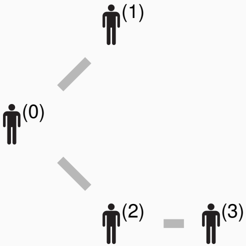
Eight(ish) challenges in (Bayesian) phylodynamics, 11 years later: revisiting Frost *et al.* (2015)

Andrew Magee

April 14, 2026

DDLS Fellow and MIMS Group Leader
Department of Molecular Biology
Umeå University

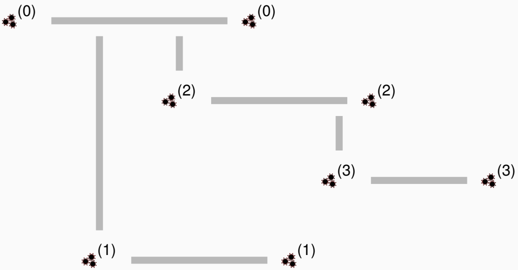
Why phylodynamics works



Nonphylogenetic infection trees:

- Infections are nodes.
- Transmissions are edges.
- Direction of transmission is known.

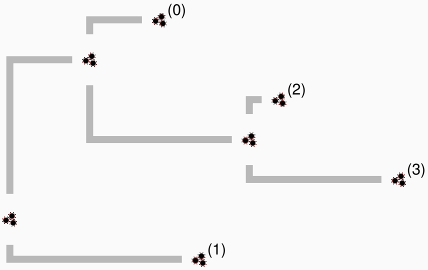
Why phylodynamics works



Phylogenetic transmission trees:

- Infections are edges.
- Transmissions are nodes.
- Direction of transmission is known.

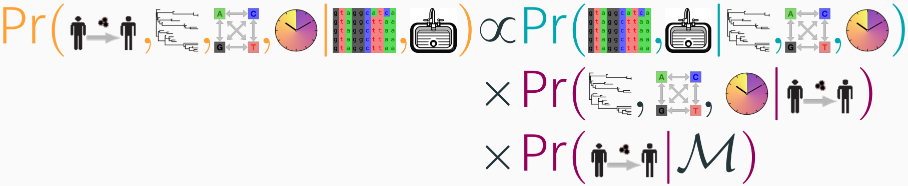
Why phylodynamics works

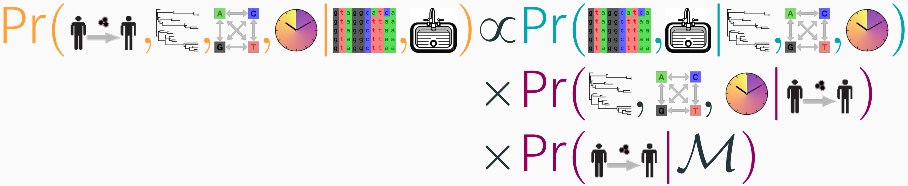


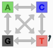
Standard phylogenetic trees:

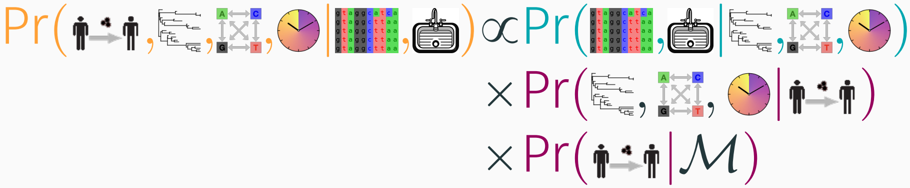
- Infections are edges.
- Transmissions are nodes.
- Direction of transmission is unknown.

BAYESIAN PHYLODYNAMICS

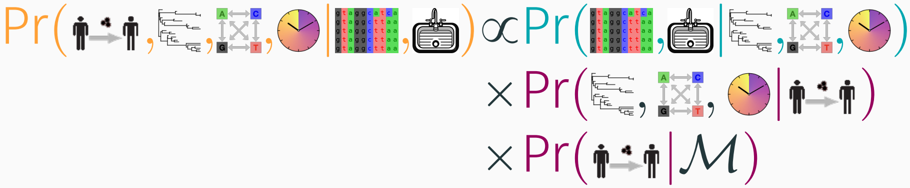




- The posterior distribution contains information about focal and nuisance parameters.
- Focal: infection process, .
- Nuisance: tree, , substitution model, , clock model, .

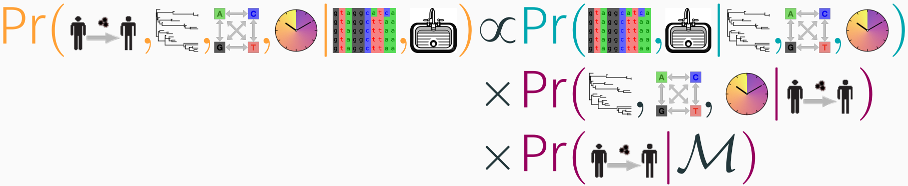


- The prior distribution is where most phylodyamics lives.
- The phylodynamic prior, Pr(|$\mathcal{M}$).
- The phylodynamic likelihood, Pr().



- The prior distribution is where most phylodyamics lives.
- The phylodynamic prior, .
- The phylodynamic likelihood, .

Bayesian phylodynamics



- The likelihood bridges the tree and the genomic data.
- The phylogenetic likelihood, 
- The likelihood of any data, , for the phylodynamic model, such as geographic locations or sampling times.

Eight(ish) challenges

Frost et al. (2015) list eight challenges to the field:

- Sampling patterns.
- Model realism.
- Stochastic effects.
- Population inhomogeneity.
- Recombination and reassortment.
- Phenotypic information.
- Multiscale modeling.
- Scalability and efficiency.

Eight(ish) challenges

Frost et al. (2015) list eight challenges to the field:

- Sampling biases.
- Model adequacy.
- Stochastic effects.
- Population inhomogeneity.
- Recombination and reassortment.
- Data integration.
- Multiscale modeling.
- Scalability and efficiency.
- Realtime inference.

Sampling biases

Frost et al. (2015): temporal imbalance is problematic, especially for birth-death models.

Progress:

- Preferentially-sampled coalescents (Karcher et al. 2016).
- Time-varying birth-death models.
    - Let the sampling rate vary arbitrarily (Gavryushkina et al. 2014).
    - Bayesian regularization of the sampling rate (Magee et al. 2020).

Remaining challenges and open questions:

- Birth-death model nonidentifiability: MacPherson et al. (2022).
- Spatiotemporal variation.

Sampling biases

Frost et al. (2015): geographic imbalance is problematic, especially for phylogeographic models.

Progress:

- Less than one might hope, but see case study.

Remaining challenges and open questions:

- Lack of data availability about sampling intensity.
- What do the models actually assume?

Model adequacy

Frost et al. (2015):

- Our clock models are simplistic.
- Epidemiology and evolution are intertwined in unmodeled ways.
- We lack tractable models of trees under selection.

Progress:

- Flexible clock models (Bletsa et al. 2019; Membrebe et al. 2019).
- "Nonparametric" models for selection (Barido-Sottani et al. 2020; Maliet et al. 2019).

Remaining challenges and open questions:

- Immune-dependent models are hard and require data we lack.
- For most purposes, selective models are lacking.

Model adequacy

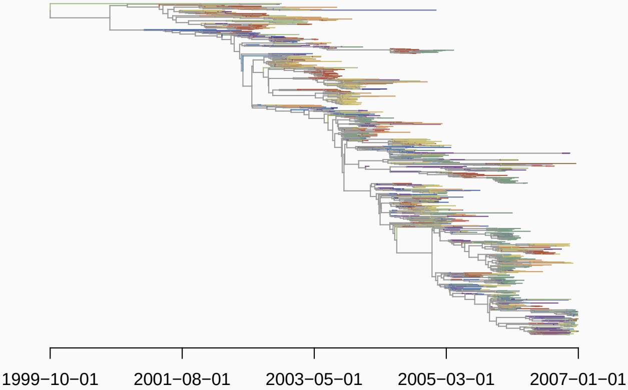
Realism for realism's sake is a trap.

- There are different kinds of wrongness.
  - Delineate focal and nuisance parameters.
  - Embrace (good) approximations (for the analysis regime).
- Don't fear the "non"s.
  - Nonmechanistic.
  - Nonparametric.
- The model just has to be good enough to answer your questions.

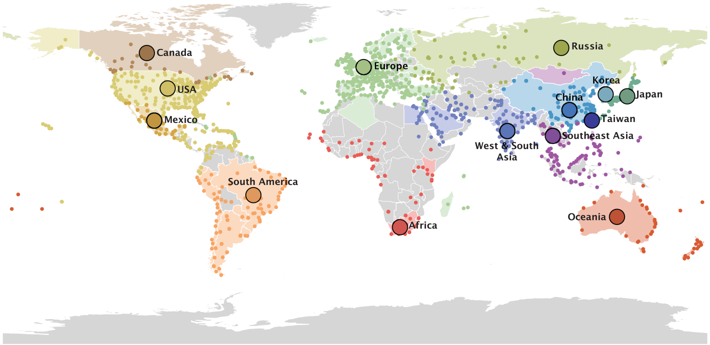
Case study: model adequacy and sampling biases

Understanding the dynamics of influenza A virus dispersal.

Understanding the dynamics of influenza A virus dispersal.



Understanding the dynamics of influenza A virus dispersal.



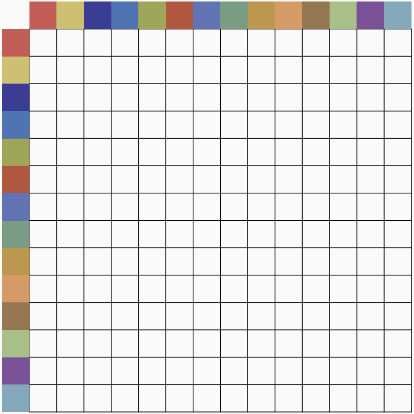
Case study: model adequacy and sampling biases

Understanding the dynamics of influenza A virus dispersal.

- Lemey et al. (2014) considered 22 potential drivers of spread.
- Across all analyses, air travel volume is most consistent predictor.
- Air travel was rarely the only supported predictor.
  - Possibly due to unmodeled sampling biases.

Case study: model adequacy and sampling biases

Understanding the dynamics of influenza A virus dispersal.



Case study: model adequacy and sampling biases

Understanding the dynamics of influenza A virus dispersal.

$$Q$$

GLM substitution models with random-effects.

$$\log(Q_{ij}) = \sum_k X_{ijk}\beta_k + \epsilon_{ij}$$

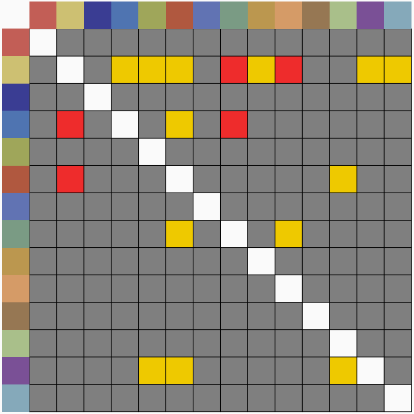
- Q_{ij} describes the instantaneous rate of spread from i to j.
- X_{ij} are observed covariates of spread.
- $\boldsymbol{\beta}$ describe how the covariates affect spread.
- $\boldsymbol{\epsilon}$ add flexibility.

Case study: model adequacy and sampling biases

Understanding the dynamics of influenza A virus dispersal.

- Keep only best-supported covariate in GLM: air travel.
- Use random-effects to ask: what does air travel miss?

Understanding the dynamics of influenza A virus dispersal.



Unimportant, equivocal support, unequivocal support.

Stochastic effects

Frost et al. (2015):

- Most models use deterministic population dynamics.
- Linear birth-death models assume exponential growth.
- Nonlinear and mechanistic models are lacking.

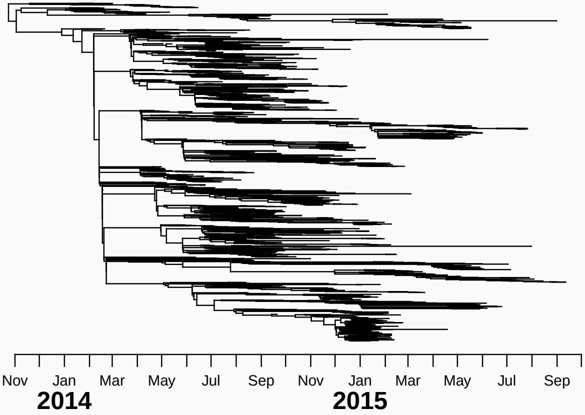
Progress:

- Locally linear birth-death models (Shao et al. 2025; Magee et al. 2020).
- Adding mechanistic components to existing models (Tang et al. 2023).
- Particle filtering for fitting a wide range of models (King et al. 2025).

Remaining challenges and open questions:

- When can we ignore stochastic effects?
- Is there a happy medium?

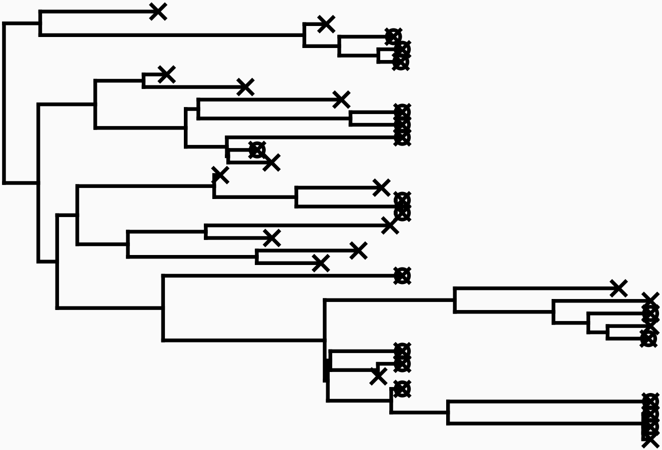
Border closures and transmission, 2014 West African Ebola epidemic.



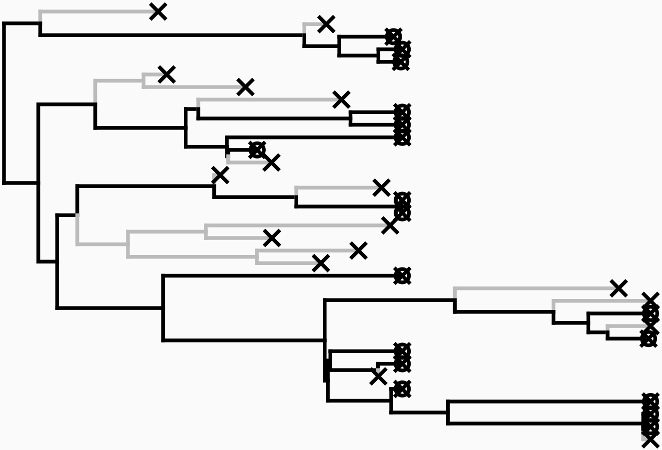
Locally linear phylogenetic birth-death models, (Shao et al. 2024; Magee et al. 2020; Gavryushkina et al. 2014).

- New infections arise at time-varying birth rate ⊏(t).
- Individuals become noninfectious at time-varying death rate →×(t).
- Samples are taken at time-varying rate -o-(t).
  - Some proportion of these become noninfectious, →⊠(t).
- Effective reproduction number: $R_e(t) =$ ⊏$(t)/($→×$(t) +$ →⊠$(t))$

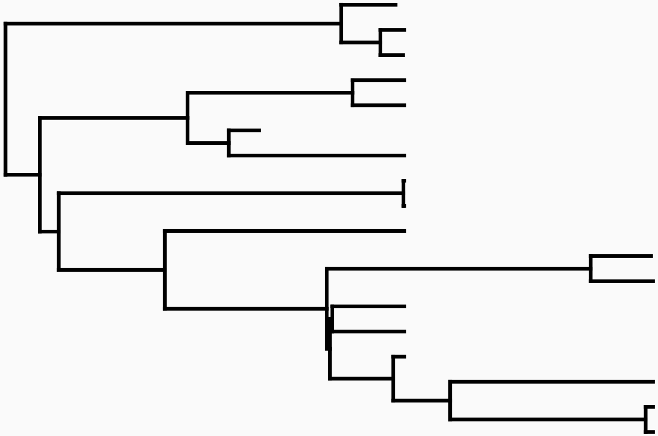
Locally linear phylogenetic birth-death models, (Shao et al. 2024; Magee et al. 2020; Gavryushkina et al. 2014).



Locally linear phylogenetic birth-death models, (Shao et al. 2024; Magee et al. 2020; Gavryushkina et al. 2014).



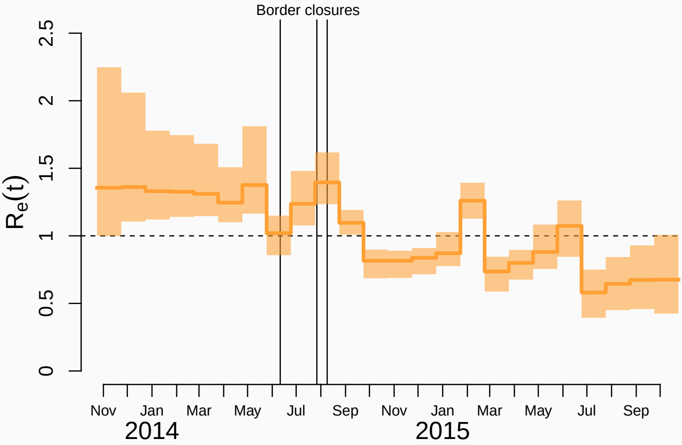
Locally linear phylogenetic birth-death models, (Shao et al. 2024; Magee et al. 2020; Gavryushkina et al. 2014).



Locally linear phylogenetic birth-death models, (Shao et al. 2024; Magee et al. 2020; Gavryushkina et al. 2014).

- Simplifying assumption: rates are piecewise constant.
    - Monthly scale.
    - Duration of infection roughly 15-17 days (Van Kerkhove et al. 2015).

Border closures and transmission, 2014 West African Ebola epidemic.



Population inhomogeneity

Frost et al. (2015): populations are not panmictic.

Progress: the structured coalescent

- Tractable approximations (De Maio et al. 2015; Müller et al. 2017).
- Efficient inference (Shao et al. 2025; Vaughan et al. 2014).

Remaining challenges and open questions:

- These models are intensely demanding.
- Multi-type phylogenetic birth-death models exist, but are unused.

Population inhomogeneity

Frost et al. (2015): high variance in contacts, *i.e.* superspreading, is challenging.

Progress:

- Λ- and Beta-coalescents, (Hoscheit et al. 2019; Zhang et al. 2025).
- Lineage-dependent birth-death models, (Barido-Sottani et al. 2020; Maliet et al. 2019).

Remaining challenges and open questions:

- Computational intensity of lineage-dependent birth-death models.
- What can we actually infer with these models?
- Does sampling intensity affect their necessity?

Recombination and reassortment

Frost et al. (2015): recombination is hard and oft ignored.

Progress:

- Inference of ancestral recombination graphs (ARGs) (Müller et al. 2022).
- Efficient data structures (Ralph et al. 2020).

Remaining challenges and open questions:

- Treespace is massive and gnarly. ARGspace is bigger and gnarlier.

Data integration

Frost et al. (2015): phenotype can be informative about transmission and ought to be modeled.

Progress:

- Linking lab measurements to phylogenies via experimental codon models (Hilton et al. 2018; Bloom 2014).
- Linking case data and phylogenies (Gupta et al. 2020).
- Much more, see Hassler et al. (2023) for a review.

Remaining challenges and open questions:

- Modular, comprehensive, user-friendly options.

Multiscale modeling

Frost et al. (2015): branching events in a population-level phylogeny are not actually transmission events.

Progress:

- Joint inference of phylogenetic and transmission tree (Klinkenberg et al. 2017).
- Bottleneck models (Hall et al. 2019).
- Quantifying different between and within host dynamics, (Vrancken et al. 2017).

Remaining challenges and open questions:

- Tree within a tree models are demanding.
- How bad of an approximation is it to ignore that?

Scalability and efficiency

Frost et al. (2015): big phylogenies pose big problems.

Progress:

- Parsimony(-based approximations) (Turakhia et al. 2021; De Maio et al. 2023).
- Gradient-based inference (Ji et al. 2020; Magee et al. 2024).
- Non-traditional tree inference (Zhang et al. 2024)

Remaining challenges and open questions:

- Parsimony likely only works well in the near-perfect regime (Wertheim et al. 2022).
- Trees are still the bottleneck and stubbornly resistant to alternative inference techniques.

Case study: scalability and efficiency

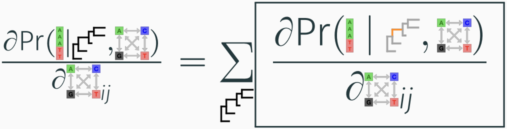
A tale of two substitution models (Magee et al. 2024).

Case study: scalability and efficiency

A tale of two substitution models (Magee et al. 2024).

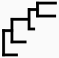
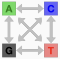
- Inferring substitution dynamics and a 583-sequence tree of SARS-CoV-2 (Pekar et al. 2022).
- Inferring phylogeographic dispersal among 1441 sequences of influenza A virus (H3N2) (Lemey et al. 2014).

Approximate substitution model gradients



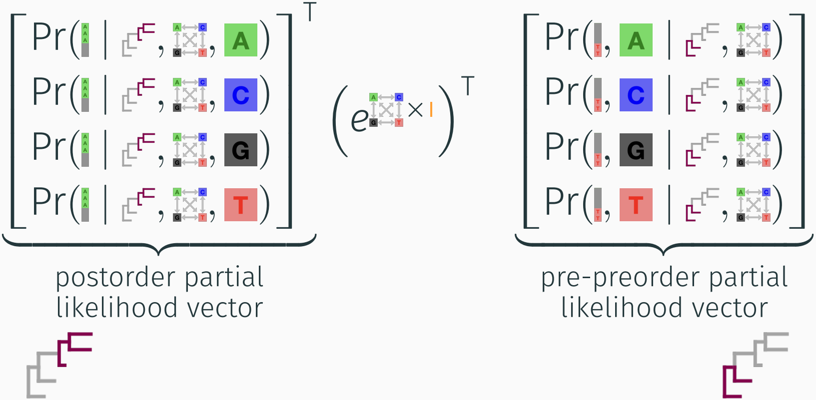
An aside on partial likelihoods.

Pr(

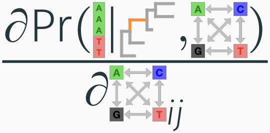
| , 

)

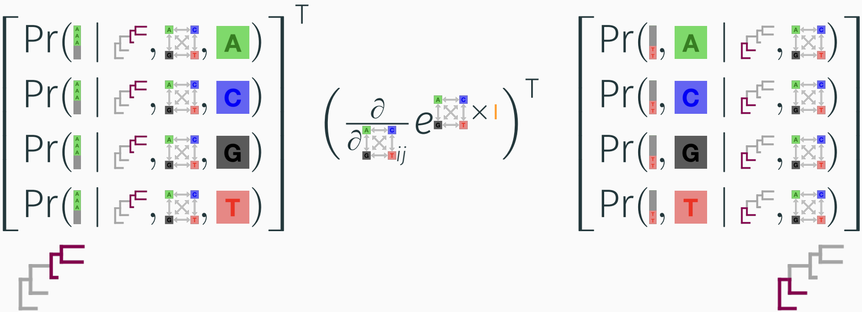
An aside on partial likelihoods.



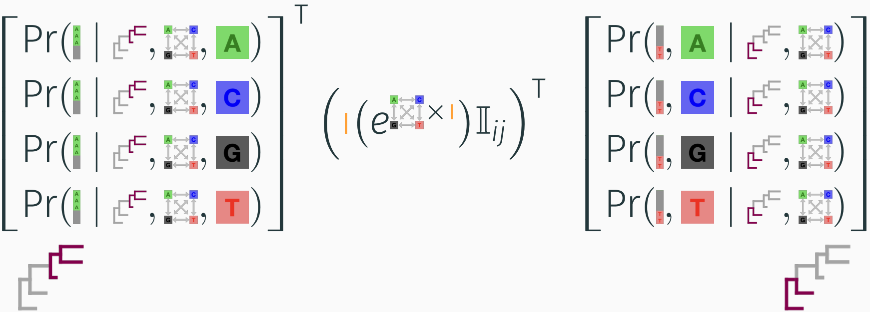
The gradient on a single branch.



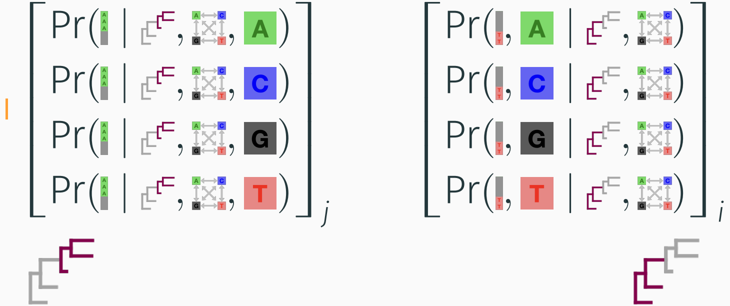
The gradient on a single branch.



The (approximate) gradient on a single branch.



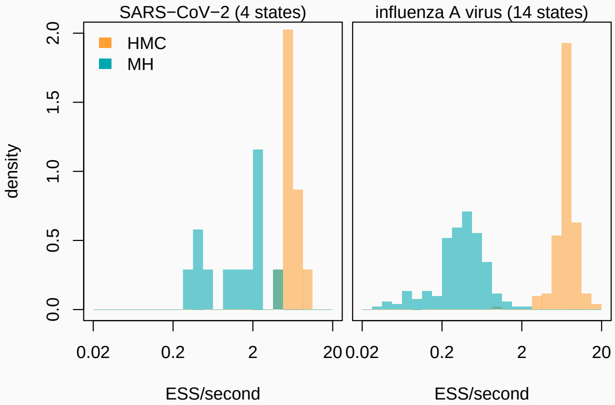
The (approximate) gradient on a single branch.



Inference using our approximate substitution gradient is highly efficient.

- Our approximation ($\mathcal{O}(ns^2)$) enables 1M MCMC samples in 1 hour for both analyses.
- With naïve gradients ($\mathcal{O}(ns^5)$), equivalent analyses take:
    - Influenza A virus phylogeography: 3 months.
    - SARS-CoV-2 substitution dynamics: 3 days.
- Without gradients, equivalent analyses take:
    - Influenza A virus phylogeography: 1.4 days.
    - SARS-CoV-2 substitution dynamics: 15 hours.

Inference using our approximate substitution gradient is highly efficient.



What does this approximate gradient actually get us?

- For high dimensional models, maybe a lot.
  - Days to hours can the difference between infeasible and feasible.
- For low-dimensional models, not so much.
  - DNA substitution models aren't really a bottleneck.
  - Original Lemey et al. (2014) used simpler models which converged faster.
- Tree inference is costly.
  - SARS-CoV-2 joint analysis takes about a week.

Realtime inference

Key challenges

- Necessary infrastructure is extensive.
- Data incompleteness is time-varying.
- Models need to be sufficiently predictive.
- Tree inference is still slow in the general case.

Some successes

- The UShER SARS-CoV-2 tree.
- NextStrain.

Open questions:

- Can we do realtime Bayesian phylogenetic inference?
- How can we reward key providers of bioinformatics infrastructure?

Conclusions

Lots of progress since 2015.

- More and better models.
- Greatly expanded data integration.
- Major improvements in scalability and efficiency.

Many important models fall between

- Computationally burdensome.
- Essentially intractable.

Fast Bayesian inference of phylogenies remains a holy grail.

Where do we go from here?

Tractable general solutions may be more robust than realistic.

- Non- or semi-mechanistic.
- Non- or semi-parametric.

Tractable specific solutions may

- Leverage clever mathematical tricks (Bryant et al. 2026).
- Likely rely on clever (or even dumb) approximations.

Some problems may need to be solved at other levels.

- Better surveillance systems beat better sampling models.

References

- Barido-Sottani, Joëlle, Timothy G Vaughan, and Tanja Stadler (2020). "A multitype birth–death model for Bayesian inference of lineage-specific birth and death rates." In: Systematic Biology 69.5, pp. 973–986.
- Bletsa, Magda et al. (2019). "Divergence dating using mixed effects clock modelling: An application to HIV-1." In: Virus Evolution 5.2, vez036.
- Bloom, Jesse D (2014). "An experimentally determined evolutionary model dramatically improves phylogenetic fit." In: Molecular Biology and Evolution 31.8, pp. 1956–1978.

Bryant, David, Celine Scornavacca, and David Swofford (2026). “LvD: A New Algorithm for Computing the Likelihood of a Phylogeny.” In: arXiv preprint arXiv:2601.19064.

De Maio, Nicola et al. (2015). “New routes to phylogeography: a Bayesian structured coalescent approximation.” In: PLoS Genetics 11.8, e1005421.

De Maio, Nicola et al. (2023). “Maximum likelihood pandemic-scale phylogenetics.” In: Nature Genetics 55.5, pp. 746–752.

Frost, Simon DW et al. (2015). “Eight challenges in phylodynamic inference.” In: Epidemics 10, pp. 88–92.

Gavryushkina, Alexandra et al. (2014). “Bayesian inference of sampled ancestor trees for epidemiology and fossil calibration.” In: PLoS Computational Biology 10.12, e1003919.

Gupta, Ankit et al. (2020). “The probability distribution of the reconstructed phylogenetic tree with occurrence data.” In: Journal of Theoretical Biology 488, p. 110115.

- Hall, Matthew D and Caroline Colijn (2019). “Transmission trees on a known pathogen phylogeny: Enumeration and sampling.” In: Molecular Biology and Evolution 36.6, pp. 1333–1343.
- Hassler, Gabriel W et al. (2023). “Data integration in Bayesian phylogenetics.” In: Annual review of statistics and its application 10.1, pp. 353–377.
- Hilton, Sarah K and Jesse D Bloom (2018). “Modeling site-specific amino-acid preferences deepens phylogenetic estimates of viral sequence divergence.” In: Virus Evolution 4.2, vey033.
- Hoscheit, Patrick and Oliver G Pybus (2019). “The multifurcating skyline plot.” In: Virus Evolution 5.2, vez031.
- Ji, Xiang et al. (2020). “Gradients do grow on trees: a linear-time O(N)-dimensional gradient for statistical phylogenetics.” In: Molecular biology and evolution 37.10, pp. 3047–3060.

Karcher, Michael D et al. (2016). "Quantifying and mitigating the effect of preferential sampling on phylodynamic inference." In: PLoS Computational Biology 12.3, e1004789.

King, Aaron A, Qianying Lin, and Edward L Ionides (2025). "Exact phylodynamic likelihood via structured Markov genealogy processes." In: ArXiv, arXiv–2405.

Klinkenberg, Don et al. (2017). "Simultaneous inference of phylogenetic and transmission trees in infectious disease outbreaks." In: PLoS Computational Biology 13.5, e1005495.

Lemey, Philippe et al. (2014). "Unifying viral genetics and human transportation data to predict the global transmission dynamics of human influenza H3N2." In: PLoS Pathogens 10.2, e1003932.

MacPherson, Ailene et al. (2022). "Unifying phylogenetic birth–death models in epidemiology and macroevolution." In: Systematic Biology 71, pp. 172–189.

- Magee, Andrew F et al. (2020). "Locally adaptive Bayesian birth-death model successfully detects slow and rapid rate shifts." In: PLoS computational biology 16.10, e1007999.
- Magee, Andrew F et al. (2024). "Random-effects substitution models for phylogenetics via scalable gradient approximations." In: Systematic Biology 73.3, pp. 562–578.
- Maliet, Odile, Florian Hartig, and Hélène Morlon (2019). "A model with many small shifts for estimating species-specific diversification rates." In: Nature Ecology & Evolution 3.7, pp. 1086–1092.
- Membrebe, Jade Vincent et al. (2019). "Bayesian inference of evolutionary histories under time-dependent substitution rates." In: Molecular Biology and Evolution 36.8, pp. 1793–1803.
- Müller, Nicola F, Kathryn E Kistler, and Trevor Bedford (2022). "A Bayesian approach to infer recombination patterns in coronaviruses." In: Nature communications 13.1, p. 4186.

Müller, Nicola F, David A Rasmussen, and Tanja Stadler (2017). "The structured coalescent and its approximations." In: Molecular Biology and Evolution 34.11, pp. 2970–2981.

Pekar, Jonathan E et al. (2022). "The molecular epidemiology of multiple zoonotic origins of SARS-CoV-2." In: Science 377.6609, pp. 960–966.

Ralph, Peter, Kevin Thornton, and Jerome Kelleher (2020). "Efficiently summarizing relationships in large samples: a general duality between statistics of genealogies and genomes." In: Genetics 215.3, pp. 779–797.

Shao, Yucai et al. (2024). "Scalable gradients enable Hamiltonian Monte Carlo sampling for phylodynamic inference under episodic birth-death-sampling models." In: PLOS Computational Biology 20.3, e1011640.

Shao, Yucai et al. (2025). "Parallel algorithms for phylogenetic inference under a structured coalescent approximation." In: bioRxiv.

- Tang, Mingwei et al. (2023). “Fitting stochastic epidemic models to gene genealogies using linear noise approximation.” In: The annals of applied statistics 17.1, p. 1.
- Turakhia, Yatish et al. (2021). “Ultrafast Sample placement on Existing tRees (UShER) enables real-time phylogenetics for the SARS-CoV-2 pandemic.” In: Nature Genetics 53.6, pp. 809–816.
- Van Kerkhove, Maria D et al. (2015). “A review of epidemiological parameters from Ebola outbreaks to inform early public health decision-making.” In: Scientific data 2.1, p. 150019.
- Vaughan, Timothy G et al. (2014). “Efficient Bayesian inference under the structured coalescent.” In: Bioinformatics 30.16, pp. 2272–2279.
- Vrancken, Bram, Marc A Suchard, and Philippe Lemey (2017). “Accurate quantification of within-and between-host HBV evolutionary rates requires explicit transmission chain modelling.” In: Virus Evolution 3.2, vex028.

- Wertheim, Joel O, Mike Steel, and Michael J Sanderson (2022). "Accuracy in near-perfect virus phylogenies." In: Systematic Biology 71.2, pp. 426–438.
- Zhang, Cheng and Frederick A Matsen IV (2024). "A variational approach to Bayesian phylogenetic inference." In: Journal of Machine Learning Research 25.145, pp. 1–56.
- Zhang, Julie and Julia A Palacios (2025). "Multiple merger coalescent inference of effective population size." In: Philosophical Transactions of the Royal Society B: Biological Sciences 380.1919.