Supplement: Collaborative estimation and evaluationof SARS-CoV-2 variant nowcasting in the U.S.

2025-11-11

# Additional figures

## Data landscape



Fig. S1 Distribution of sequence counts compared to population size in the U.S. A. Total number of sequences collected and submitted by the final evaluation date for each state during the period from September 2024 to June 2025. Color indicates whether the location’s sequences are among the top 90th percentile of all sequences submitted nationally B. Population size for each jurisdiction, ordered by total number of sequences submitted. Color indicates whether the population size is among the top 90th percentile of the total population.



Fig. S2 Example final sequence data from CA

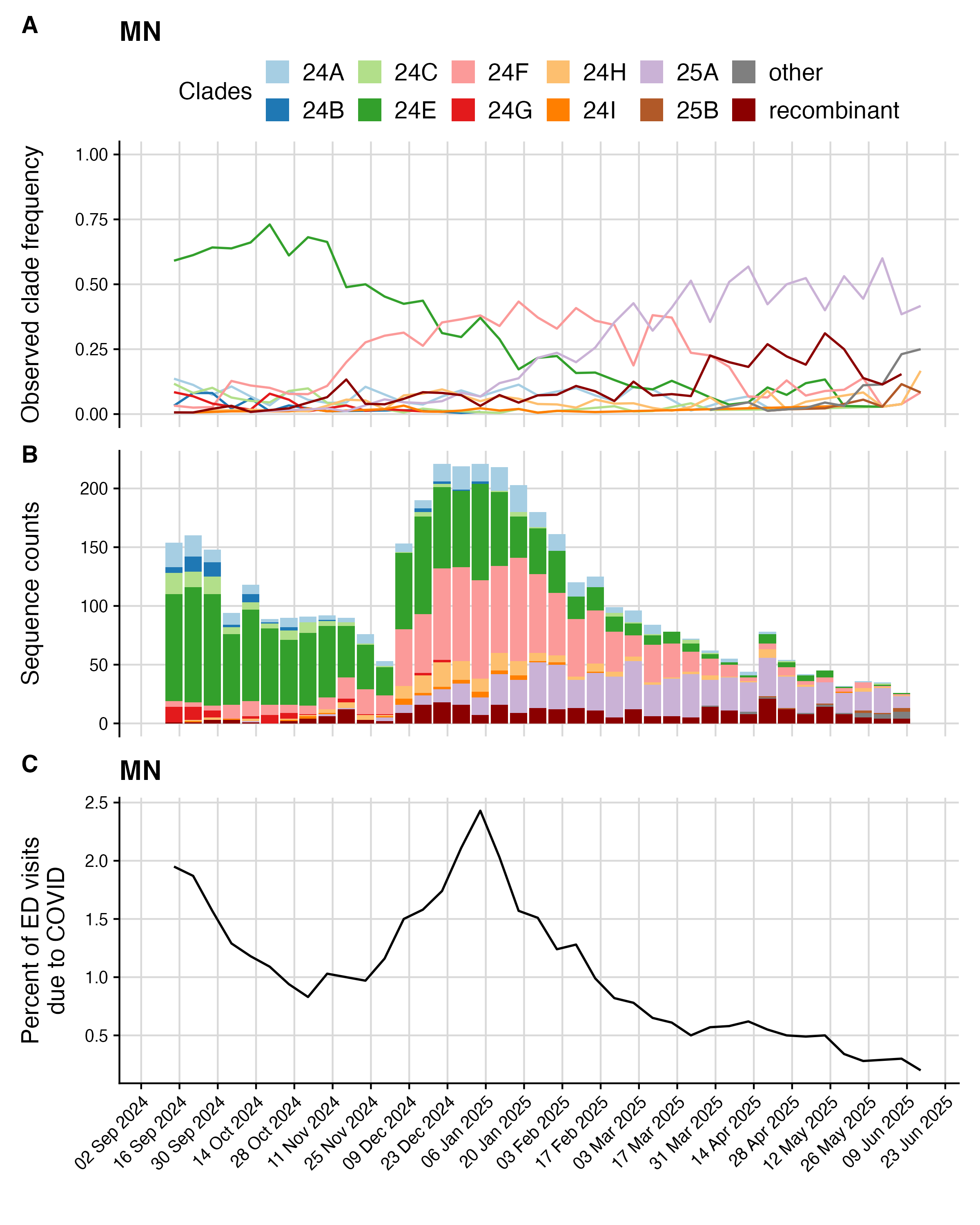


Fig. S3 Example final sequence data from IL



Fig. S4 Example final sequence data from MN

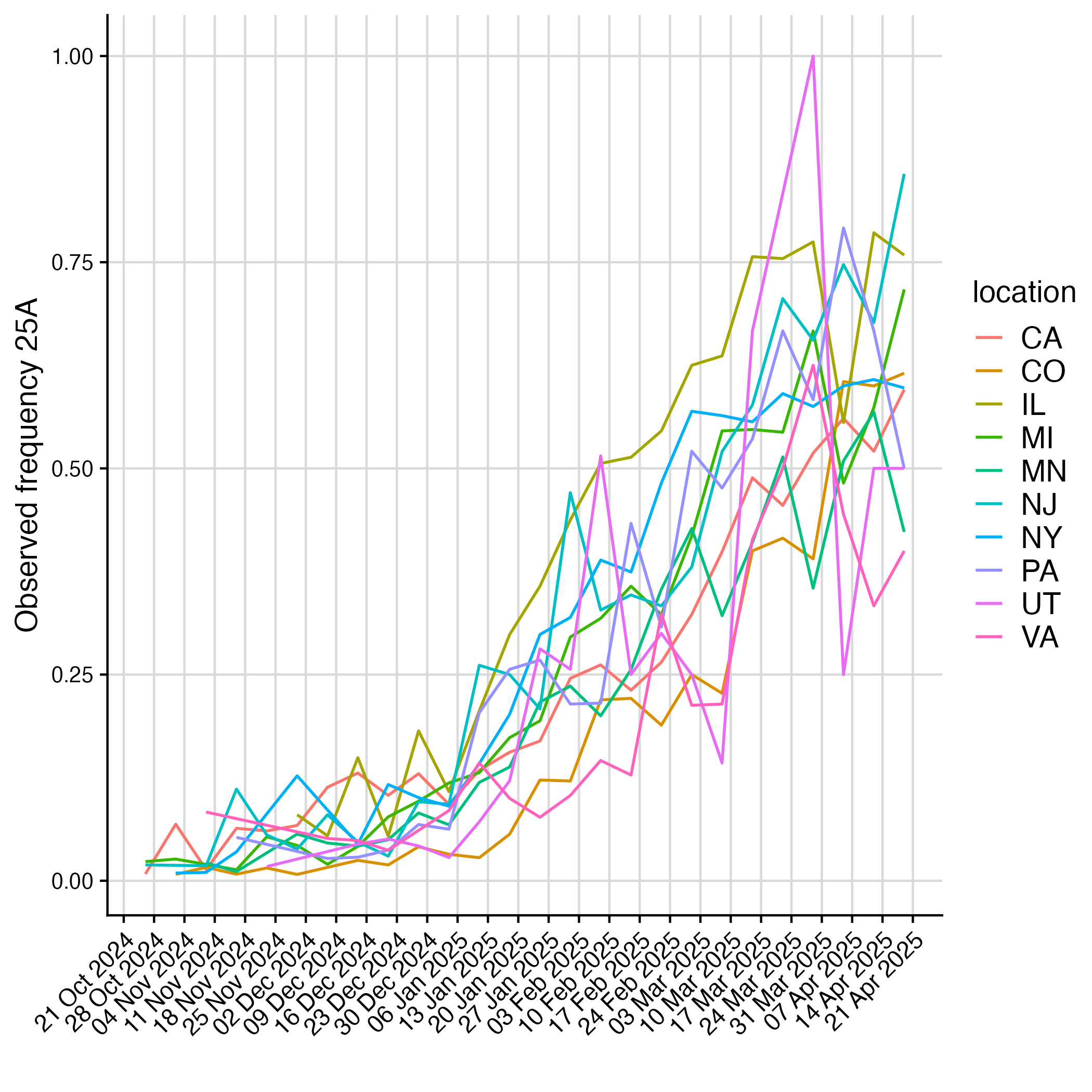


Fig. S5 Different dynamics of 25A emergence

## Model nowcasts

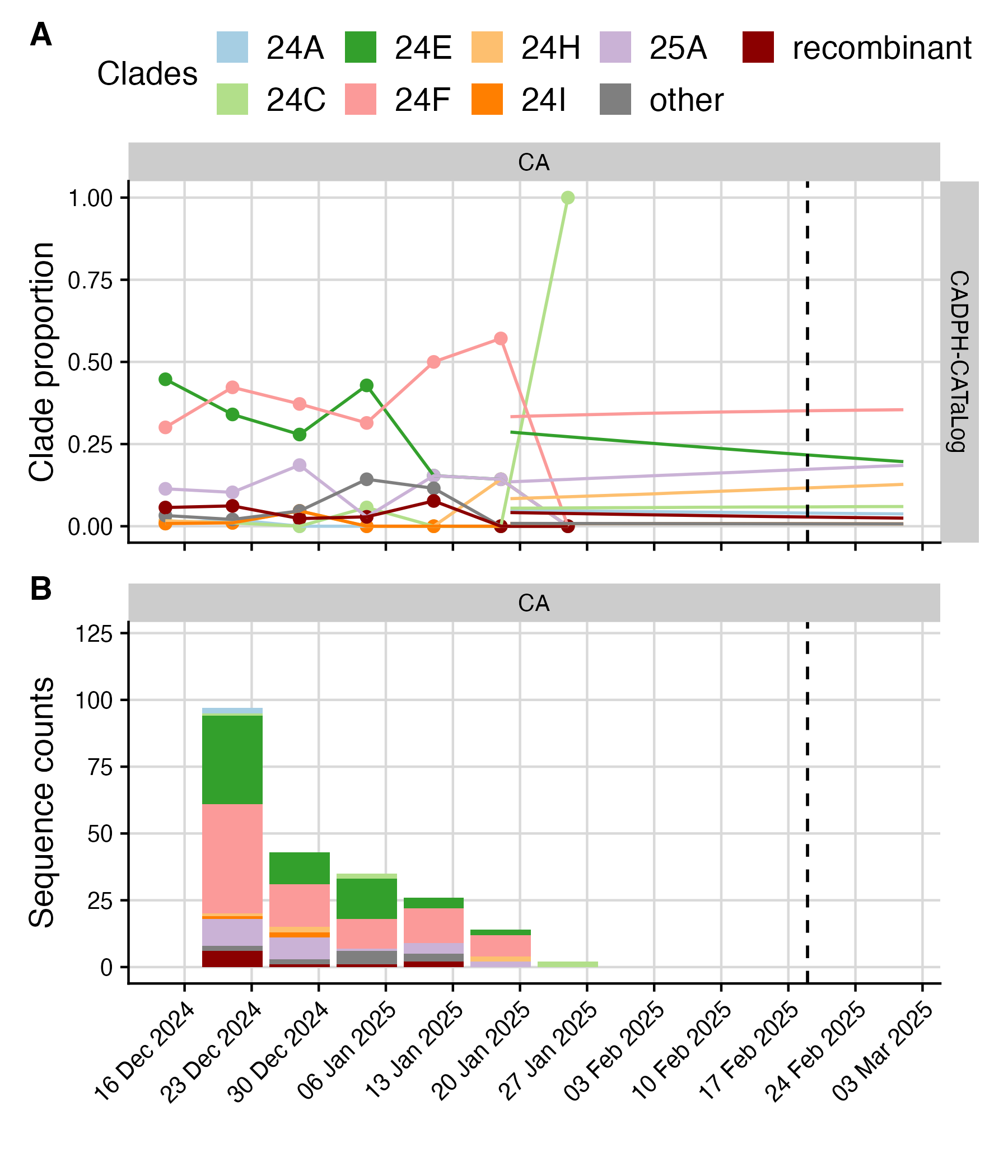


Fig. S6 Model nowcasts for the CADPH-CATaLog model from February 19th, 2025.

## Submission metadata

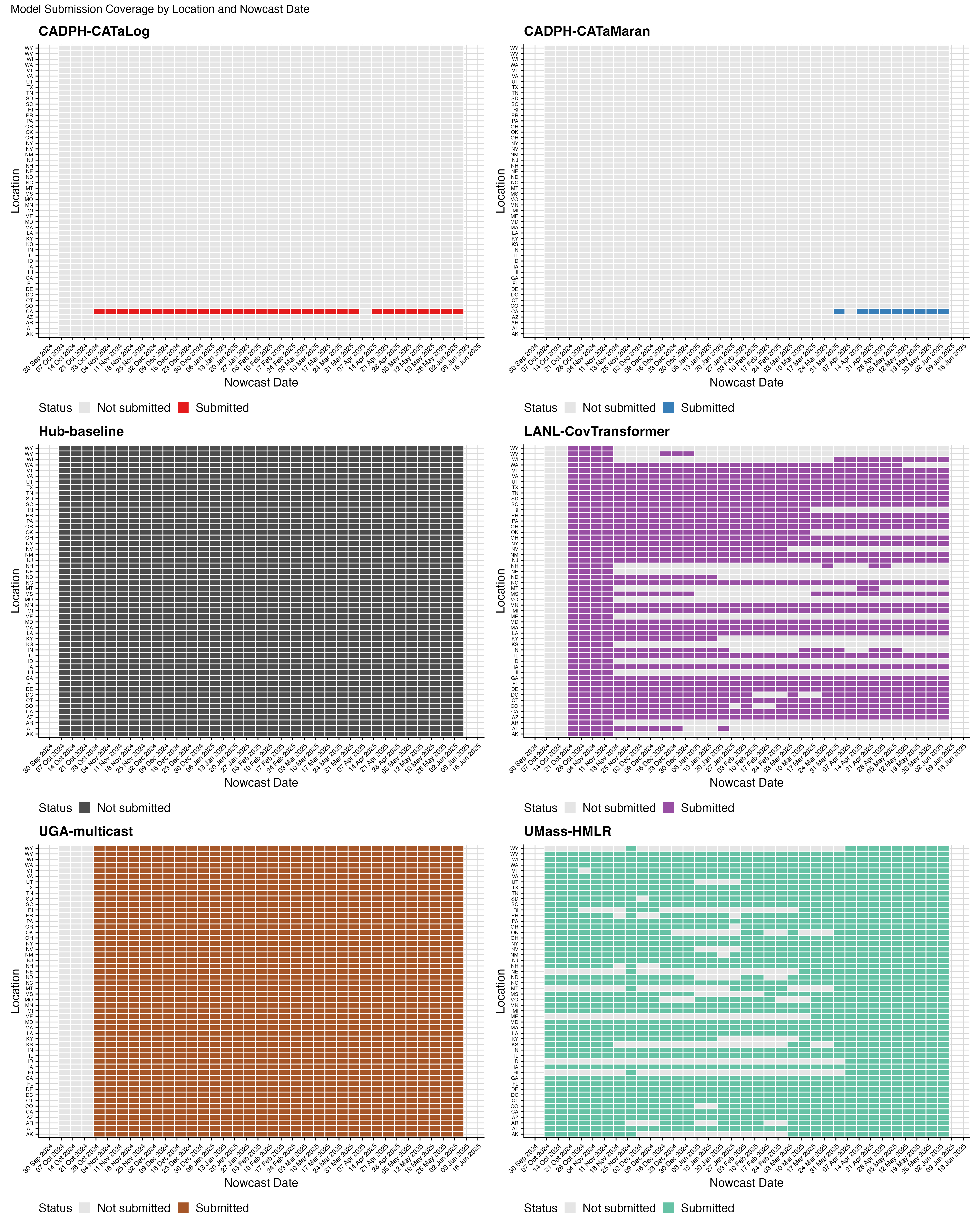


Fig. S7 Model submission coverage by location and nowcast date for each participating model. Tiles colored by model-specific colors indicate dates and locations where the model submitted a forecast. Gray tiles indicate missing submissions.

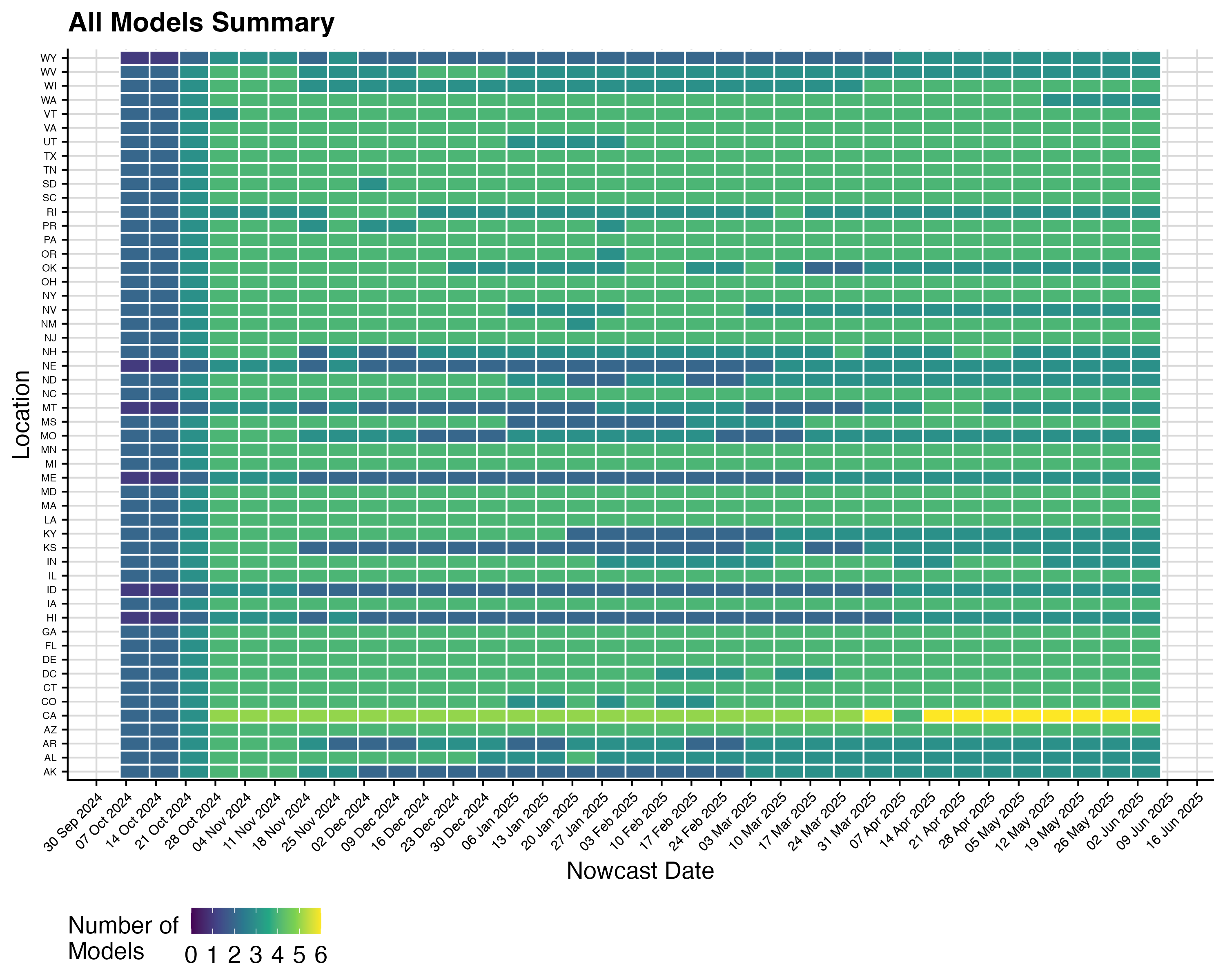


Fig. S8 Summary of model submission coverage across all models. Color indicates the number of models that submitted forecasts for each location and nowcast date combination, ranging from 0 (no models) to 6 (all models).

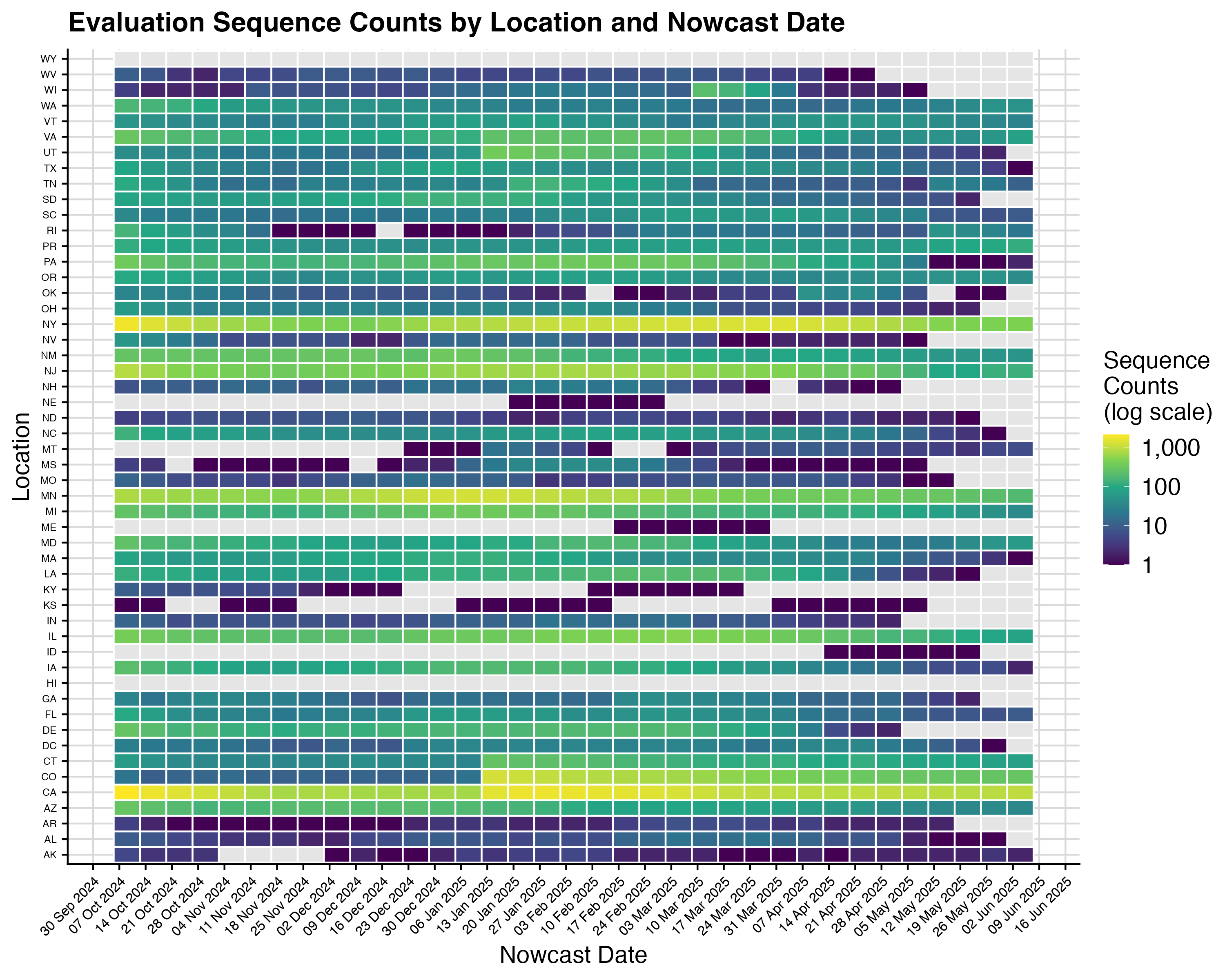


Fig. S9 Evaluation sequence counts by location and nowcast date. Color indicates the total number of sequences available in the evaluation data (log scale) for each location and nowcast date combination. Gray tiles indicate nowcast date-location pairs with no sequences across all horizons.

## Additional results

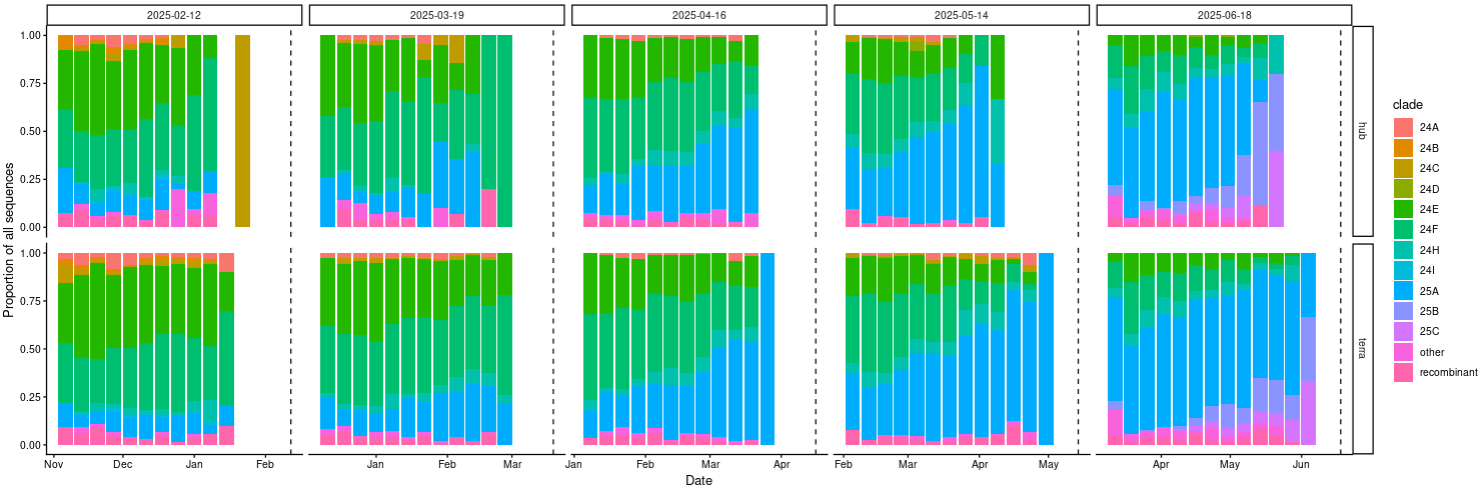


Fig. S10 Comparison of the observed clade proportions by sequence collection week across nowcast dates (columns) in the NCBI GenBank data used by the Hub (top row) and the California specific data source for sequencing called terra (bottom row). Colors indicate clade, dashed line indicates the nowcast date. In most weeks, the California specifc dataset terra has observed clade proportions for a more recent week than the NCBI GenBank dataset for California.

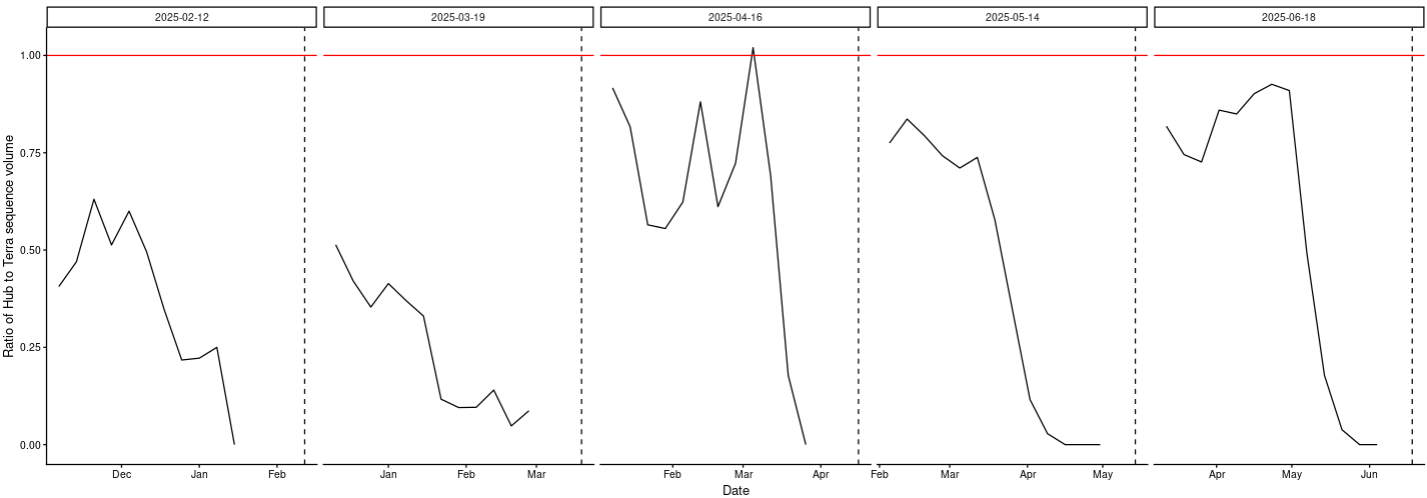


Fig. S11 Ratio of the number of sequences available in the NCBI GenBank dataset for Californi compared to the California-specific dataset across collection dates for a subset of nowcast dates (colums). Patterns indicate that even further back in time, the GenBank dataset at most makes up between 50% and 90% of the California specific data volume, with a particular drop off closer to the nowcast date.

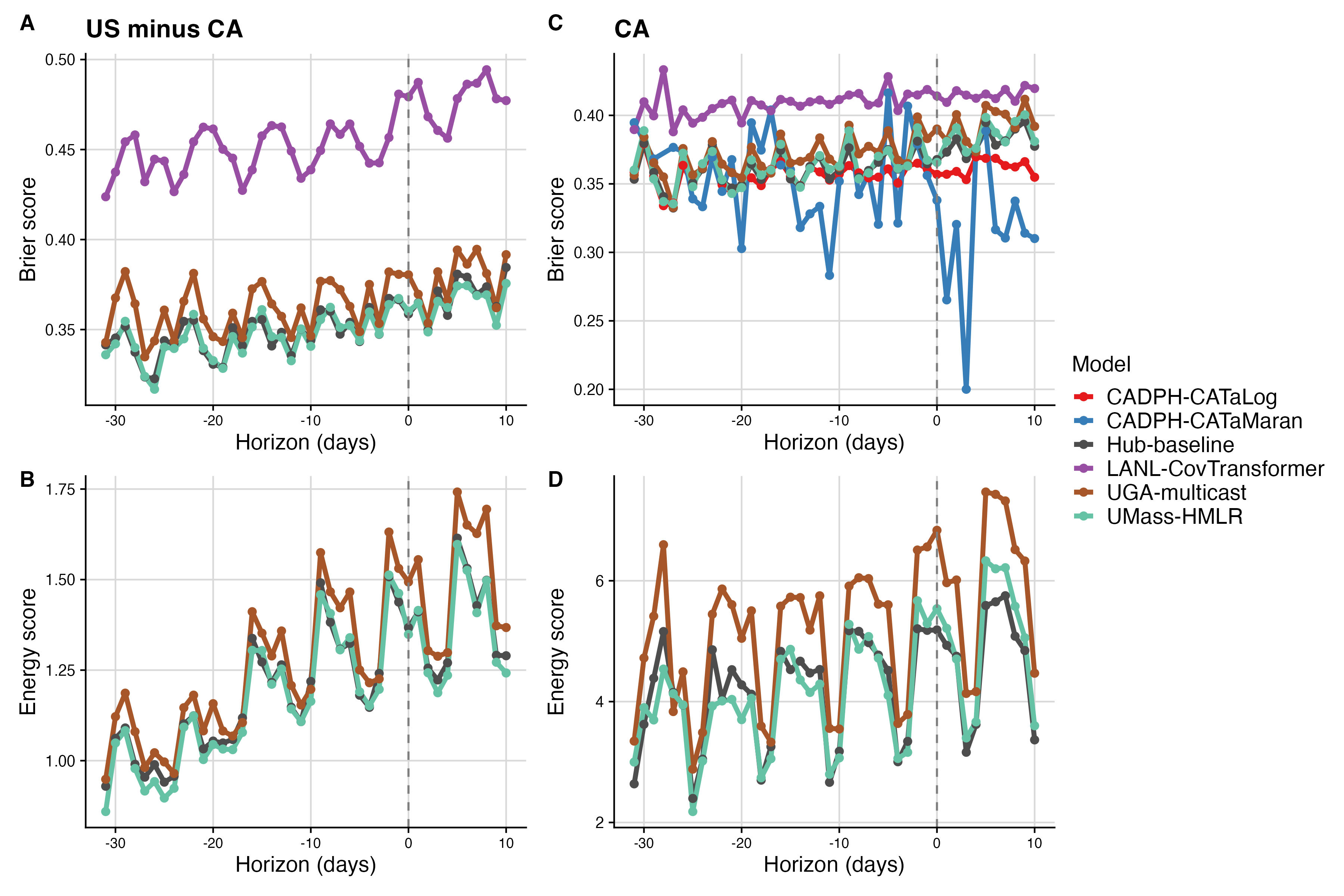


Fig. S12 Absolute Brier (top) and energy (bottom) in the U.S. excluding California (left) and California (right).

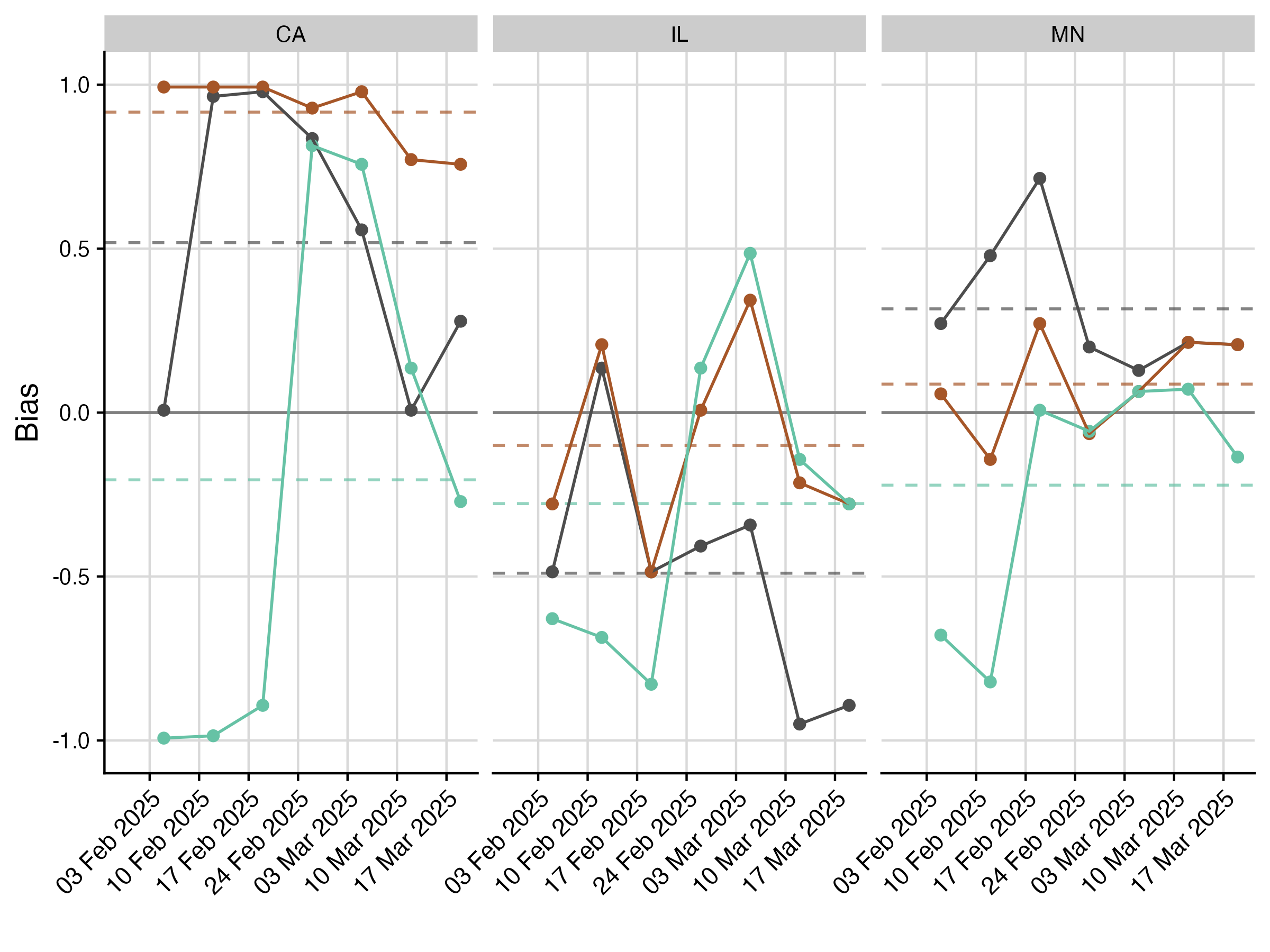


Fig. S13 Bias over time for three example states in the US during the 25A emergence.