performed before Feb 3, 2021 we used the VirSNiP SARS-CoV-2 Spike del+501 (TIB MOLBIOL) assay. The sampling varied between French regions and we excluded from the analysis regions with less than 100 tests. 936 tests were also removed because the sampling region was missing.

These tests have probes with 3 targets: a control one in the virus N gene, the \textDelta 69-70 deletion, and the N501Y mutation. For Alpha variants, both the deletion and the mutation are present. For Beta or Gamma variants, only the N501Y is detected. For VirSNiP assay, it is based on 501 and 69/70 fragments amplified and analyzed with a melting curve using mutation-specific probes, as described earlier (Haim-Boukobza et al., 2021). As indicated in Haim-Boukobza et al. (2021), the test specificity was confirmed internally using next-generation sequencing.

The main cofactors in the analysis were the assay used, the patient age, the sampling date, the sampling region, and the sampling facility (hospitals or city screening).

For 119,708 ID SOLUTION tests, we also analyse the cycle threshold value (Ct) of the virus control gene of the assay. Ct values greater than 30 were ignored because they may provide unreliable results regarding the variant-specific probes LoD (Limit of Detection). Indeed, the latter are located in the S gene, which tends to exhibit higher Ct values than the N gene (Alizon et al., 2021).

Linear model for the Ct analysis

We performed a type I error for the analysis-of-variance. Our response variable was the Ct value. The main covariate of interest was the strain and it could take 4 values (Alpha, Beta or Gamma, wild type, or other). The other covariates were the age, the sampling facility (hospital or city), the sampling date, and the geographical region. We also considered an interaction between sampling region and date. We used a type-I analysis of variance (ANOVA) and added the strain covariate last. The motivation for this is that with the sequential assumption of the summing of the squares (type I method), the order in which
the covariates are tested matters, and, in the case of an uneven sampling, the last one in the list is less likely to be significant. Therefore, our assumption decreases the risk of erroneously attributing observed variance to a variant effect.

We used a F-test to determine whether the addition of the strain effect statistically improved the explanation of the data.

**Generalised linear model to correct for variant sampling bias**

As indicated in (Haim-Boukobza et al., 2021), for a given variant category (Alpha or Beta/Gamma) we first perform a generalised linear model with a binomial error distribution where the variable of interest is the binary variant variable (with values ‘variant’ or ‘wild type’) and the explanatory variables are the sampling date, the sampling region, and the individual age. We also include an interaction between sampling region and date. We then use the residuals of this model to infer the transmission advantage of the variant.

**Logistic growth fitting**

We used the fitted values of a GLM model applied to the data after removing samples from hospitals (the sampling location effect was also obviously removed from the model) to perform the inference of a two-parameter logistic growth kinetic curve: \( f(t) = \left(1 + e^{-\rho(t-\tau)}\right)^{-1} \), where \( f(t) \) is the frequency of the variants in the new infections at time \( t \), \( \rho \) is the relative growth rate of the variants and \( \tau \) is the time at which \( f \) reaches 1/2. This method is indeed more appropriate to deal with temporal auto-correlation biases in proportion time series (Davies et al., 2021; Volz et al., 2021).

The parameter estimation was performed using the `drc` package in R both at the national and the regional level (for regions with at least 1,000 samples). The confidence intervals of the fitted curves rely on those of the estimated relative growth rate.

The unitless estimated transmission advantage is expressed in terms of multiplicative gain in reproduction number with respect to that of the wild type, such that \( R_{\text{variant}} = (1 + \text{ETA}) \times R_{\text{wild type}} \). Its calculation was made by solving the Euler-Lotka equation \( (R_{\text{variant}} \int_0^\infty e^{-\rho w(t)} dt = 1) \) assuming a serial interval \( w \) following a Weibull distribution with a mean and SD of 4.8 and 2.3 days (Nishiura et al., 2020) and a constant \( R_{\text{wild type}} \) equal to 1. The confidence interval rely on those of the estimated relative growth rate.

The estimate of the frequency of variant on Mar 12, 2021, was done by first estimating the proportion \( p_x \) of a given variant \( x \) compared to the wild type (while ignoring the other variant \( y \)) and second performing
the same analysis to look at the proportion $p_x^T$ of wild type and $x$ compared to the whole population ($x$ plus $y$ plus wild type). The frequency of variant $X$ was then obtained as $p_x \times p_x^T$.

Supplementary figures

Figure S1: Estimated proportion of new infections caused by A) wild type, B) Alpha variant, and C) Beta or Gamma variants on Mar 12, 2021, in French regions. Regions with insufficient sampling are in white.

Figure S2: Distribution of the residual values of the multiple regression linear model.
Figure S3: **Estimating the transmission advantage of the Alpha variant over the wild type strain.** The dots indicate the GLM-fitted values values and the line is the output of the logistic growth model estimation. The top figures indicate the estimated transmission advantage of the Alpha variant (with respect to the wild type reproduction number) and its 95%-confidence interval. The x-axis shows the date (month-day format).

Figure S4: **Estimating the transmission advantage of the Beta or Gamma variants over the wild type strain.** See Figure S3 for details.