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^{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 Δ69-70 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: $f(t) = \left(1 + \mathrm{e}^{-\rho(t-\tau)}\right)^{-1}$, where $f(t)$ is the frequency of the variants in the new infections at time t, ρ 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 (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 date of reaching half proportion of new infections (τ).

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 $\mathcal{R}_{\text{variant}} = (1 + \text{ETA}) \times \mathcal{R}_{\text{wild type}}$. Its calculation was made by solving the Euler-Lotka equation ($\mathcal{R}_{\text{variant}} \int_0^\infty \mathrm{e}^{-\rho t} w(t)\, \mathrm{d}t = 1$) 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 $\mathcal{R}_{\text{wild type}}$ 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 $p_{\mathrm{x}} \times p_{\mathrm{x}}^T$.

Supplementary figures

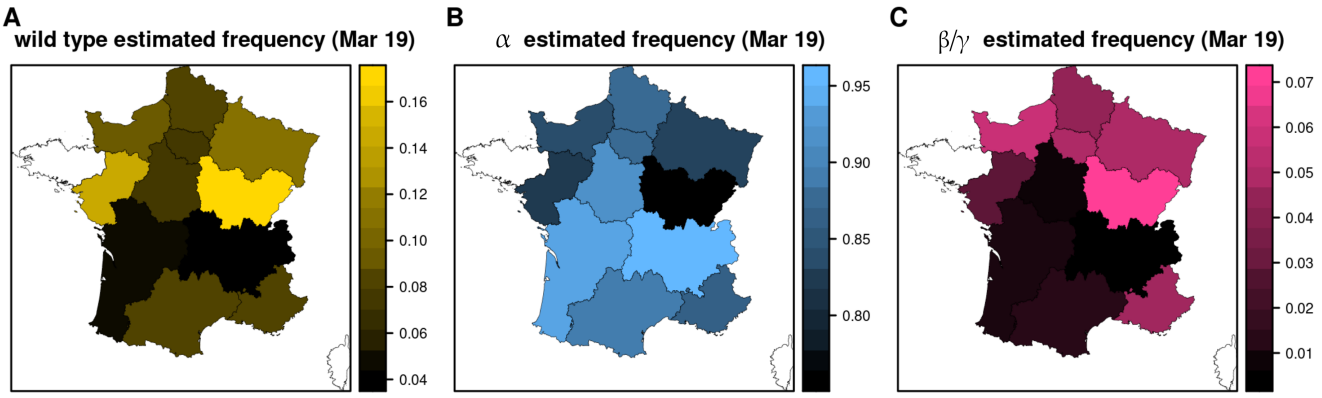

Figure S1: **Estimated proportion of new infections caused by A) wild type, B) Alpha variant, and C) Beta or Gamma variants on Mar 12, 2021, in French regions.** Regions with insufficient sampling are in white.

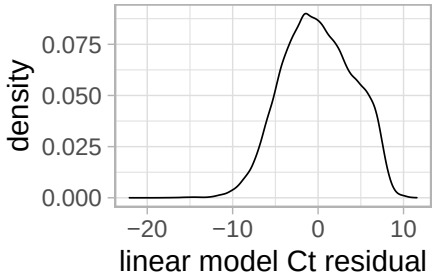

Figure S2: **Distribution of the residual values of the multiple regression linear model.**

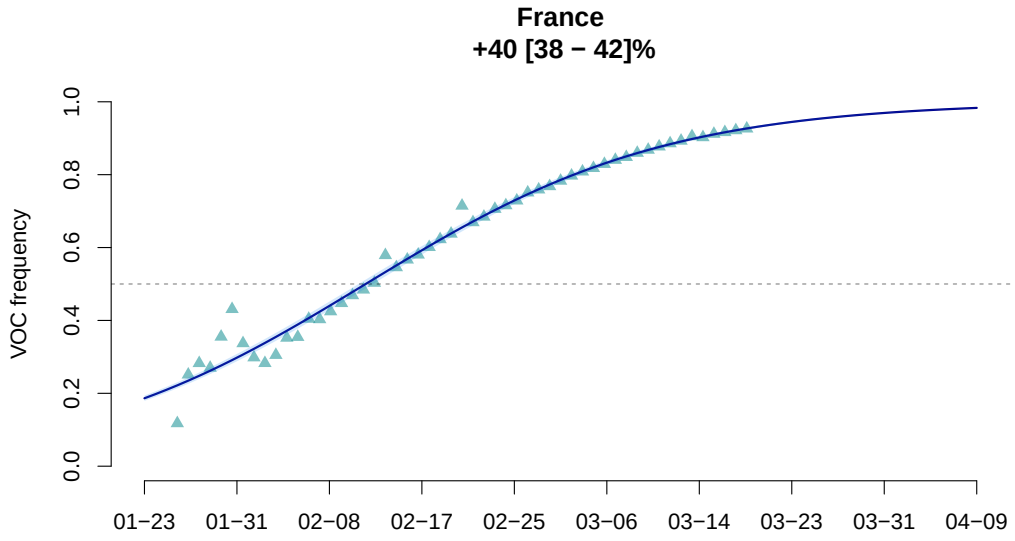

Figure S3: **Estimating the transmission advantage of the Alpha variant over the wild type strain.** The dots indicate the GLM-fitted values values and the line is the output of the logisitic growth model estimation. The top figures indicate the estimated transmission advantage of the Alpha variant (with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

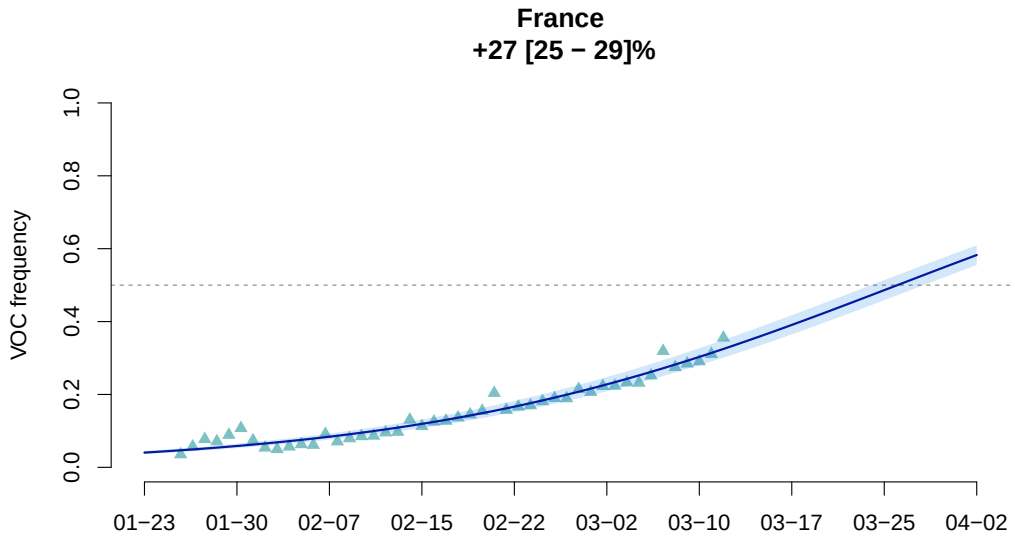

Figure S4: **Estimating the transmission advantage of the Beta or Gamma variants over the wild type strain.** See Figure S3 for details.

Additional bibliography

Nishiura H, Linton NM, Akhmetzhanov AR. Serial interval of novel coronavirus (COVID-19) infections. International Journal of Infectious Diseases. 2020 Apr;93:284-286. Available from: https://www.ijidonline.com/article/S1201-9712(20)30119-3/abstract.