# Introduction to Markov genealogy processes

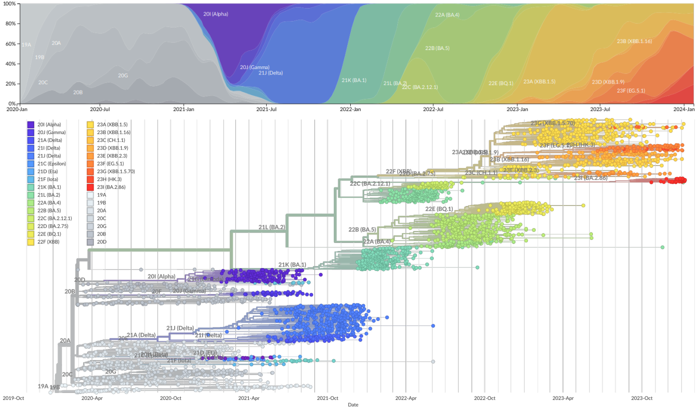
Aaron A. King

March 4, 2024

# Outline

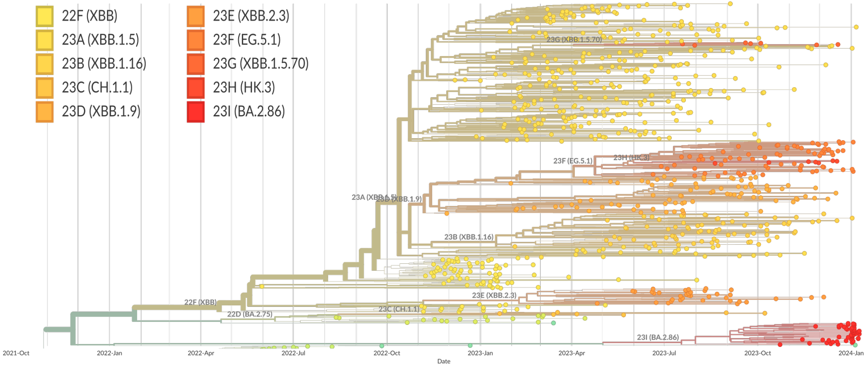
1. Context
   - Example: emerging variants
   - Phylodynamics
   - Problems of phylodynamics
2. Population process
   - Examples
   - Formalization
3. Genealogy process
   - Genealogies
   - Induced genealogy process
4. Pruned and obscured genealogies

# Example: surveillance for emerging SARS-CoV-2 variants



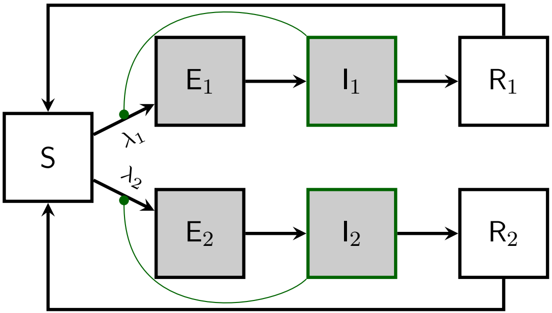
nextstrain.org (Hadfield *et al.*, 2018)

# Example: surveillance for emerging SARS-CoV-2 variants



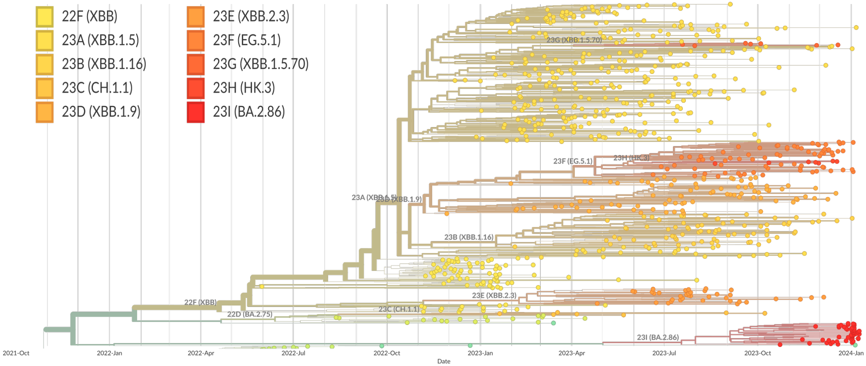
`nextstrain.org` (Hadfield *et al.*, 2018)

# Example: surveillance for emerging SARS-CoV-2 variants



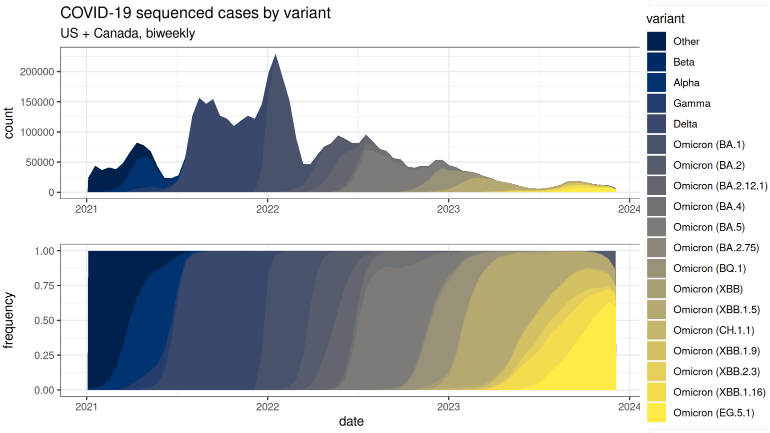
$$\lambda_1 = \beta_1 \frac{I_1}{N} \qquad \lambda_2 = \beta_2 \frac{I_2}{N}$$

# Example: surveillance for emerging SARS-CoV-2 variants



`nextstrain.org` (Hadfield *et al.*, 2018)

# Example: surveillance for emerging SARS-CoV-2 variants



(Mathieu *et al.*, 2020)

# What is phylodynamics?

Broadly:

Phylodynamics is the project of inferring

*determinants of epidemic spread*

using

*genomic data from pathogen samples*.

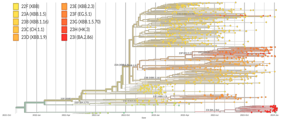
In this talk:

Phylodynamics means using

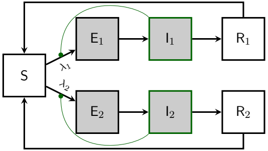
*genomic data*

to infer

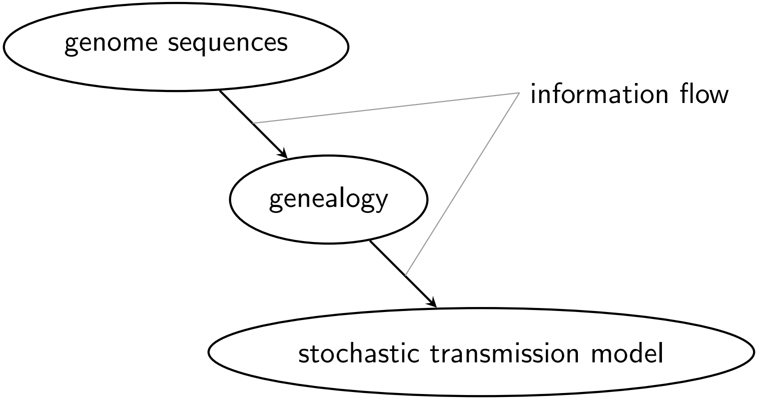
*stochastic dynamic transmission models*.



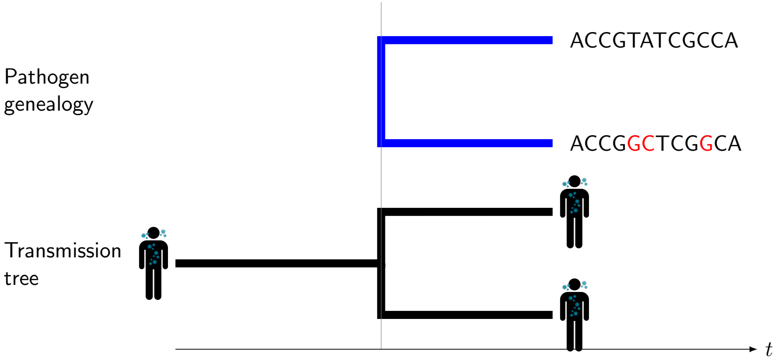
⇓



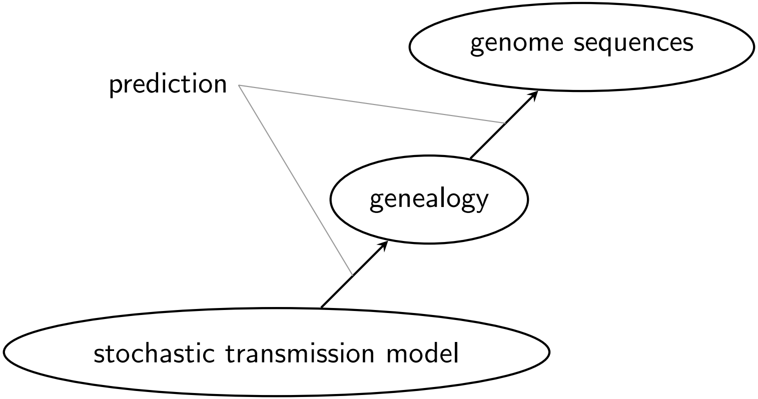
# Core problems of phylodynamics



# Core problems of phylodynamics



# Core problems of phylodynamics



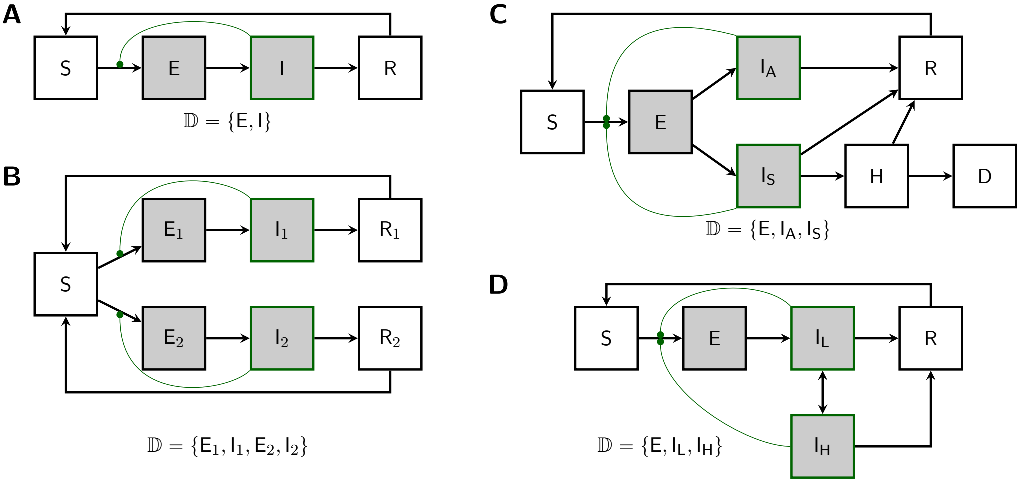
# Overview

- We show how a given population process induces a unique genealogy process.
- *Pruning* and *obscuration* project a genealogy onto observable data.

# Outline

1. Context
   - Example: emerging variants
   - Phylodynamics
   - Problems of phylodynamics
2. Population process
   - Examples
   - Formalization
3. Genealogy process
   - Genealogies
   - Induced genealogy process
4. Pruned and obscured genealogies

# Population process



A: $\mathbb{D} = \{\mathrm{E}, \mathrm{I}\}$

B: $\mathbb{D} = \{\mathrm{E}_1, \mathrm{I}_1, \mathrm{E}_2, \mathrm{I}_2\}$

C: $\mathbb{D} = \{\mathrm{E}, \mathrm{I}_\mathrm{A}, \mathrm{I}_\mathrm{S}\}$

D: $\mathbb{D} = \{\mathrm{E}, \mathrm{I}_\mathrm{L}, \mathrm{I}_\mathrm{H}\}$

# Population process

- Non-explosive Markov jump process, $\mathbf{X}_t \in \mathbb{X}$, $t \in \mathbb{R}_+$: the *population process*.
- Initial-state distribution, $p_0$:

$$\text{Prob}\left[\mathbf{X}_0 \in \mathcal{E}\right] = \int_{\mathcal{E}} p_0(x)\,\mathrm{d}x$$

- Jump rates: $\alpha(t, x, x') =$ rate of jump $x \to x'$

$$\alpha(t, x, x') \geqslant 0, \qquad \int_{\mathbb{X}} \alpha(t, x, x')\,\mathrm{d}x' < \infty$$

- Multiple events at each jump are allowed.

# Population process

Kolmogorov forward equation (KFE):

If

$$\frac{\partial w}{\partial t}(t,x) = \int w(t,x')\,\alpha(t,x',x)\,\mathrm{d}x' - \int w(t,x)\,\alpha(t,x,x')\,\mathrm{d}x'$$

and

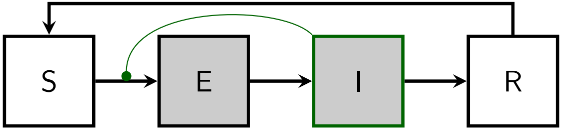
$$w(0,x) = p_0(x)$$

then

$$\int_{\mathcal{E}} w(t,x)\,\mathrm{d}x = \text{Prob}\left[\mathbf{X}_t \in \mathcal{E}\right].$$

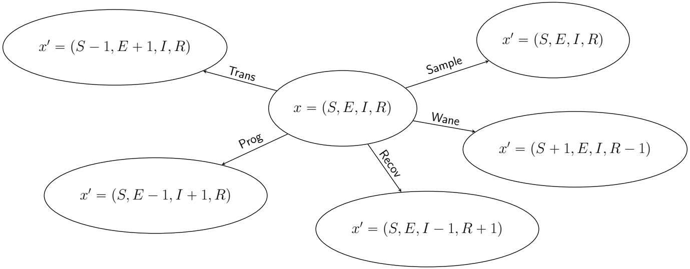
KFE is sometimes called the *master equation* for $\mathbf{X}_t$.

# Population process



$$\frac{\partial w}{\partial t}(t,x) = \int w(t,x')\,\alpha(t,x',x)\,\mathrm{d}x' - \int w(t,x)\,\alpha(t,x,x')\,\mathrm{d}x'$$

# Population process



$x' = (S-1, E+1, I, R)$

Trans

$x' = (S, E, I, R)$

Sample

$x = (S, E, I, R)$

Wane

$x' = (S+1, E, I, R-1)$

Prog

Recov

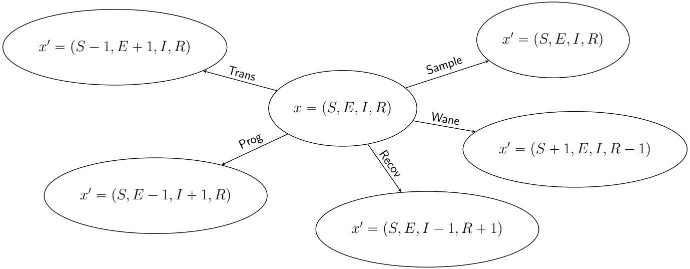
$x' = (S, E-1, I+1, R)$

$x' = (S, E, I-1, R+1)$

$\mathbb{U} = \{\text{Trans}, \text{Prog}, \text{Recov}, \text{Wane}, \text{Sample}\}$

$$\frac{\partial w}{\partial t}(t,x) = \sum_{u \in \mathbb{U}} \left\{ \int w(t,x')\, \alpha_u(t,x',x)\, \mathrm{d}x' - \int w(t,x)\, \alpha_u(t,x,x')\, \mathrm{d}x' \right\}$$

# Population process



$x = (S, E, I, R)$

Trans

$x' = (S-1, E+1, I, R)$

Sample

$x' = (S, E, I, R)$

Wane

$x' = (S+1, E, I, R-1)$

Prog

$x' = (S, E-1, I+1, R)$

Recov

$x' = (S, E, I-1, R+1)$

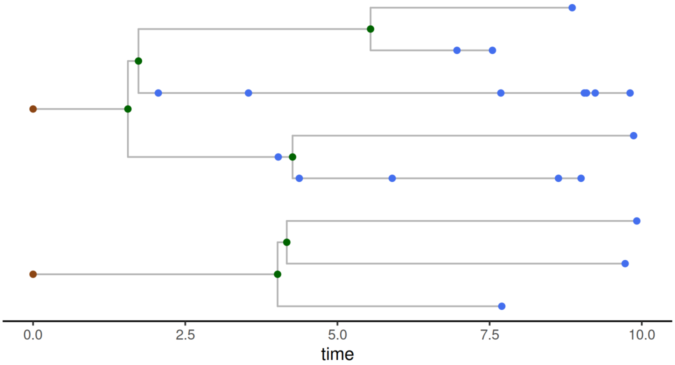
$\mathbb{U} = \{\text{Trans}, \text{Prog}, \text{Recov}, \text{Wane}, \text{Sample}\}$

$$\begin{aligned}\frac{\partial w}{\partial t}(t, S, E, I, R) = {} & \frac{\beta(t)\,(S+1)\,I}{N}\,w(t, S+1, E-1, I, R) - \frac{\beta(t)\,S\,I}{N}\,w(t, S, E, I, R) \\ & + \sigma\,(E+1)\,w(t, S, E+1, I-1, R) - \sigma\,E\,w(t, S, E, I, R) \\ & + \gamma\,(I+1)\,w(t, S, E, I+1, R-1) - \gamma\,I\,w(t, S, E, I, R) \\ & + \omega\,(R+1)\,w(t, S-1, E, I, R+1) - \omega\,R\,w(t, S, E, I, R)\end{aligned}$$

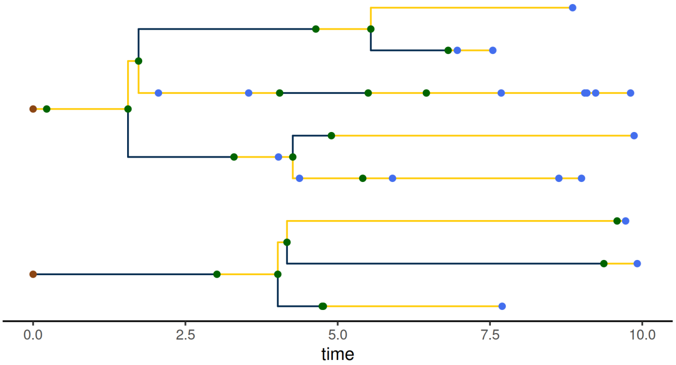
# Outline

1. Context
   - Example: emerging variants
   - Phylodynamics
   - Problems of phylodynamics
2. Population process
   - Examples
   - Formalization
3. Genealogy process
   - Genealogies
   - Induced genealogy process
4. Pruned and obscured genealogies

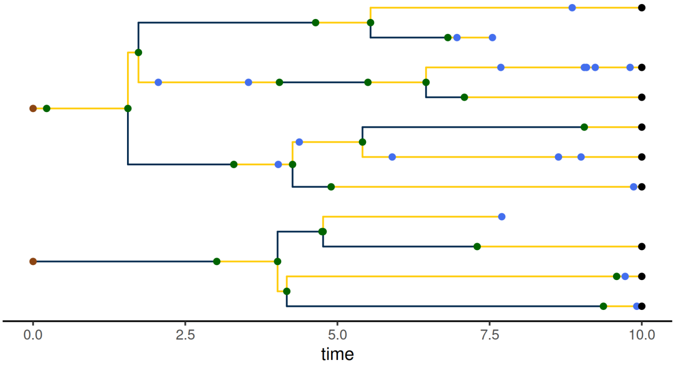
# What is a genealogy?



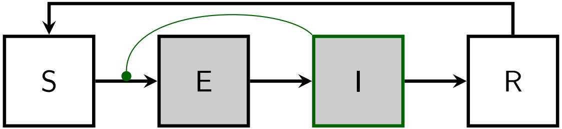
# What is a genealogy?



# What is a genealogy?

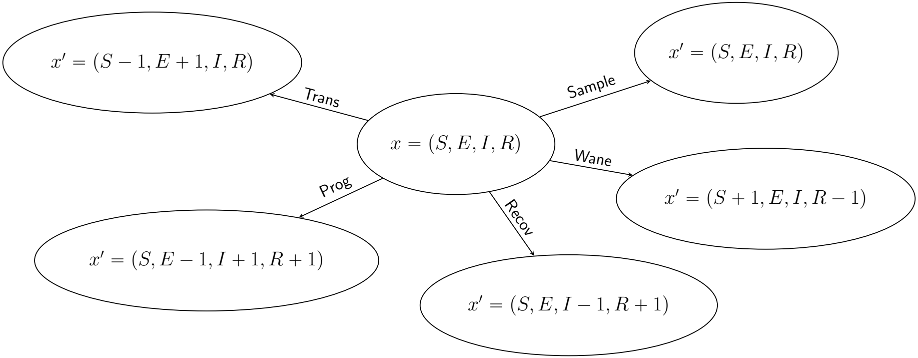


# Event types



$$\mathbb{D} = \{\mathsf{E}, \mathsf{I}\}$$

# Event types



$x = (S, E, I, R)$

Trans

$x' = (S-1, E+1, I, R)$

Sample

$x' = (S, E, I, R)$

Wane

$x' = (S+1, E, I, R-1)$

Prog

$x' = (S, E-1, I+1, R+1)$

Recov

$x' = (S, E, I-1, R+1)$

$\mathbb{U} = \{\text{Trans}, \text{Prog}, \text{Recov}, \text{Wane}, \text{Sample}\}$

# Event types

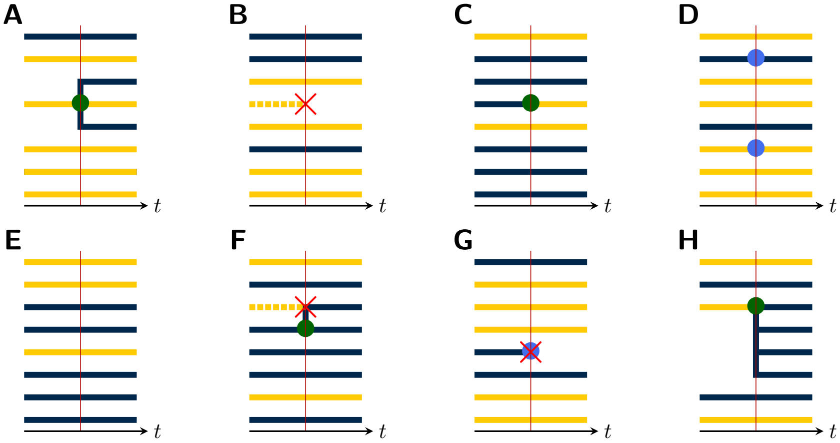
If we write

$$\alpha(t, x, x') = \sum_{u \in \mathbb{U}} \alpha_u(t, x, x'),$$

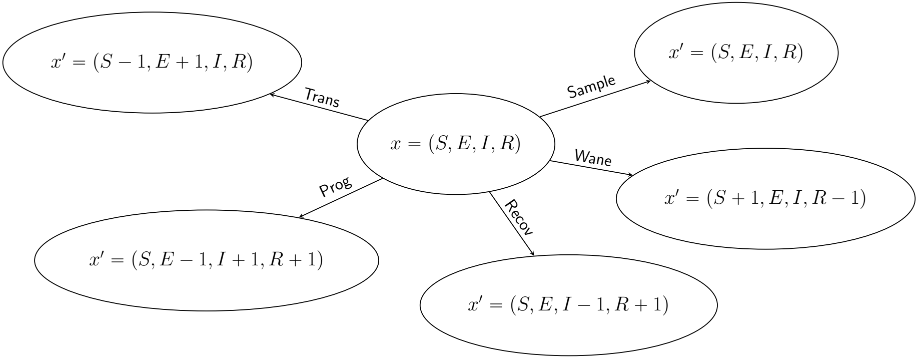
the KFE becomes

$$\frac{\partial w}{\partial t}(t, x) = \sum_u \int w(t, x')\, \alpha_u(t, x', x)\, \mathrm{d}x' - \sum_u \int w(t, x)\, \alpha_u(t, x, x')\, \mathrm{d}x'$$

# Event types



# Event types



$$\mathbb{U} = \{\text{Trans}, \text{Prog}, \text{Recov}, \text{Wane}, \text{Sample}\}$$

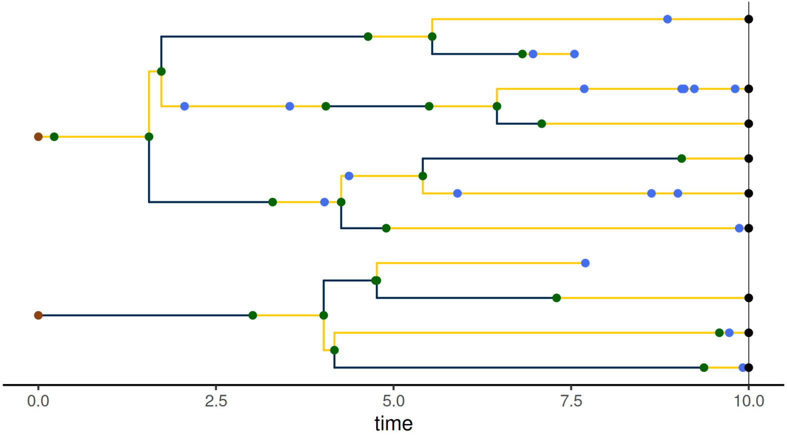
# A population process induces a genealogy process

- $\mathbf{G}_t$ is a stochastic process on the space of genealogies.
- The map $\mathbf{X} \mapsto \mathbf{G}$ is random.
- **Key assumption:** Lineages within a deme are *exchangeable*.
  There is no more structure than is implied by the population process.
- Simulation code on `github.com/kingaa/phylopomp`
- Animations at `https://kingaa.github.io/manuals/phylopomp/vignettes/`

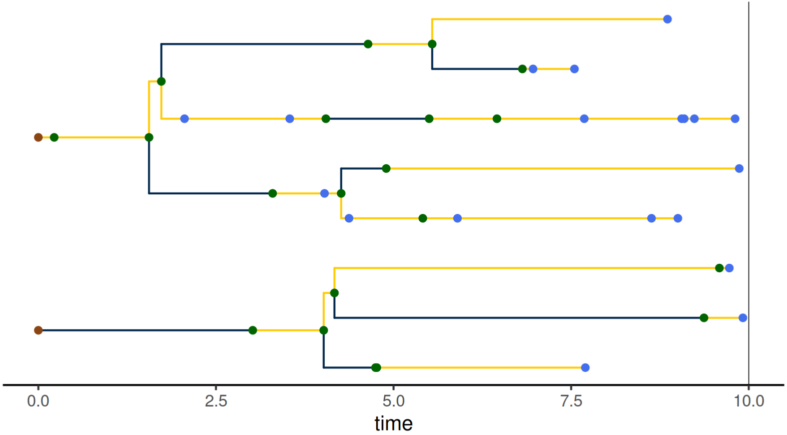
# Outline

1. Context
   - Example: emerging variants
   - Phylodynamics
   - Problems of phylodynamics
2. Population process
   - Examples
   - Formalization
3. Genealogy process
   - Genealogies
   - Induced genealogy process
4. Pruned and obscured genealogies

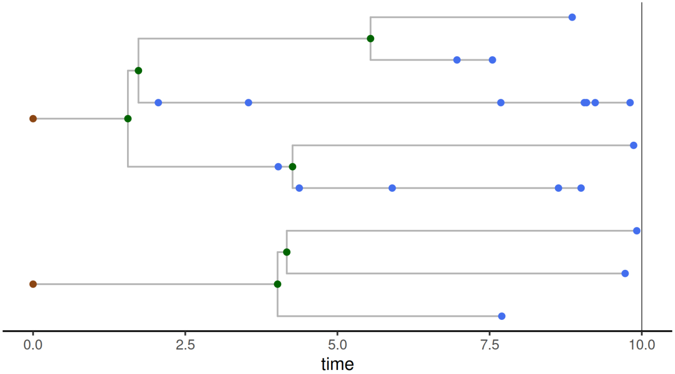
# Full genealogy



# Pruned genealogy



# Obscured genealogy



An obscured genealogy is specified by $(\mathrm{T}, \mathrm{Z})$.

# Summary

- A discretely structured Markov population process uniquely induces a genealogy-valued Markov process.
- The likelihood of an observed genealogy satisfies a nonlinear filtering equation, which can be efficiently computed via Feynman-Kaç (sequential Monte Carlo) algorithms.
- In principle, these results liberate us to entertain models that more closely match our biological questions, without less hindrance from inference methodology.

# References

Hadfield J, Megill C, Bell SM, Huddleston J, Potter B, Callender C, Sagulenko P, Bedford T, Neher RA (2018). “Nextstrain: real-time tracking of pathogen evolution.” *Bioinformatics*, **34**(23), 4121–4123. doi: 10.1093/bioinformatics/bty407.

King AA, Nguyen D, Ionides EL (2016). “Statistical inference for partially observed Markov processes via the R package pomp.” *J Stat Softw*, **69**(12), 1–43. doi: 10.18637/jss.v069.i12.

Mathieu E, Ritchie H, Rodés-Guirao L, Appel C, Giattino C, Hasell J, Macdonald B, Dattani S, Beltekian D, Ortiz-Ospina E, Roser M (2020). “Coronavirus pandemic (COVID-19).” *Our World in Data [Online resource]*. URL https://ourworldindata.org/coronavirus.

# References II

Vaughan TG, Leventhal GE, Rasmussen DA, Drummond AJ, Welch D, Stadler T (2019). "Estimating epidemic incidence and prevalence from genomic data." *Mol. Biol. Evol.*, **36**, 1804–1816. `doi: 10.1093/molbev/msz106`.

# License, acknowledgments, and links

- The materials build on previous versions of this course and related courses.
- Licensed under the Creative Commons Attribution-NonCommercial license. Please share and remix non-commercially, mentioning its origin.
- Produced with R version 4.3.2 and **pomp** version 5.6.
- Compiled on March 4, 2024.

Back to Lesson

R codes for this lesson