Supporting information: Fitness models provide accurate short-term forecasts of SARS-CoV-2 variant frequency

Eslam Abousamra[1,2,†,*], **Marlin Figgins**[1,3,†] & **Trevor Bedford**[1,2,4]

[1]Vaccine and Infectious Disease Division, Fred Hutchinson Cancer Center, Seattle, WA, USA, [2]Department of Epidemiology, University of Washington, Seattle, WA, USA, [3]Department of Applied Mathematics, University of Washington, Seattle, WA, USA, [4]Howard Hughes Medical Institute, Seattle, WA, USA, [†]These authors contributed equally to this work., [*]To whom correspondence should be addressed: eabousam@uw.edu

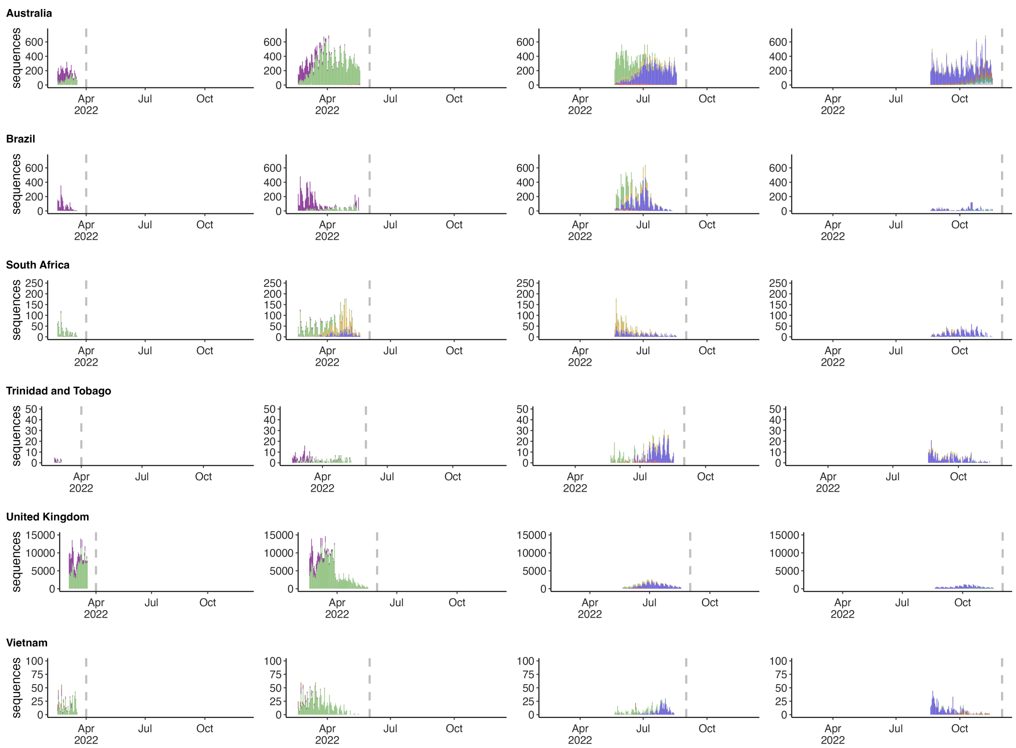

Figure S1. Reconstructing available data sets for Australia, Brazil, South Africa, Trinidad and Tobago, the United Kingdom, and Vietnam. (A) Variant sequence counts categorized by Nextstrain clade at 4 different analysis dates.

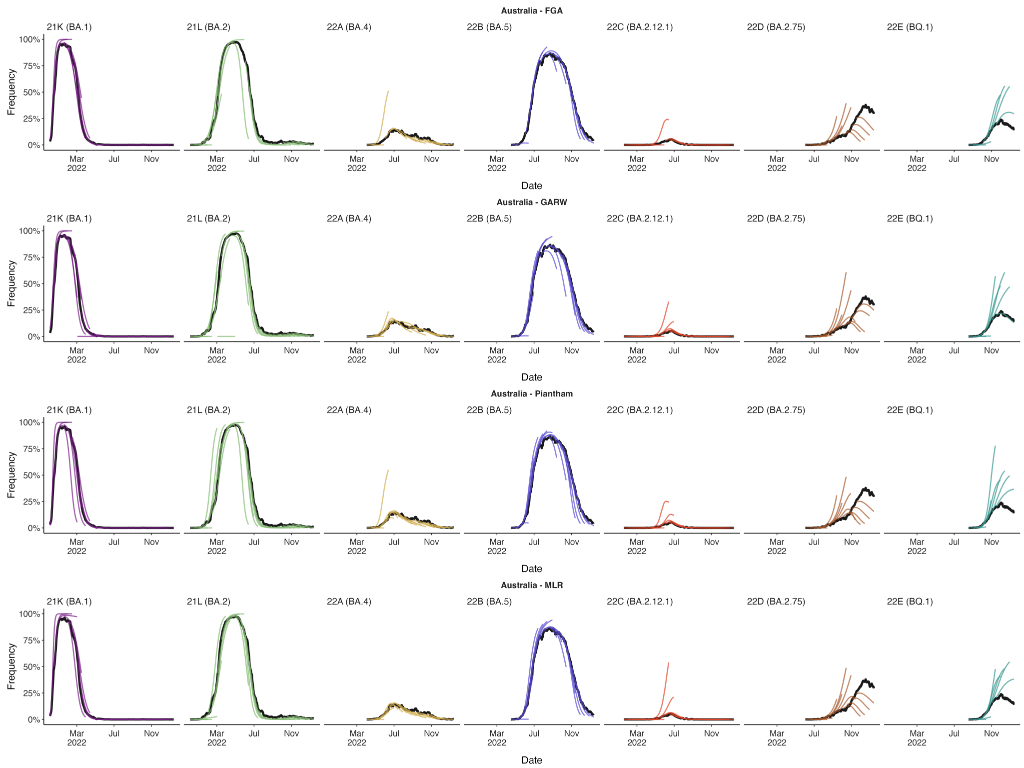

Figure S2. Reconstructing predictions for Australia (A) +30 day frequency forecasts for variants in bimonthly intervals using the MLR model for Australia. Each forecast trajectory is shown as a different colored line. Retrospective smoothed frequency is shown as a thick black line.

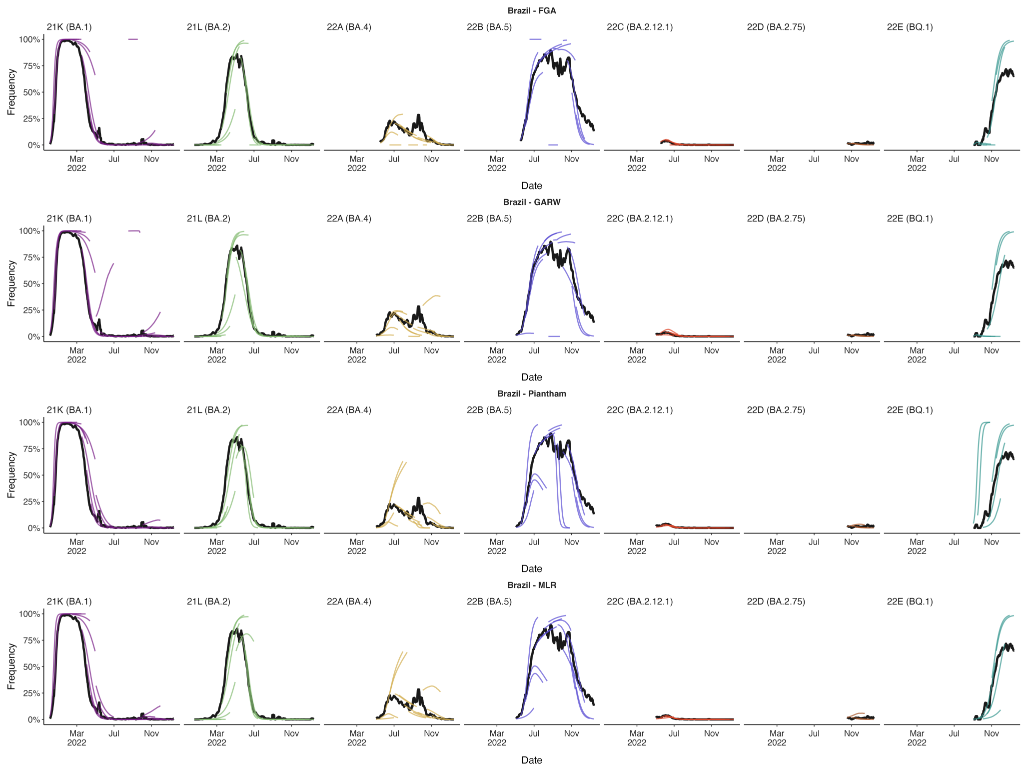

Figure S3. Reconstructing predictions for Brazil (A) +30 day frequency forecasts for variants in bimonthly intervals using the MLR model for Brazil. Each forecast trajectory is shown as a different colored line. Retrospective smoothed frequency is shown as a thick black line.

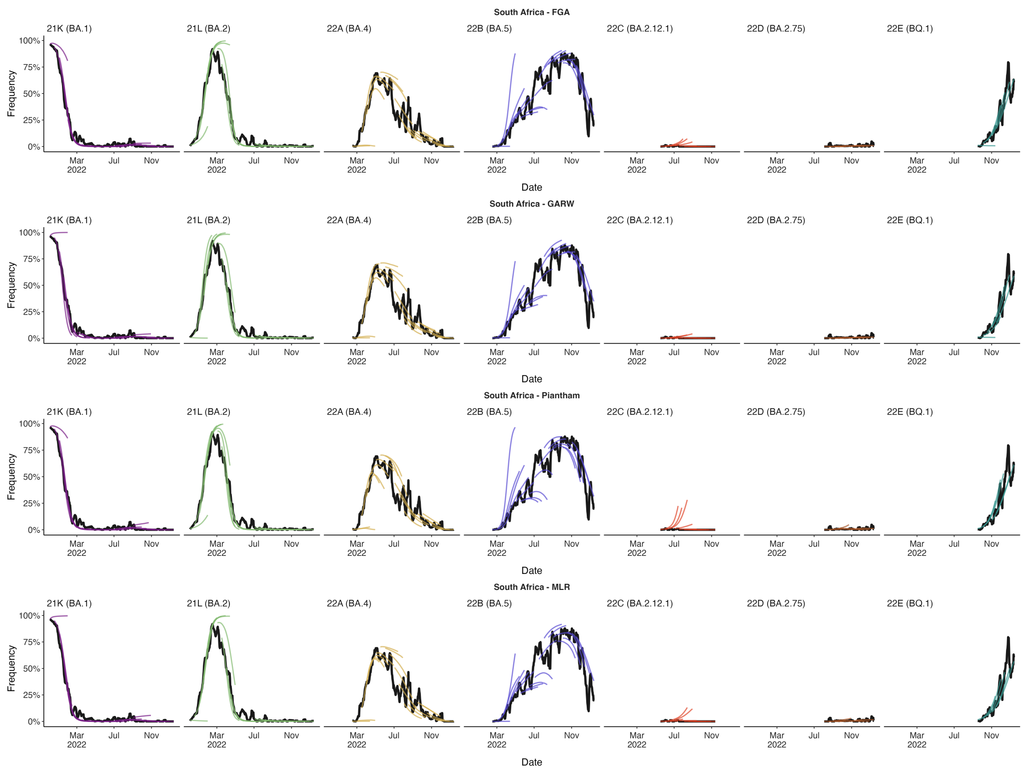

Figure S4. Reconstructing predictions for South Africa (A) +30 day frequency forecasts for variants in bimonthly intervals using the MLR model for South Africa. Each forecast trajectory is shown as a different colored line. Retrospective smoothed frequency is shown as a thick black line.

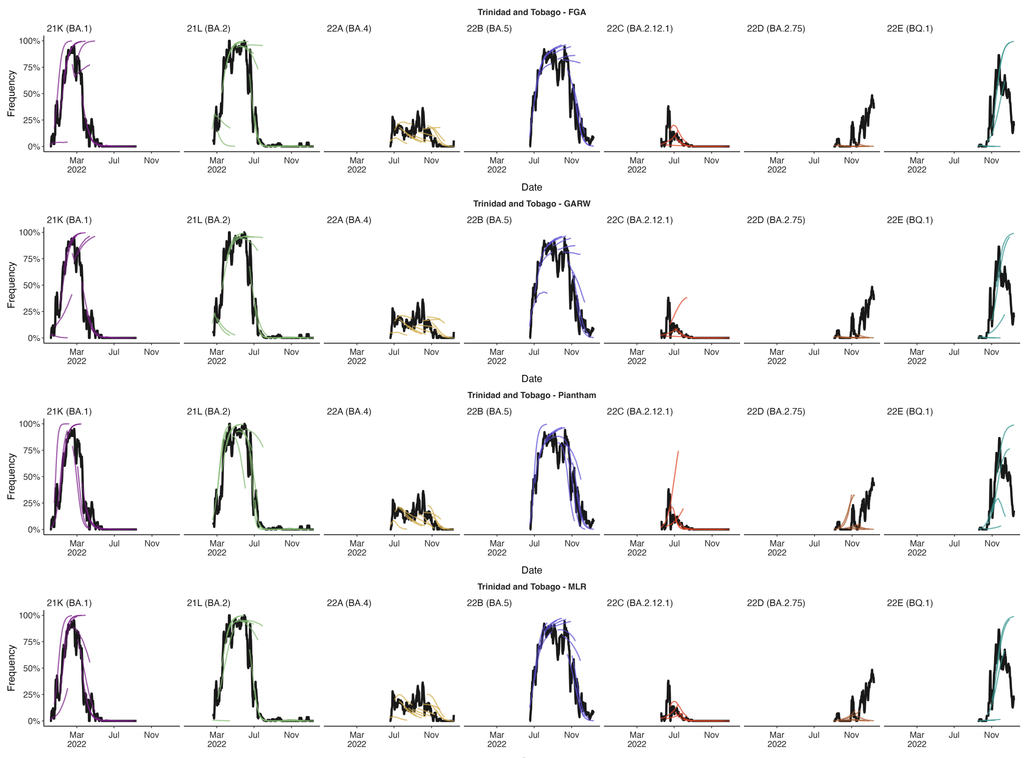

Figure S5. Reconstructing predictions for Trinidad and Tobago (A) +30 day frequency forecasts for variants in bimonthly intervals using the MLR model for Trinidad and Tobago. Each forecast trajectory is shown as a different colored line. Retrospective smoothed frequency is shown as a thick black line.

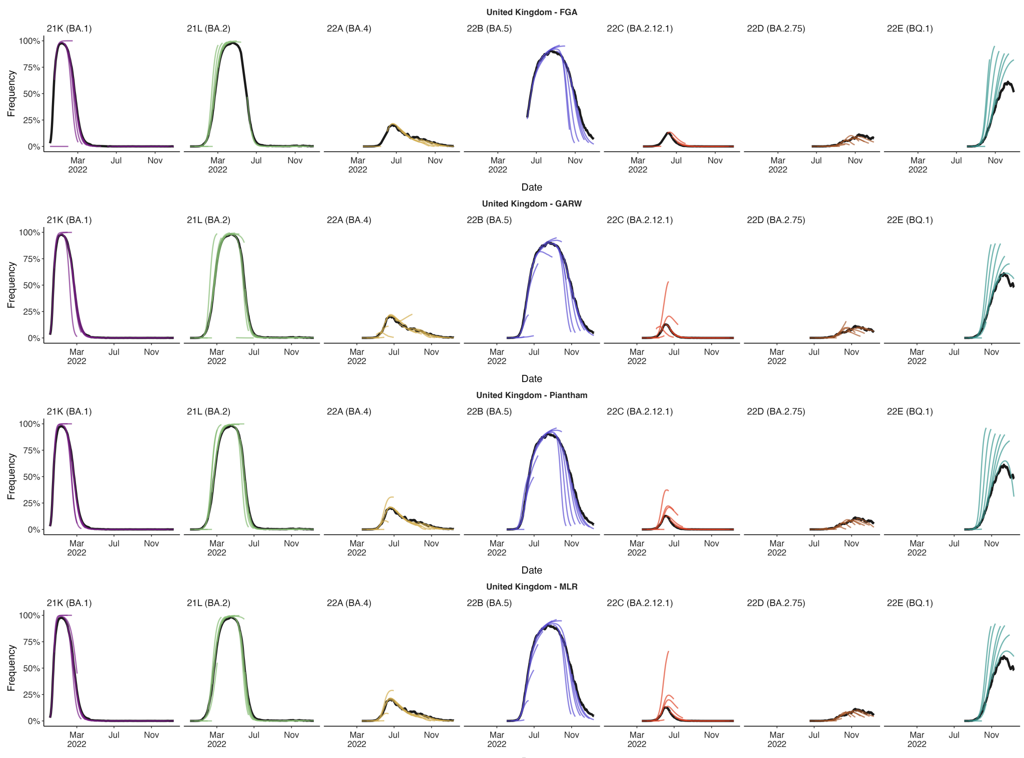

Figure S6. Reconstructing predictions for United Kingdom (A) +30 day frequency forecasts for variants in bimonthly intervals using the MLR model for United Kingdom. Each forecast trajectory is shown as a different colored line. Retrospective smoothed frequency is shown as a thick black line.

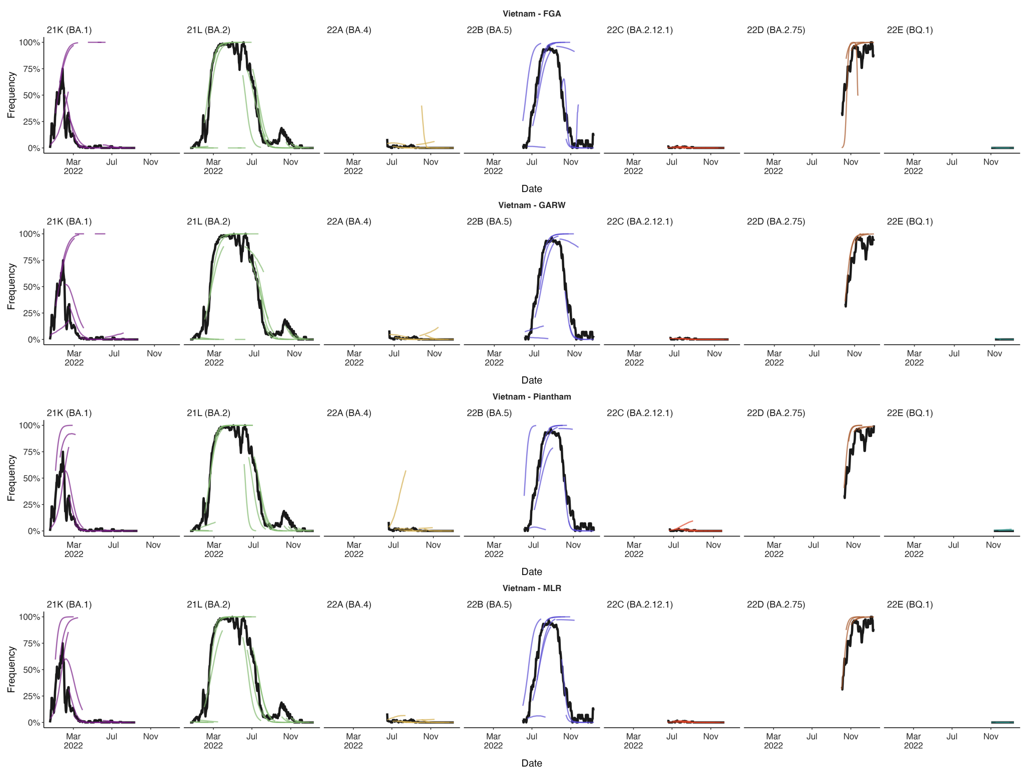

Figure S7. Reconstructing predictions for Vietnam (A) +30 day frequency forecasts for variants in bimonthly intervals using the MLR model for Vietnam. Each forecast trajectory is shown as a different colored line. Retrospective smoothed frequency is shown as a thick black line.

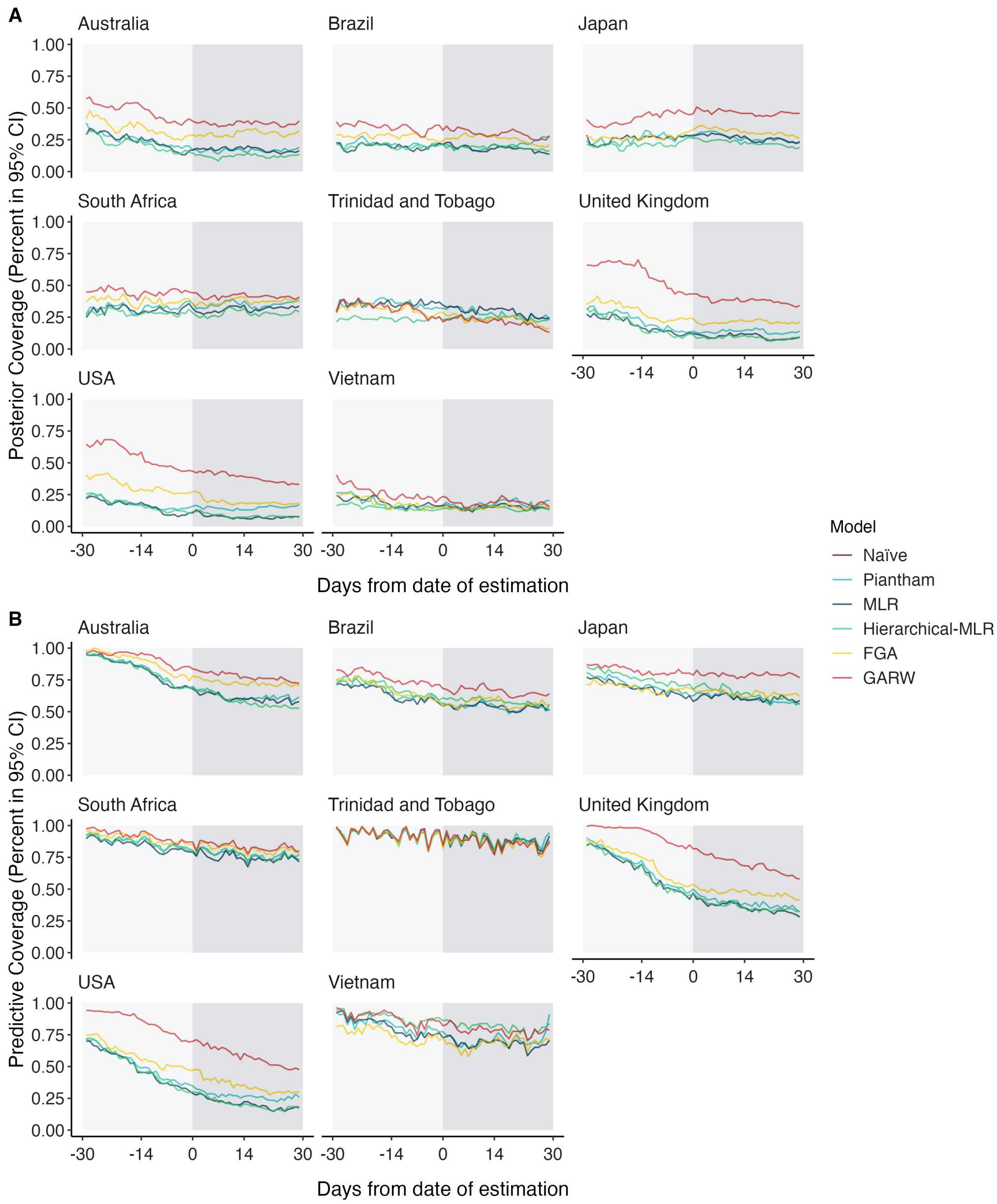

Figure S8. Posterior and predictive coverage for estimates across countries and models (A) The proportion of estimates lying within the 95% confidence intervals (CIs) of posterior latent frequencies across lag times (-30,-30). (B) The proportion of estimates lying within the 95% confidence intervals (CIs) of posterior predictive sample frequencies across lag times (-30,-30). We generate the posterior predictive sample frequencies by sampling random counts for each variant using their posterior latent frequencies conditioning on the total sequences being those observed retrospectively.

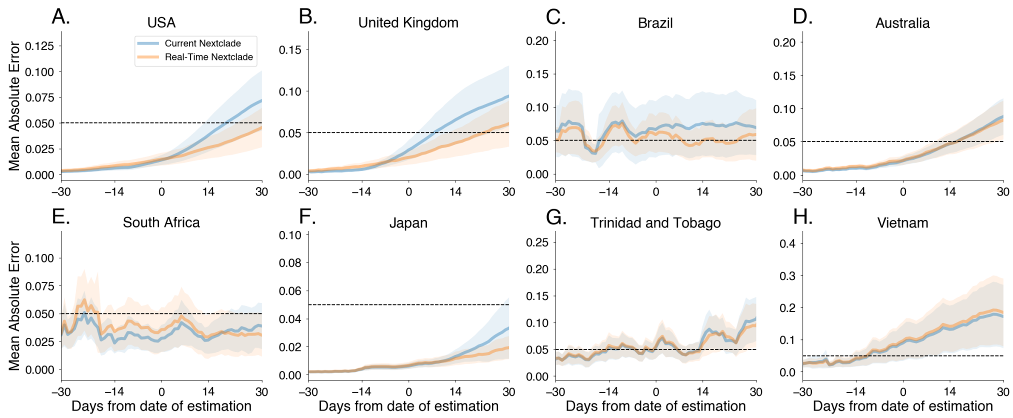

Figure S9. Comparing the accuracy of short-term forecast models under retrospective vs real-time clade assignments. (A-H) Mean absolute error for MLR as a function of days since date of estimation, starting from 30 day hindcasts to 30 days forecasts. Intervals shown have width of two standard errors of the mean. We compare retrospective Nextstrain clade assignments made today ('Current Nextclade') to Nextstrain clade assignments available in Oct 2022 ('Real-time Nextclade'). We find that errors are qualitatively similar regardless of Nextclade version with errors being potentially higher for the current Nextclade version.

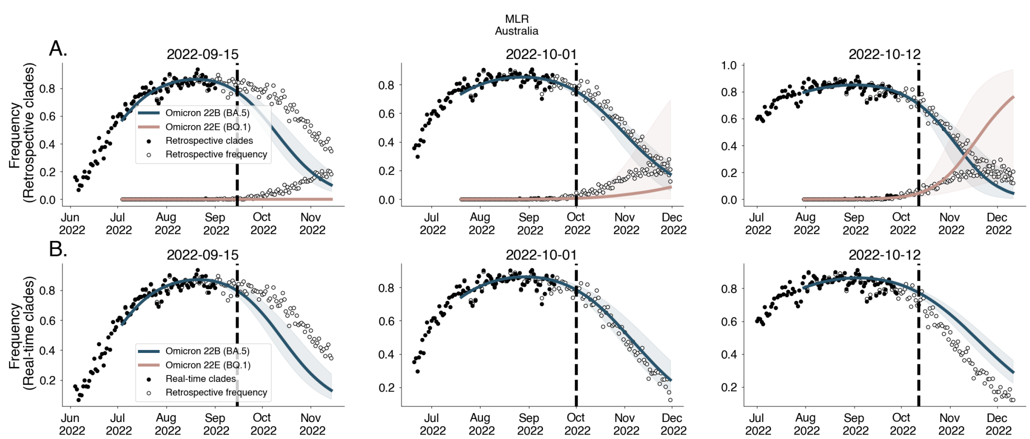

Figure S10. Forecasts for Australia using clade designations under retrospective vs real-time clade assignments Forecasts from MLR fit to data generated using retrospective Nextstrain clade designations ('Current Nextclade') (A) and Nextstrain clade assignments available in Oct 2022 ('Real-time Nextclade') (B).

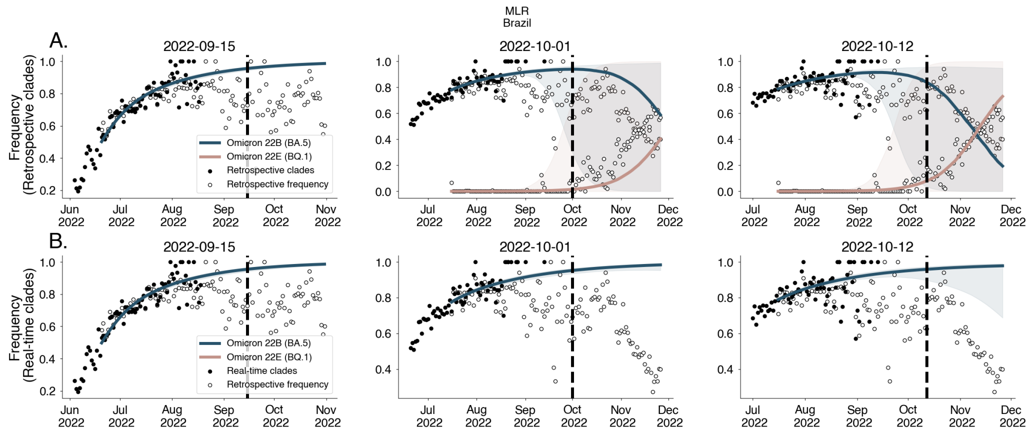

Figure S11. Forecasts for Brazil using clade designations under retrospective vs real-time clade assignments Forecasts from MLR fit to data generated using retrospective Nextstrain clade designations ('Current Nextclade') (A) and Nextstrain clade assignments available in Oct 2022 ('Real-time Nextclade') (B).

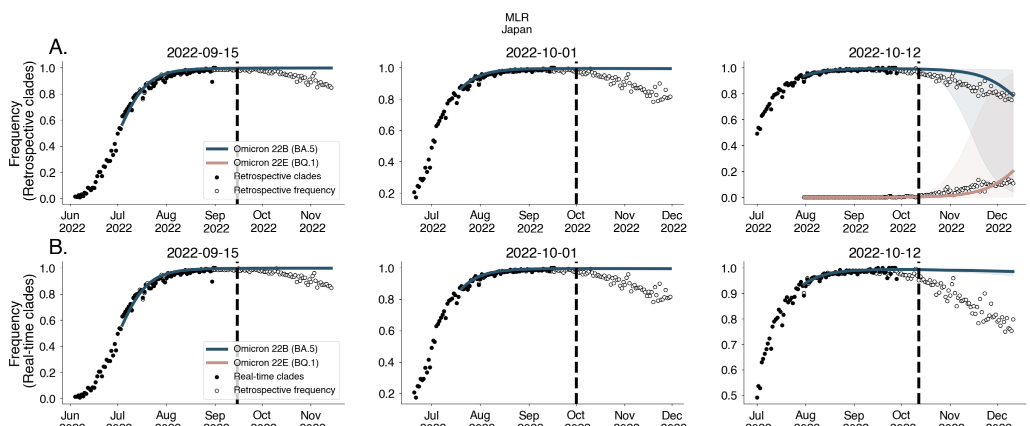

Figure S12. Forecasts for Japan using clade designations under retrospective vs real-time clade assignments Forecasts from MLR fit to data generated using retrospective Nextstrain clade designations ('Current Nextclade') (A) and Nextstrain clade assignments available in Oct 2022 ('Real-time Nextclade') (B).

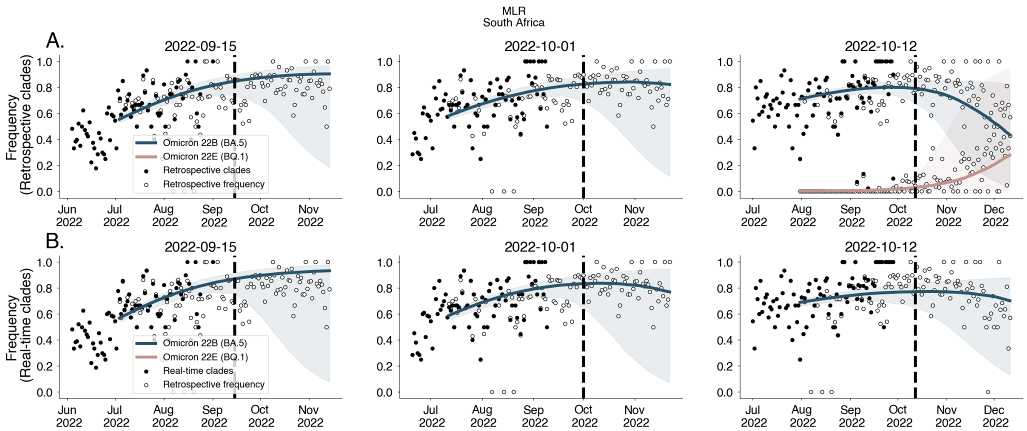

Figure S13. Forecasts for South Africa using clade designations under retrospective vs real-time clade assignments Forecasts from MLR fit to data generated using retrospective Nextstrain clade designations ('Current Nextclade') (A) and Nextstrain clade assignments available in Oct 2022 ('Real-time Nextclade') (B).

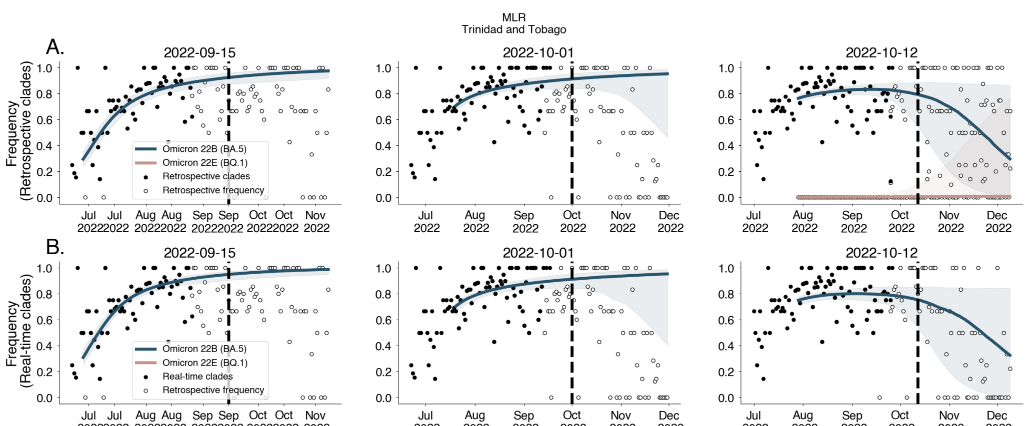

Figure S14. Forecasts for Trinidad and Tobago using clade designations under retrospective vs real-time clade assignments Forecasts from MLR fit to data generated using retrospective Nextstrain clade designations ('Current Nextclade') (A) and Nextstrain clade assignments available in Oct 2022 ('Real-time Nextclade') (B).

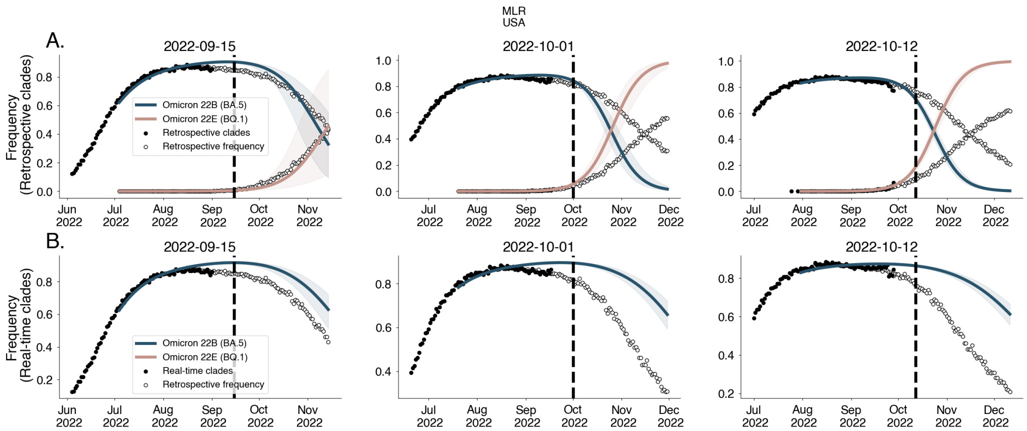

Figure S15. Forecasts for United States using clade designations under retrospective vs real-time clade assignments Forecasts from MLR fit to data generated using retrospective Nextstrain clade designations ('Current Nextclade') (A) and Nextstrain clade assignments available in Oct 2022 ('Real-time Nextclade') (B).

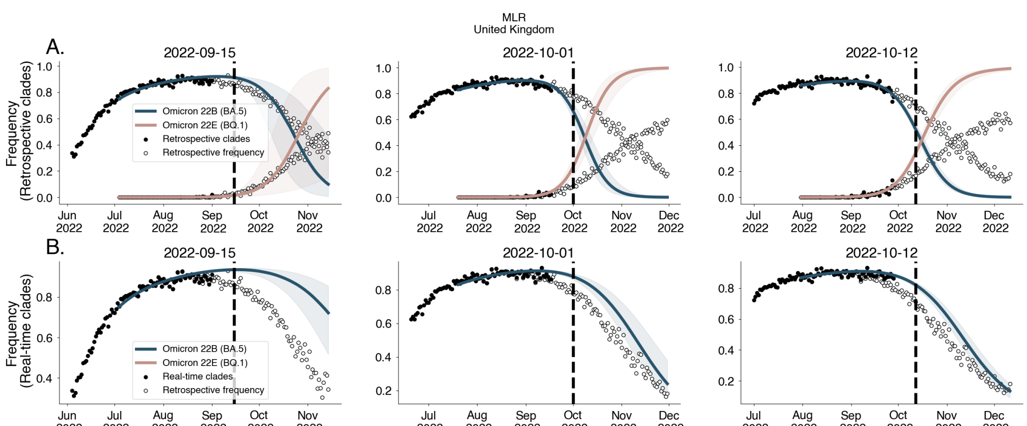

Figure S16. Forecasts for United Kingdom using clade designations under retrospective vs real-time clade assignments Forecasts from MLR fit to data generated using retrospective Nextstrain clade designations ('Current Nextclade') (A) and Nextstrain clade assignments available in Oct 2022 ('Real-time Nextclade') (B).

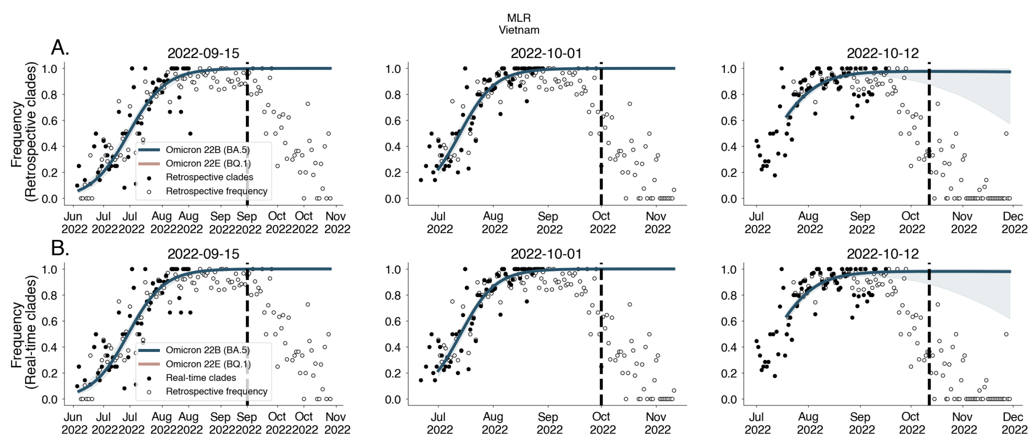

Figure S17. Forecasts for Vietnam using clade designations under retrospective vs real-time clade assignments Forecasts from MLR fit to data generated using retrospective Nextstrain clade designations ('Current Nextclade') (A) and Nextstrain clade assignments available in Oct 2022 ('Real-time Nextclade') (B).