

Estimating the strength of selection for new SARS-CoV-2 variants

Supporting Information

Christiaan H. van Dorp^{1,*}, Emma E. Goldberg^{1,*}, Nick Hengartner^{1,2}, Ruian Ke^{1,2}, and
Ethan O. Romero-Severson^{1,2,†}

¹Theoretical Biology and Biophysics (T-6), Los Alamos National Laboratory,
Los Alamos NM, USA

²New Mexico Consortium, Los Alamos NM, USA

*These authors contributed equally to this work

†Correspondence: eoromero@lanl.gov

March 29, 2021

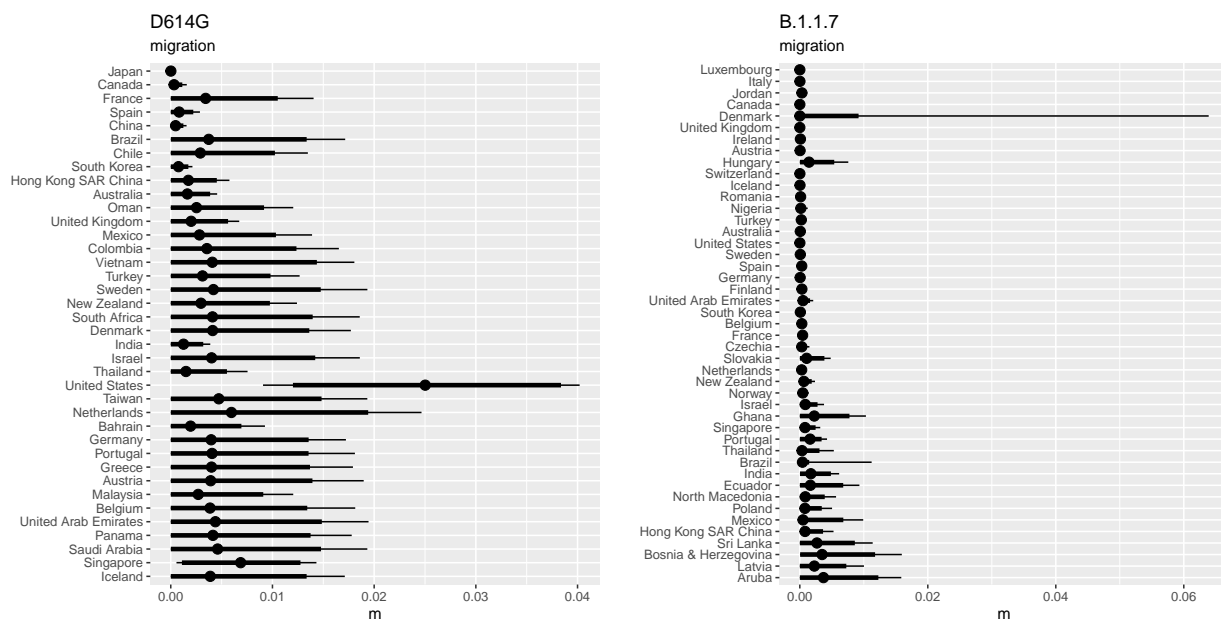


Figure S1: Migration fractions for each country, from the population genetic model. Points mark the median, and thick and thin lines are 90% and 95% credible intervals (CrI), respectively. Corresponding estimates of selection are in Fig. 3.

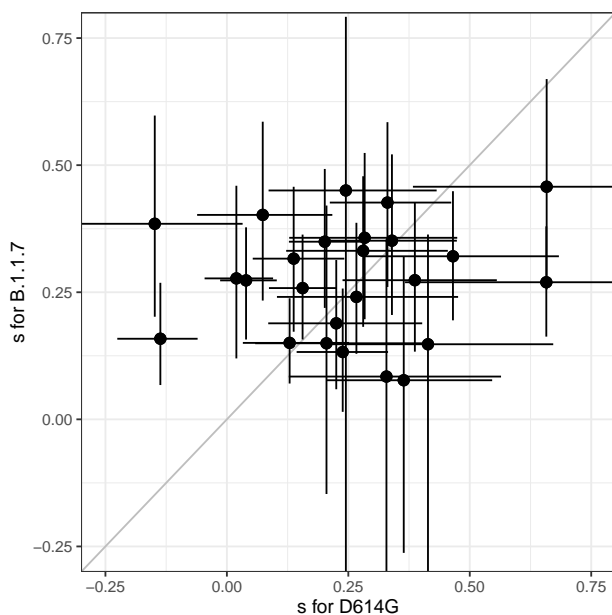


Figure S2: Estimates of the selection coefficient for the two variants, across countries. Data are as in Fig. 3, with one point for each country in which both variants were analyzed. Cross-bars show the 90% CrIs.

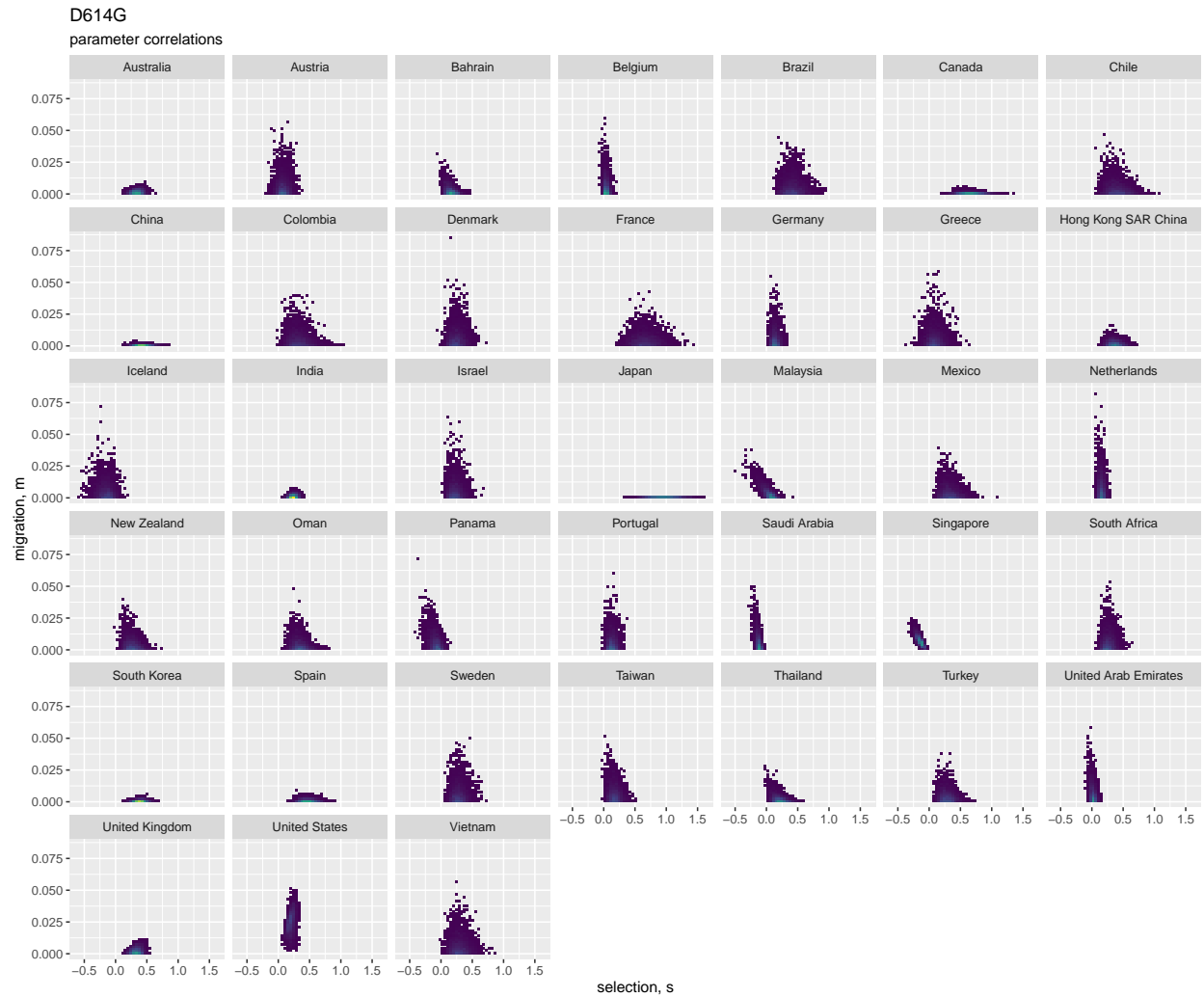


Figure S3: Correlations between the selection and migration parameters in D614G, from the population genetic model. The scatter plots represent samples from the joint posterior distribution of s and m .

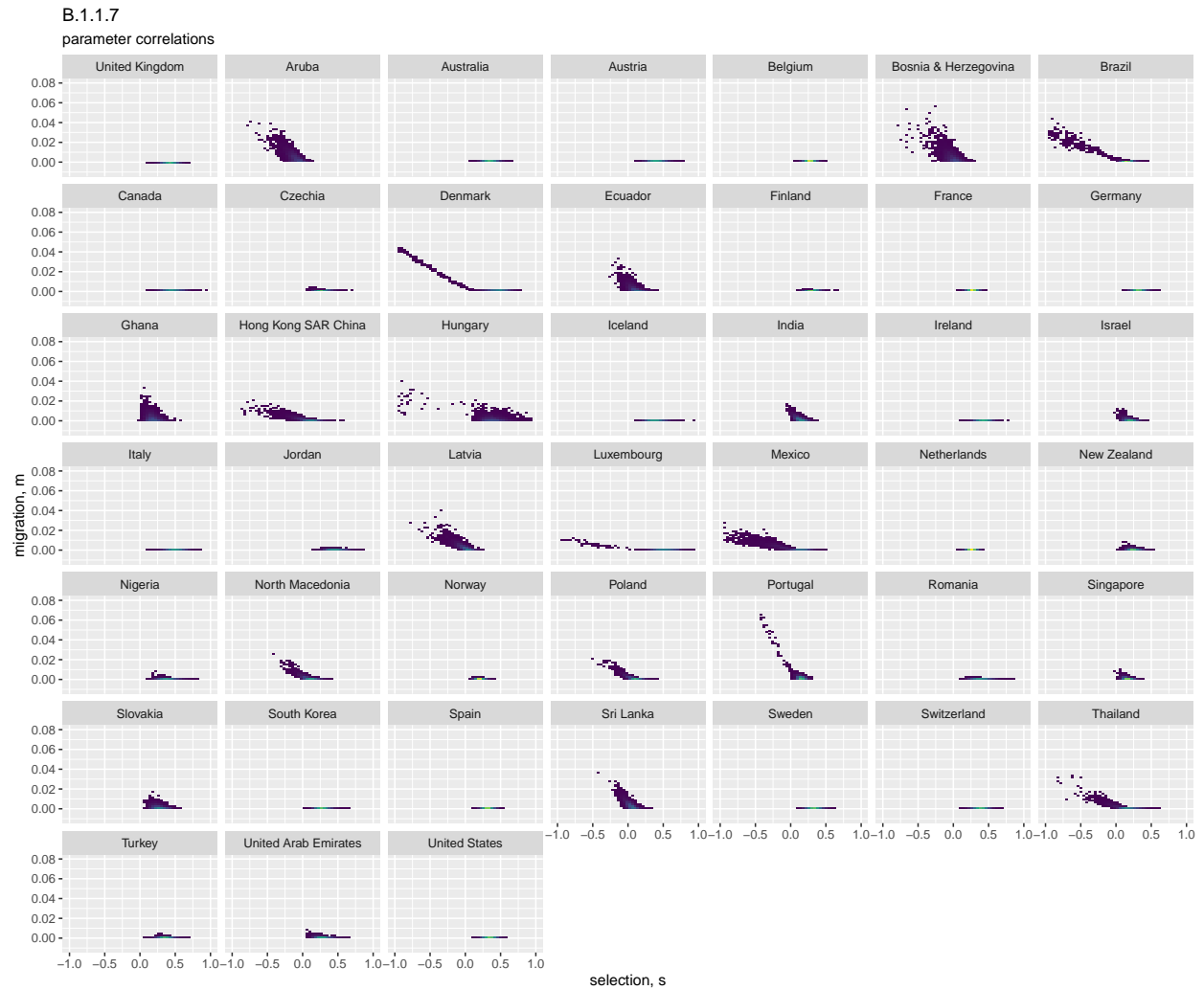


Figure S4: Correlations between the selection and migration parameters in B.1.1.7, from the population genetic model. The scatter plots represent samples from the joint posterior distribution of s and m .

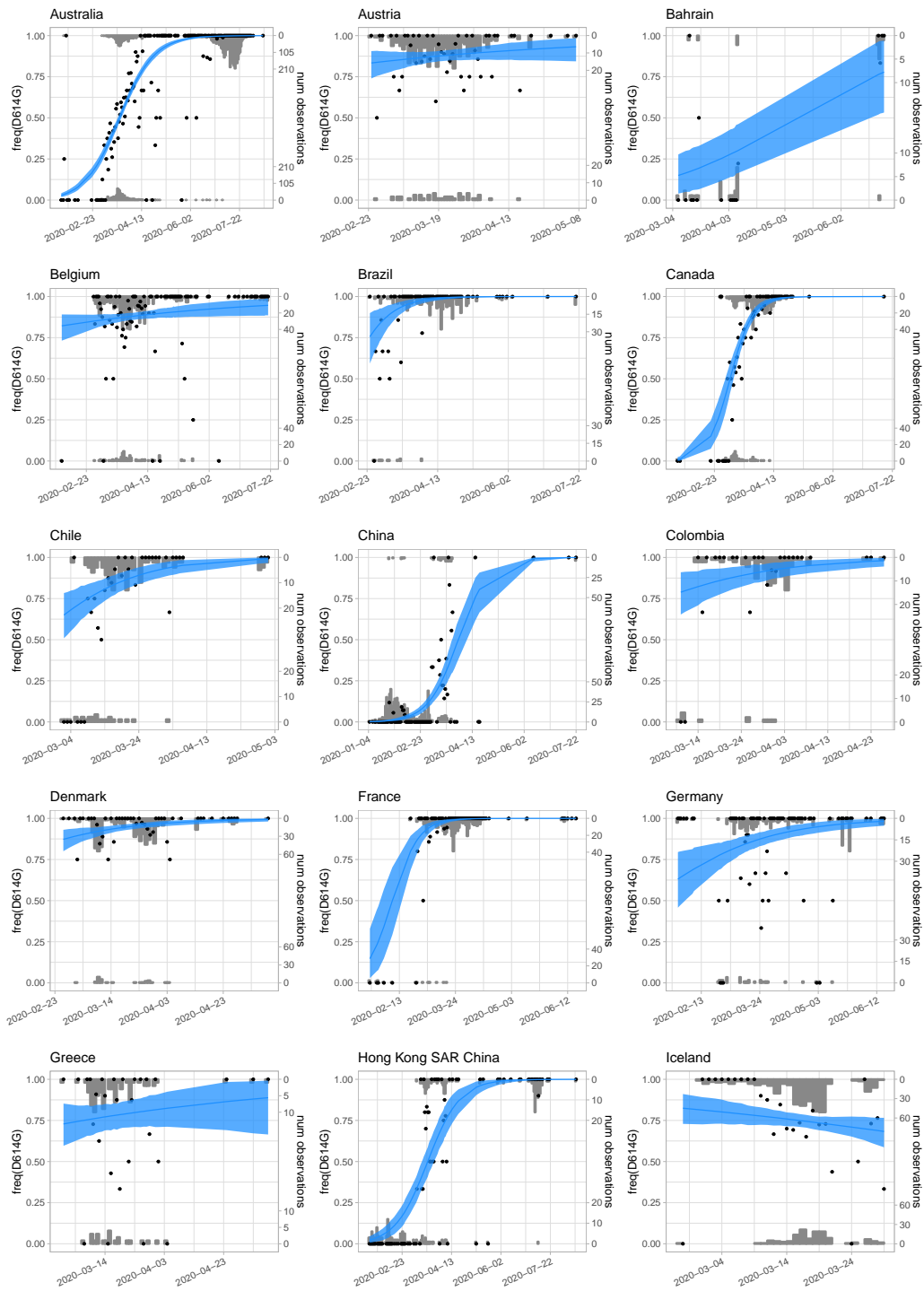


Figure S5: Fits from the population genetic model to the daily variant data for D614G. Gray bars show histograms of numbers of observations of D614G (upper axis) and all other variants (lower axis) each day. Black dots are the daily frequencies. Blue shows the frequencies predicted from the model, with 95% credible intervals shaded.

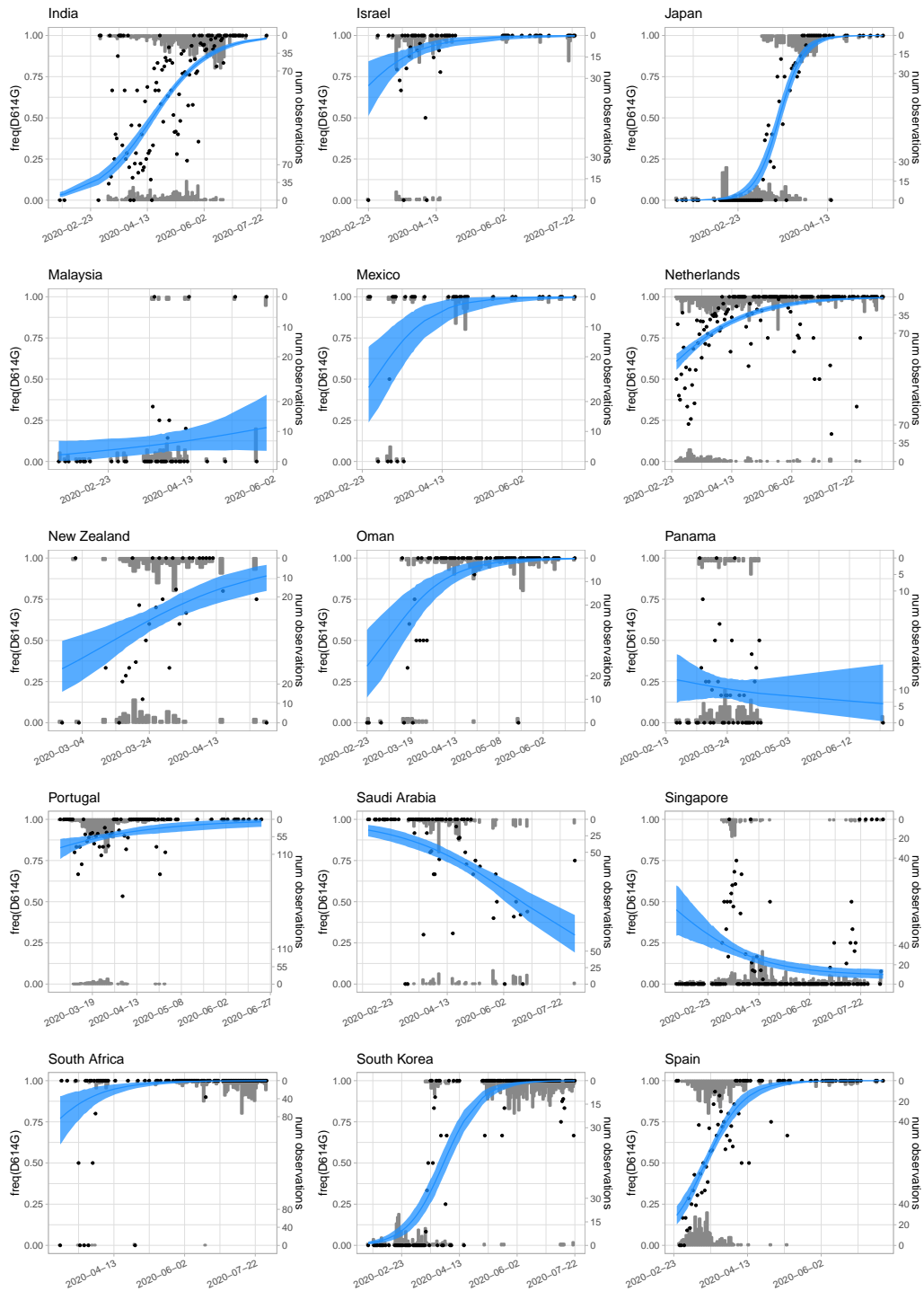


Figure S5: (continued)

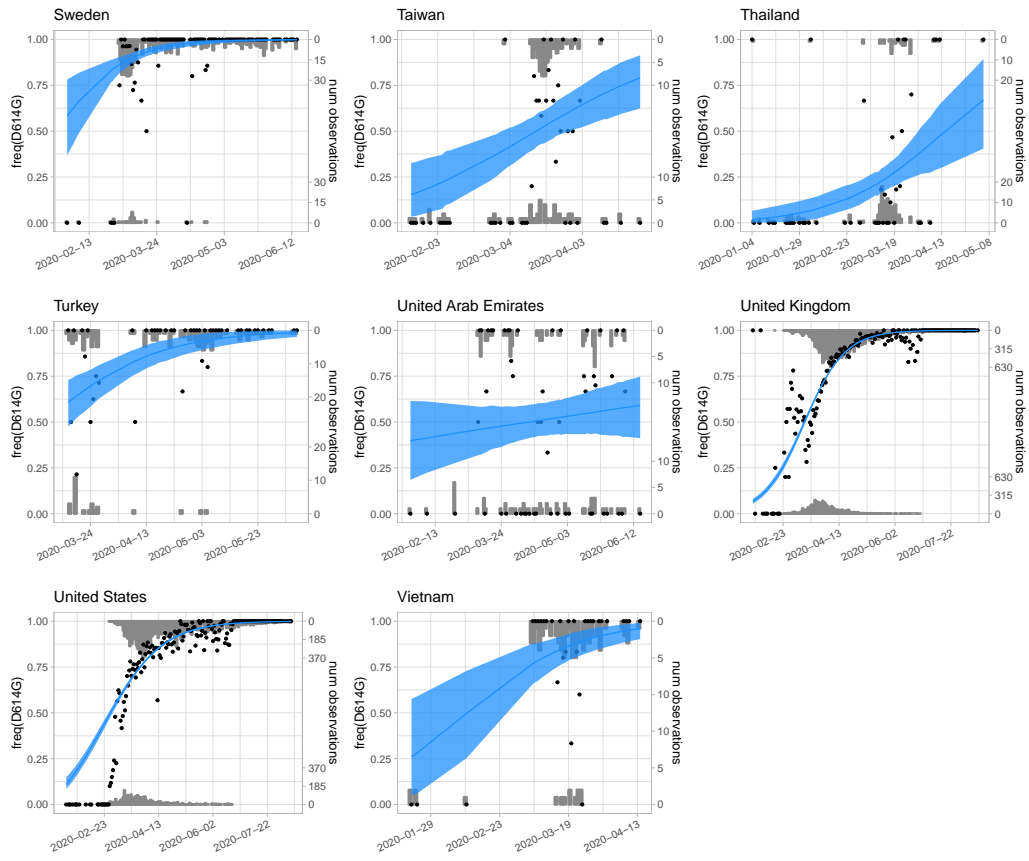


Figure S5: (continued)

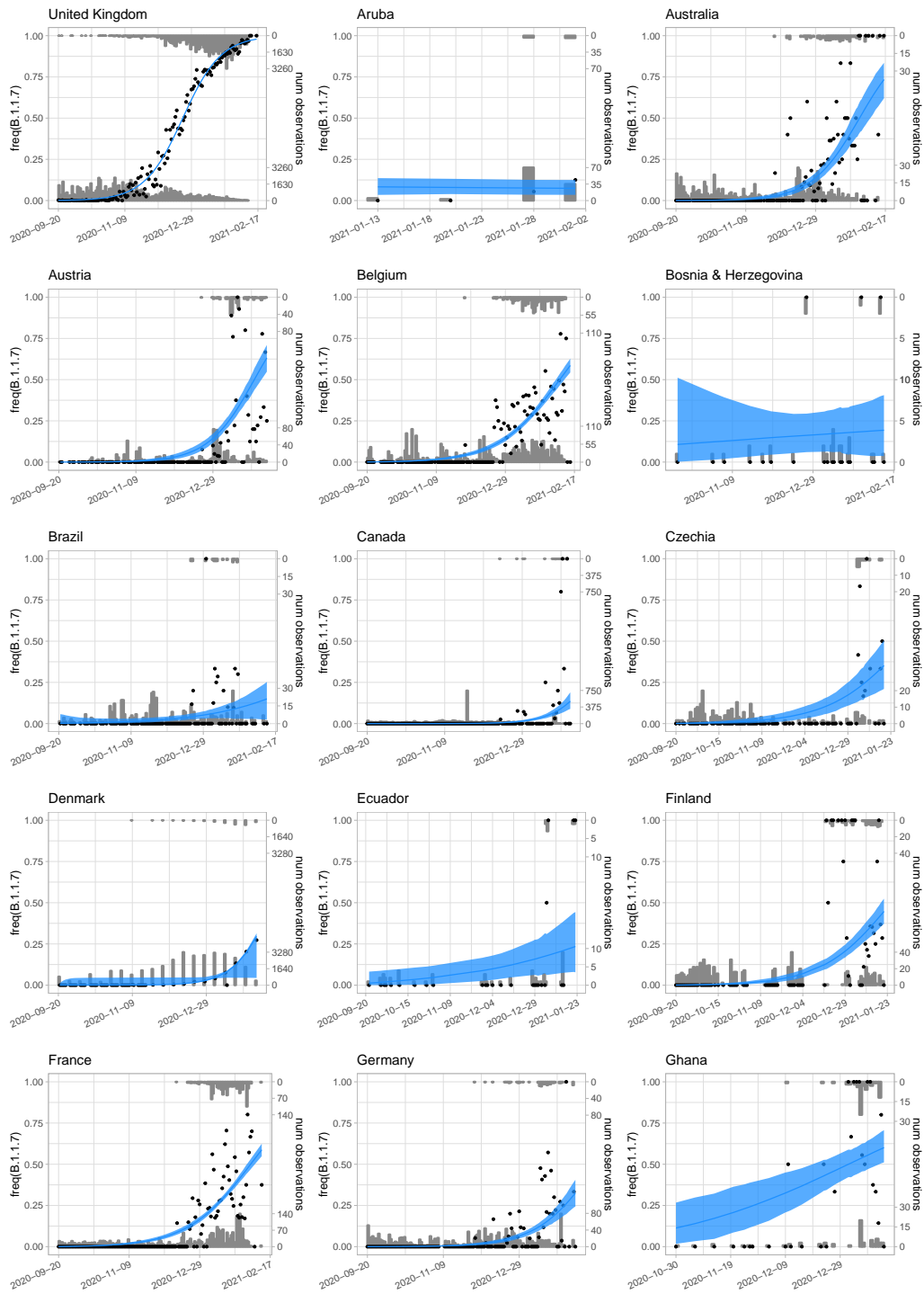


Figure S6: Fits from the population genetic model to the daily variant data for B.1.1.7. Gray bars show histograms of numbers of observations of B.1.1.7 (upper axis) and all other variants (lower axis) each day. Black dots are the daily frequencies. Blue shows the frequencies predicted from the model, with 95% credible intervals shaded.

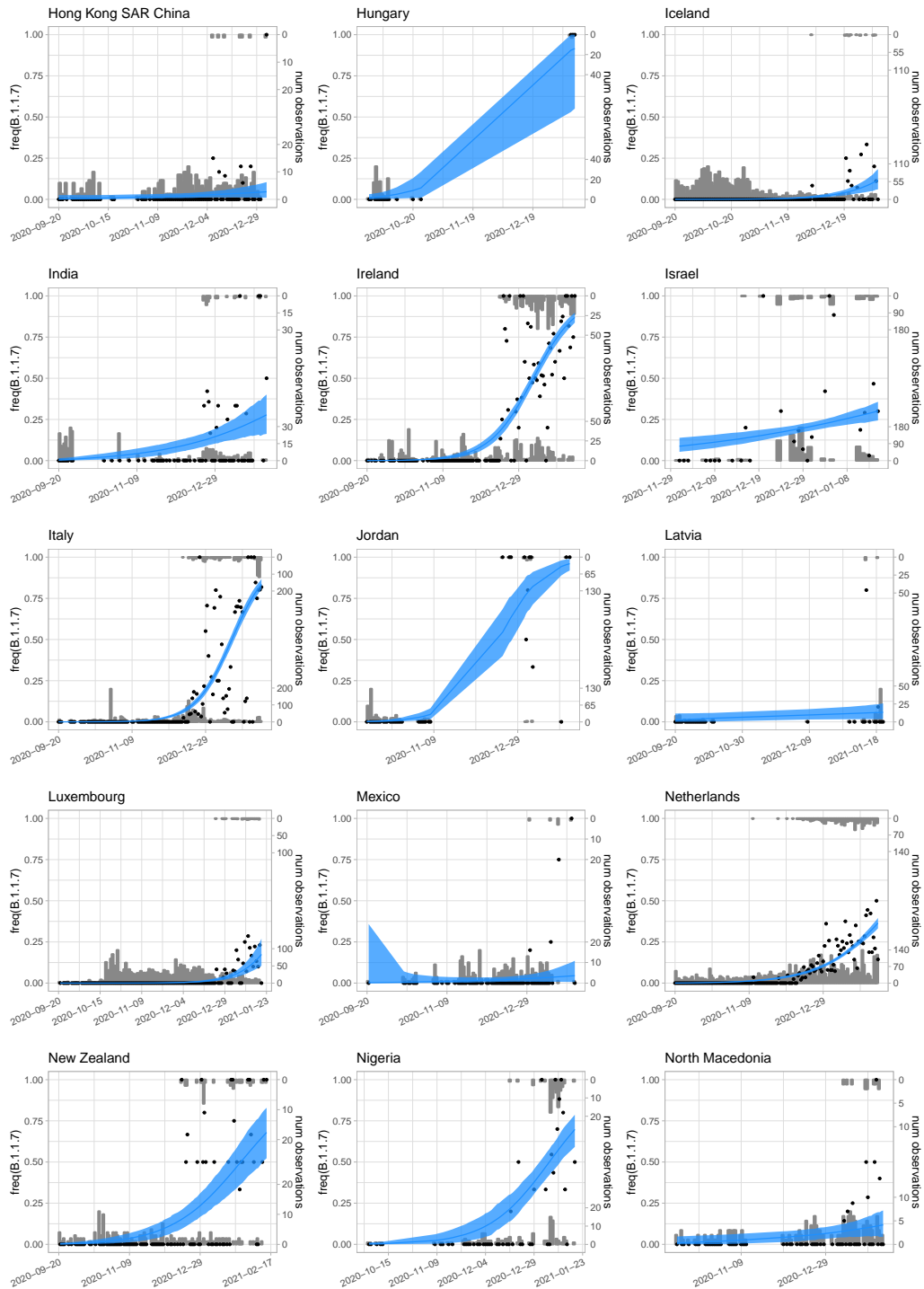


Figure S6: (continued)

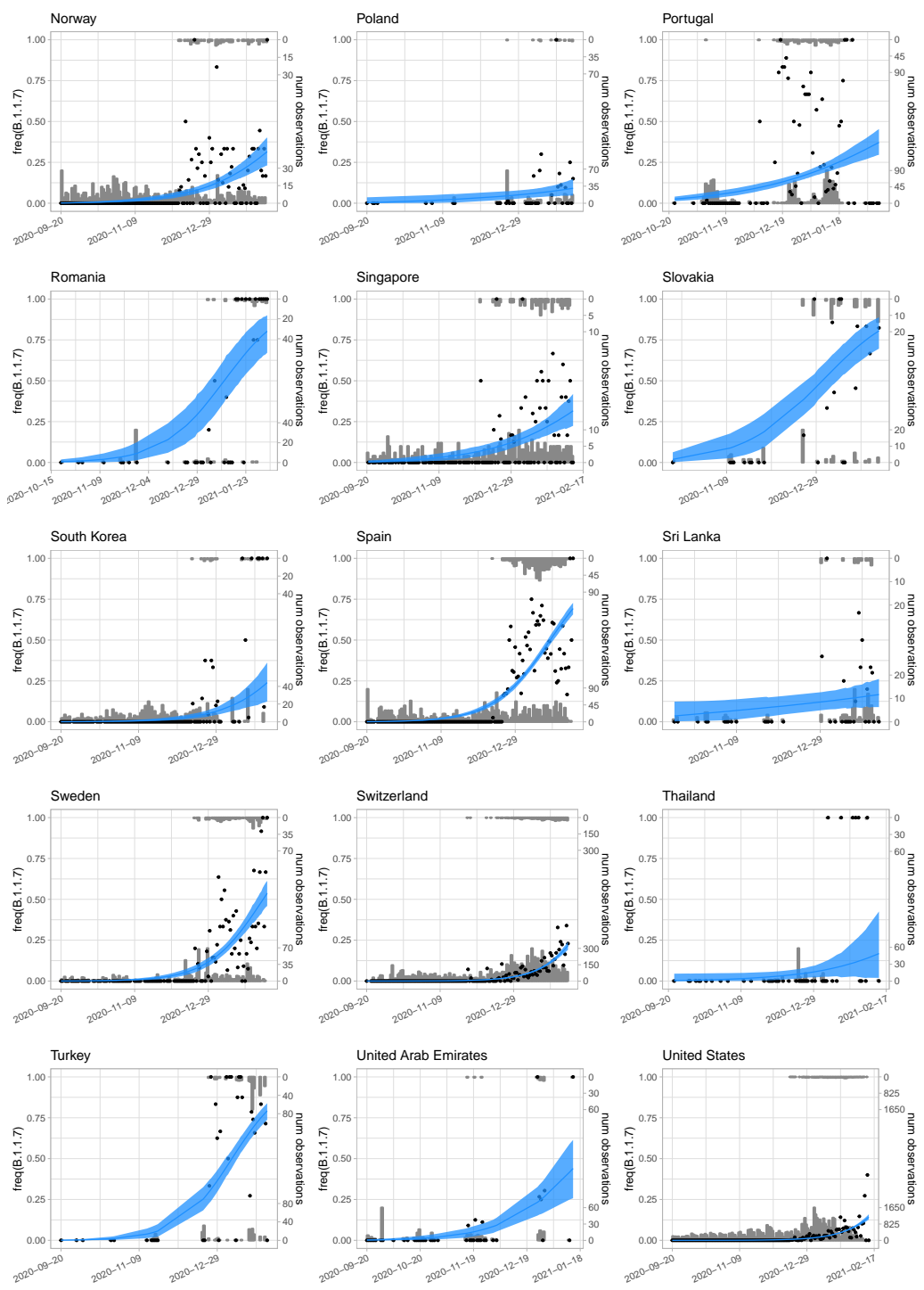


Figure S6: (continued)

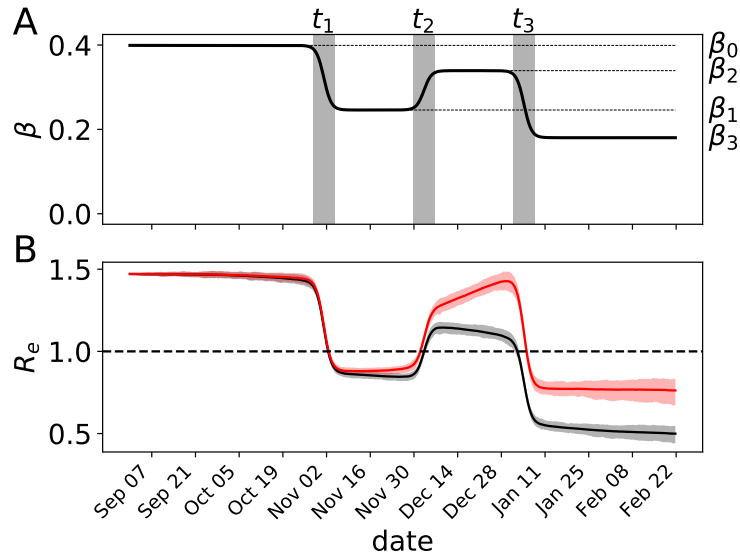


Figure S7: Inferred effective reproduction number in the UK. A. The infection rate β is time-dependent to model government responses to the pandemic. The uptake-period of changes in restrictions is shown in gray. Parameters used for this example are from the UK and variant B.1.1.7. B. Using the reconstructed trajectories of the fitted stochastic epidemic model, the instantaneous effective reproduction number (R_e) is calculated as $R_e = \tilde{\beta}/(\gamma + \nu)S/N$. The lighter bands indicate the 2.5 - 97.5 percentiles of the filtered trajectories. The thick lines indicate the mean. The effective reproduction number for the wild-type is shown in black ($\tilde{\beta} = \beta$). The red curve indicates the weighted average of the effective reproduction numbers of the wild-type and mutant ($\tilde{\beta} = \beta(1 + sI_{mt}/I)$). The threshold value of $R_e = 1$ secondary infections per infected individual is indicated by the horizontal, dashed line.

D614G				B117		
symbol	domain	sd	time domain	domain	sd	time domain
s	$[-1, \infty)$	$2 \cdot 10^{-3}$	-	.	.	.
β_0	$[0, \infty)$	$2 \cdot 10^{-3}$	$(-\infty, t_1]$.	.	.
β_1	$[0, \infty)$	$2 \cdot 10^{-3}$	$[t_1, t_2]$.	.	.
β_2	$[0, \infty)$	$2 \cdot 10^{-3}$	$[t_2, \infty)$.	.	$[t_2, t_3]$
β_3	-	-	-	$[0, \infty)$	$2 \cdot 10^{-3}$	$[t_3, \infty)$
t_1	$[60, 110]$	0.05	$(t_1 - 30, t_1 + 30)$	$[280, 320]$.	.
t_2	$[160, 210]$	0.05	$(t_2 - 30, t_2 + 30)$	$[310, 360]$.	.
t_3	-	-	-	$[340, 380]$	0.05	$(t_3 - 30, t_3 + 30)$
p_0	$[0, 1]$	10^{-2}	$\{t_0\}$.	$3 \cdot 10^{-4}$.
λ_0	-	-	-	$[0, \infty)$	10^{-4}	-
ε	$[7 \cdot 10^{-7}, 10^{-4}]$	10^{-6}	$\{t_0\}$	$[10^{-5}, 5 \cdot 10^{-3}]$	$2 \cdot 10^{-5}$.
r_D	$[10, 10^3]$	0.1	-	.	.	.
r_F	$[10, 10^3]$	0.1	-	.	.	.
τ	$[0, 0.1]$	10^{-4}	-	.	.	.

Table S1: Specifications of the iterated particle filter algorithm. The “sd” columns specify the standard deviation used for the parameter random walk ($\sqrt{\Sigma_{ii}}$). The “time domain” columns specify the time window during the simulation in which the parameter is allowed to be perturbed. A dot (.) indicates that the same value is used for both variants and a dash (-) that the value is not applicable.