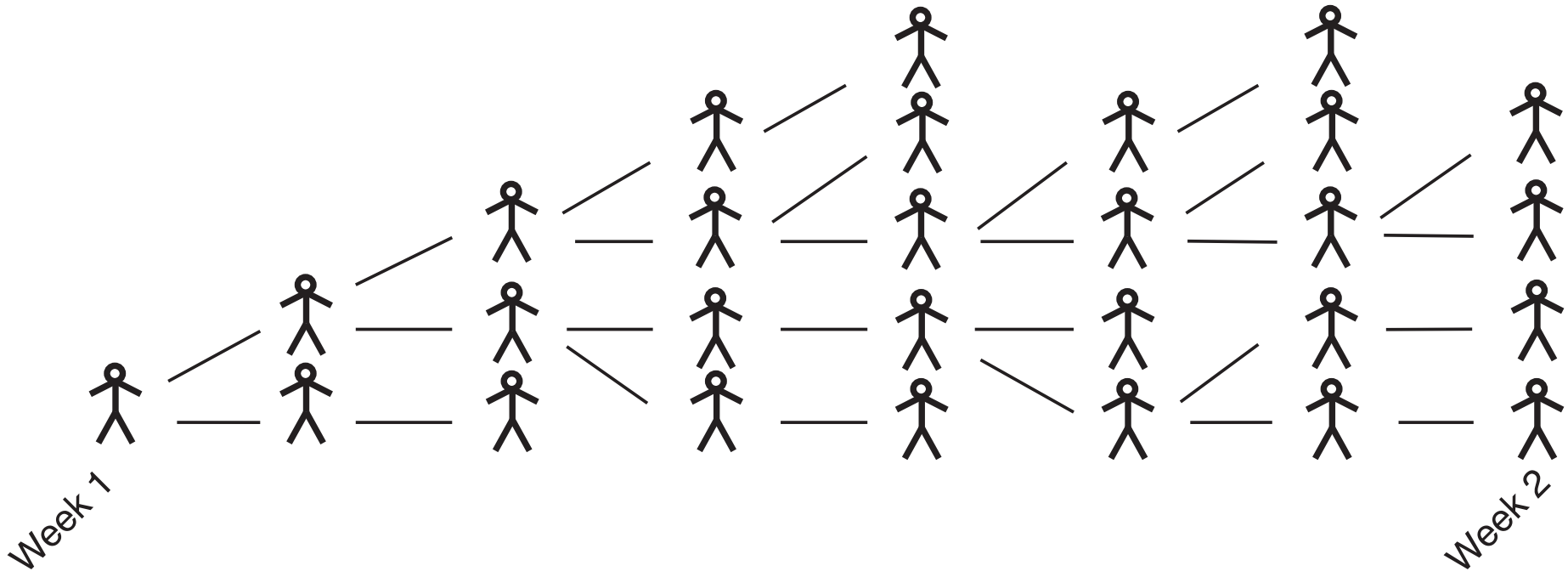
Accounting for population structure

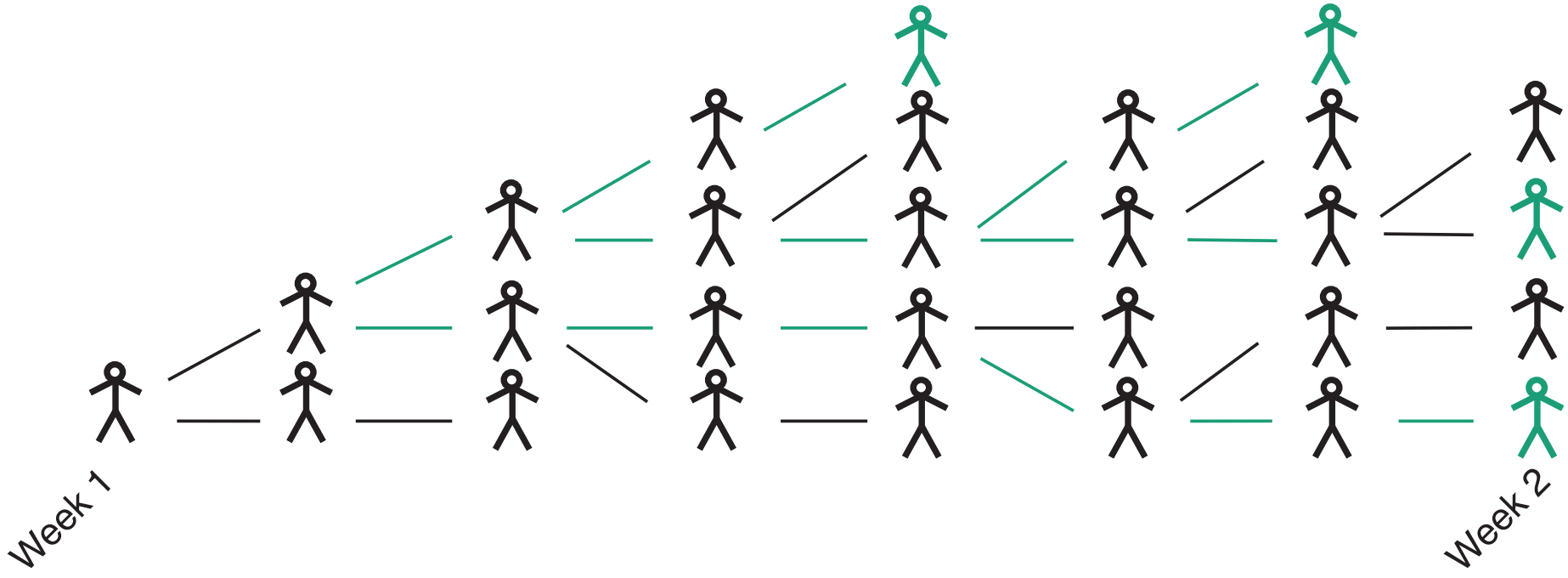
Nicola F. Müller

e-mail: nicola.mueller@ucsf.edu

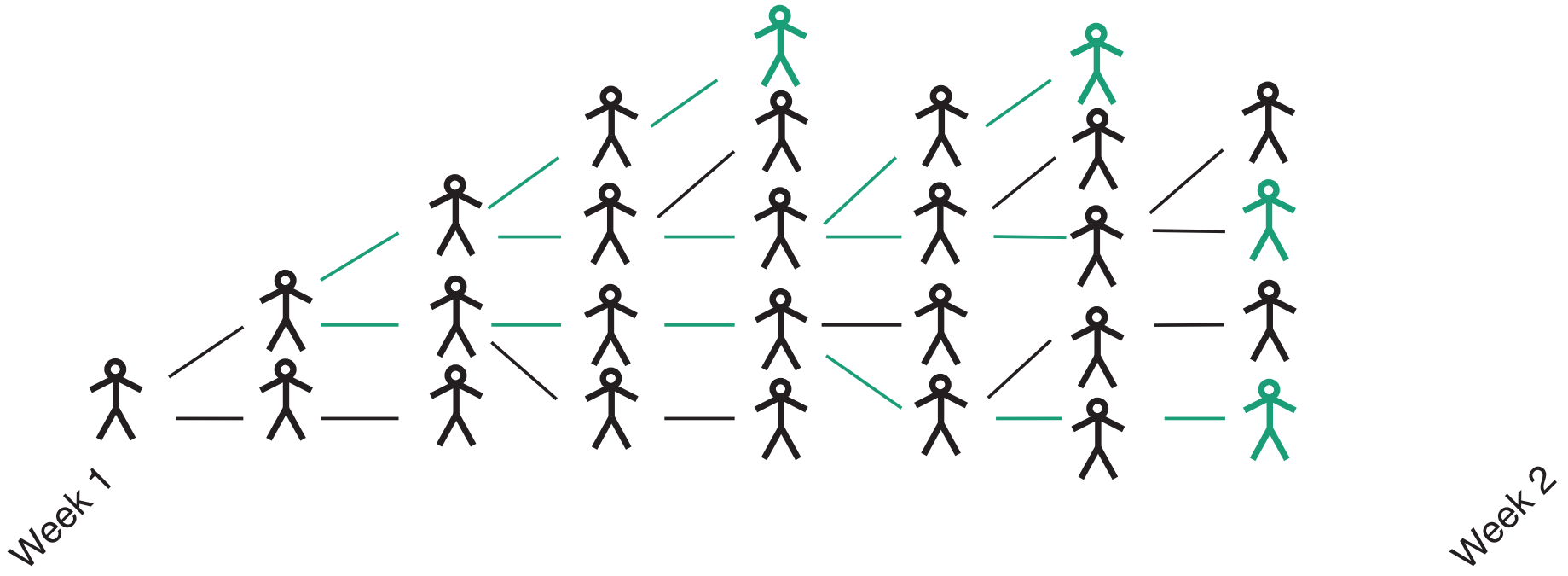
Genomics allows tracking the shared ancestry.



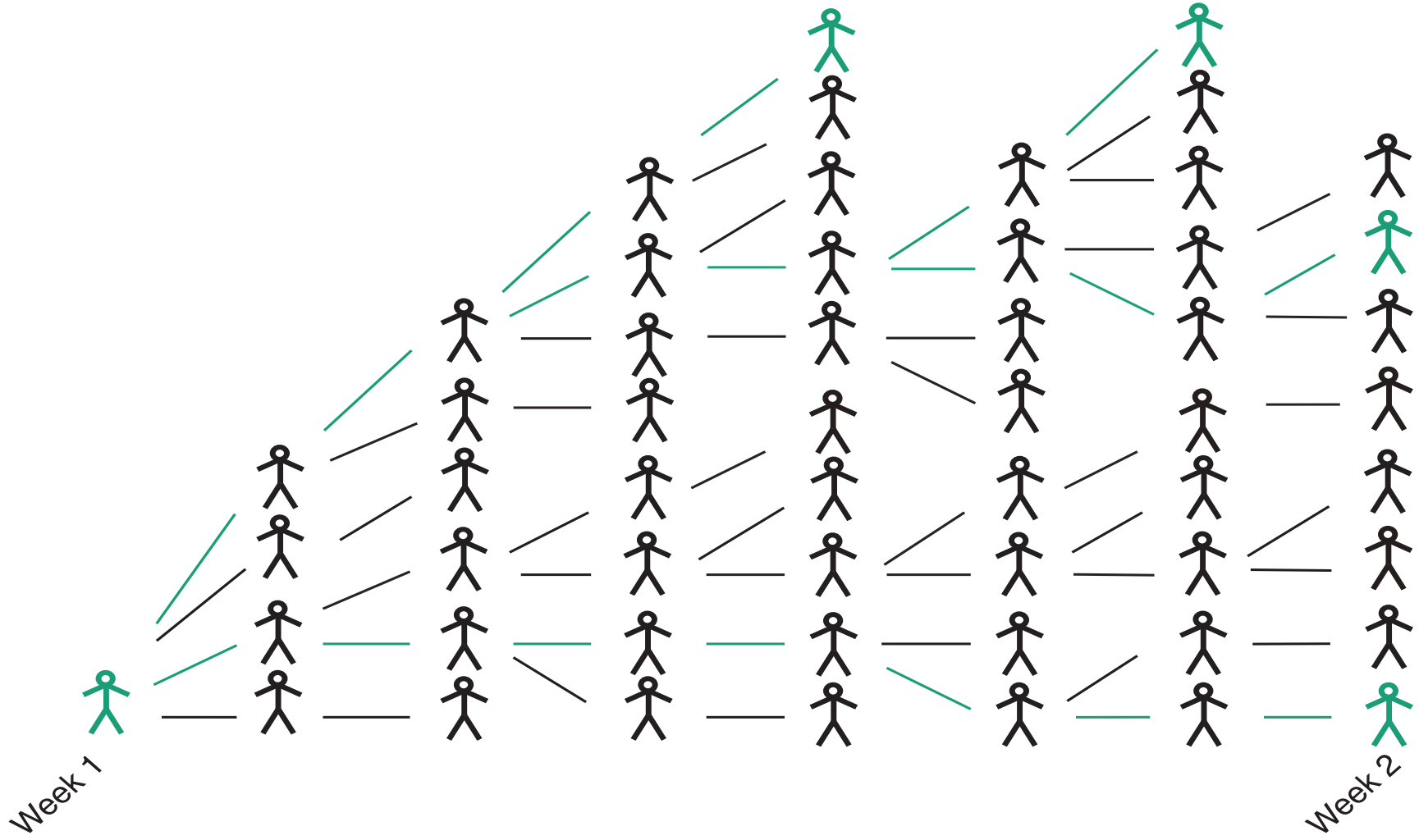
If we track the shared ancestry of the 4 sampled individuals, we will reach a common ancestor more quickly, the smaller the Ne.



What happens if we have the same number of infected individuals, but higher turnover?



What happens if we have a higher number of infected individuals and higher turnover?

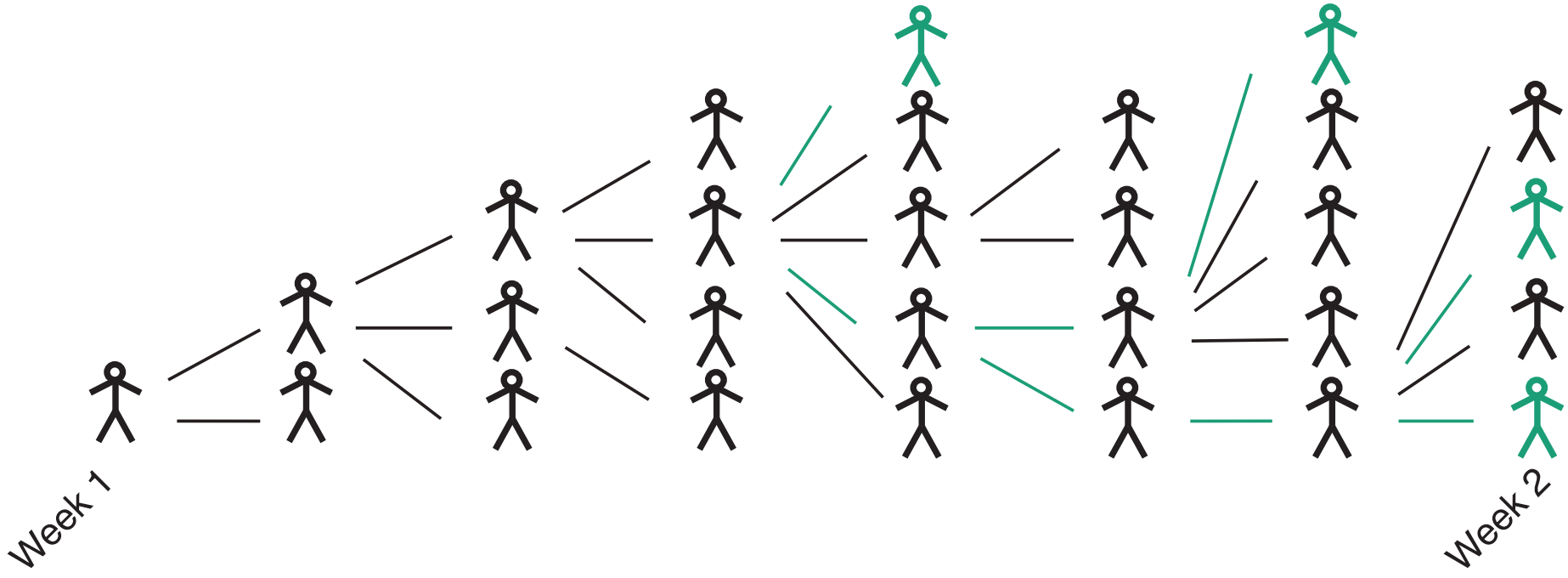


The effective population size (Ne) can be denoted as a function of epi-parameters for SIR models

$$Ne(t) = \frac{I(t)}{\theta \frac{S(t)}{N}}$$

Volz, E. M. (2012). "*Complex Population Dynamics and the Coalescent Under Neutrality. Genetics*," 190(1), 187-201. https://doi.org/10.1534/genetics.111.134627

What happens if we have a skewed offspring distribution?

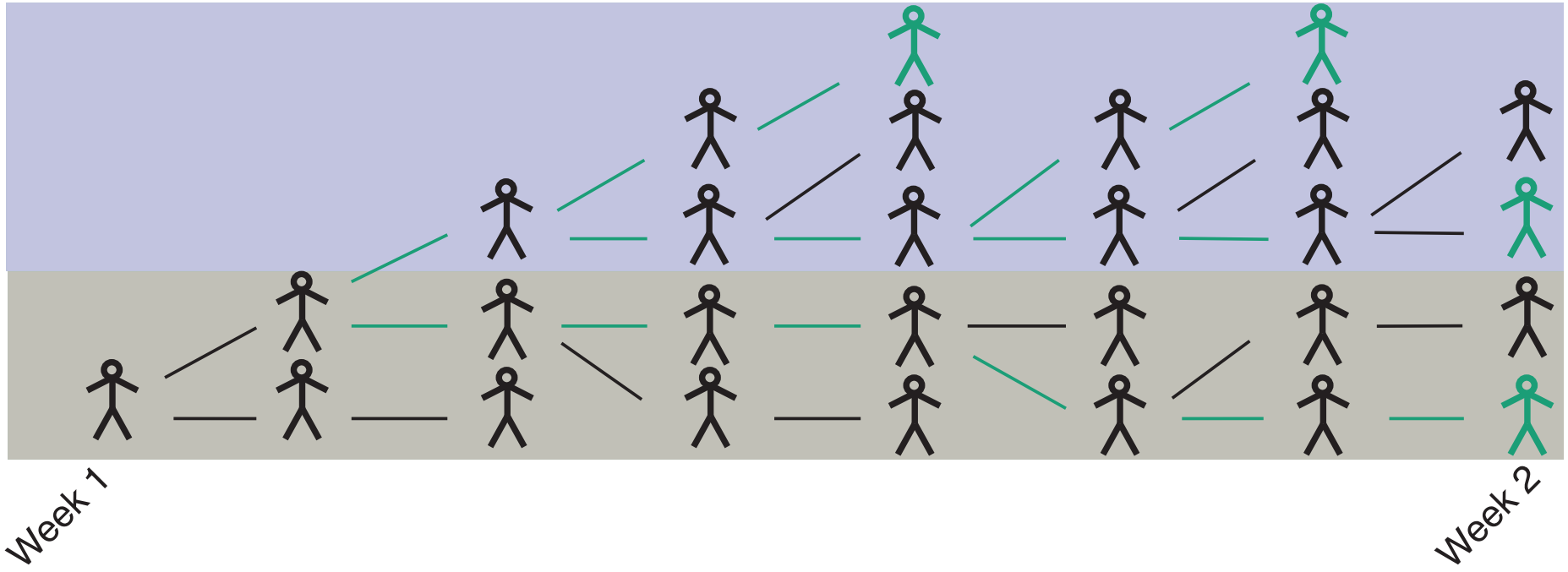


We can also account for some offspring distributions in the Ne

$$N_e(t) = \frac{I(t)}{\theta\left(1+\frac{1}{k}\right)}$$

Li, L. M., Grassly, N. C., & Fraser, C. (2017). *"Quantifying Transmission Heterogeneity Using Both Pathogen Phylogenies and Incidence Time Series."* Molecular Biology and Evolution, 34(11), 2982-2995. https://doi.org/10.1093/molbev/msx195

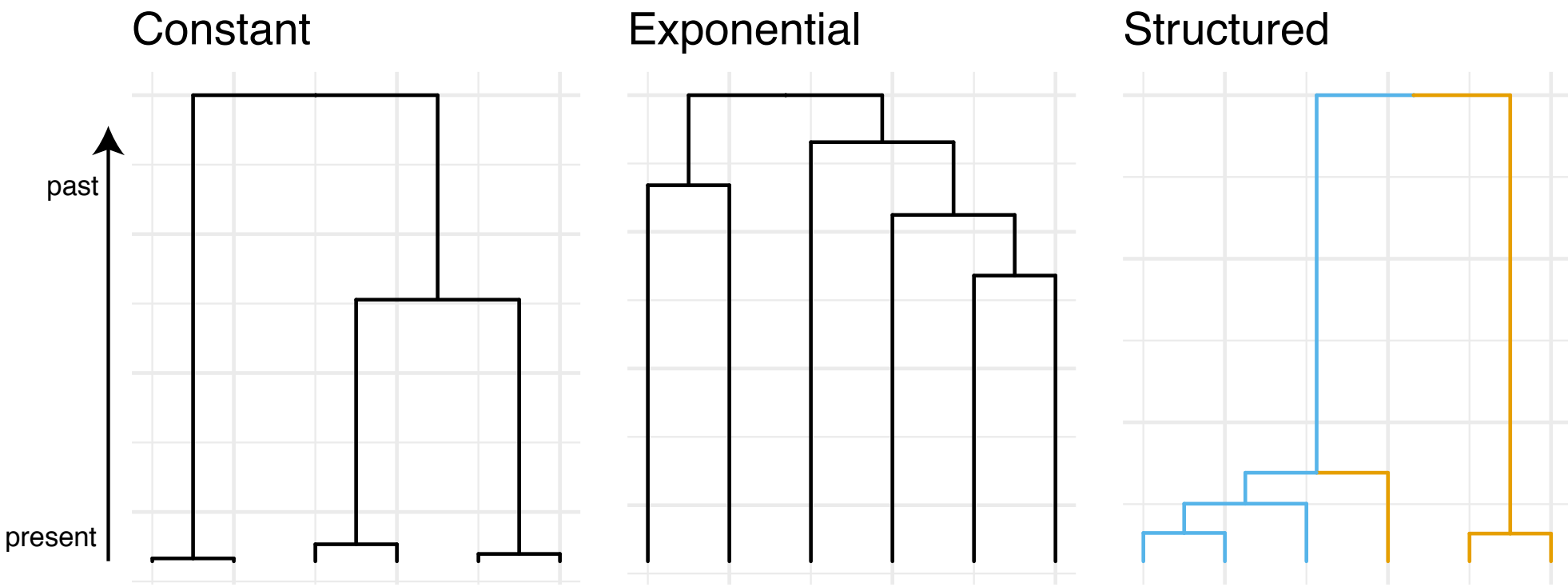
What happens if there is population structure?



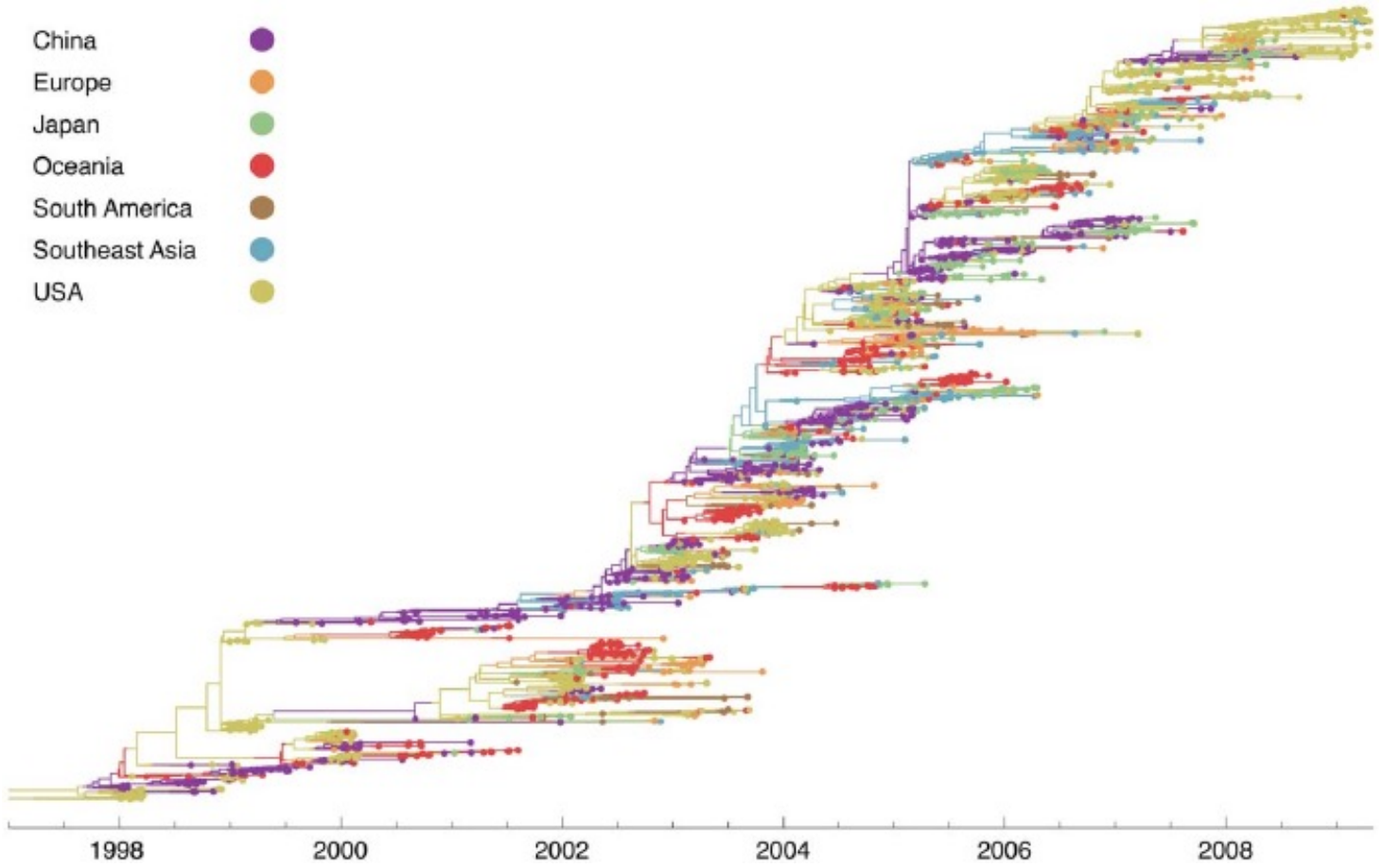
Once there is population structure, the meaning of the effective population size is not obvious

$$N_e(t) = ?$$

Phylogenetic trees are formed by population processes and contain information about them

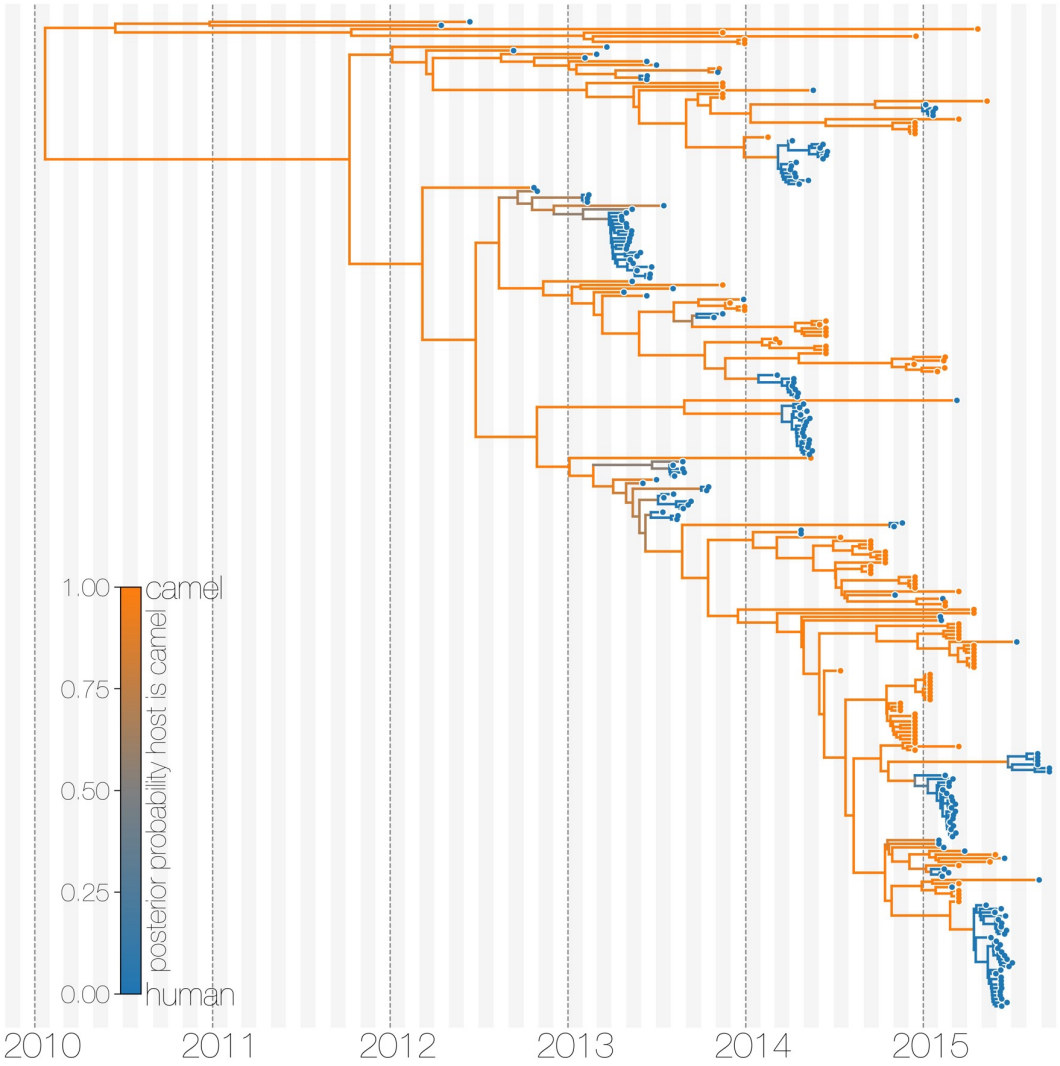


H3N2 sequences show spatial clustering



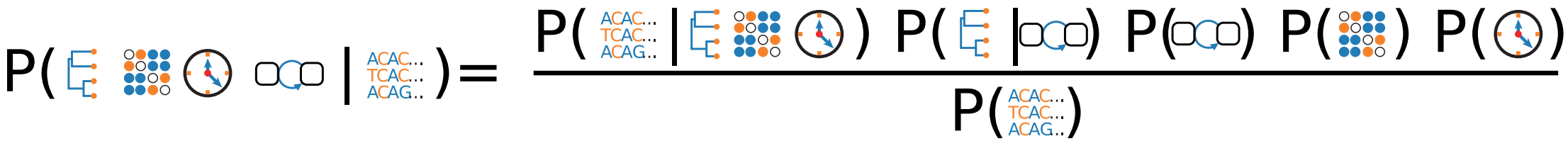
Bedford et al., 2010, *PLoS Path.*

MERS sequences cluster by host of isolation



Dudas et al., 2018, *eLife*

In the posterior probability, modelling the population dynamics happens in the tree prior.



DuPlessis, L. et al. (2015), Trends in Microbiology

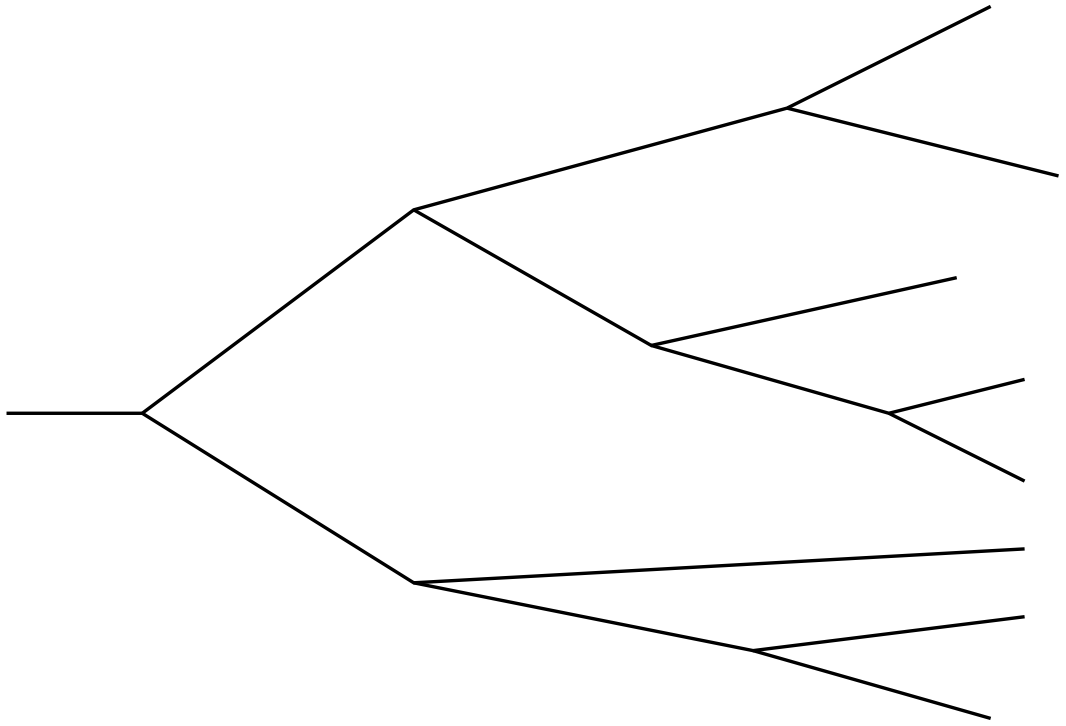
Models to account for population structure:

- **DTA:** Assumes structure to be a neutral trait evolving “on top” of a phylogenetic tree (Lemey 2009).

- **Structured Coalescent:** Models how lineages coalesce within and migrate between discrete locations.

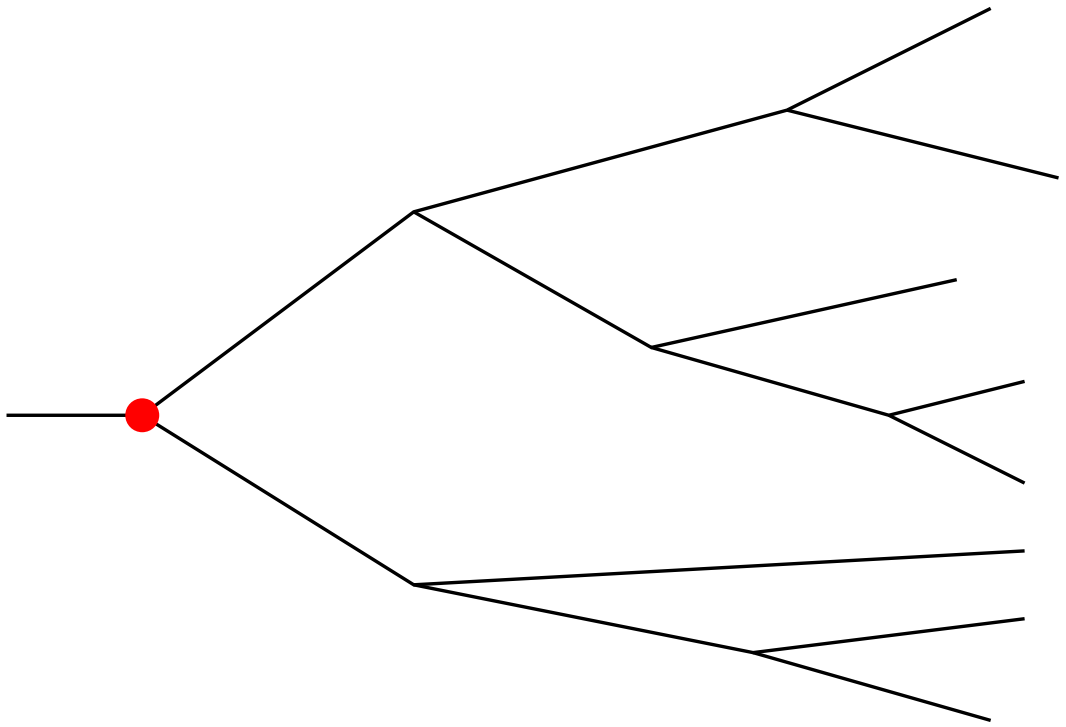
- **Multi-type Birth-death models:** Models how lineages give birth to other lineage in the same and migrate between them.

Tree was generated by a different process (e.g. a coalescent process).



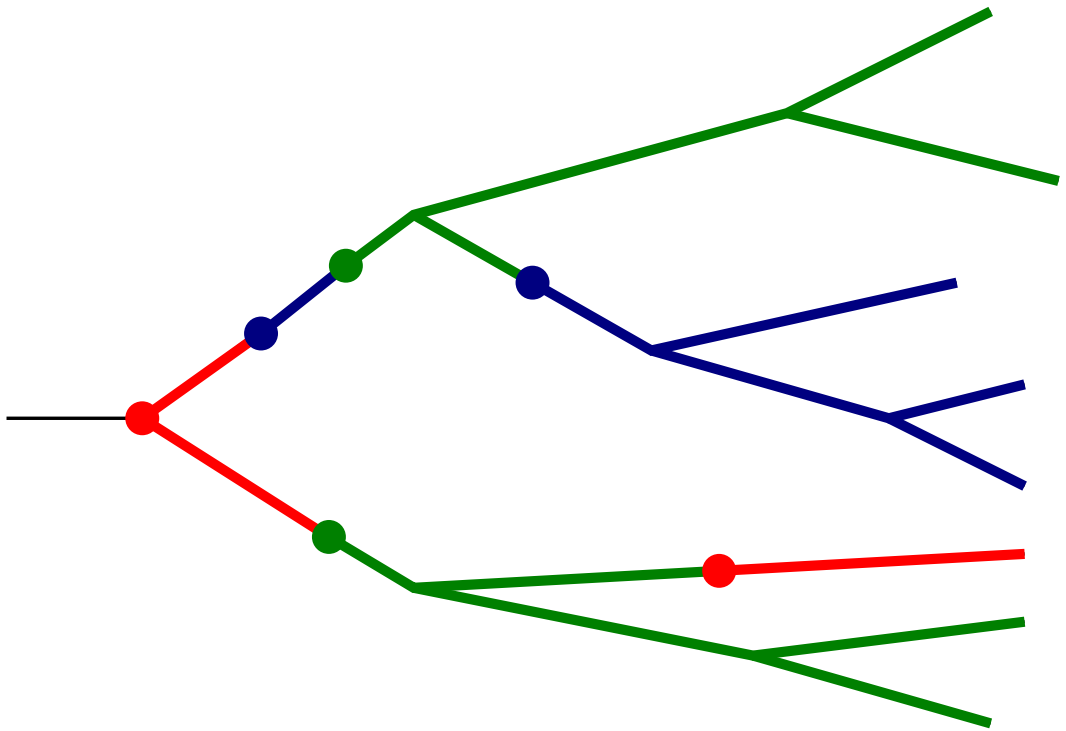
Slide from Timothy Vaughan

Root of the tree is in some location (here red)



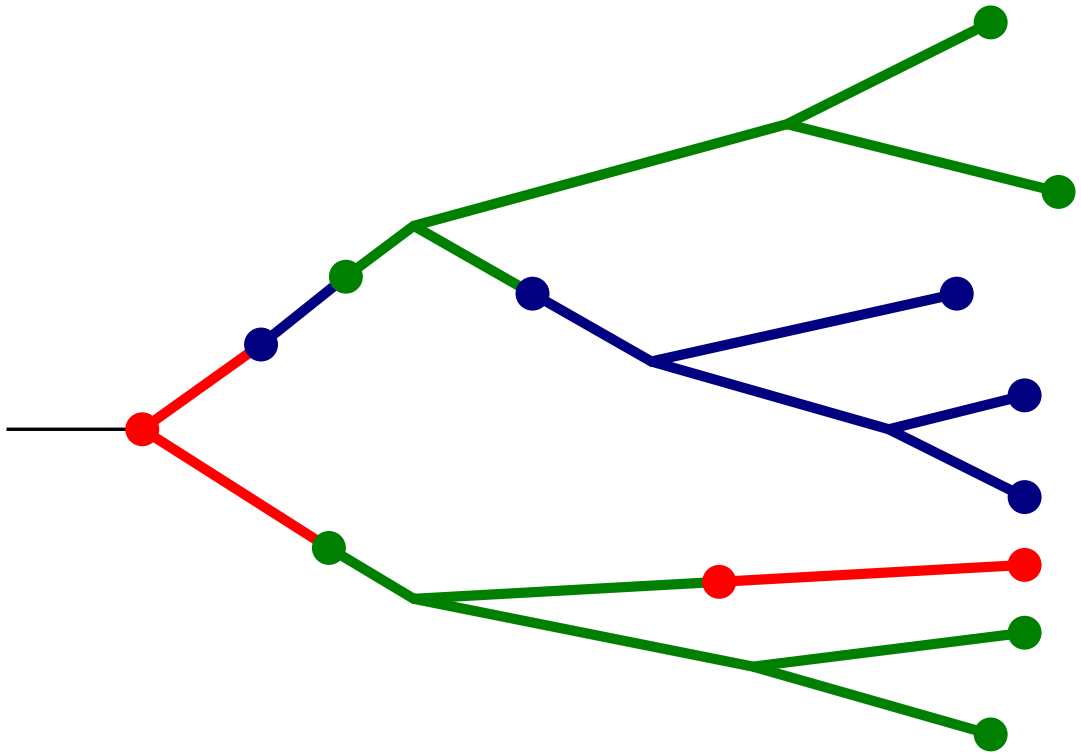
Slide from Timothy Vaughan

Then models the migration process on top of this tree from the root to the tips



Slide from Timothy Vaughan

Sampling locations are considered Data.



Slide from Timothy Vaughan

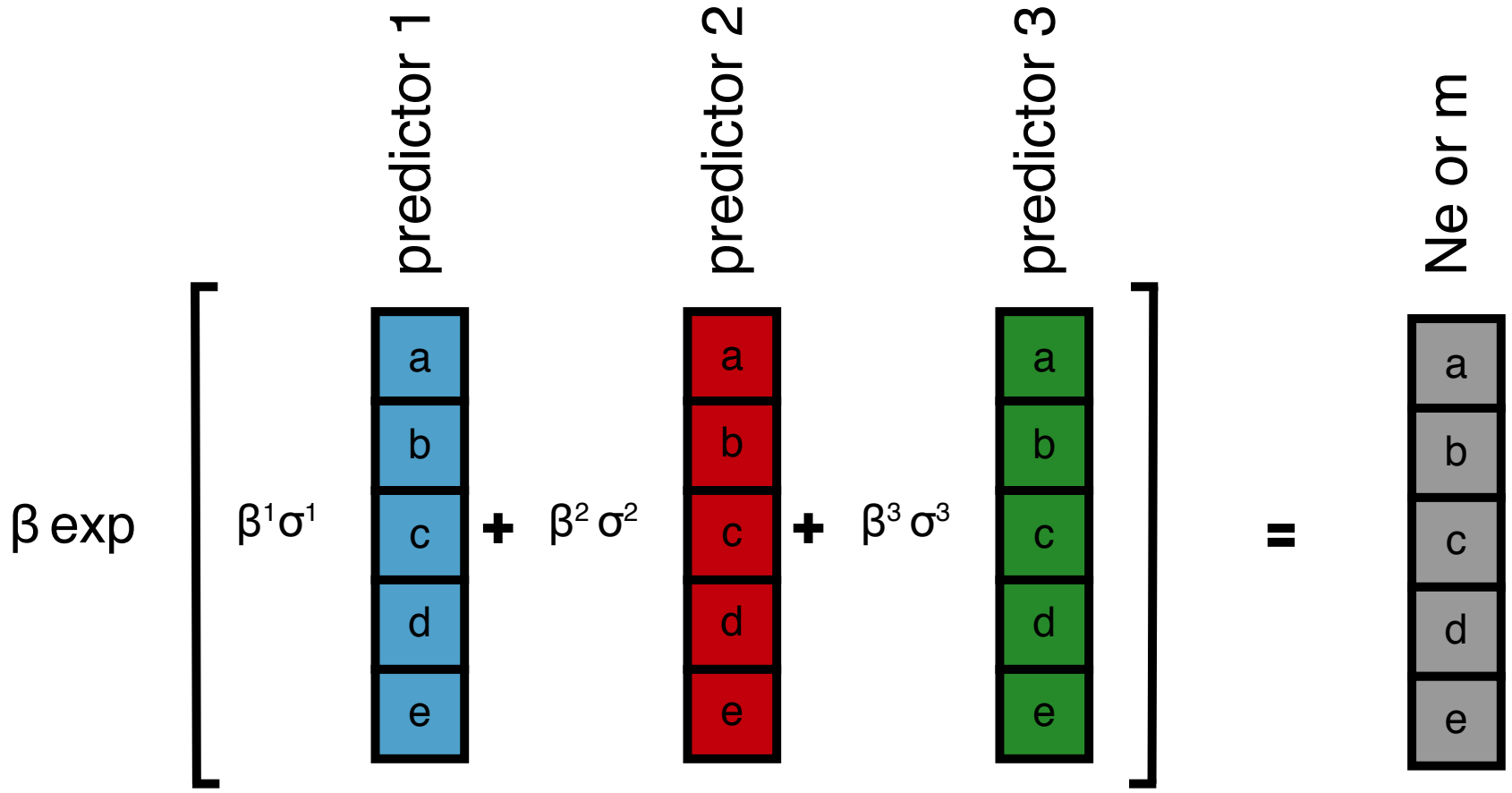
Discrete trait analyses treat geography the same way as substitutions are treated

- Some number of states instead of the 4 nucleotides are considered, but only one position. The number of states corresponds to discrete locations.

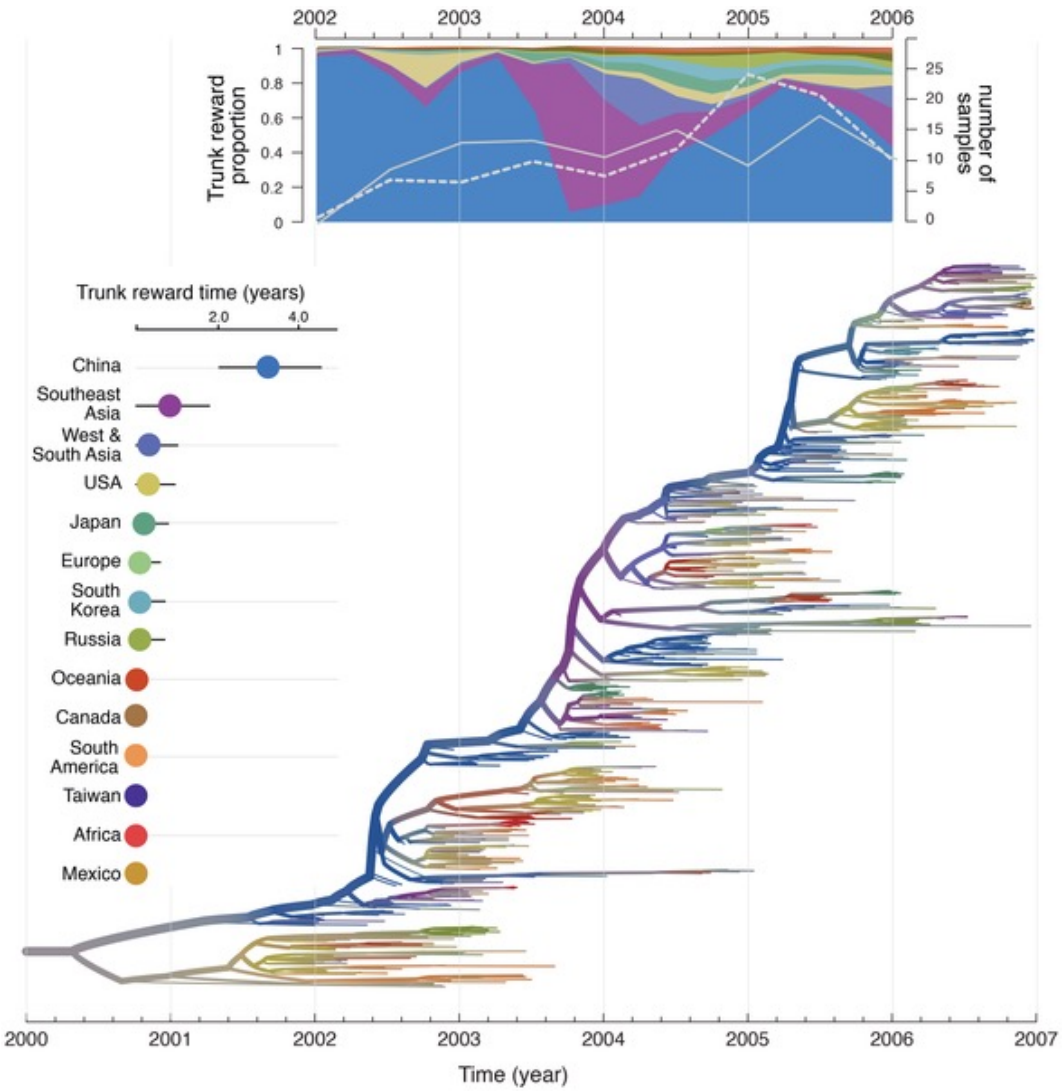
- Some transition model between states (can be asymmetric) which are the migration rates.

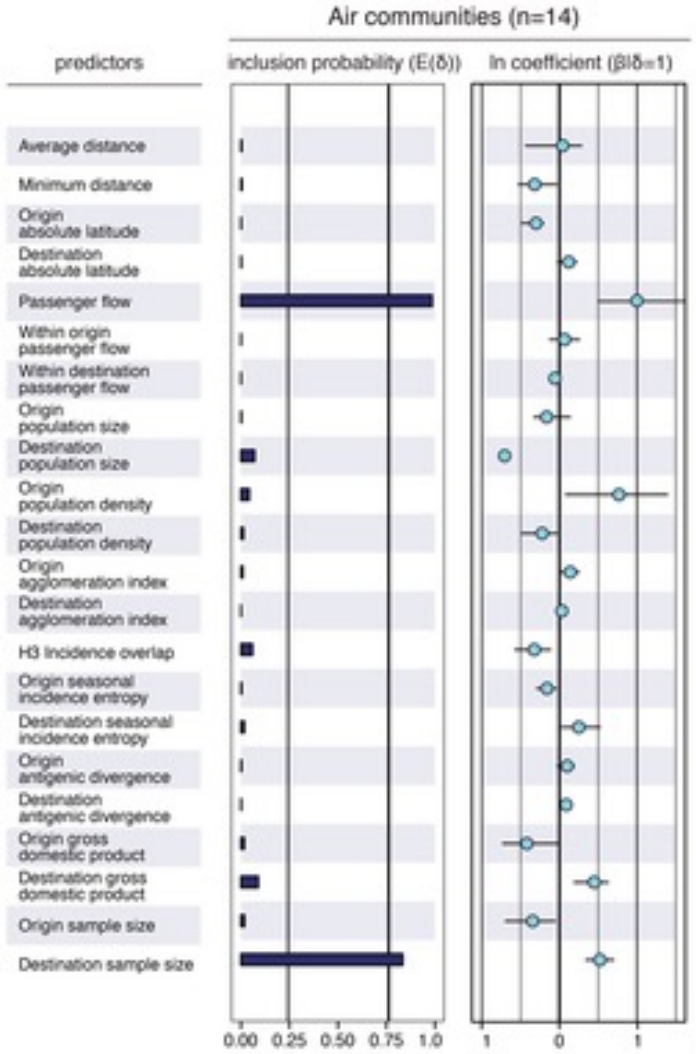
- Is not a tree prior, as such, a separate tree prior is needed.

Migration rates can be informed using outside data in a GLM approach



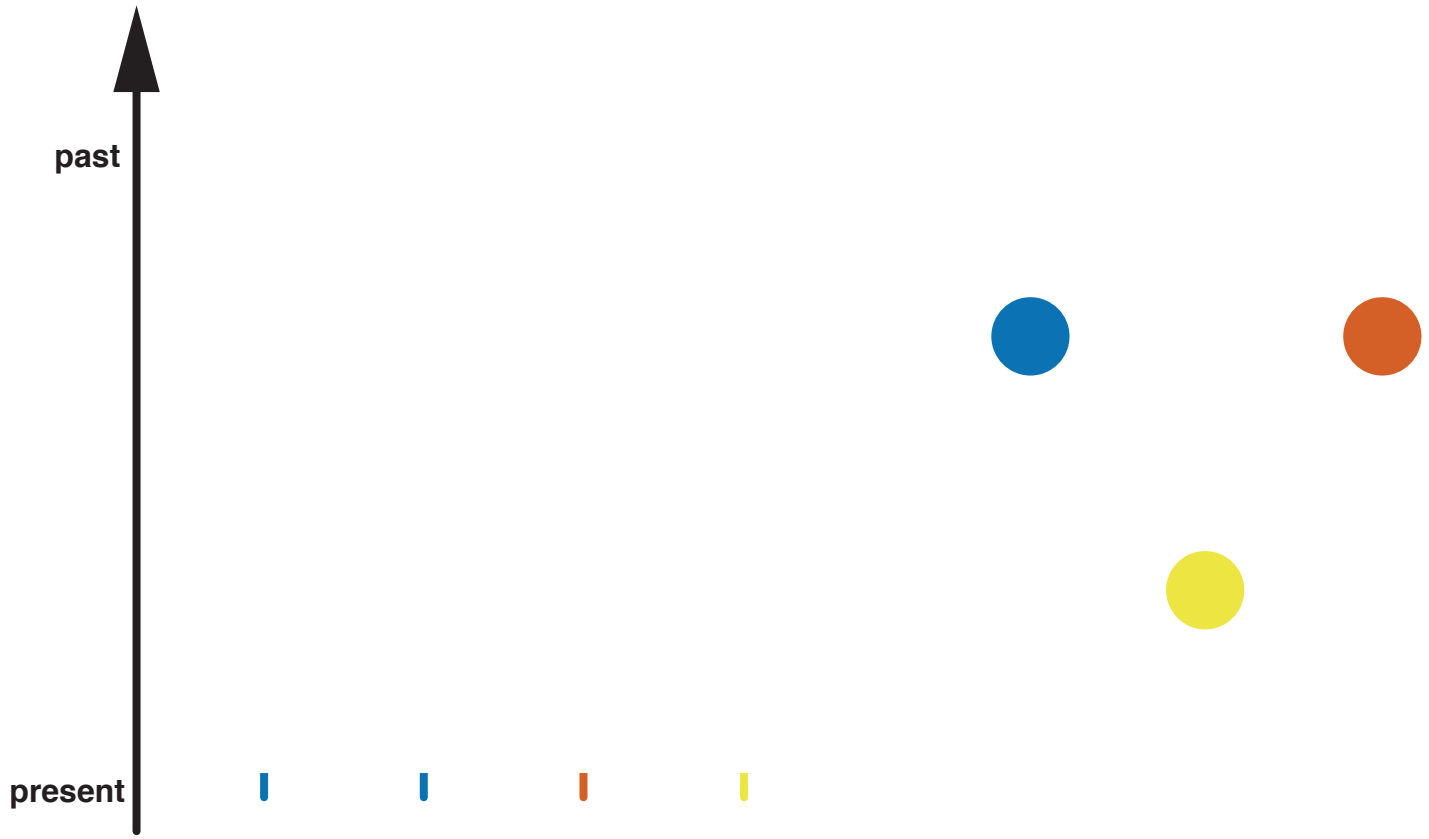
Migration rates in the DTA model can be informed by transportation Data



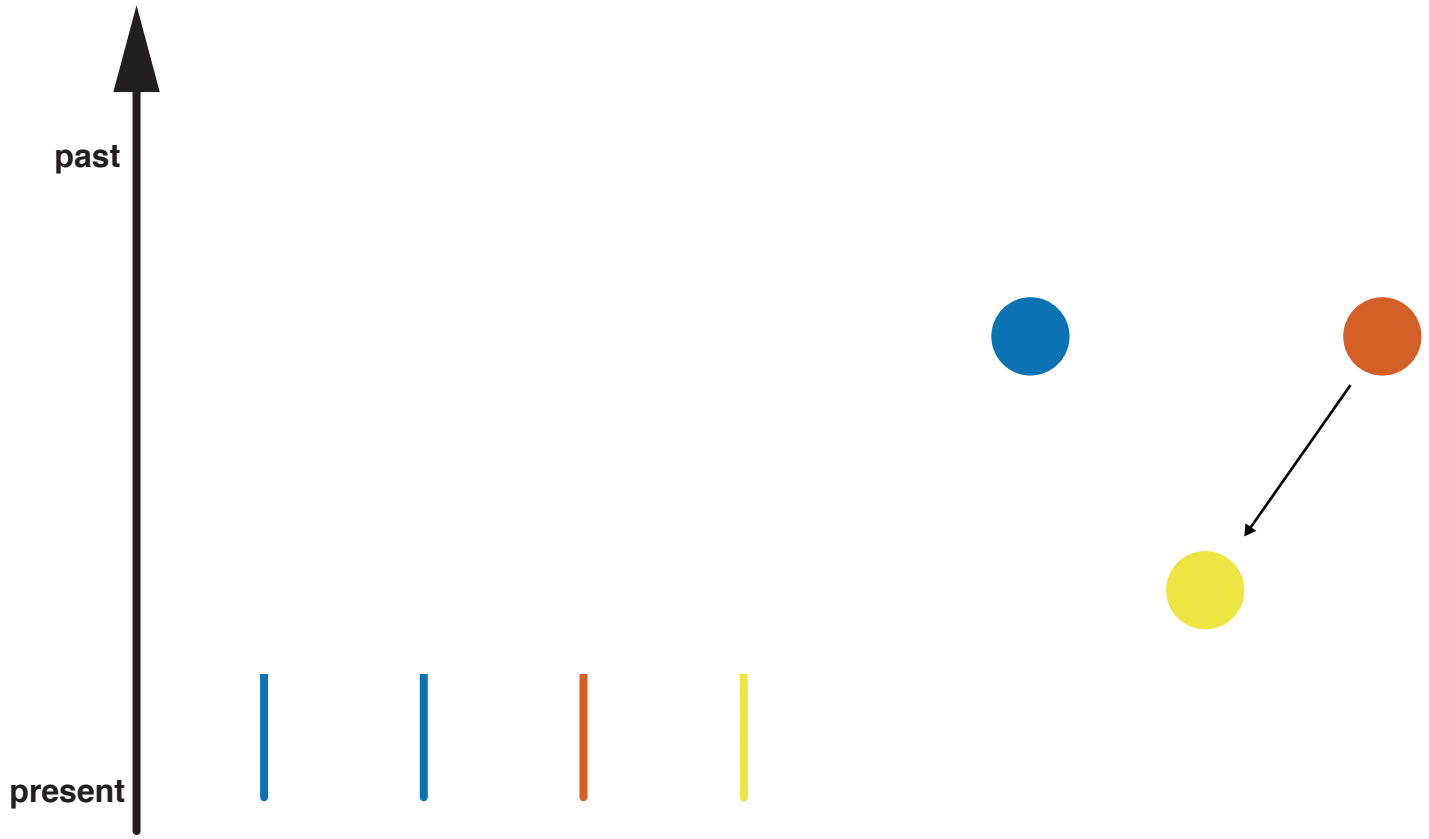


Lemey et al., 2014, *PLoS Path.*

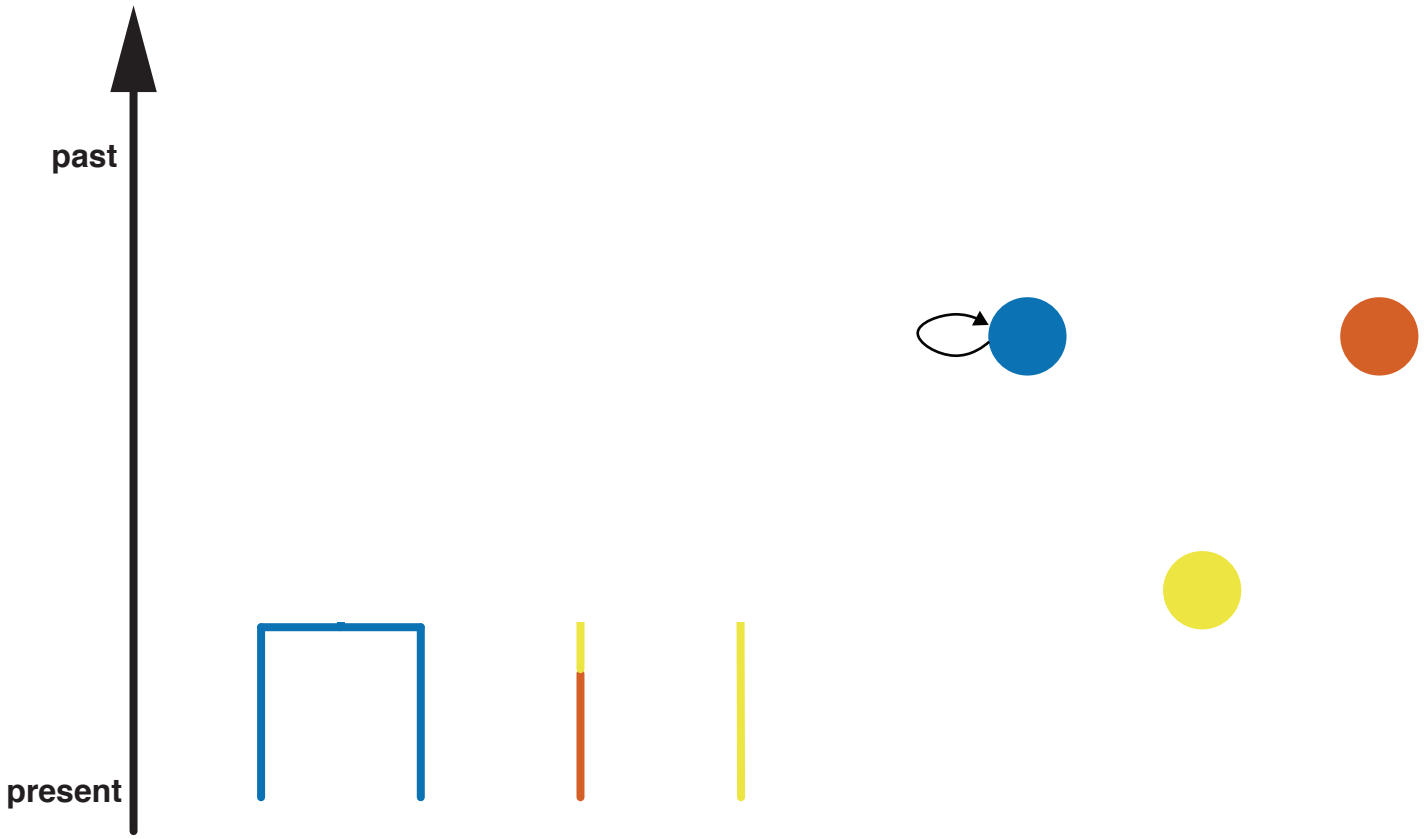
We track the ancestry of 4 viruses isolated from 4 people in 3 locations from present to past



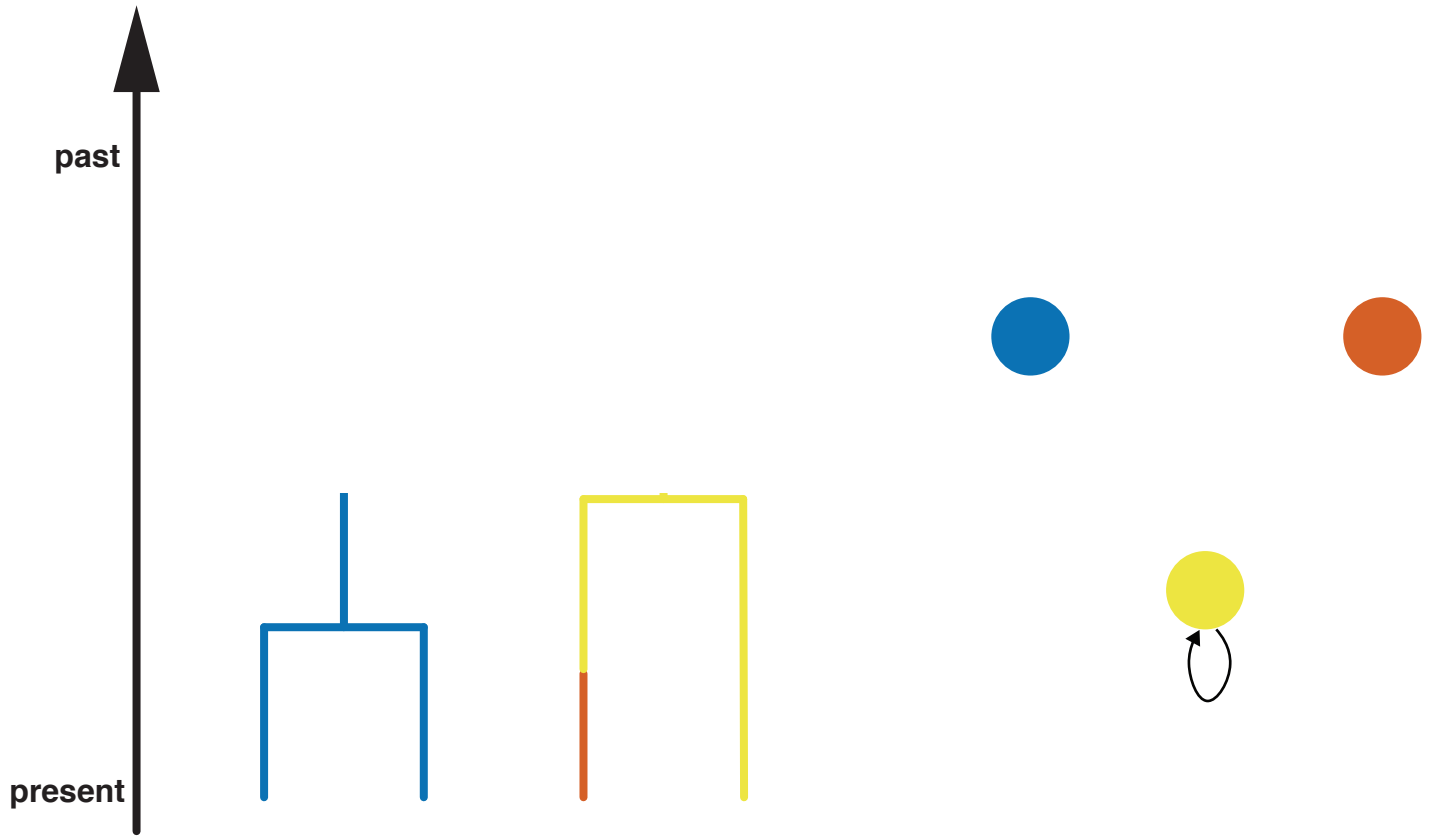
Lineages can have originated from a different location at a rate given by the migration rate.



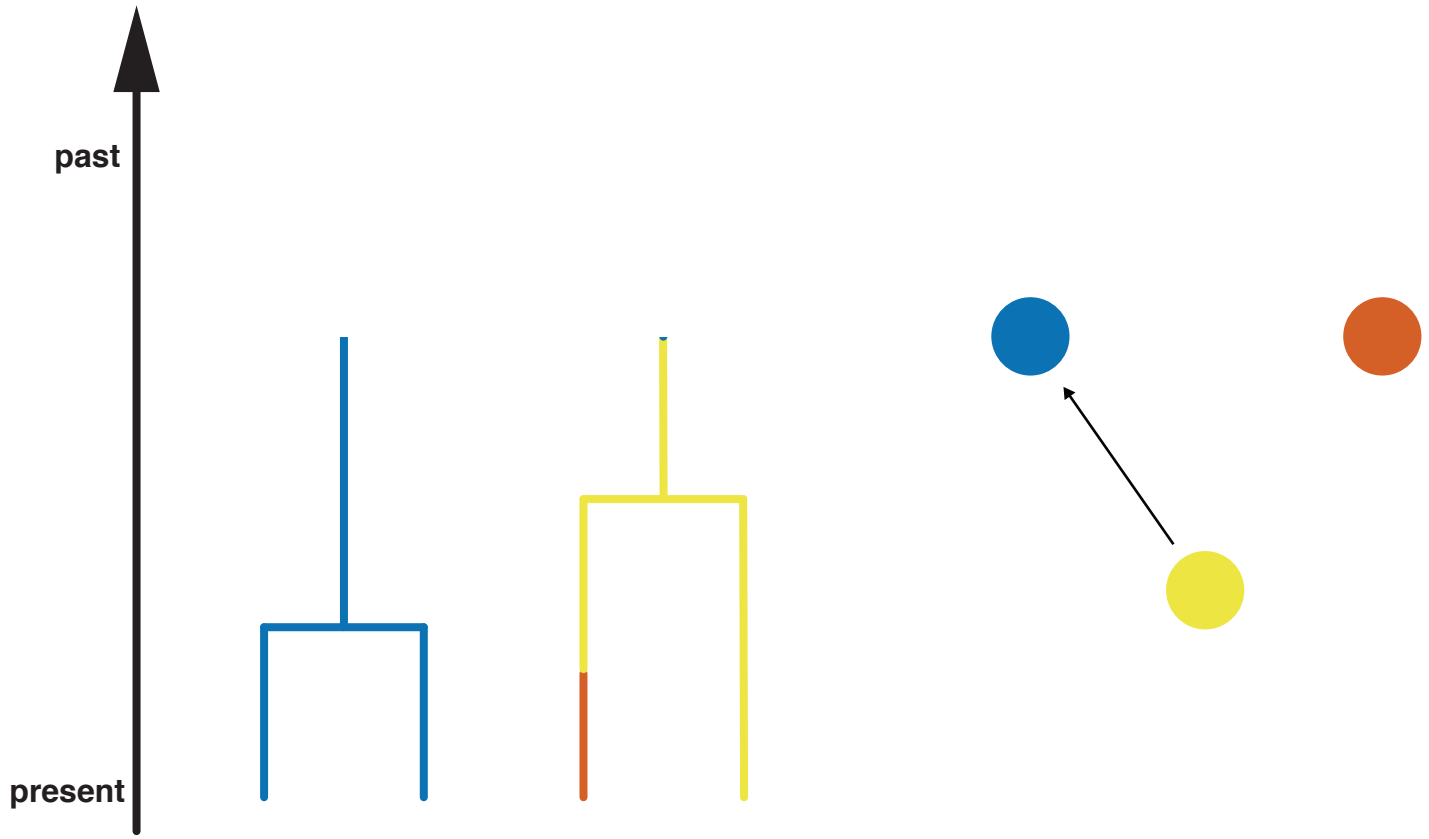
Two lineages can share a common ancestor at a rate inverse proportional to the Ne of location blue



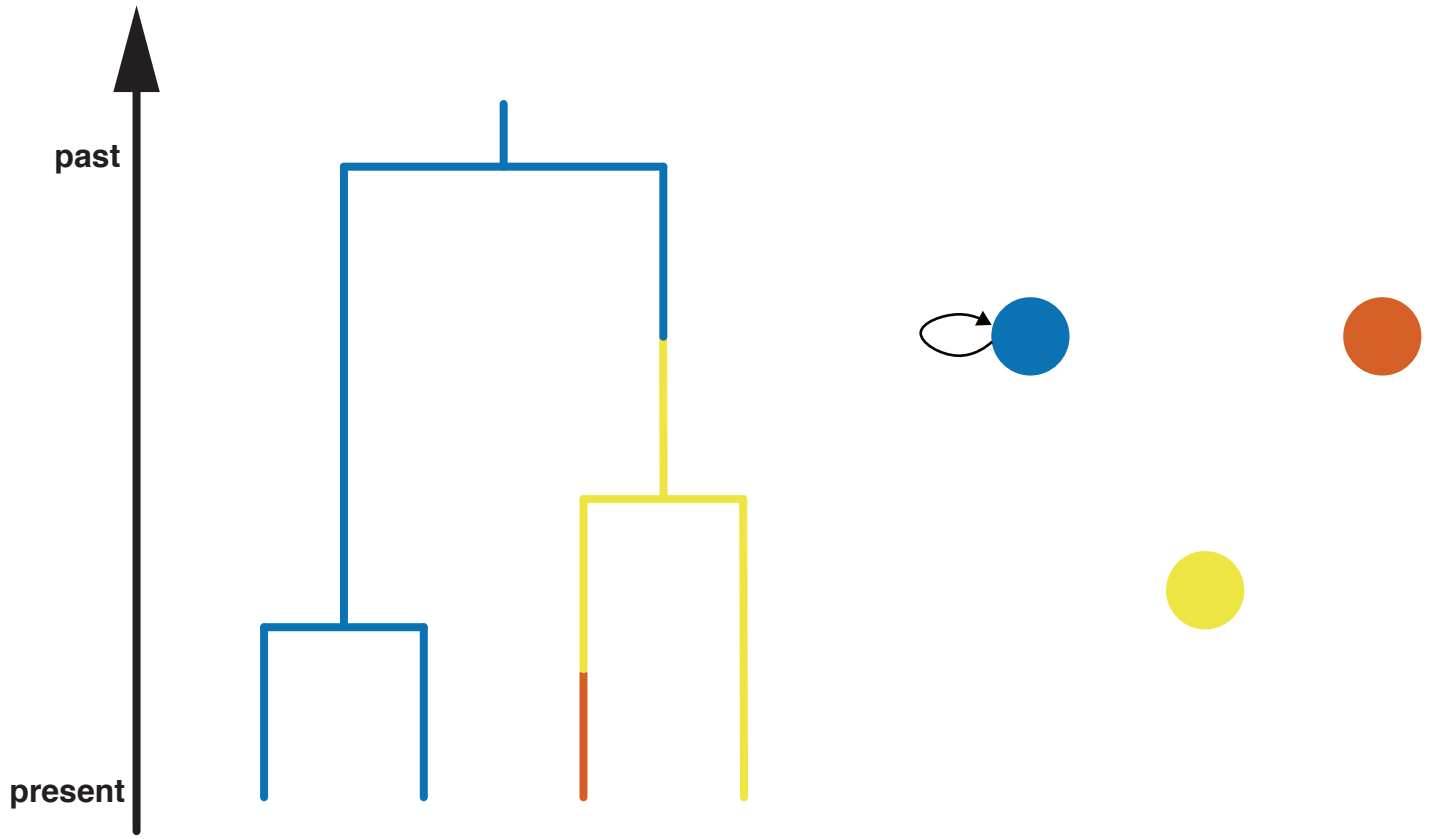
Two lineages can share a common ancestor at a rate inverse proportional to the Ne of location yellow



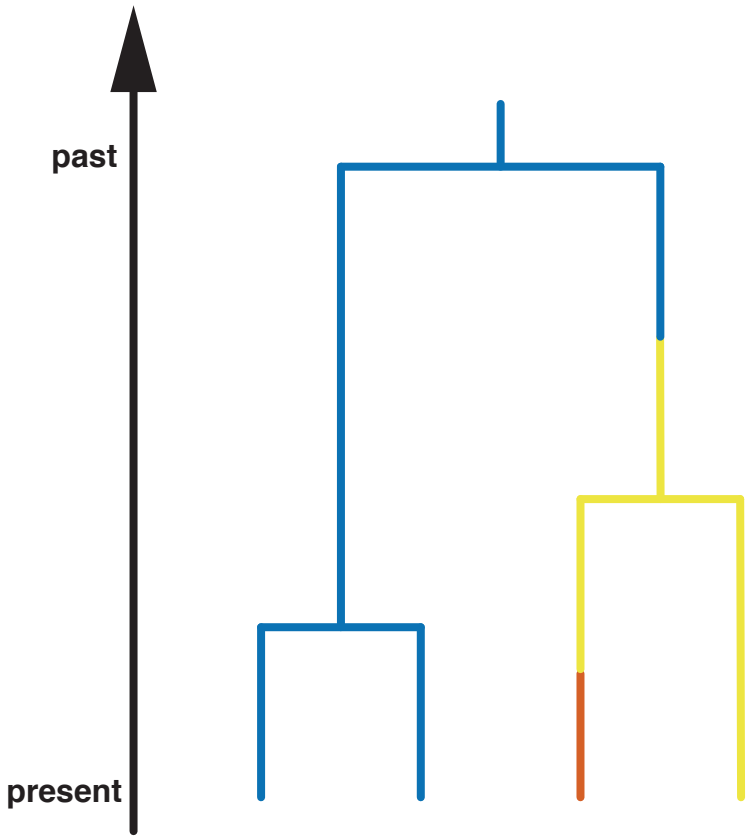
Lineages can have originated from a different location at a rate given by the migration rate.

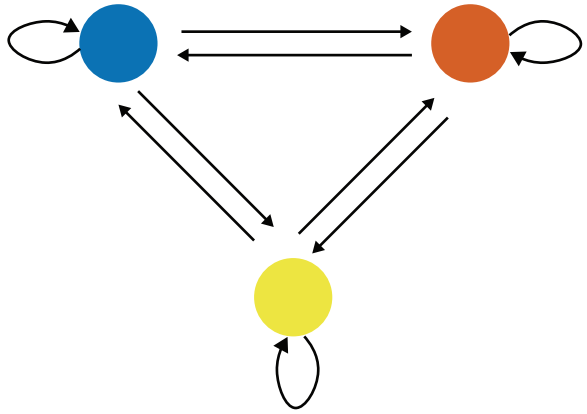


And then coalesce again in location blue, reaching the root of the tree

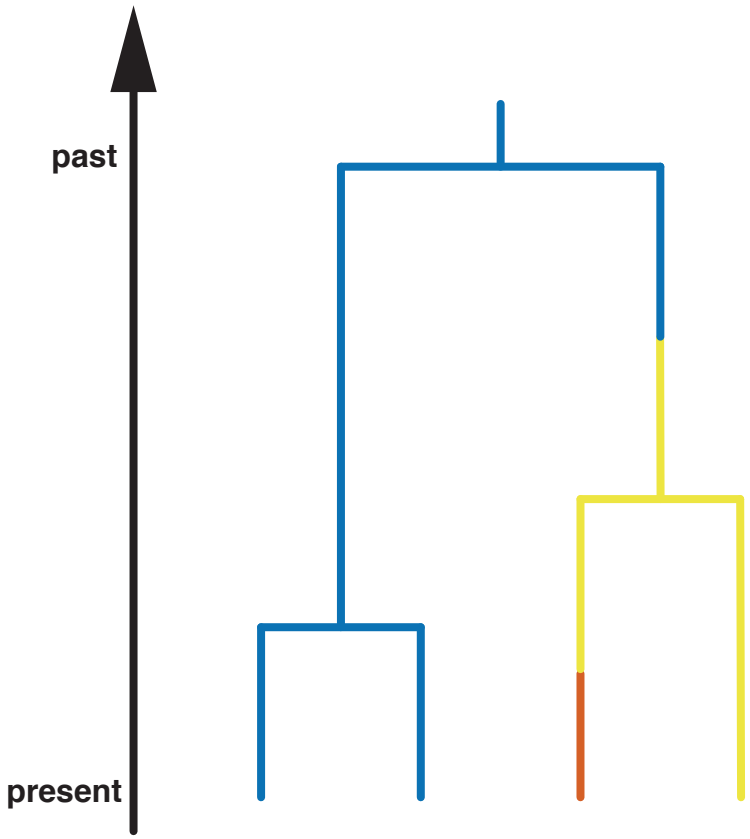


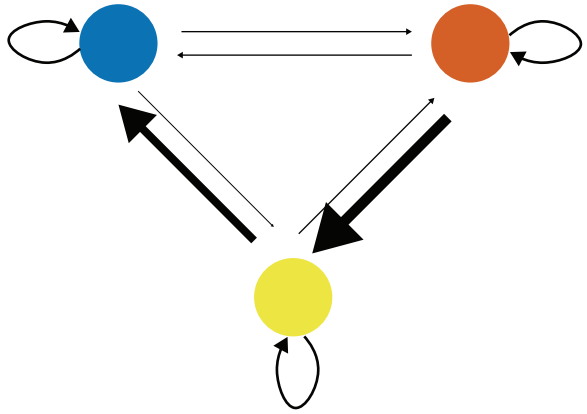
The structured coalescent models how lineages coalesce within and migrate between locations



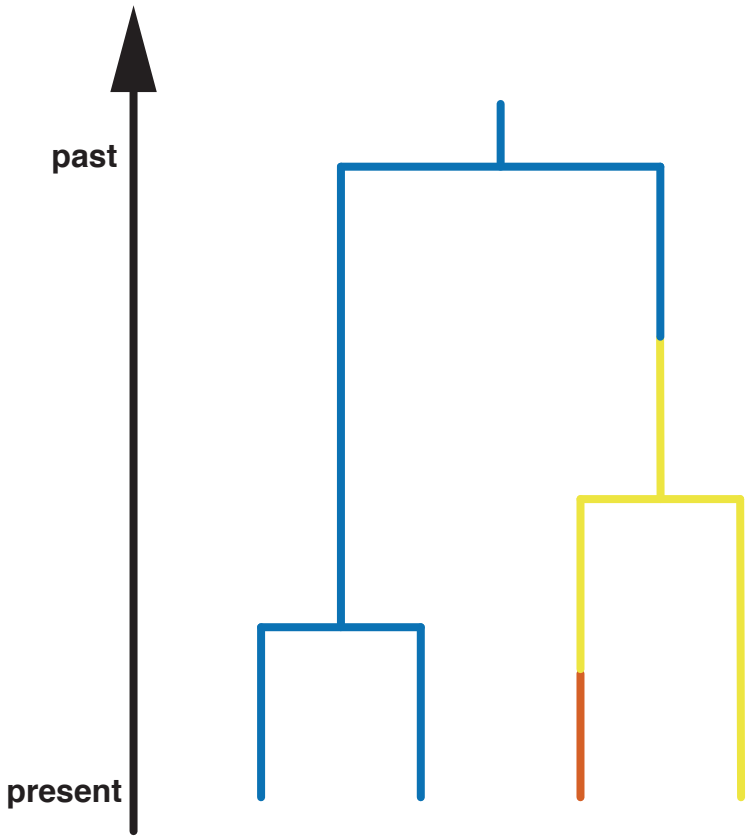


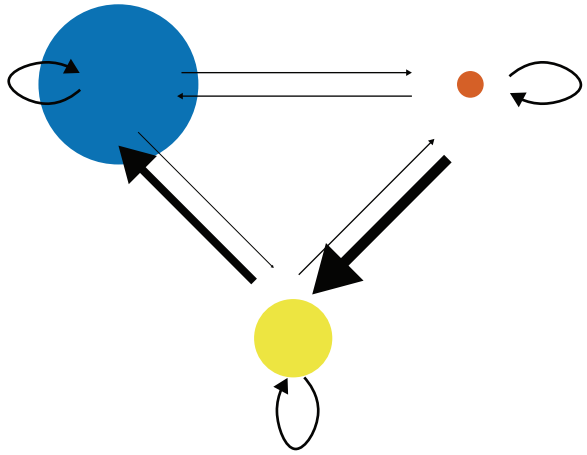
This allows us to estimate how migration rates between locations differ



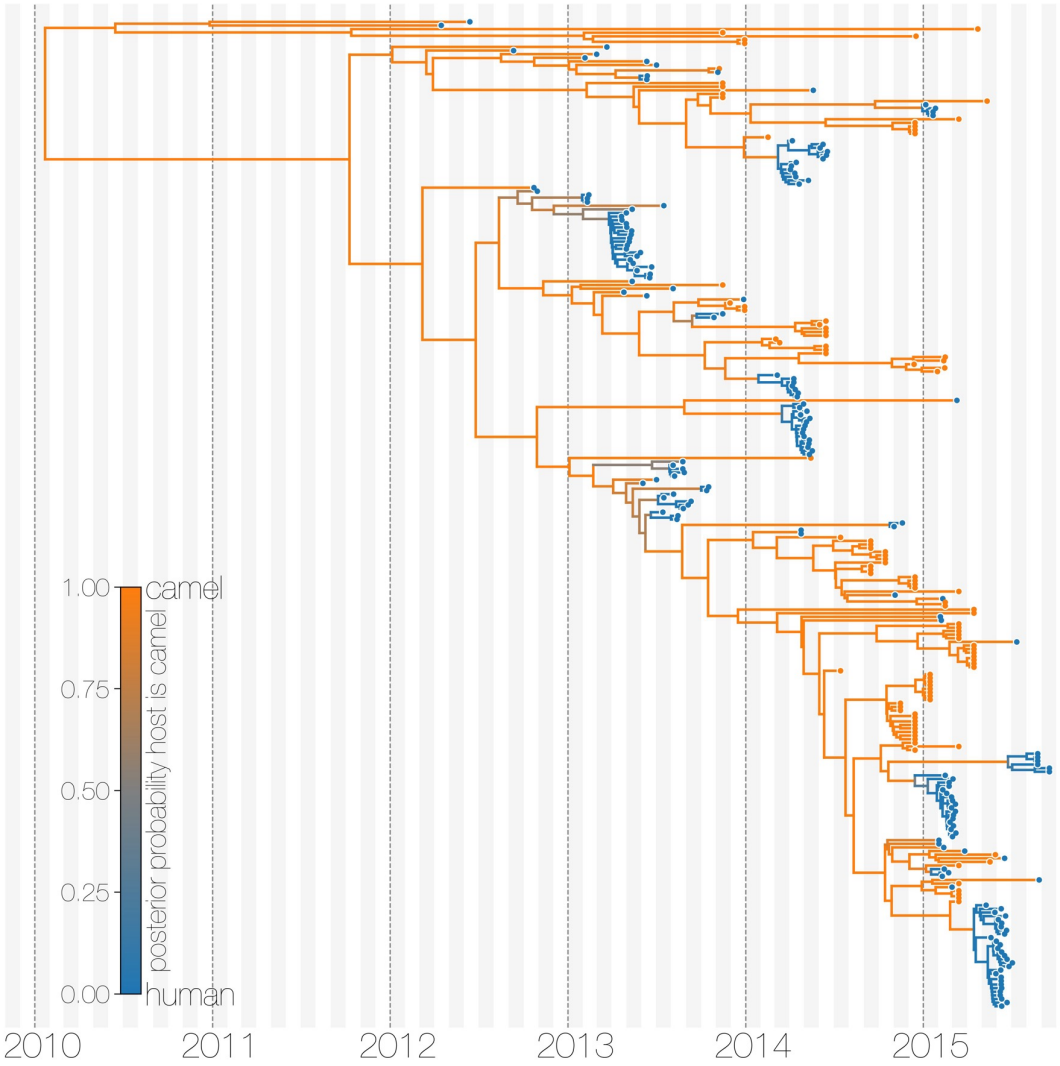


Or how the number of infected individuals differs between location



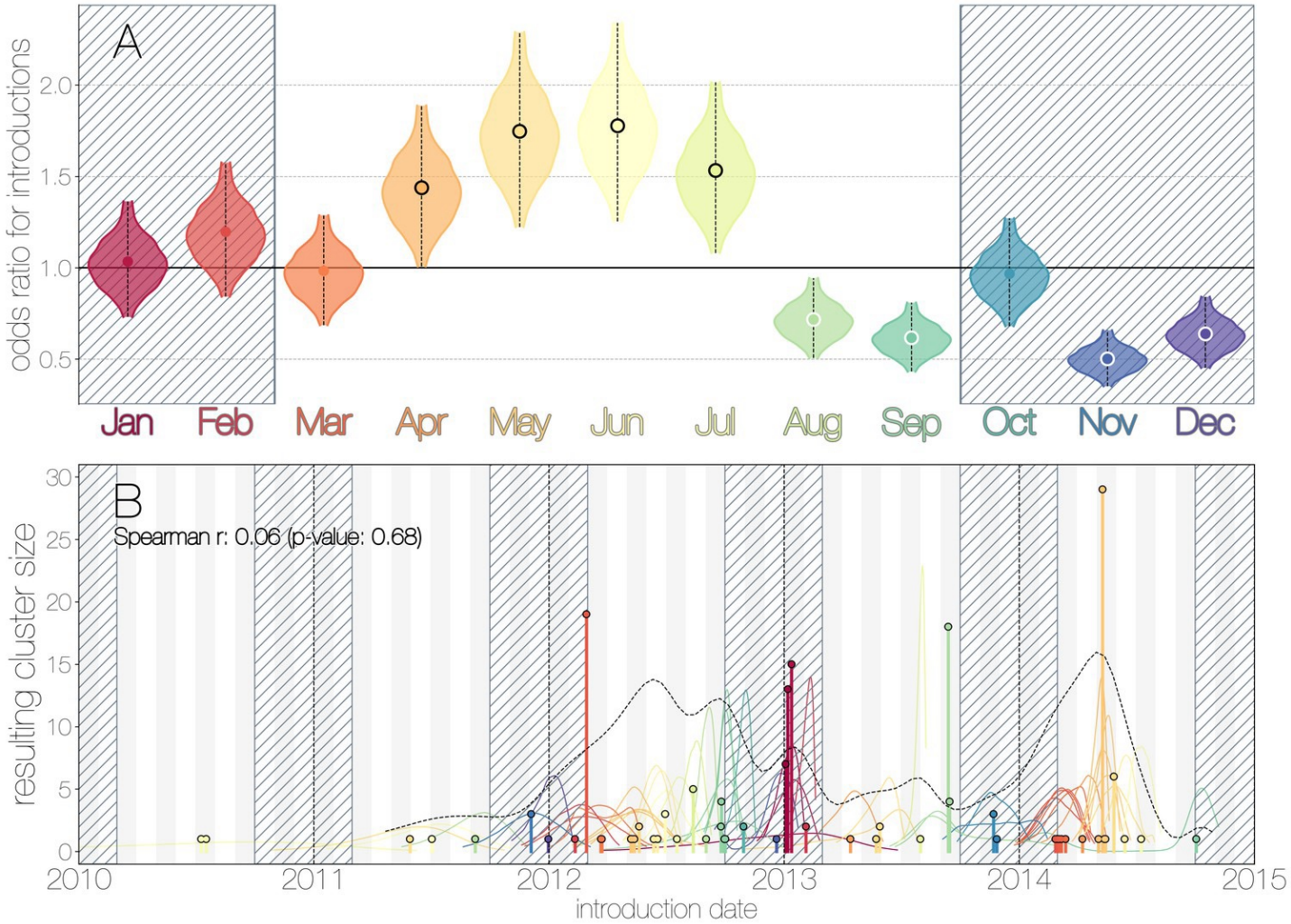


MERS sequences cluster by host of isolation



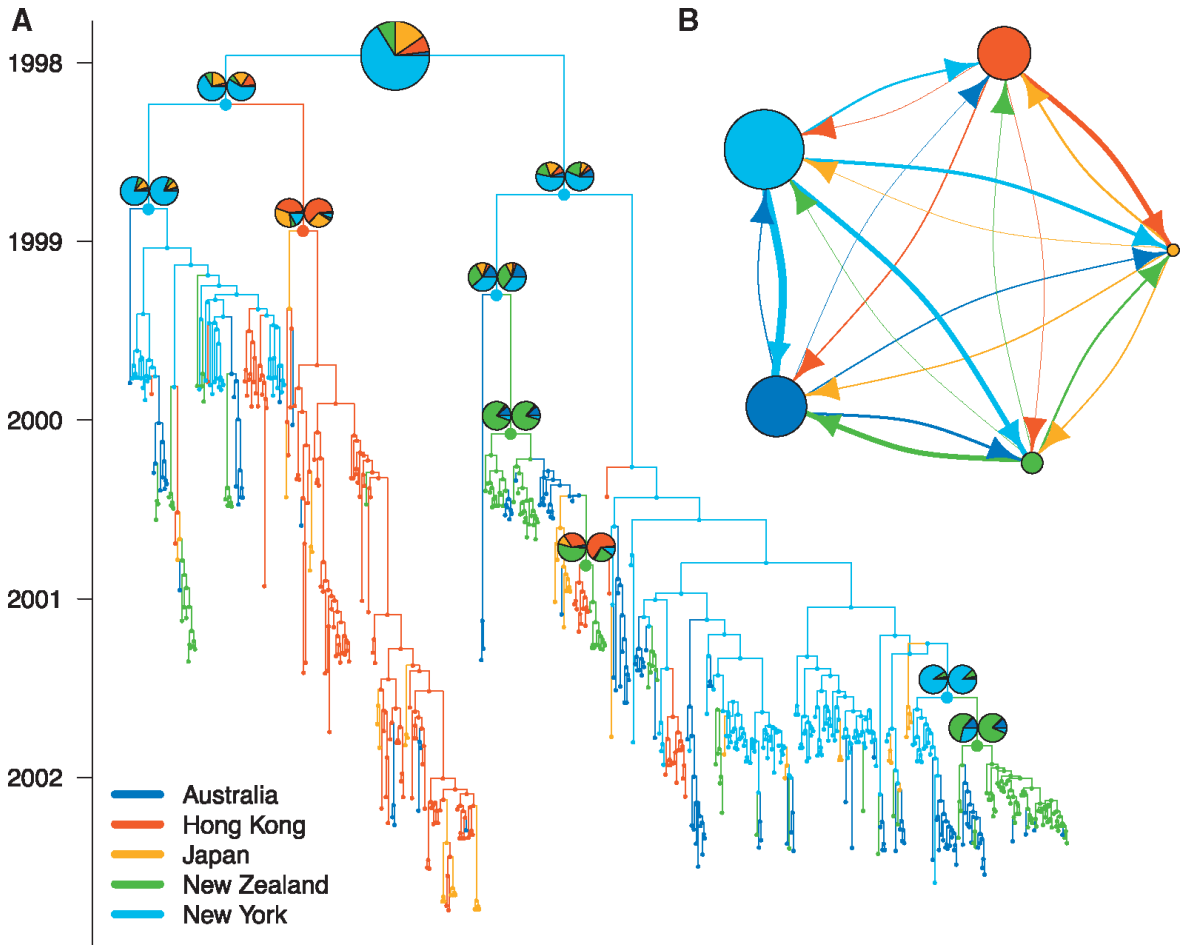
Dudas et al., 2018, *eLife*

Seasonal trends in camel to human transmissions



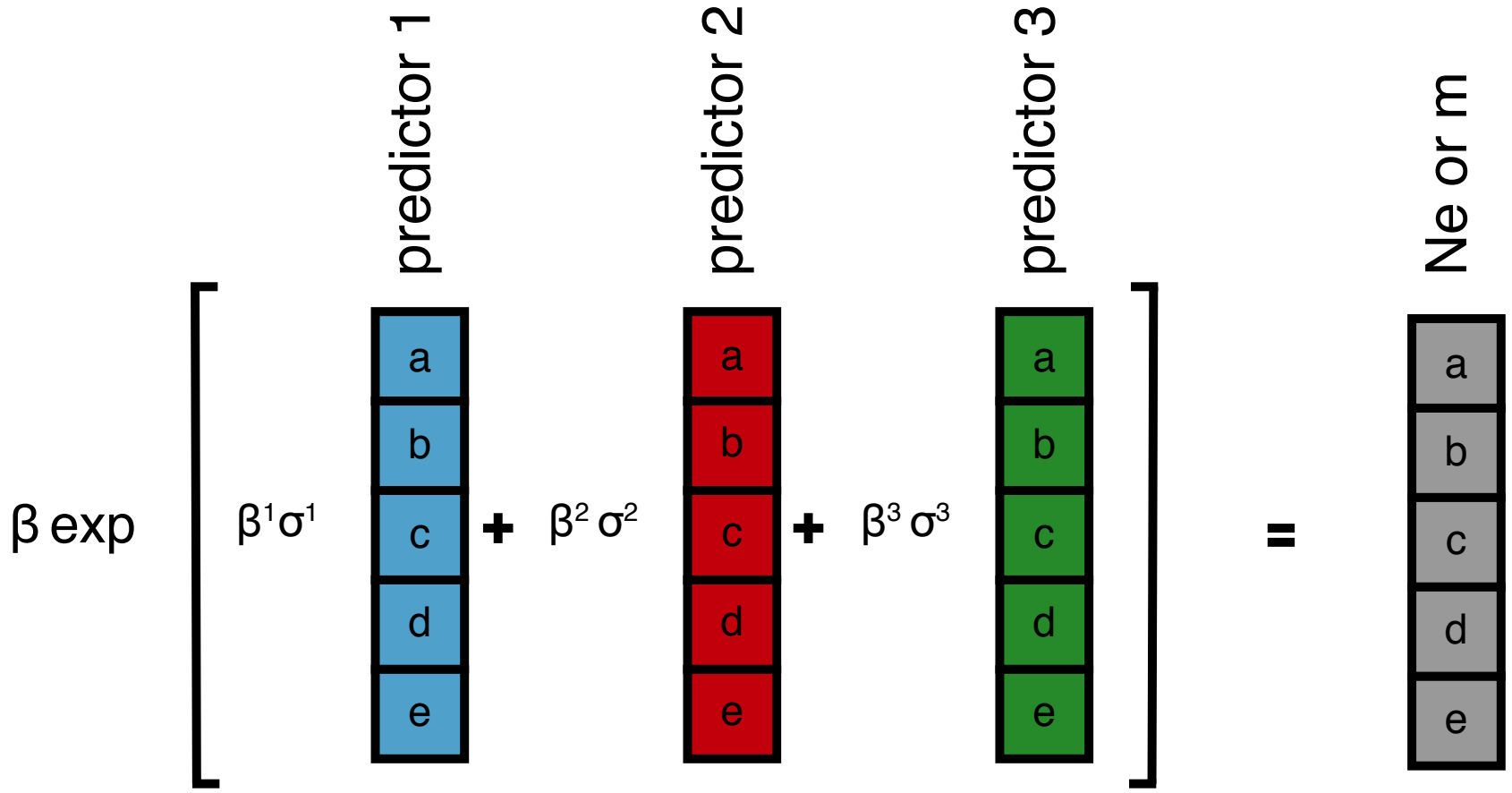
Dudas et al., 2018, *eLife*

MASCOT approximates the structured coalescent to dramatically improve efficiency



Müller et al., 2018, *Bioinf.*

Migration rates and effective population sizes can be informed using outside data in a GLM approach



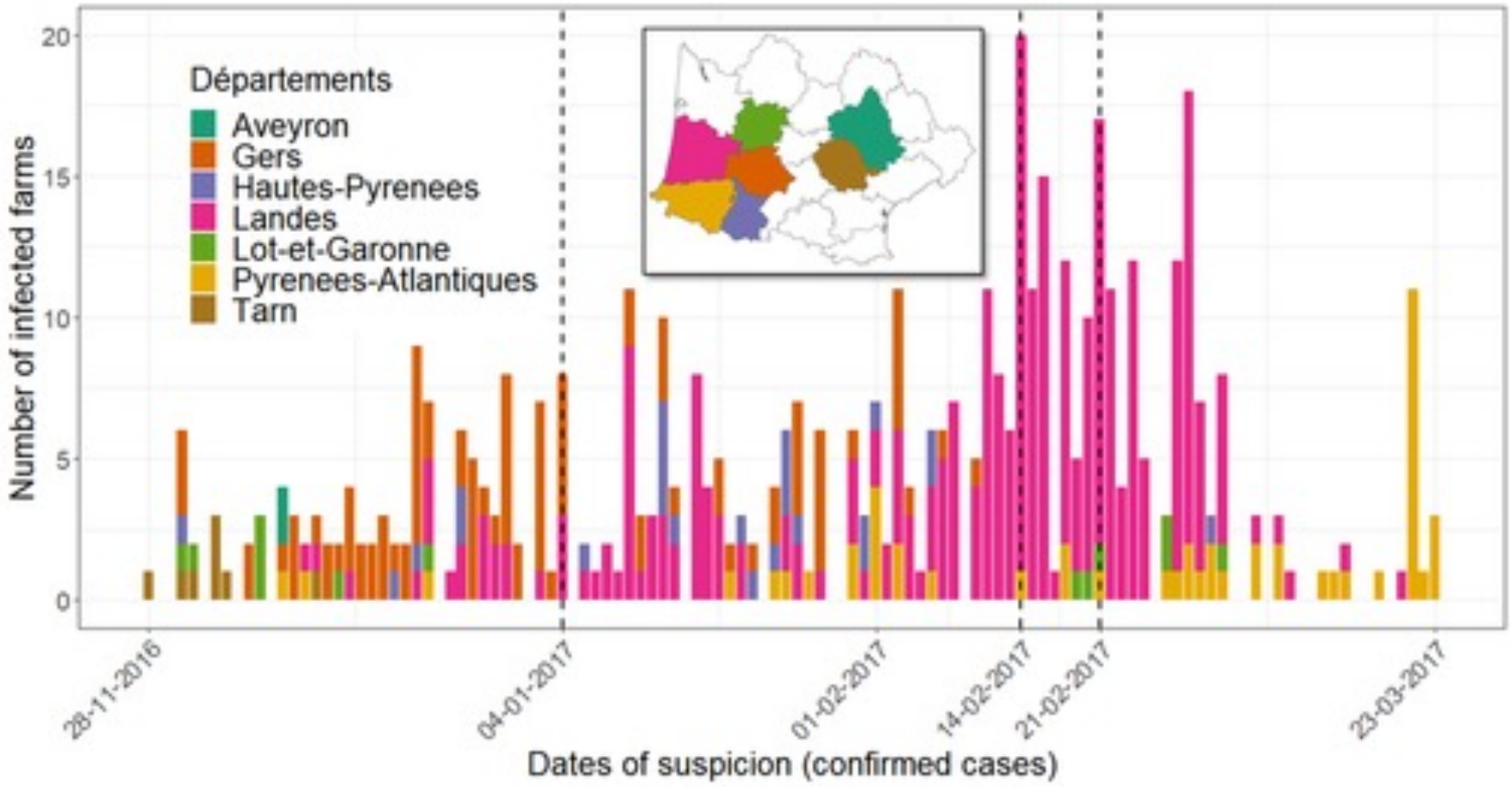
In 2016 and 2017, there was an Influenza A/H5N8 outbreak in France

French Foie Gras Makers Fear the Worst as Bird Flu Toll Rises

The bird flu epidemic in southwest France, home to most duck and foie gras producers, has led to more than three million poultry being killed.

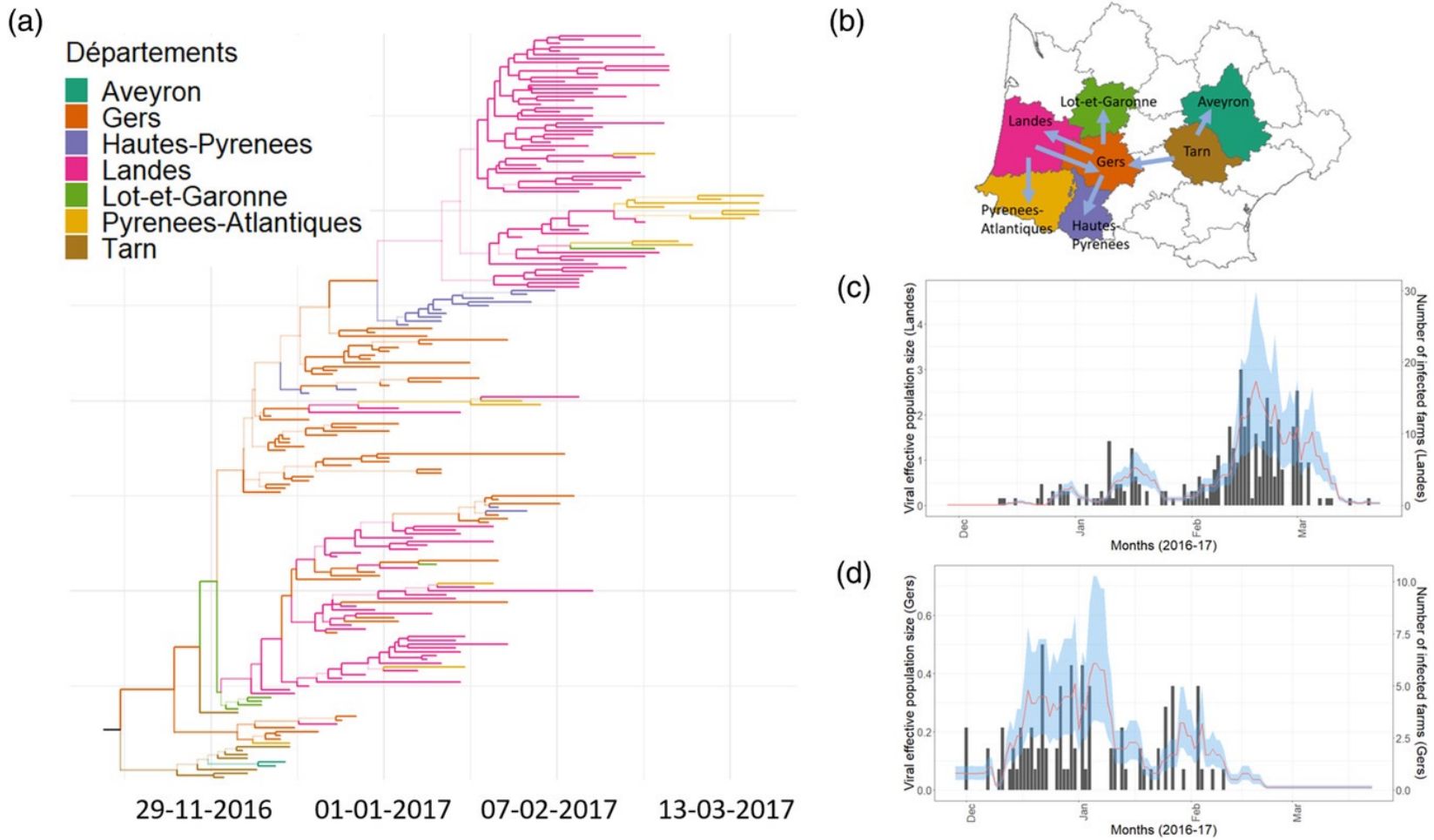
https://www.nbcnews.com/news/world/french-foie-gras-makers-fear-worst-bird-flu-toll-rises-n721276

Cases first appeared in the department (state) of tarn and H5N8 was soon detected in other regions as well



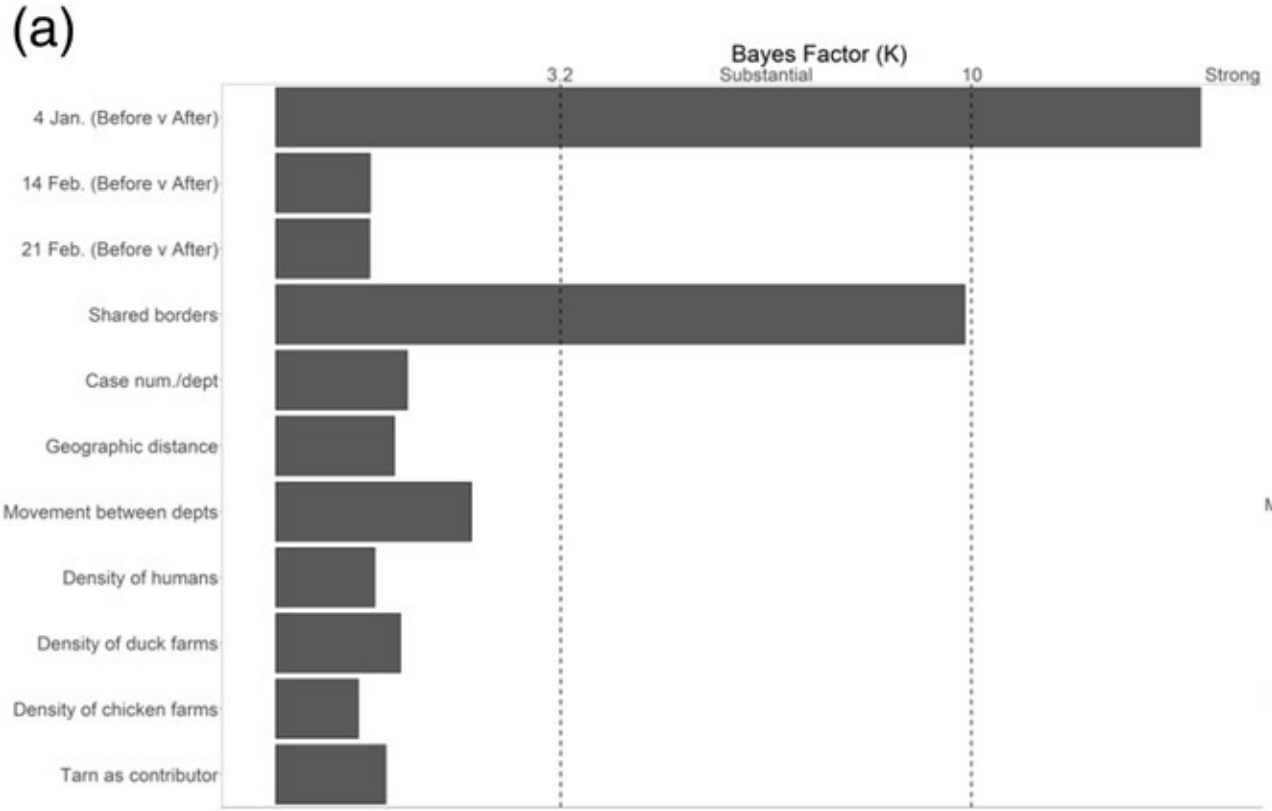
Chakraborty et al., *2022, Transboundary and Emerging Diseases*

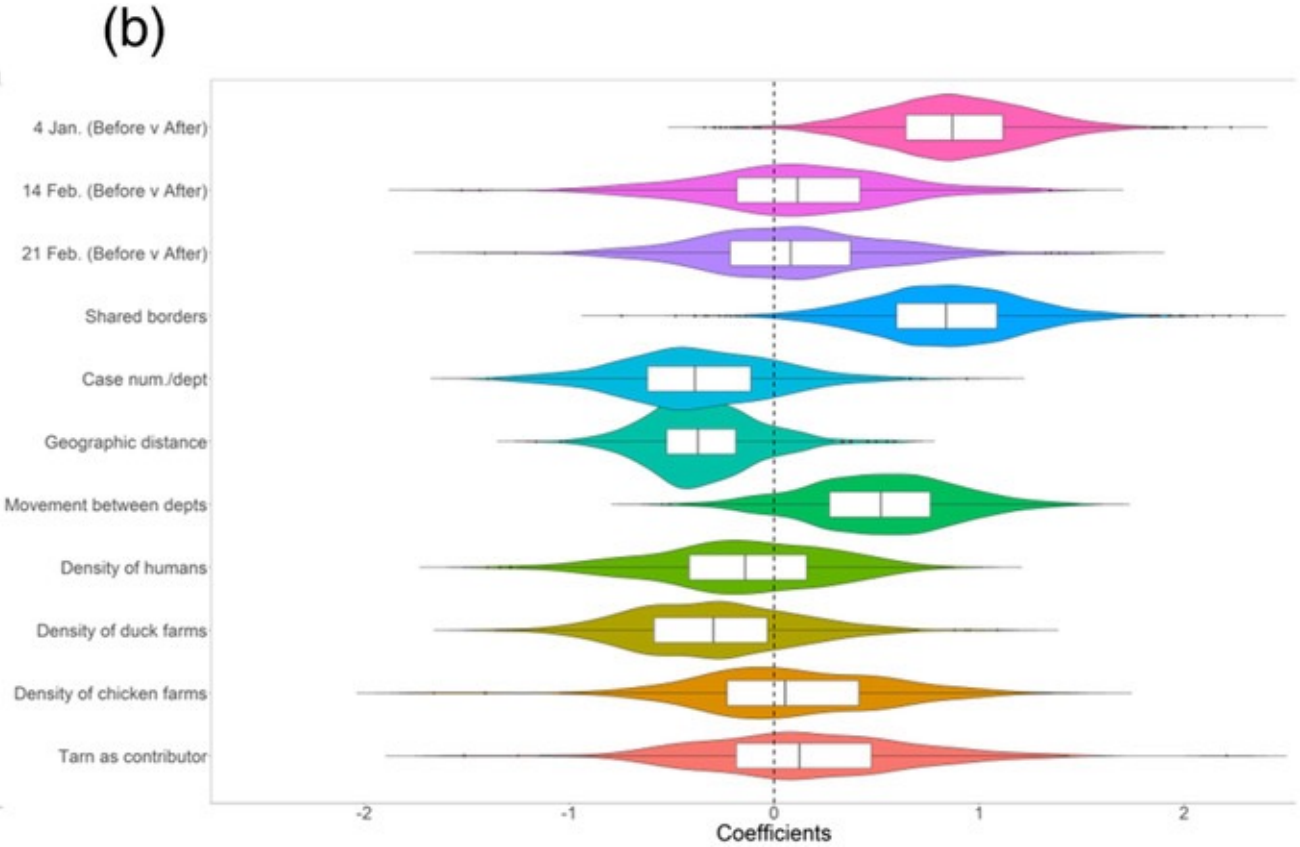
MASCOT-GLM reconstructs Tarn to be the most likely source, consistent with case data



Chakraborty et al., *2022, Transboundary and Emerging Diseases*

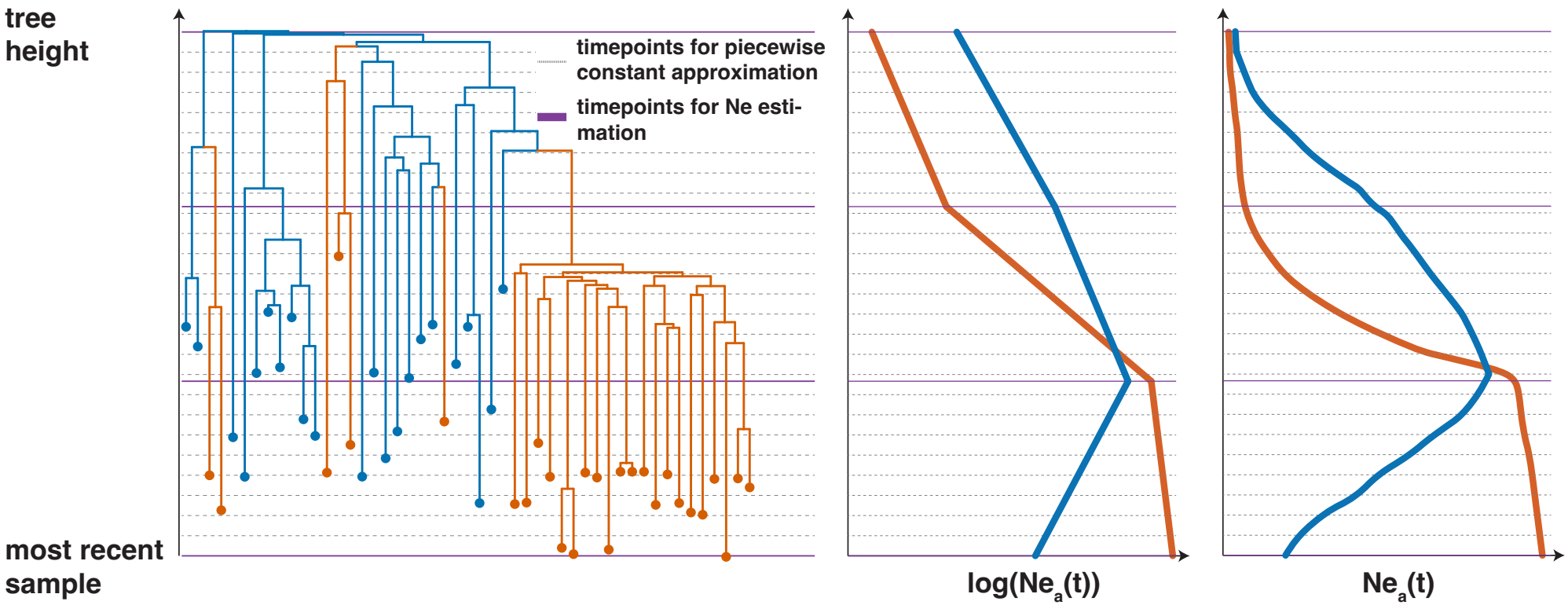
Evidence that the initial culling of birds reduced migration between location





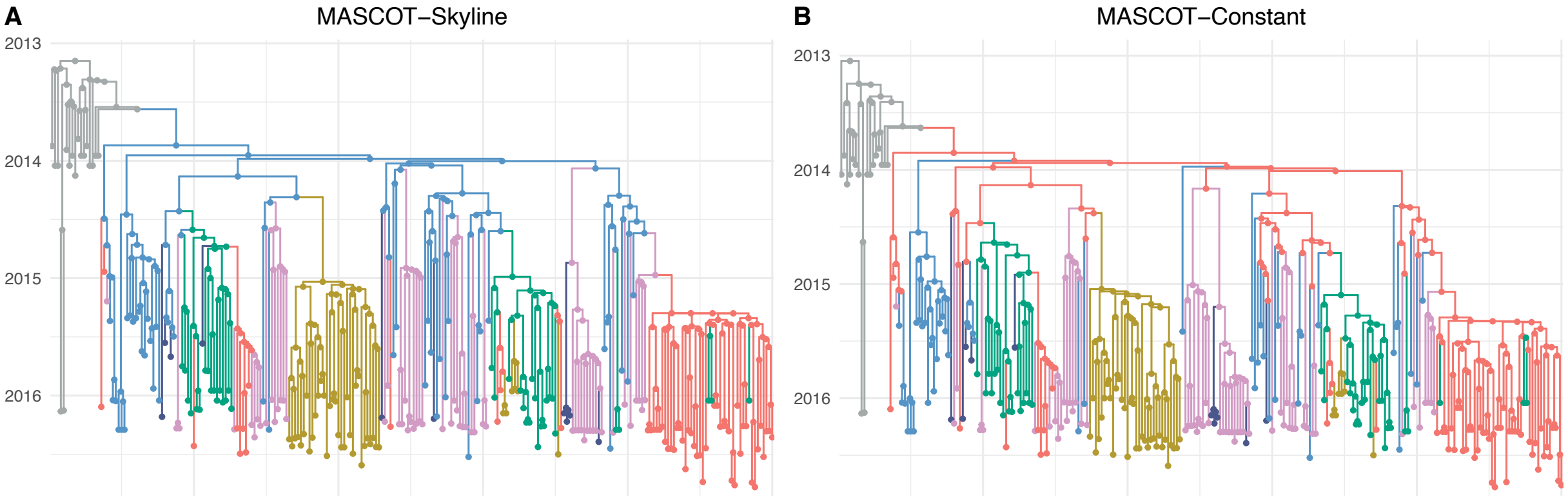
Chakraborty et al., *2022, Transboundary and Emerging Diseases*

Effective population size dynamics can be modelled using MASCOT-Skyline



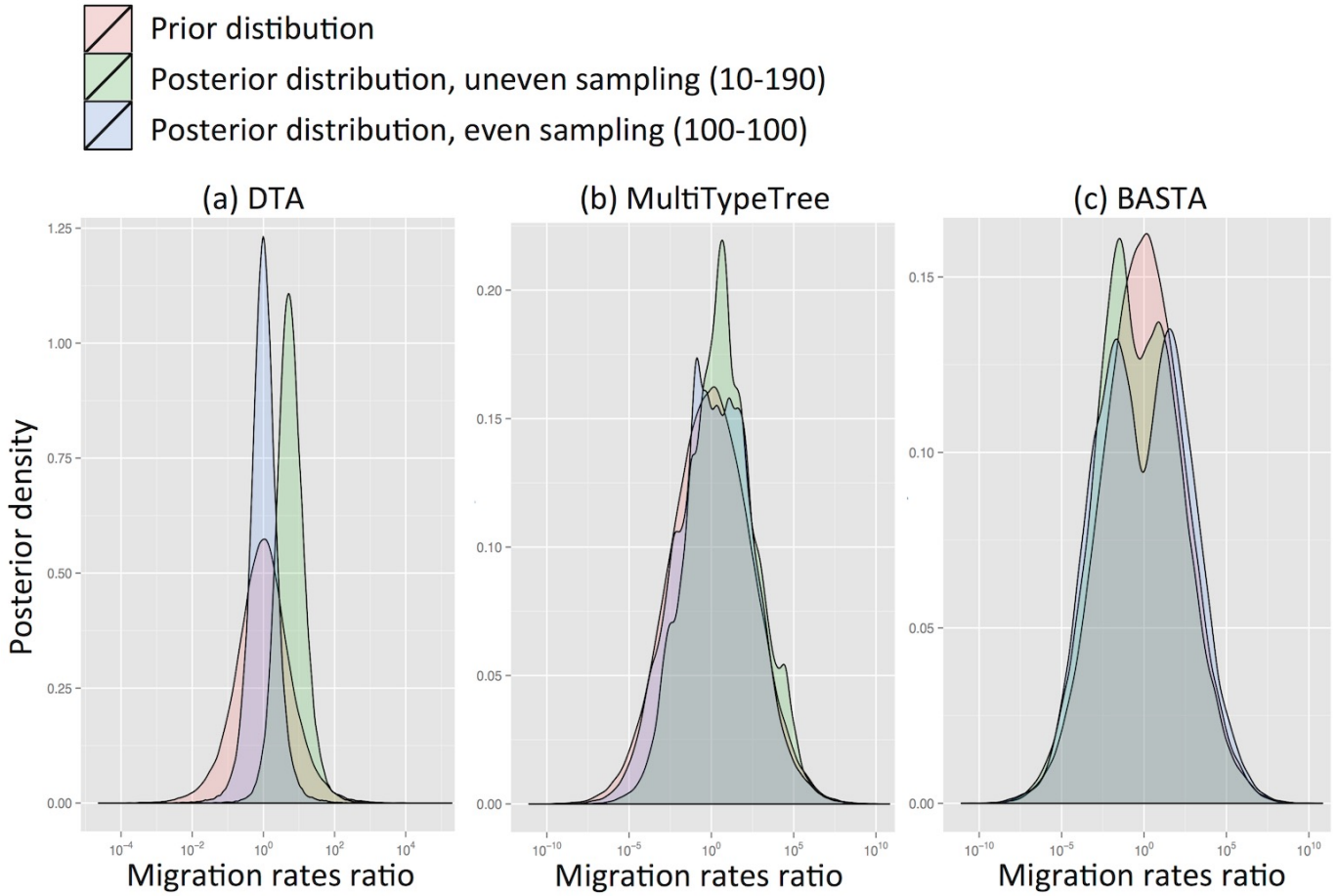
Müller *et al.* BioRxiv, 2024

Accounting for population dynamics in structured coalescent can impact ancestral state reconstruction



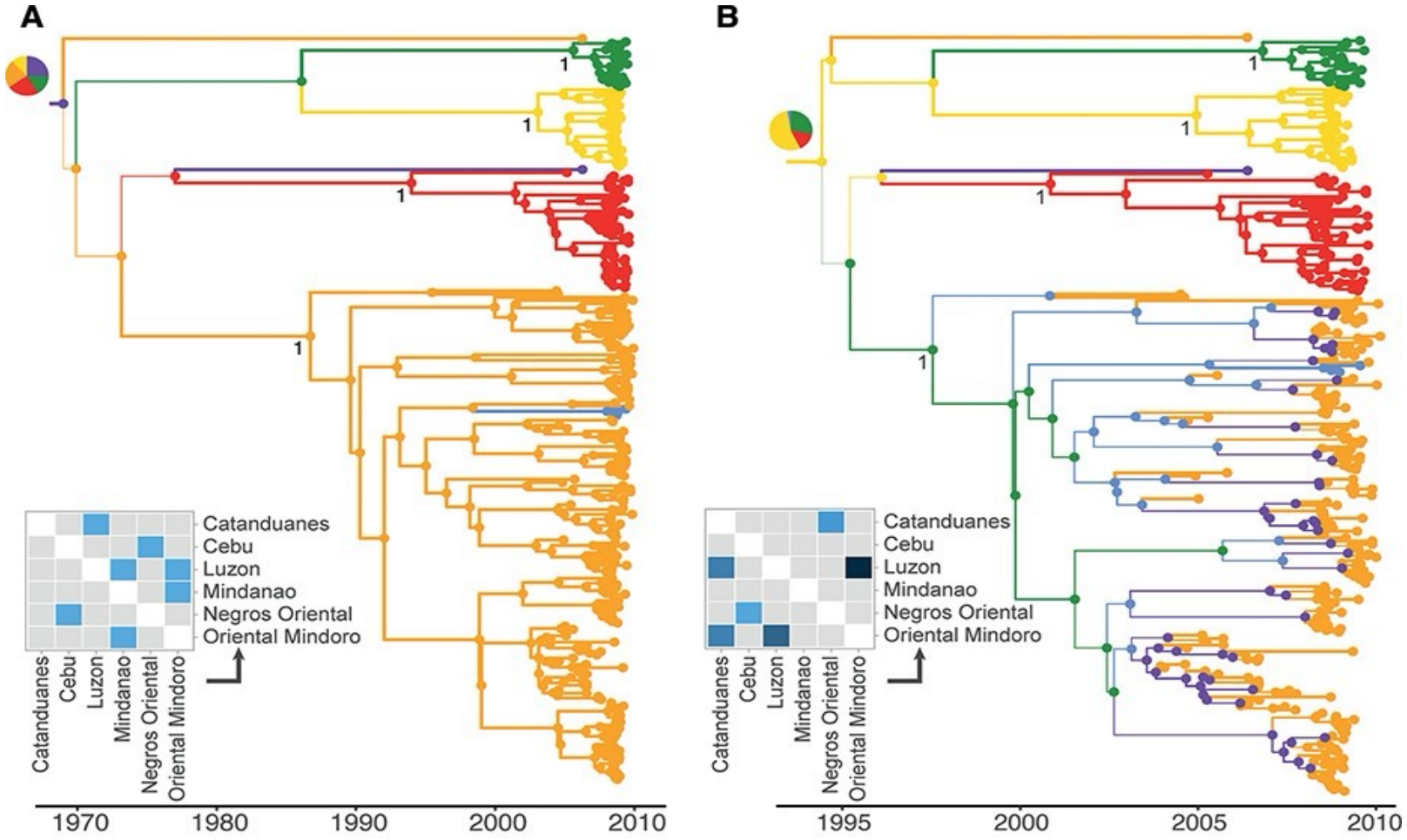
Müller *et al.* BioRxiv, 2024

Biases in Structured phylodynamics



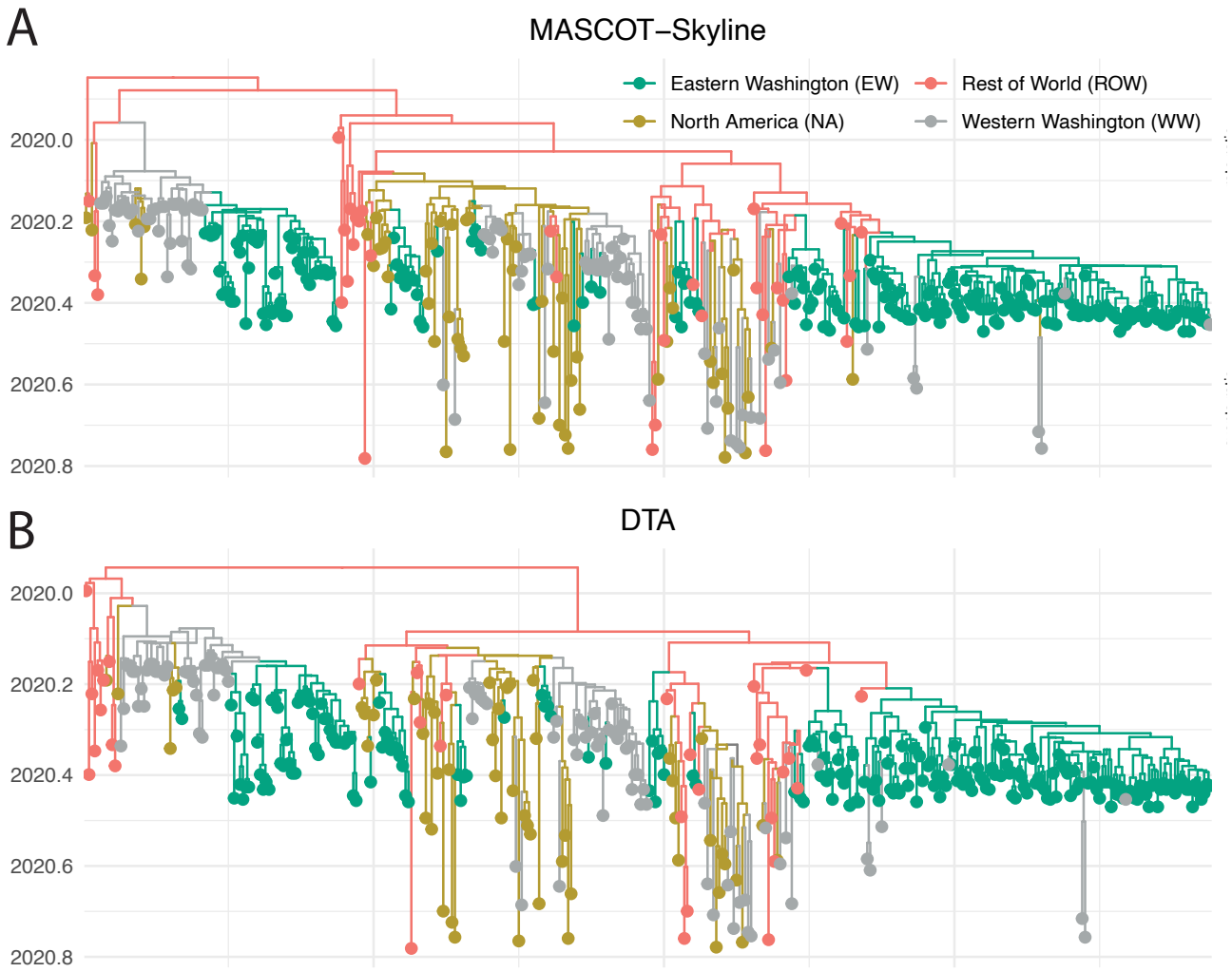
DeMaio *et al.* PLoS Biology, 2015

DTA typically returns a rather parsimonious ancestral state reconstruction



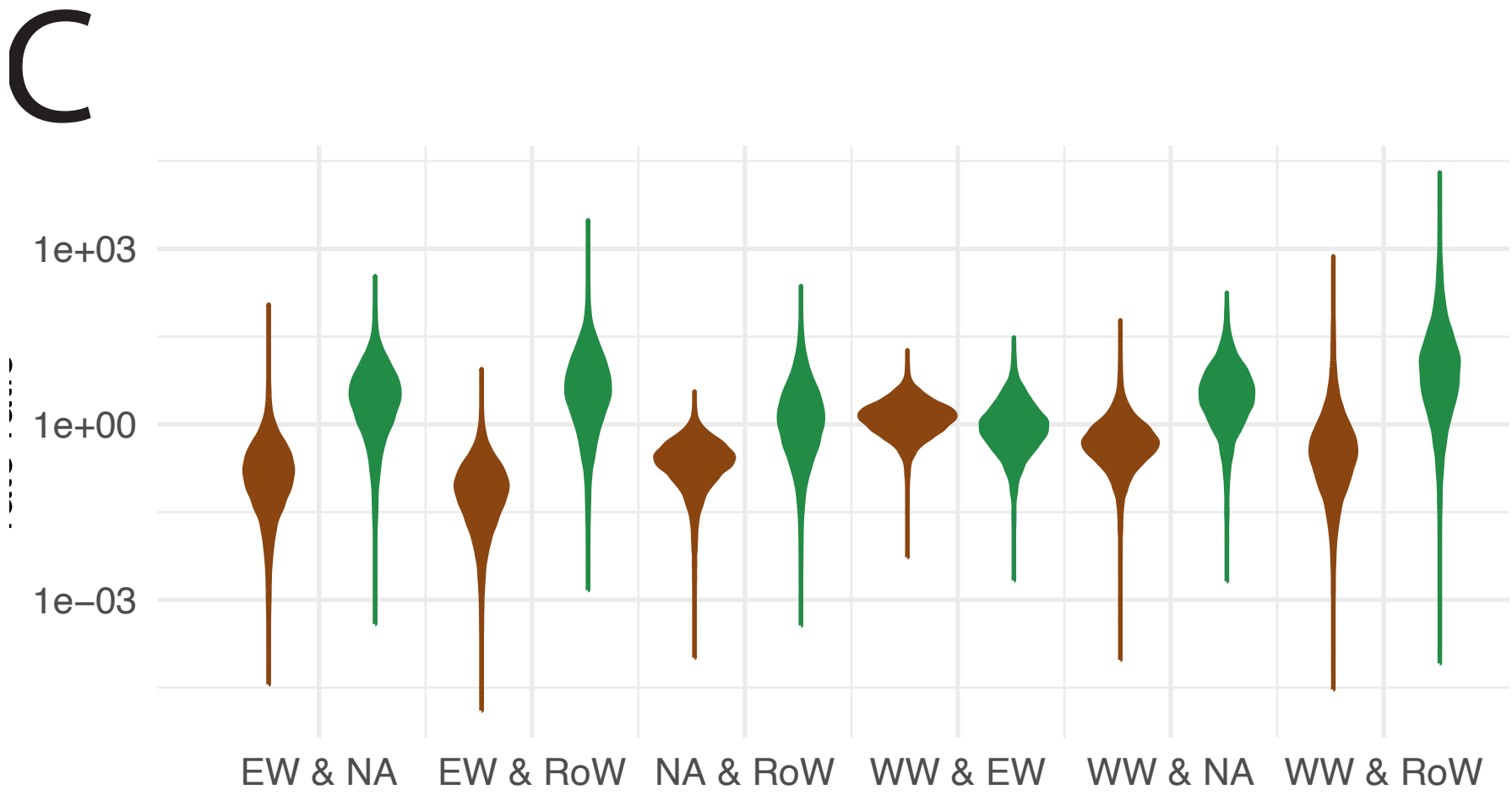
Layan *et al.* Virus Evolution, 2023

Different methods can interpret very similar observations differently



Müller *et al.* BioRxiv, 2024

Different methods can interpret very similar observations differently

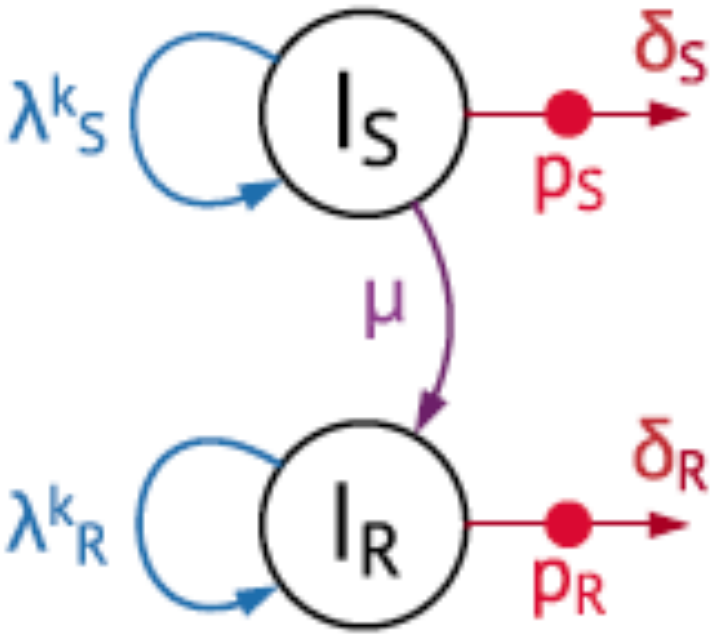


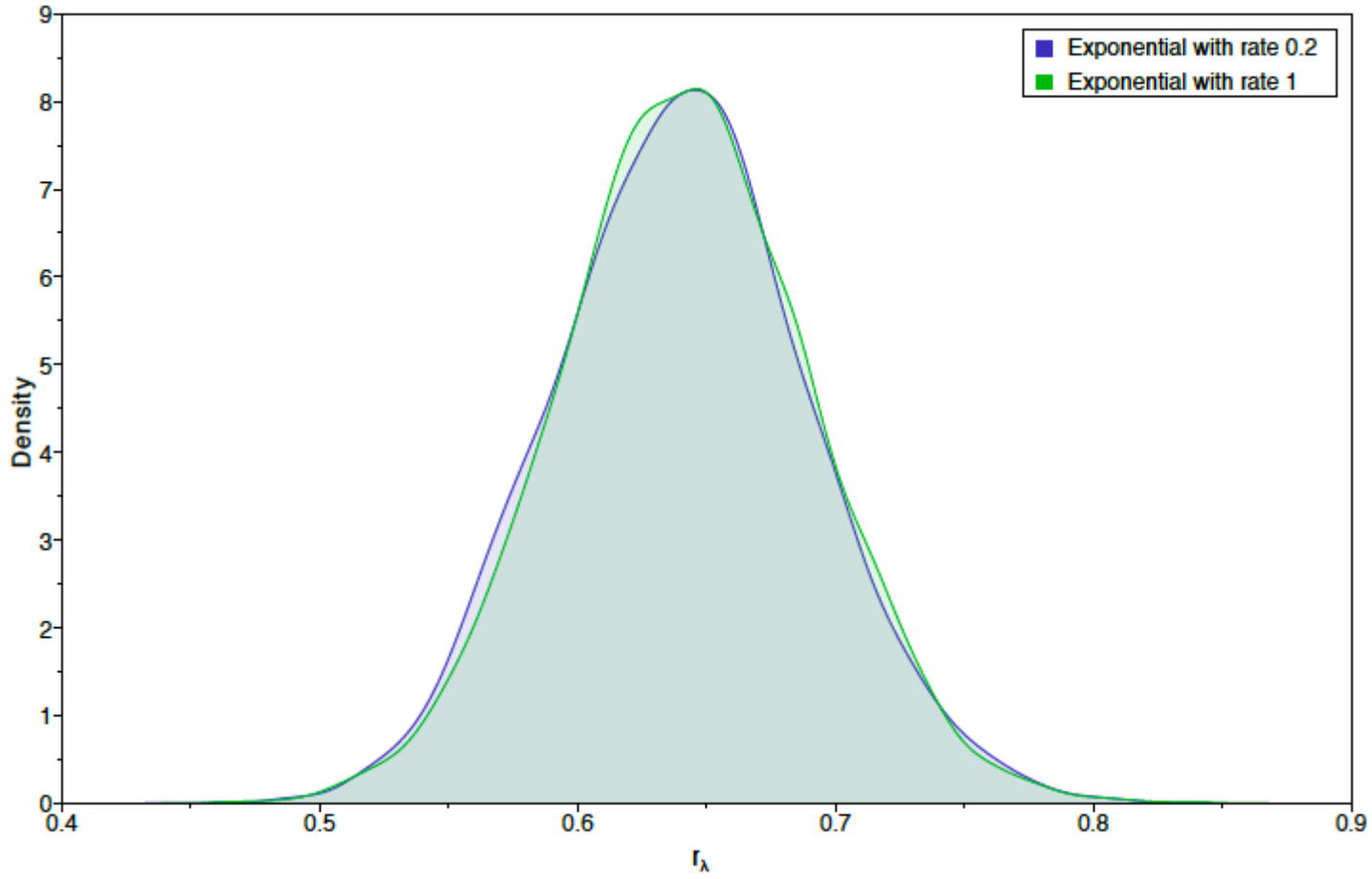
Müller *et al.* BioRxiv, 2024

Variations

- Typed or untyped trees, i.e. trees including the migration history or not (BDMM)
- Fixed types at the tips (BDMM) or estimated (MSBD)
- Extended multi-type birth-death model with birth rate between demes $\lambda_{i,j}$ $\forall$ $i \neq j$ (BDMM, XML only)

Example: tuberculosis



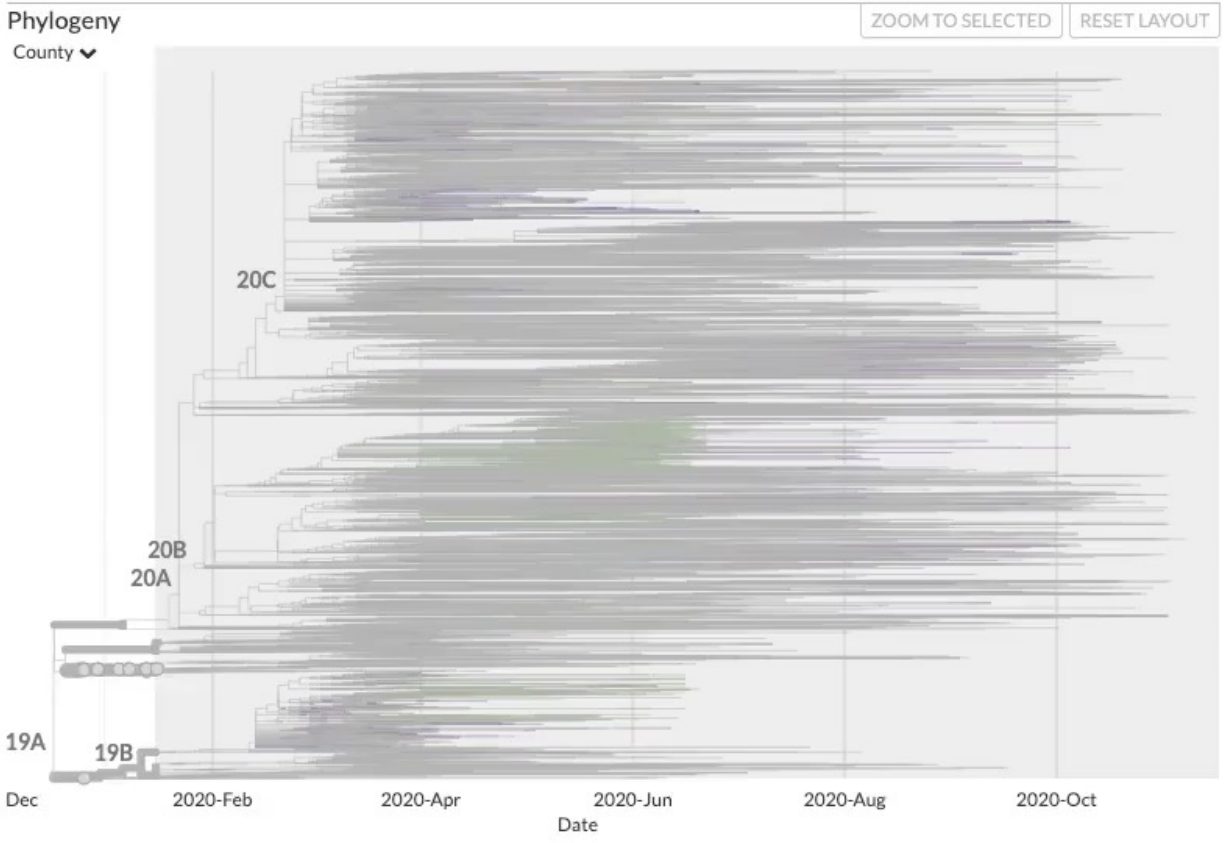


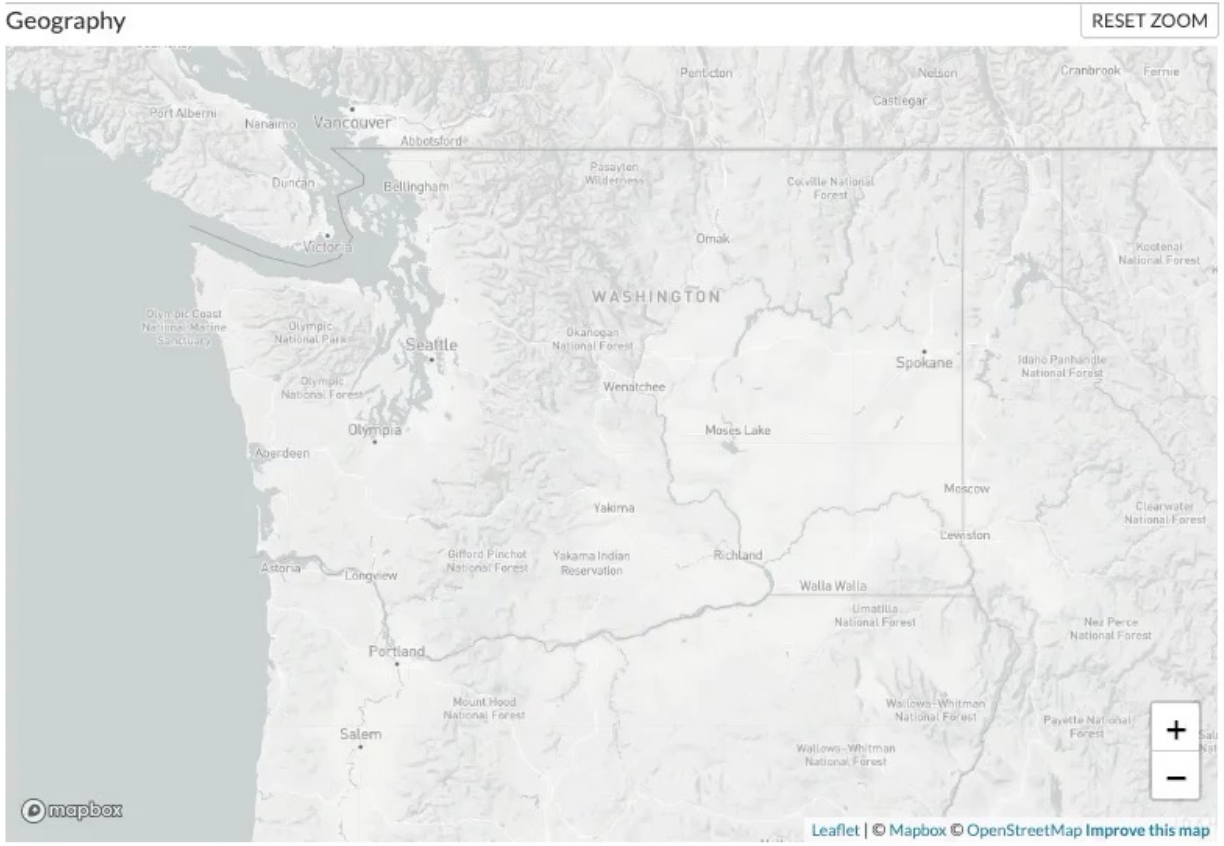
Pečerska *et al.*, 2021, Epidemics

Pros / Cons

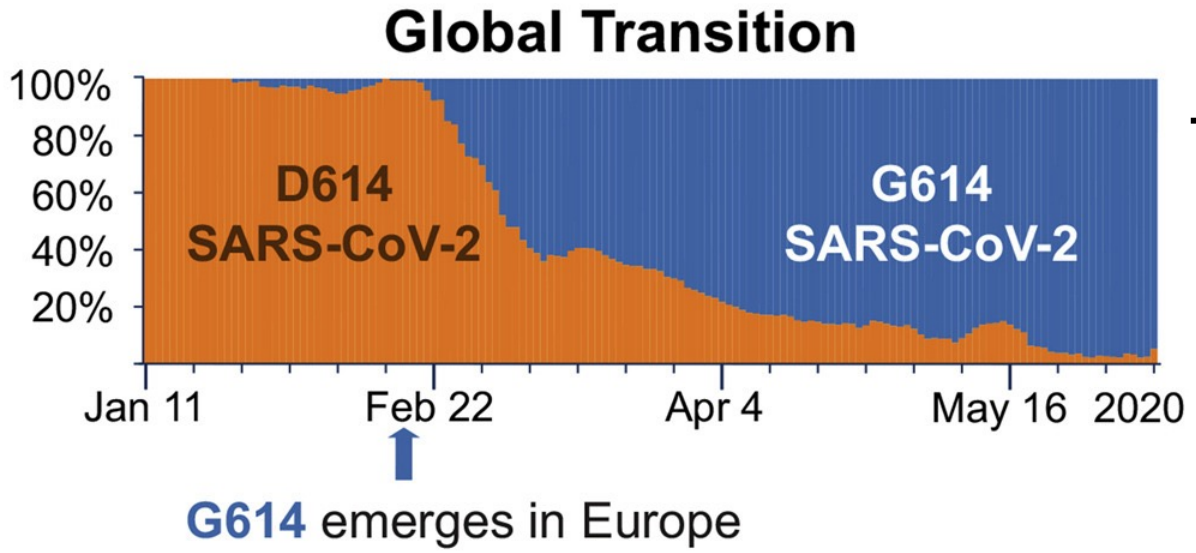
- Pro: parameter estimates map to biologically meaningful quantities (R_0, length of infectious period, etc)
- Con: computational cost (BDMM) or approximations (MSBD)
- Con: sensitive to unmodelled sampling discrepancies

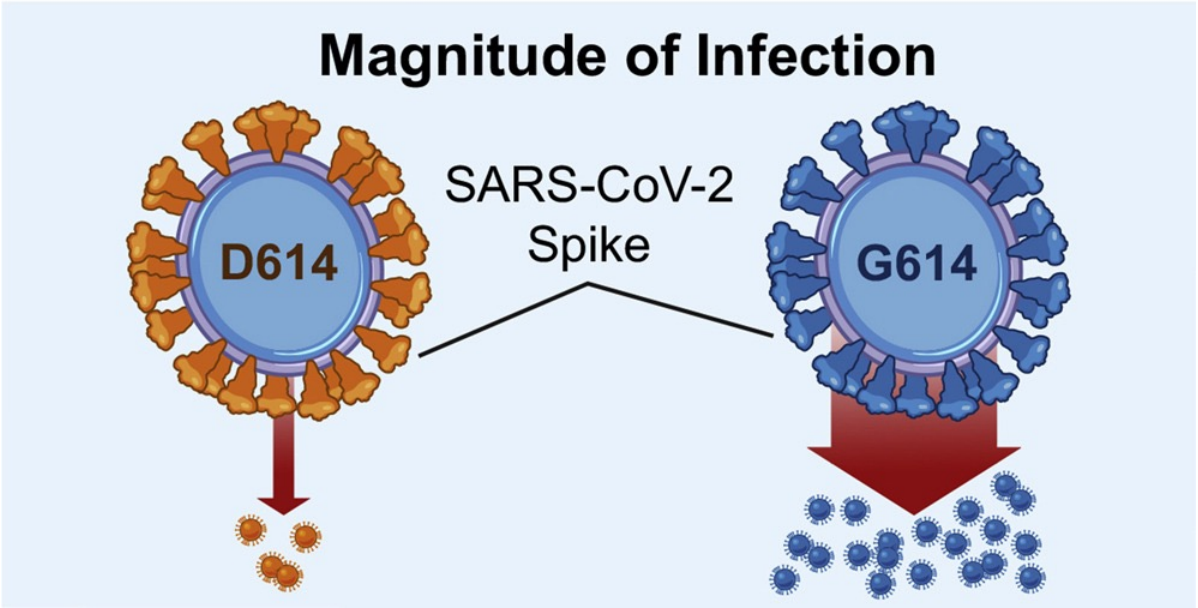
Phylogenetic tree of SARS-CoV-2 sequences isolated in Washington State reveal multiple introductions.





https://nextstrain.org/groups/blab/ncov/wa-phylodynamics?c=county

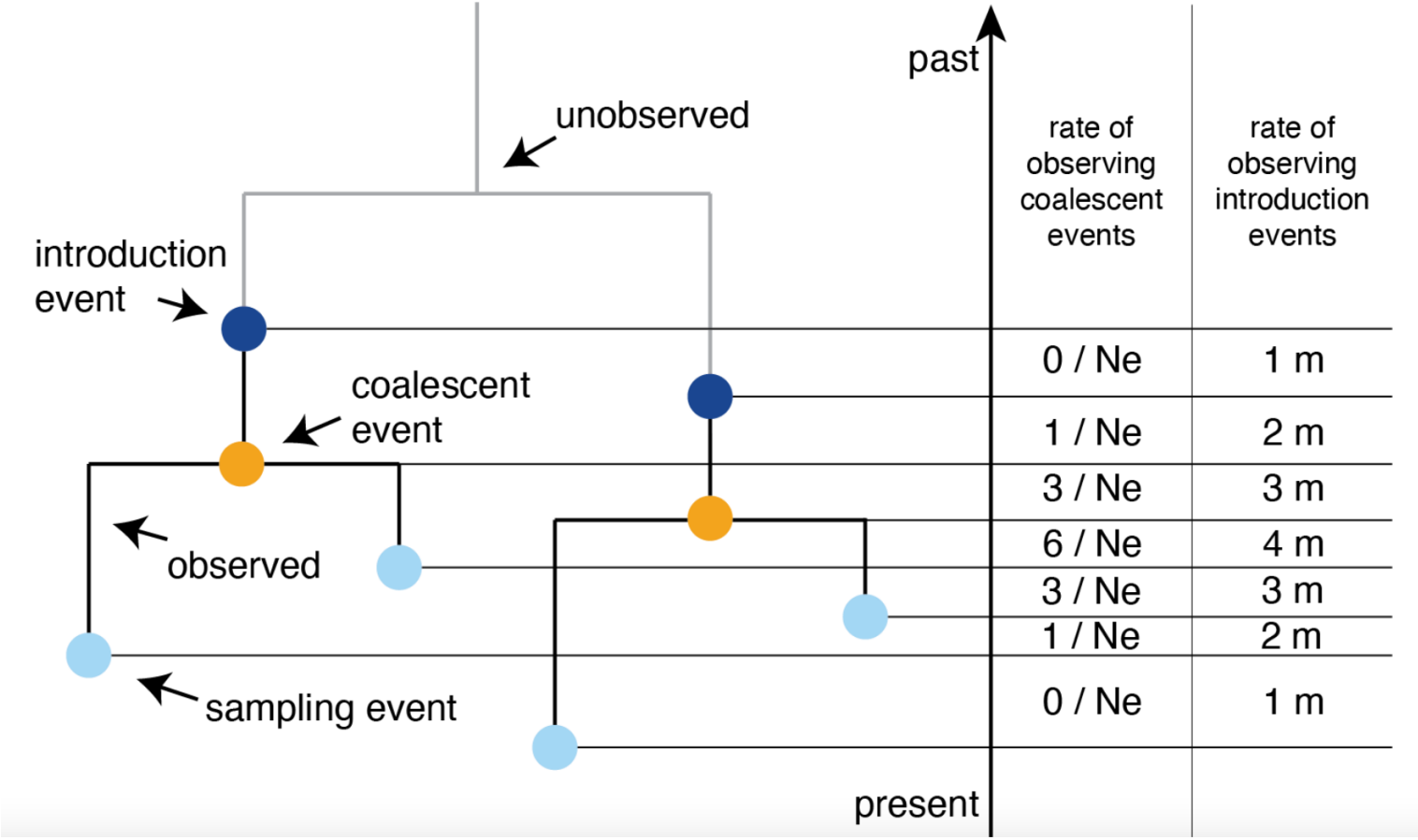




The two spike variants 614D and 614G were shown to differ in viral load and presumably transmission rate.

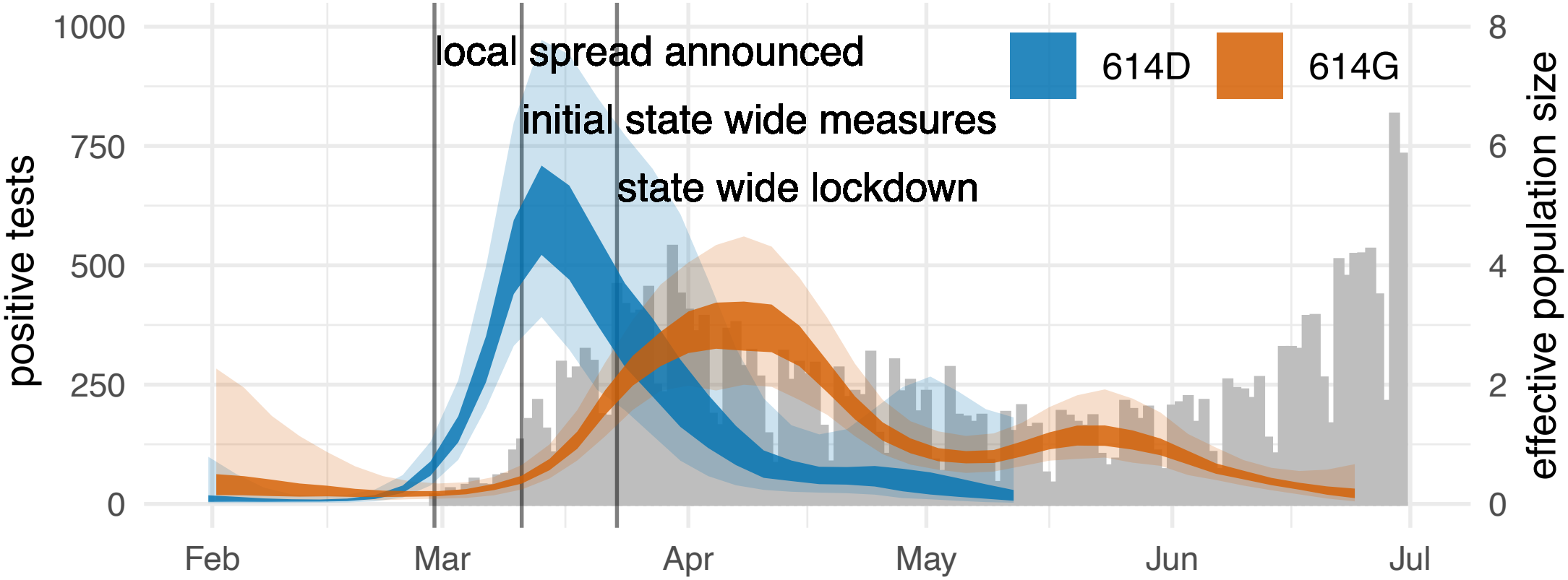
Korber et al. (2020), *Cell*

Migration histories can be conditioned on instead of estimated



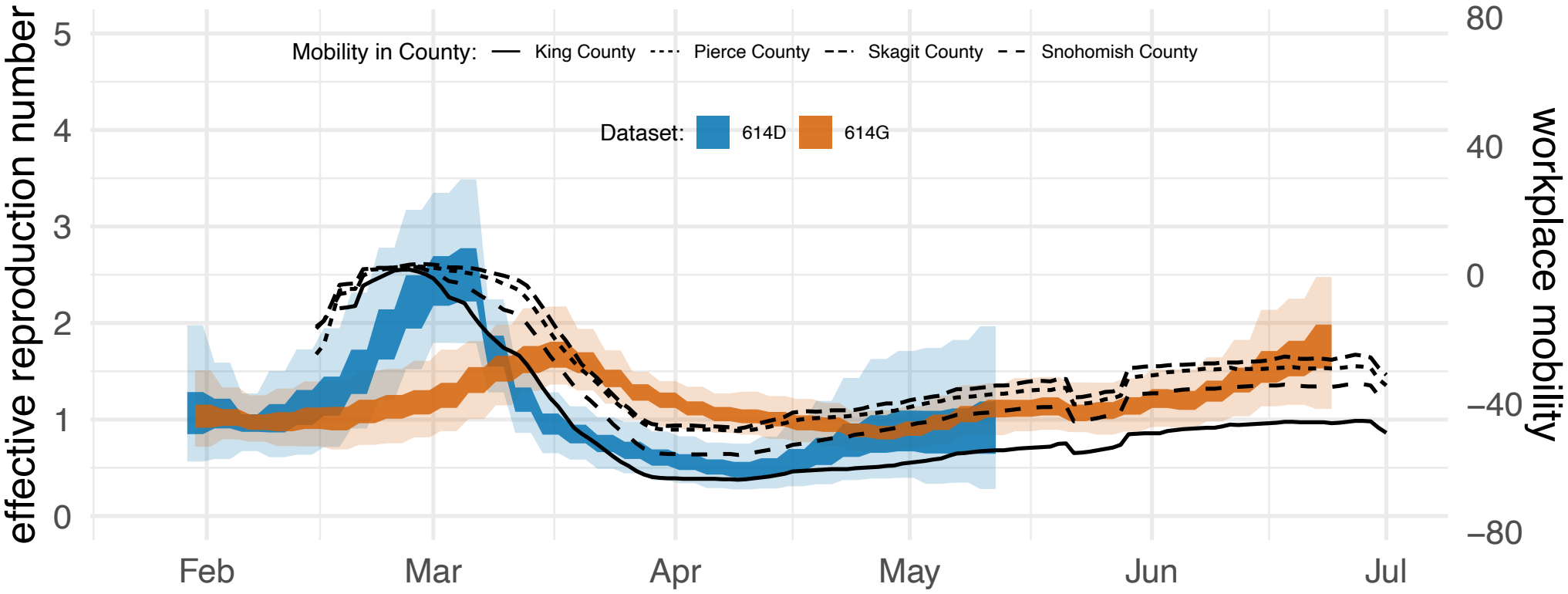
Müller, Wagner, Frazar and Roychoudhury et al. (2021), *BioRxiv*

Reconstructing the effective population size of the spike variants reveals different temporal patterns



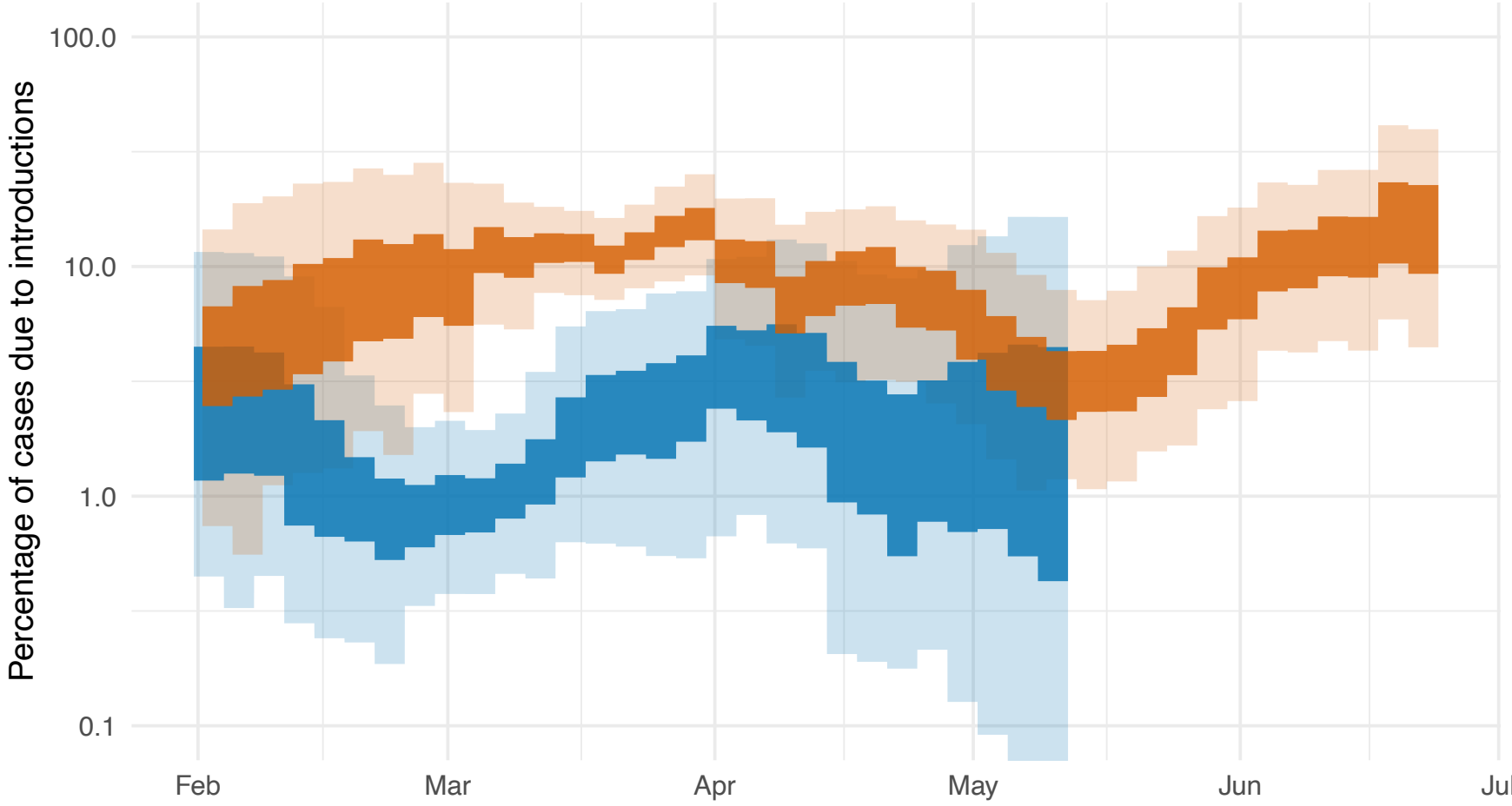
Müller, Wagner, Frazar and Roychoudhury et al. (2021), *Sci. Trans. Med.*

Changes in the reproduction rate of different spike variants reflect different mobility trends in the state



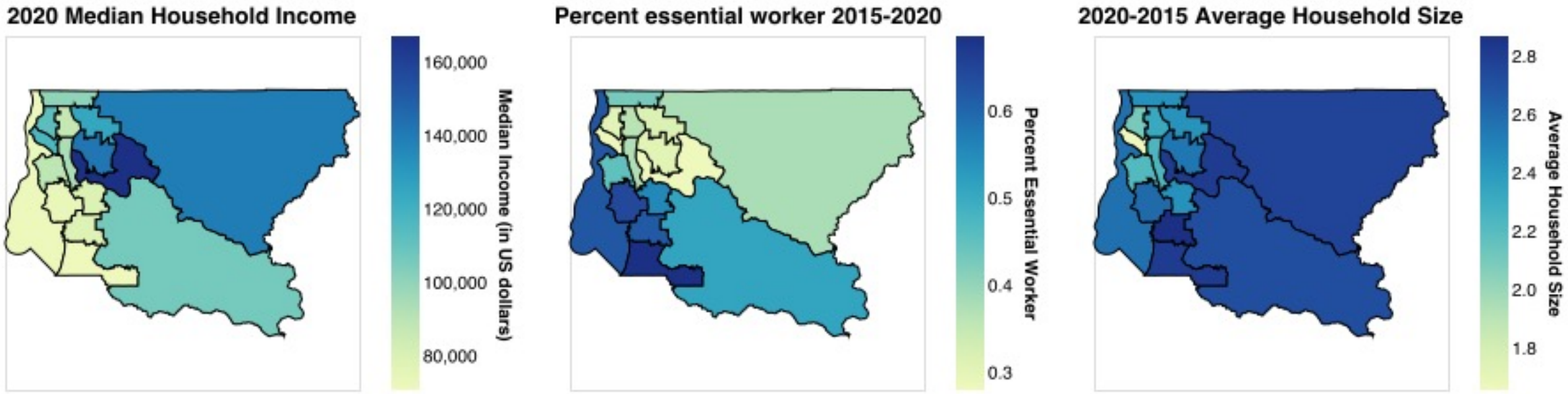
Müller, Wagner, Frazar and Roychoudhury et al. (2021), *Sci. Trans. Med.*

Introductions played different roles in the spread of different SARS-CoV-2 spike variants



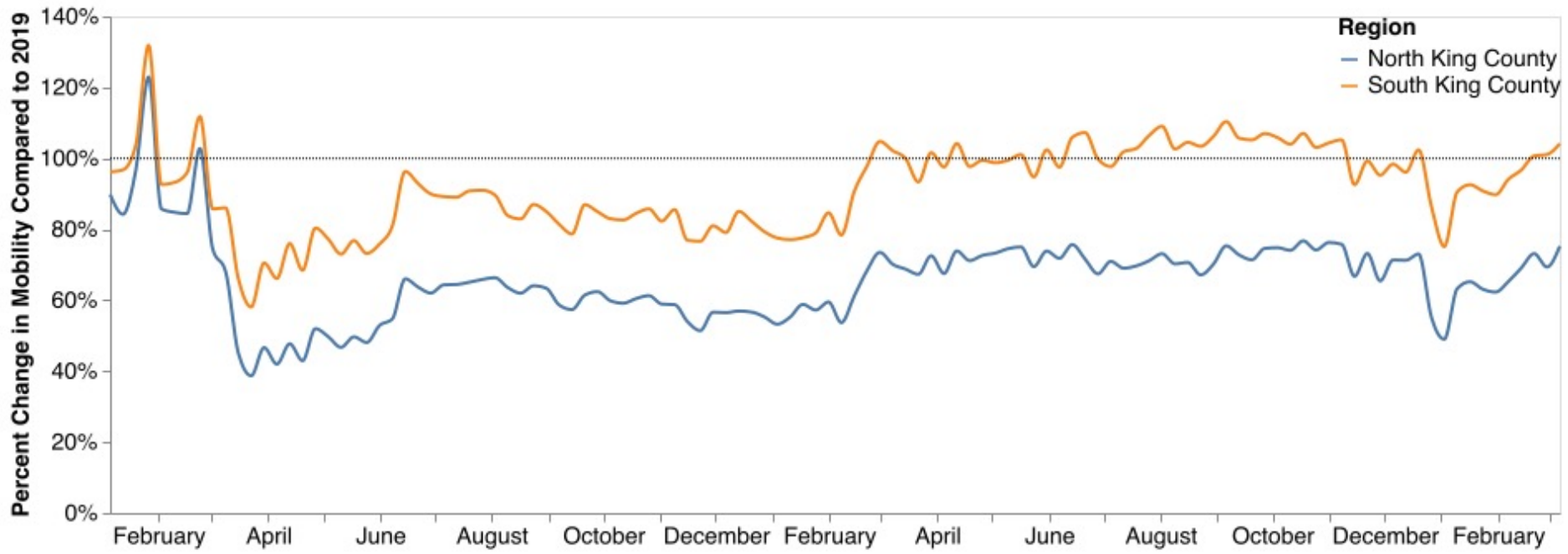
Müller, Wagner, Frazar and Roychoudhury et al. (2021), *Sci. Trans. Med.*

Higher incomes in North King County, fewer essential workers, smaller households (work led by Miguel Paredes)



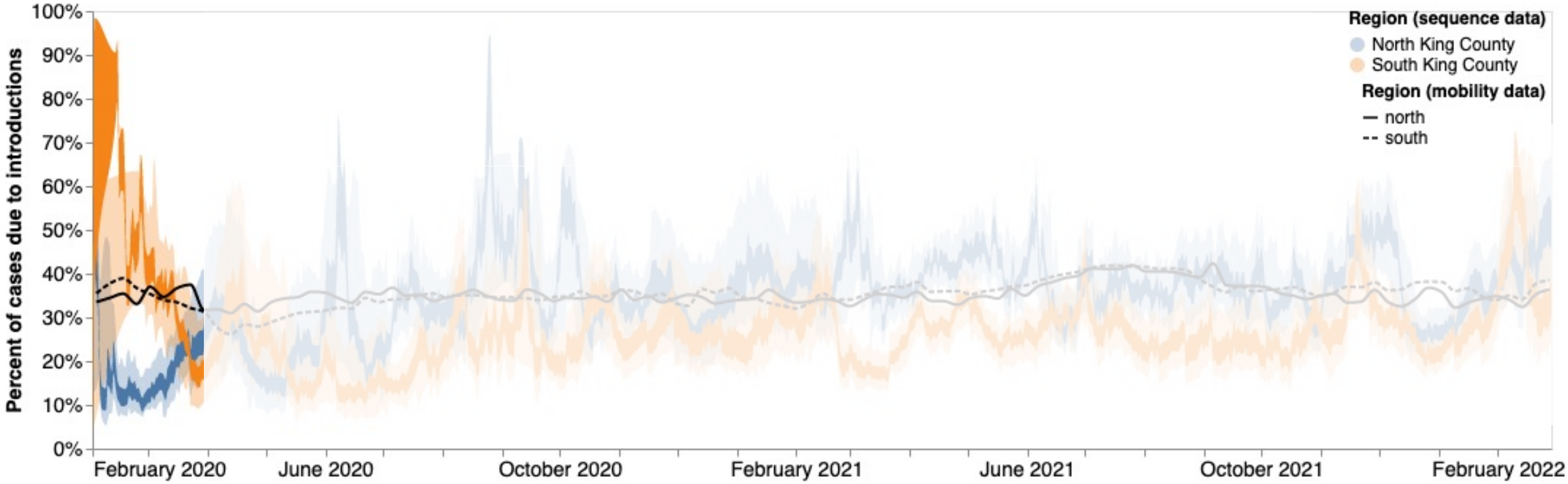
M. I. Paredes, …, **N. F. Müller***, T. Bedford*, "*Local-Scale phylodynamics reveal differential community impact of SARS-CoV-2 in metropolitan US county*" medRxiv. 2022
**joint supervision*

Stronger and more sustained reductions in mobility in North King County



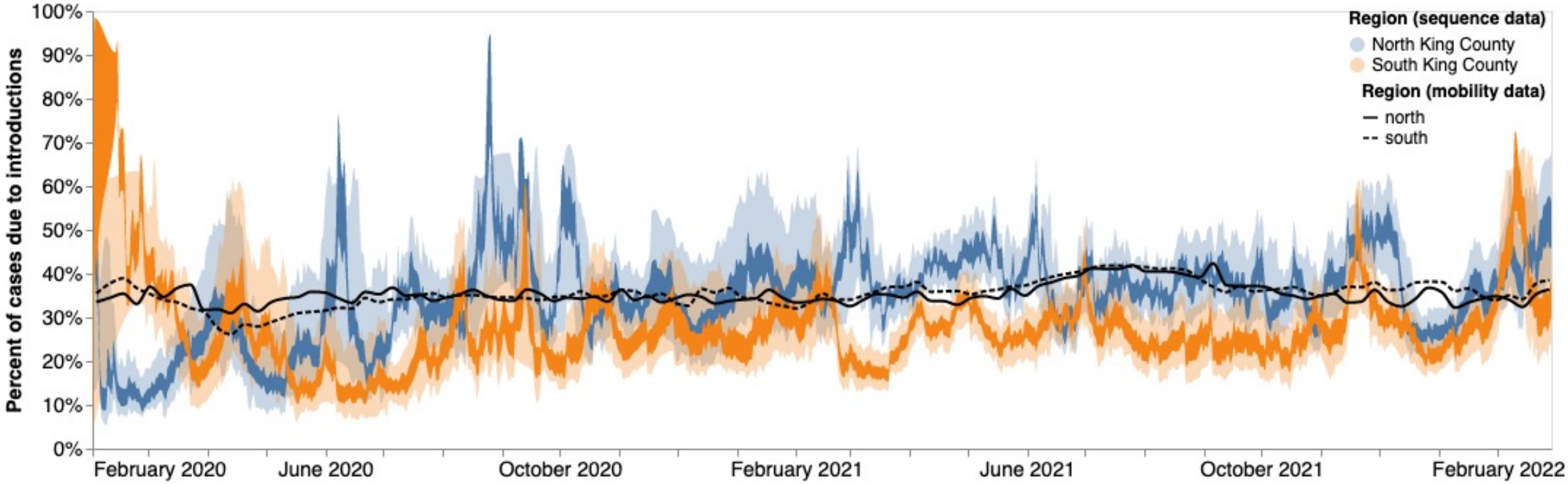
M. I. Paredes, …, **N. F. Müller***, T. Bedford*, "*Local-Scale phylodynamics reveal differential community impact of SARS-CoV-2 in metropolitan US county*" medRxiv. 2022
**joint supervision*

Pre measures, local spread dominated in North, while introductions dominated in South



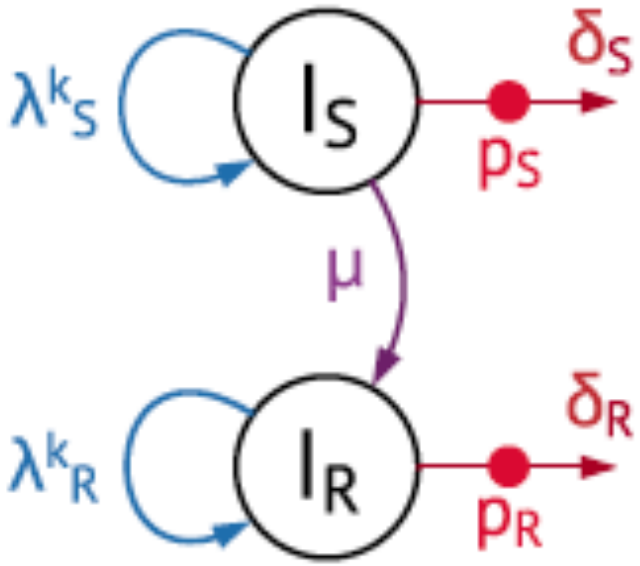
M. I. Paredes, …, **N. F. Müller***, T. Bedford*, "*Local-Scale phylodynamics reveal differential community impact of SARS-CoV-2 in metropolitan US county*" medRxiv. 2022
**joint supervision*

After measures around 30-40% of cases in North were from introductions, while only 20-30% of cases in South were



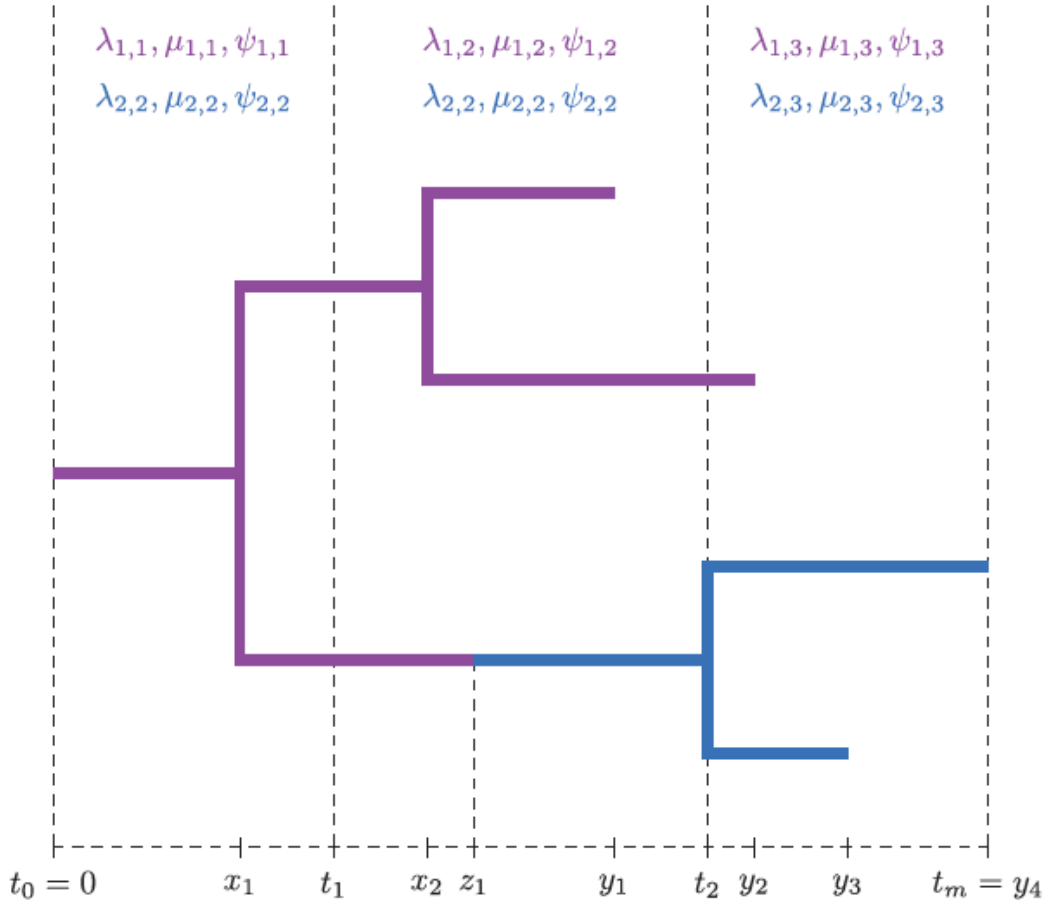
M. I. Paredes, …, **N. F. Müller***, T. Bedford*, "*Local-Scale phylodynamics reveal differential community impact of SARS-CoV-2 in metropolitan US county*" medRxiv. 2022
**joint supervision*

The multi-type birth-death process models how lineages give birth, die, are sampled or jump between states.



Multi-type birth-death process

- N types, i = 1,…,N
- N birth rates λ_i
- N death rates μ_i
- N sampling rates ψ_i
- Migration rates $m_{i,j}$ $\forall$ $i \neq j$



Kühnert *et al.* MBE, 2016

QUESTIONS?

Some reading material

- Paper on DTA: https://doi.org/10.1371/journal.pcbi.1000520
- How do different structured coalescent methods compare: https://academic.oup.com/mbe/article-abstract/34/11/2970/3896419
- Birth-death models with migration: https://doi.org/10.1093/molbev/msw064
- GLM for DTA: https://doi.org/10.1371/journal.ppat.1003932
- GLM for MASCOT: https://doi.org/10.1093/ve/vez030