Detecting Rapid Spread of SARS-CoV-2 Variants, France, January 26–February 16, 2021

Stéphanie Haim-Boukobza, Bénédicte Roquebert, Sabine Trombert-Paolantoni, Emmanuel Lecorche, Laura Verdurme, Vincent Foulongne, Christian Selinger, Yannis Michalakis, Mircea T. Sofonea,[1] Samuel Alizon[1]

Variants of severe acute respiratory syndrome coronavirus 2 raise concerns regarding the control of coronavirus disease epidemics. We analyzed 40,000 specific reverse transcription PCR tests performed on positive samples during January 26–February 16, 2021, in France. We found high transmission advantage of variants and more advanced spread than anticipated.

Since the end of 2020, at least 3 strains, or "variants," of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) bearing a high number of mutations have been associated with rapid epidemic spread in the United Kingdom (lineage B.1.1.7) (*1*), South Africa (lineage B.1.351) (*2*), and Brazil (lineage P.1) (*3*). Because of their increased transmissibility (*4*; E. Volz et al., unpub. data, https://www.medrxiv.org/content/10.1101/2020.12.30.20249034v2; A.S. Walker et al., unpub. data, https://www.medrxiv.org/content/10.1101/2021.01.13.21249721v1) and potential ability to evade host immunity (*5*; S. Cele et al., unpub. data, https://www.medrxiv.org/content/10.1101/2021.01.26.21250224v1), monitoring these variants is crucial in the context of mass vaccination.

In France, beginning February 5, 2021, every sample that tested SARS-CoV-2–positive by reverse transcription PCR (RT-PCR) underwent an additional variant-specific RT-PCR with probes targeting the Δ69–70 deletion and the N501Y mutation, both in the spike glycoprotein. Both targets are present in lineage B.1.1.7. For lineages B.1.351 and P.1, only the N501Y mutation is present. If only the Δ69–70 deletion is detected, the infection might be caused by another variant or by a wild-type strain with a deletion. Finally, if neither target is detected, the infection is considered to be caused by a wild-type strain. These tests are cheaper and easier to implement than full-genome sequencing, which enables their rapid deployment on a wide scale. We report the results of this testing program.

The Study

RT-PCR testing for SARS-CoV-2 strains was conducted using 2 assays, VirSNiP SARS-CoV-2 Spike del+501 (TIB Molbiol, https://www.tib-molbiol.de) and ID SARS-CoV-2/UK/SA Variant Triplex (ID Solutions, https://www.id-solutions.fr), which target the Δ69–70 deletion and N501Y mutation. We performed tests on 42,229 positive samples collected during January 26–February 16, 2021, from 40,777 persons from 12 regions in France. Most samples came from the general population, and 3,323 (7.9%) samples came from hospitals. For the 1,397 patients for whom multiple tests were performed, only the first test was considered. We only included data from persons 5–80 years of age to minimize the weight of preschool children and persons living in aged-care facilities in our analysis. Finally, we removed persons whose age or testing region was unknown. This study was approved by the Institutional Review Board of the CHU of Montpellier and is registered at ClinicalTrials.gov (identifier NCT04738331).

Overall, we analyzed 35,208 SARS-CoV-2–positive samples from the same number of persons (Appendix 1 Table 1, https://wwwnc.cdc.gov/EID/article/27/5/21-0397-App1.xlsx). Results of 6,702 (19%) variant-specific RT-PCR tests were uninterpretable, mainly because of an insufficient amplification of the control, which targets the SARS-CoV-2

Author affiliations: Cerba Laboratory, Saint Ouen L'Aumône, France (S. Haim-Boukobza, B. Roquebert, S. Trombert-Paolantoni, E. Lecorche, L. Verdurme); CHU de Montpellier, Montpellier, France (V. Foulongne); IRD, Montpellier (C. Selinger); CNRS, Montpellier (Y. Michalakis, S. Alizon); Université de Montpellier, Montpellier (M.T. Sofonea)

DOI: https://doi.org/10.3201/eid2705.210397

[1]These authors contributed equally to this article.

N gene. These results were treated as missing in the analyses. Given that most of the variants were B.1.1.7 (Appendix 2 Figure 1, https://wwwnc.cdc.gov/EID/article/27/5/21-0397-App2.pdf), we grouped all samples bearing the N501Y mutation into a broader class of variant-positive.

We used a generalized linear model (GLM) to analyze the binary strain variable (with values wild-type or variant). The covariates were the patient's age, the RT-PCR kit used for variant detection, the sampling date, and the geographic region from which the sample originated (Appendix 2). By using a type-II analysis of variance, we found that all covariates except the type of RT-PCR kit to be significant (Table 1). In particular, the proportion of variants increased with date and decreased with age (Appendix 2 Figure 2) and hospital origin.

To investigate the temporal trends, we fitted a logistic growth model to the fitted values of an analogous GLM only on the data from general population samples (Appendix 2). Assuming that variations in frequencies are driven by transmission advantages, we found that variants have a 50% (95% CI 37%–64%) transmission advantage over wild-type strains (Figure 1).

The analysis of variance already showed that variant frequency varied across regions (Table 1). We performed the logistic growth fit at the local level for regions for which adequate data was available (Figure 2). The growth advantage of the variant was more pronounced in some regions. In Ile-de-France, more than half of infections already appeared to be caused by the variants by February 16, whereas in other regions, such as Burgundy, this proportion would not be reached until March 2021. However, some regions were less well represented in this analysis, which could affect local estimates.

Finally, we investigated the correlation between the increase in variant frequency among positive tests in a region and the temporal reproduction number, denoted R_t, in that same region. R_t was estimated from coronavirus disease intensive care unit admission data by using the EpiEstim method (*6*) with a serial interval from Nishiura et al. (*7*), as described in Reyné et al. (unpub. data, https://www.medrxiv.org/content/10.1101/2020.12.05.20244376v1) (Appendix 2). We used the Spearman rank correlation test and found a positive but nonsignificant trend ($\rho = 0.50$; $p = 0.09$) (Appendix 2 Figure 3).

Table 1. Risk for variant detection estimated using the general linear model in study of rapid spread of severe acute respiratory syndrome coronavirus 2 variants, France, January 26–February 16, 2021*

Covariate	OR	2.5% CI	97.5% CI
Date, per day	**1.07**	1.03	1.11
Age, per year	**0.993**	0.992	0.995
Kit 2	0.94	0.93	1.16
Nonhospital location	**1.25**	1.13	1.39

*For categorical variables, reference values are the other kit (1) and the hospital setting. Bold indicates statistical significance. OR, odds ratio.

Conclusions

We used 2 variant-specific RT-PCR tests to detect the fraction of infections caused by SARS-CoV-2 lineages B.1.1.7, B.351, and P.1 in regions in France during January 25–February 16, 2021. We did not find any significant difference between the 2 specific RT-PCR kits used, suggesting that similar data collected in France could be pooled. Our results have several practical implications.

In general, we found that many infections screened were caused by variants, especially B.1.1.7, and the trend increased over time. On the basis of our estimates, by February 16, 2021, more than half of SARS-CoV-2 infections in France could have been caused by variants, although with pronounced spatial heterogeneity. In a conservative scenario, where all uninterpretable tests were assumed to be caused by the wild type, most infections would have been caused by variants by the end of week 7 of 2021, and the estimated variants transmission advantage was 36% (95% CI 26%–48%) (Appendix 2 Figure 4).

Variant-positive samples originated from significantly younger patients, which is consistent with an earlier report (E. Volz et al., unpub. data) but contrasts with Davies et al. (*4*). Our analysis did not enable us to discriminate between epidemiologic effects (e.g., if variants' transmission chains were seeded in different populations than the wild types), sampling biases, or biologic effects. Additional data

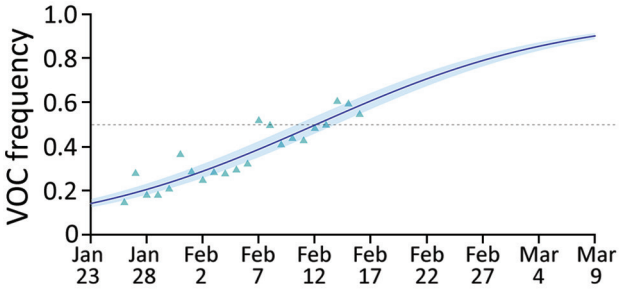


Figure 1. Estimated variants frequency kinetics in study of rapid spread of severe acute respiratory syndrome coronavirus 2 spread, France, January 26–February 16, 2021. Triangles indicate the general linear model–fitted values, line indicates output of the logistic growth model estimation, and shading indicates 95% CIs. Overall estimated transmission advantage of the variants (with respect to the wild-type reproduction number) is 50 (95% CI 38%–64%) (Appendix 2, https://wwwnc.cdc.gov/EID/article/27/5/21-0397-App2.pdf). VOC, variant of concern.

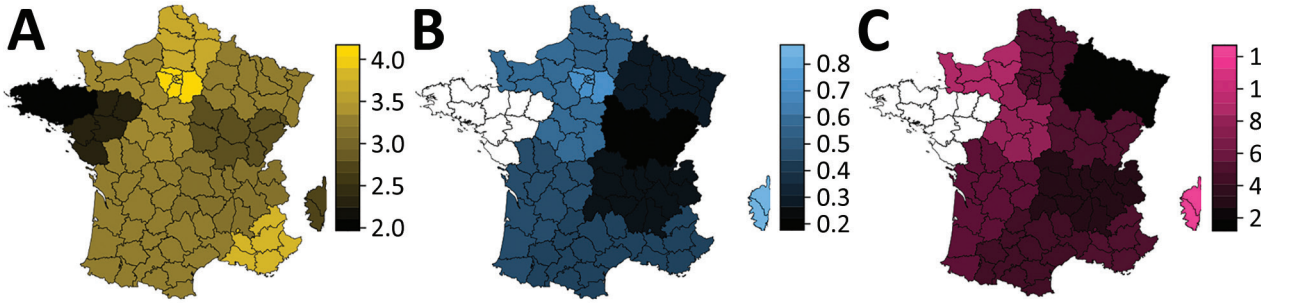


Figure 2. Regions of France in study of rapid spread of SARS-CoV-2 variants, January 26–February 16, 2021. For each region, we show the $\log_{10}$ of the number of tests analyzed (A), the estimated total variant frequency by February 16, 2021 (B), and the estimated percentage transmission advantage (of the variant strains relative to the wild-type strain (C). Additional details in Appendix 2 Figure 5 (https://wwwnc.cdc.gov/EID/article/27/5/21-0397-App2.pdf).

from RT-PCR amplification cycles could provide useful insights. Finally, earlier reports have found variant proportion to be associated with higher basic reproduction number (*4*; E. Volz et al., unpub. data). We found such a trend among regions in France, but it is not statistically significant.

A limitation of this study is that, in spite of its intensity, the sampling was performed retrospectively, which could generate biases if, for instance, transmission chains associated with variants were increasingly sampled. However, we found that samples that originated in hospitals were associated with a lower variant detection. Because testing in the general population is usually performed a week after infection and hospital admissions occur ≈2 weeks after infection (M.T. Sofonea et al., unpub. data, https://www.medrxiv.org/content/10.1101/2020.05.22.20110593v1), we expect hospital data would reflect an older state of the epidemic than screening data. RT-PCR does not have the resolution of full-genome sequencing, and other variants of concern could be underestimated or missed with this approach. However, the time scale considered and the relatively slow evolutionary rate of SARS-CoV-2 make this approach appropriate to monitor variant spread. Furthermore, next-generation sequencing performed on 48 samples showed a strong consistency with the specific RT-PCR tests (Cohen κ of 1 for the TIB Molbiol test and 0.87 or 0.88 for the ID Solutions test depending on the variant; data not shown).

These results illustrate that variant-specific RT-PCRs are an option for SARS-CoV-2 epidemic monitoring because of their affordability and rapid deployment. They also demonstrate that SARS-CoV-2 variants spread in France was faster than anticipated (L.D. Domenico et al., unpub. data, https://www.medrxiv.org/content/10.1101/2021.02.14.21251708v1), which stresses the importance of swift public health responses.

This article was preprinted at https://www.medrxiv.org/content/10.1101/2021.02.17.251927.

Acknowledgments

We thank the ETE modelling team and Florence Débarre for discussion.

Funding was provided by the CNRS, the IRD, the ANR, and the Région Occitanie (PHYEPI grant).

About the Author

Dr. Haim-Boukobza is a virologist and head of the Department of Infectious Disease at Cerba Laboratory, Saint Ouen L'Aumône, France. Her primary research interests are viral hepatitis, human papillomaviruses, and HIV, with additional interests in gut and vaginal microbiomes.

References

1. Rambaut A, Loman N, Pybus O, Barclay W, Barrett J, Carabelli A, et al. Preliminary genomic characterisation of an emergent SARS-CoV-2 lineage in the UK defined by a novel set of spike mutations. Virological. 2021 [cited 2020 Dec 9]. https://virological.org/t/preliminary-genomic-characterisation-of-an-emergent-sars-cov-2-lineage-in-the-uk-defined-by-a-novel-set-of-spike-mutations/563
2. Tegally H, Wilkinson E, Giovanetti M, Iranzadeh A, Fonseca V, Giandhari J, et al. Emergence of a SARS-CoV-2 variant of concern with mutations in spike glycoprotein. Nature. 2021; Epub ahead of print. https://doi.org/10.1038/s41586-021-03402-9
3. Faria NR, Claro IM, Candido D, Moyses Franco L, Andrade PS, Coletti TM, et al. Genomic characterisation of an emergent SARS-CoV-2 lineage in Manaus: preliminary findings. Virological. 2021 [cited 2021 Jan 12]. https://virological.org/t/genomic-characterisation-of-an-emergent-sars-cov-2-lineage-in-manaus-preliminary-findings/586
4. Davies NG, Abbott S, Barnard RC, Jarvis CI, Kucharski AJ, Munday J, et al. Estimated transmissibility and severity of novel SARS-CoV-2 variant of concern 202012/01 in England. Science. 2021 Mar 3 [Epub ahead of print]. https://doi.org/10.1126/science.abg3055

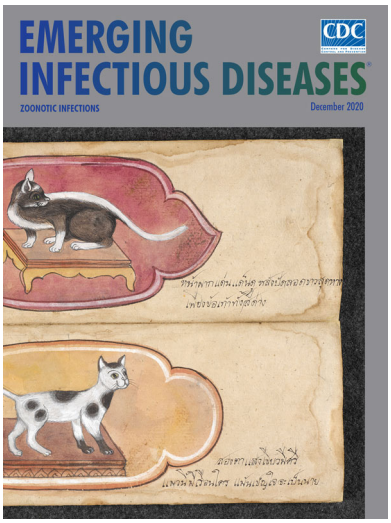
5. Thomson EC, Rosen LE, Shepherd JG, Spreafico R, da Silva Filipe A, Wojcechowskyj JA, et al.; ISARIC4C Investigators; COVID-19 Genomics UK (COG-UK) Consortium. Circulating SARS-CoV-2 spike N439K variants maintain fitness while evading antibody-mediated immunity. Cell. 2021; 184:1171–1187.e20. https://doi.org/10.1016/j.cell.2021.01.037
6. Cori A, Ferguson NM, Fraser C, Cauchemez S. A new framework and software to estimate time-varying reproduction numbers during epidemics. Am J Epidemiol. 2013;178:1505–12. https://doi.org/10.1093/aje/kwt133
7. Nishiura H, Linton NM, Akhmetzhanov AR. Serial interval of novel coronavirus (COVID-19) infections. Int J Infect Dis. 2020;93:284–6. https://doi.org/10.1016/j.ijid.2020.02.060

Address for correspondence: Samuel Alizon, MIVEGEC, CNRS, IRD, Université de Montpellier, 911 av. Agropolis, 34394 Montpellier CEDEX 5, France; email: samuel.alizon@cnrs.fr

December 2020

Zoonotic Infections

- Outbreak of Anthrax Associated with Handling and Eating Meat from a Cow, Uganda, 2018
- *Mycoplasma bovis* Infections in Free-Ranging Pronghorn, Wyoming, USA
- Control and Prevention of Anthrax, Texas, 2019
- Animal Rabies Surveillance, China, 2004–2018
- Small Particle Aerosol Exposure of African Green Monkeys to MERS-CoV as a Model for Highly Pathogenic Coronavirus Infection
- Coronavirus Disease Model to Inform Transmission-Reducing Measures and Health System Preparedness, Australia
- Genomic Epidemiology of Severe Acute Respiratory Syndrome Coronavirus 2, Colombia
- SARS-CoV-2 Seroprevalence among Healthcare, First Response, and Public Safety Personnel, Detroit Metropolitan Area, Michigan, USA, May–June 2020
- Flight-Associated Transmission of Severe Acute Respiratory Syndrome Coronavirus 2 Corroborated by Whole-Genome Sequencing
- Risk for Hepatitis E Virus Transmission by Solvent/Detergent–Treated Plasma
- Game Animal Density, Climate, and Tick-Borne Encephalitis in Finland, 2007–2017
- Trends in Population Dynamics of Escherichia coli Sequence Type 131, Calgary, Alberta, Canada 2006–2016 G. Peirano et al. 2907
- Outbreak of Haff Disease along the Yangtze River, Anhui Province, China, 2016



- Equine-Like H3 Avian Influenza Viruses in Wild Birds, Chile
- Clinical and Multimodal Imaging Findings and Risk Factors for Ocular Involvement in a Presumed Waterborne Toxoplasmosis Outbreak, Brazil
- Tuberculosis among Children and Adolescents at HIV Treatment Centers in Sub-Saharan Africa
- Human-Pathogenic Kasokero Virus in Field-Collected Ticks
- Characterization and Source Investigation of Multidrug-Resistant *Salmonella Anatum* from a Sustained Outbreak, Taiwan
- Human Monocytic Ehrlichiosis, Mexico City, Mexico
- Hantavirus Cardiopulmonary Syndrome in Canada
- Outbreaks of H5N6 Highly Pathogenic Avian Influenza (H5N6) Virus Subclade 2.3.4.4h in Swans, Xinjiang, Western China, 2020
- Differential Tropism of SARS-CoV and SARS-CoV-2 in Bat Cells
- Highly Pathogenic Avian Influenza A(H7N3) Virus in Poultry, United States, 2020
- Sensitive Detection of SARS-CoV-2–Specific Antibodies in Dried Blood Spot Samples
- Antibody Profiles According to Mild or Severe SARS-CoV-2 Infection, Atlanta, Georgia, USA, 2020
- Experimental Infection of Cattle with SARS-CoV-2
- Susceptibility of Raccoon Dogs for Experimental SARS-CoV-2 Infection
- Coyotes as Reservoirs for *Onchocerca lupi*, United States, 2015–2018
- Direct Transmission of Severe Fever with Thrombocytopenia Syndrome Virus from Domestic Cat to Veterinary Personnel
- Endovascular Infection with *Kingella kingae* Complicated by Septic Arthritis in Immunocompromised Adult Patient
- Lymphocytic Choriomeningitis Virus Infections and Seroprevalence, Southern Iraq
- Range Expansion of Bombali Virus in *Mops condylurus* Bats, Kenya
- Novel Rickettsia Species Infecting Dogs, United States
- Detection and Characterization of Bat Sarbecovirus Phylogenetically Related to SARS-CoV-2, Japan

EMERGING INFECTIOUS DISEASES

To revisit the December 2020 issue, go to:
https://wwwnc.cdc.gov/eid/articles/issue/26/12/table-of-contents

Article DOI: https://doi.org/10.3201/eid2705.210397

Detecting Rapid Spread of SARS-CoV-2 Variants, France, January 26–February 16, 2021

Appendix 2

Supplementary Methods

Generalized Linear Model (GLM)

The main model was performed by using a GLM assuming a binomial distribution, where the variable of interest was the binary variable strain (i.e., wild type or variant) and the explanatory variables were the sampling date (integer), the individual age (integer), the test kit used (boolean), the location of the sampling (boolean), and the region (factor). We further added an interaction between the region and the date. Odds ratios were computed by estimating a likelihood profile. We use a type II error for the analysis of variance given the uneven sampling between regions (the analysis of variance function in the car package of R).

Logistic Growth Fitting

We used the fitted values of the GLM model applied to the data after removing samples that came from hospitals (the sampling location effect was also obviously removed from the model) to perform the inference of a 2-parameter logistic growth kinetic curve: $f(t) = \left(1 + e^{-\sigma(t-\tau)}\right)^{-1}$, where $f(t)$ is the frequency of the variants in the new infections at time t, σ is the relative growth rate of the variants and τ 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 (*1*; E. Volz et al., unpub. data, https://www.medrxiv.org/content/10.1101/2020.12.30.20249034v2).

The parameter estimation was performed by using the drc package in R at the national and regional level (for regions with ≥1,000 samples). The CI of the fitted curves rely on those of the estimated date of reaching half proportion of new infections (τ). The unitless estimated

transmission advantage, ETA, is expressed in terms of multiplicative gain in reproduction number with respect to that of the wild type, such that $R_{variant} = (1 + ETA) * R_{wildtype}$.

Its calculation was made by solving the Euler-Lotka equation ($R_{variant} \int_0^{\infty} e^{-\sigma t} w(t) dt = 1$) assuming a serial interval *w* following a Weibull distribution with a mean and SD of 4.8 and 2.3 days (*2*) and a constant equal to 1. The CI relies on those of the estimated relative growth rate.

Reproduction Number and Variant Increase

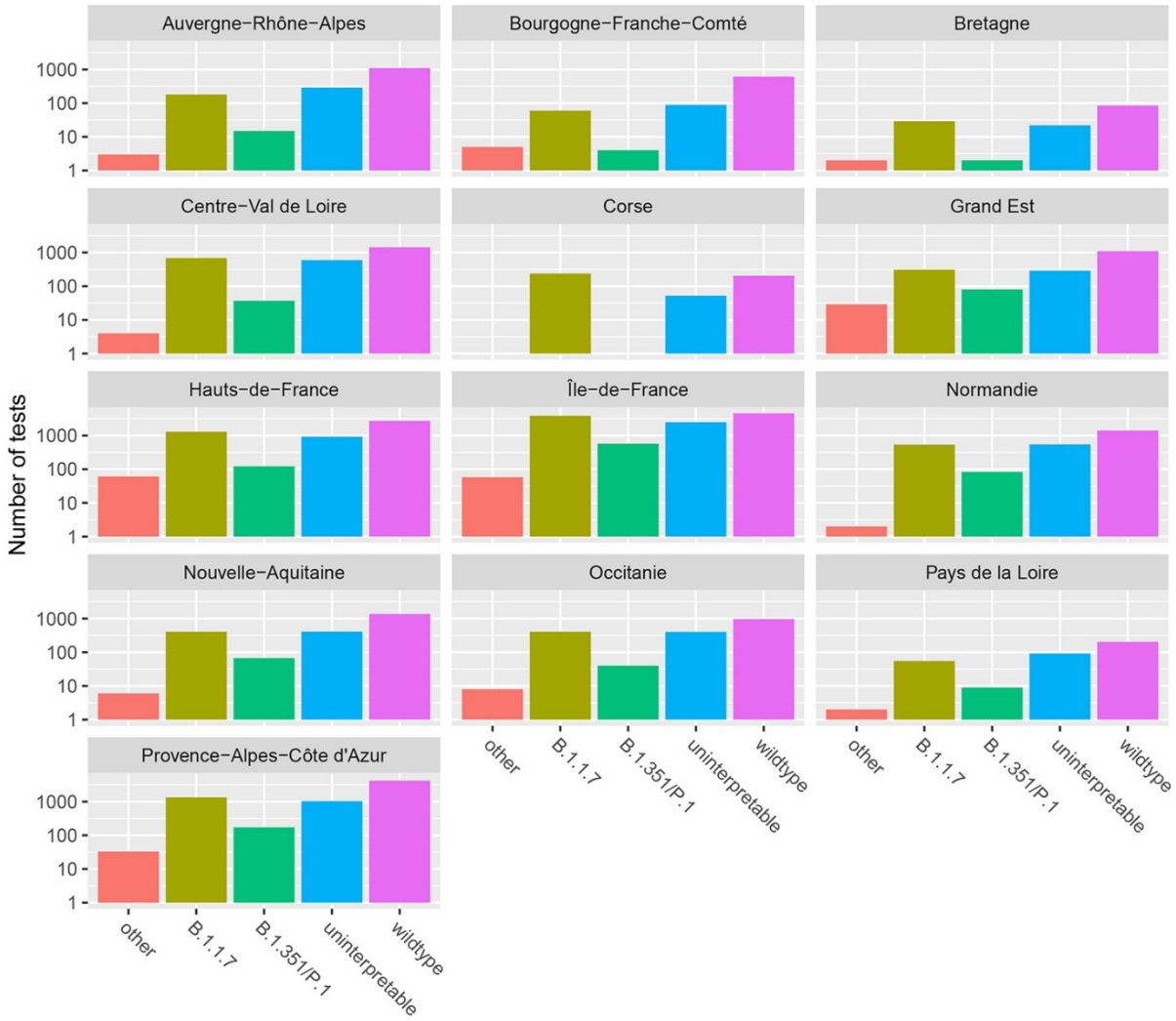
The reproduction number was estimated by using the method described in Reyné et al. (unpub. data, https://www.medrxiv.org/content/10.1101/2020.12.05.20244376v1). In brief, intensive care unit admission data was collected from https://www.data.gouv.fr/fr/datasets/donnees-hospitalieres-relatives-a-lepidemie-de-covid-19 and the reproduction number was then computed for each region in France by using the EpiEstim package in R by setting the serial interval to that reported by Nishiura et al. (*2*).

To estimate the increase in the proportion of variants among positive tests, we performed a GLM with a binomial distribution to explain the type of infection (wild type or variant) as a function of 2 factors (sampling date and individual age). This was done only on data collected outside hospital settings. The regression coefficients were used to perform a Spearman's rank correlation with the most recent time point reproduction number estimate (February 16, 2021) (Appendix 2 Figure 4).

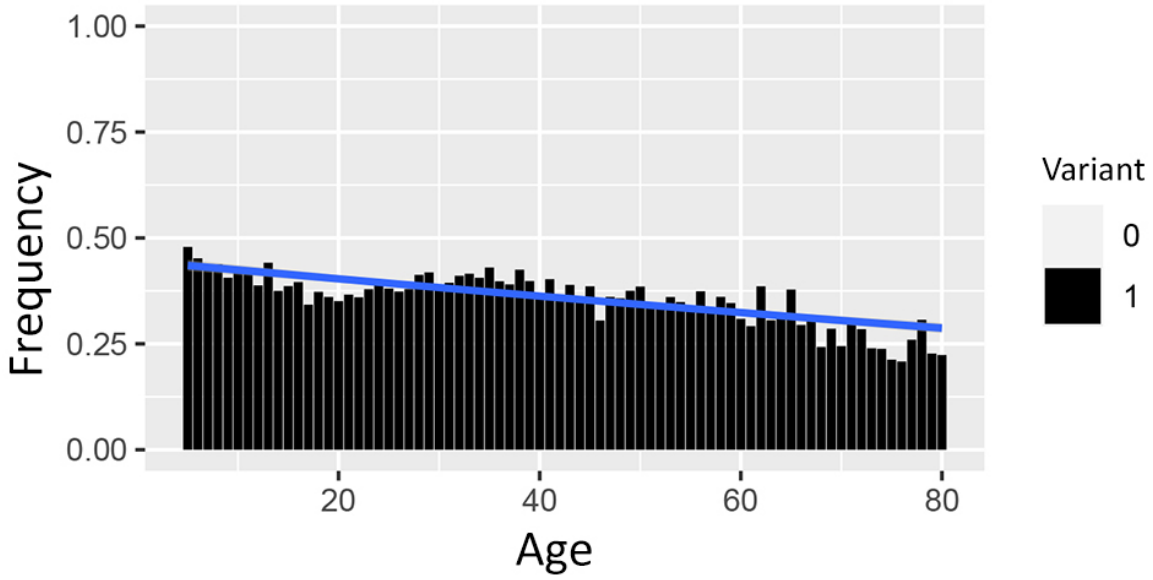
References

1. Davies NG, Abbott S, Barnard RC, Jarvis CI, Kucharski AJ, Munday J, et al. Estimated transmissibility and severity of novel SARS-CoV-2 variant of concern 202012/01 in England. Science. 2021;eabg3055 [Epub ahead of print]. https://doi.org/10.1126/science.abg3055

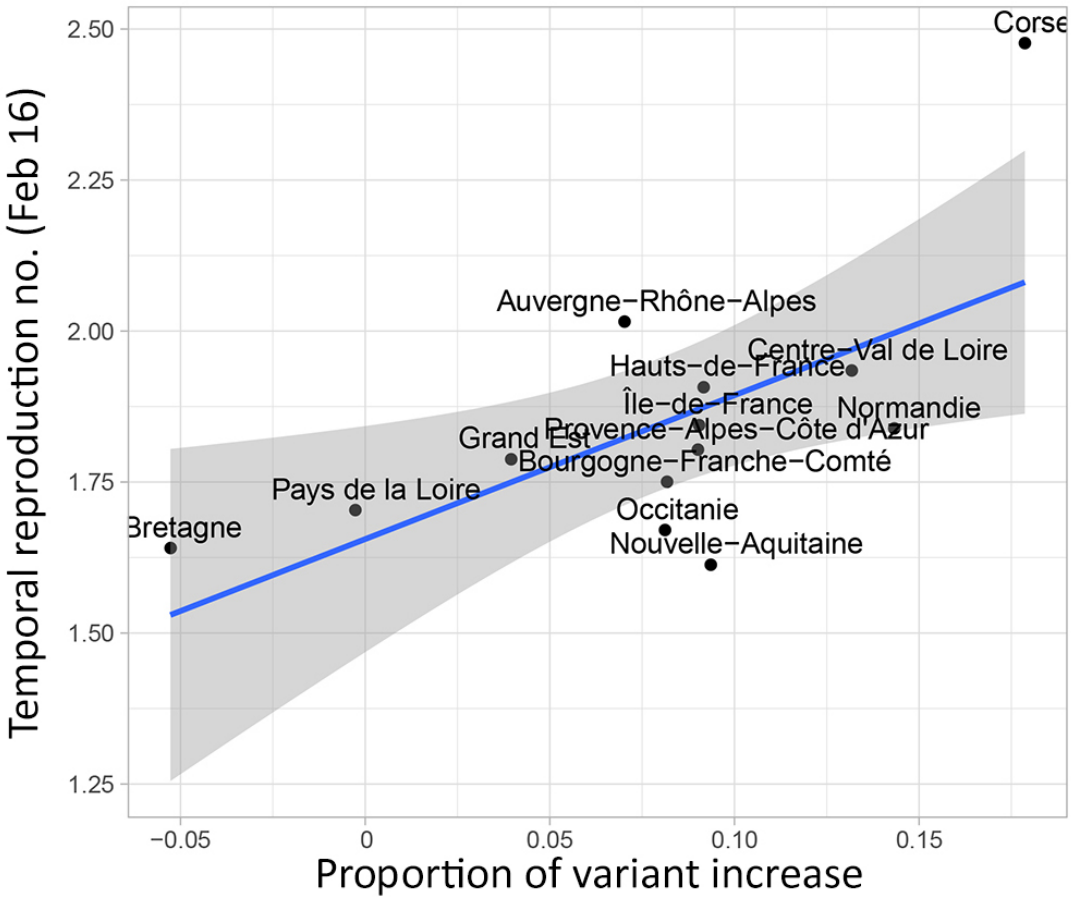
2. Nishiura H, Linton NM, Akhmetzhanov AR. Serial interval of novel coronavirus (COVID-19) infections. Int J Infect Dis. 2020;93:284–6. https://doi.org/10.1016/j.ijid.2020.02.060 PubMed



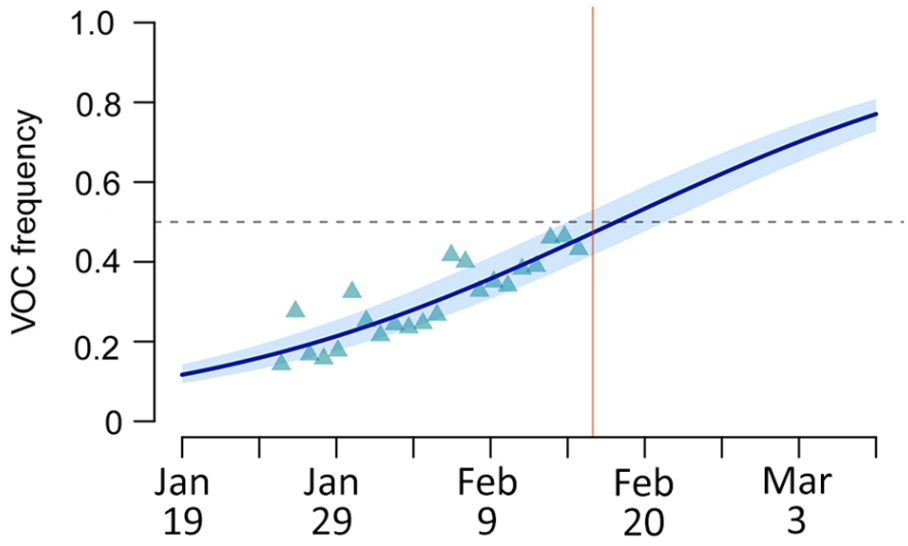
Appendix 2 Figure 1. Results of severe acute respiratory syndrome coronavirus 2 variant-specific reverse transcription PCR test results per geographic region, France, January 26–February 16, 2021.



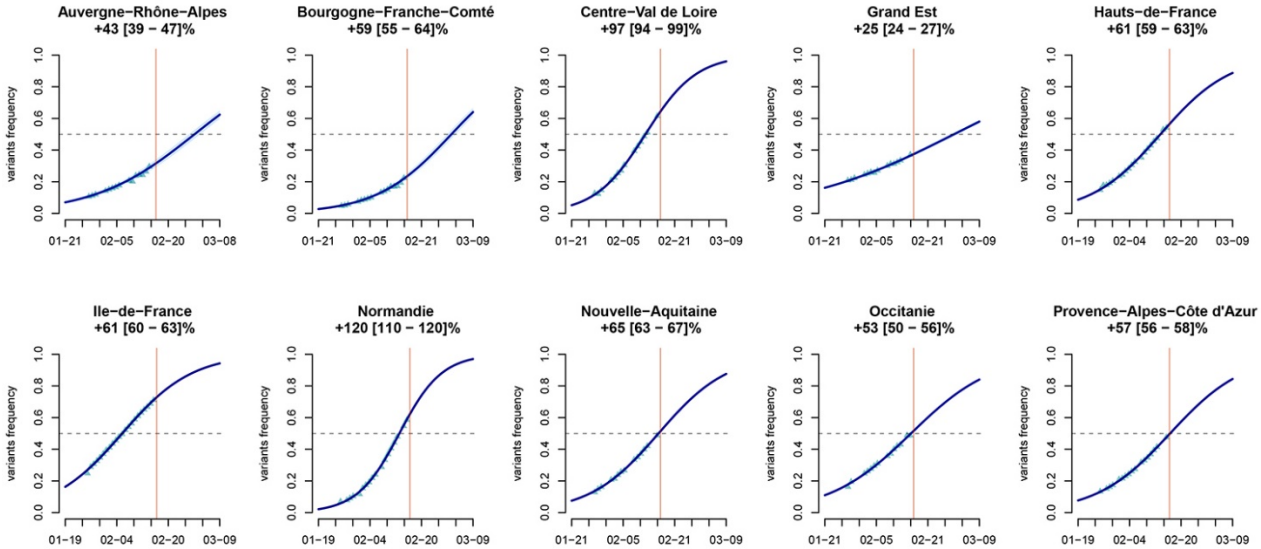
Appendix 2 Figure 2. Proportion of severe acute respiratory syndrome coronavirus 2 infections caused by variants as a function of age, France, January 26–February 16, 2021. The blue line is the trend obtained using a univariate generalized linear model. Uninterpretable tests as well as age classes <5 years or >80 years are not shown.



Appendix 2 Figure 3. Regional reproduction number (R_t) as a function of the estimated increase in severe acute respiratory syndrome coronavirus 2 variant frequency, France, January 26–February 16, 2021. The dashed line shows the output of a univariate linear model.



Appendix 2 Figure 4. Severe acute respiratory syndrome coronavirus 2 variants frequency, assuming that uninterpretable tests are all caused by wild-type strains, France, January 26–February 16, 2021. The value indicates the transmission spread advantage of the variants and its 95% CI.



Appendix 2 Figure 5. Logistic growth at the regional level in study of rapid spread of severe acute respiratory syndrome coronavirus 2 variants, France, January 26–February 16, 2021. The dots indicate the generalized linear model–fitted values, and the line is the output of the logistic growth model estimation. The vertical orange bar indicates the date of the analysis. The caption on the top of the figures indicate the estimated transmission advantage of the variants (with respect to the wild-type reproduction number) and its 95% CI.