2023 Nordita meeting

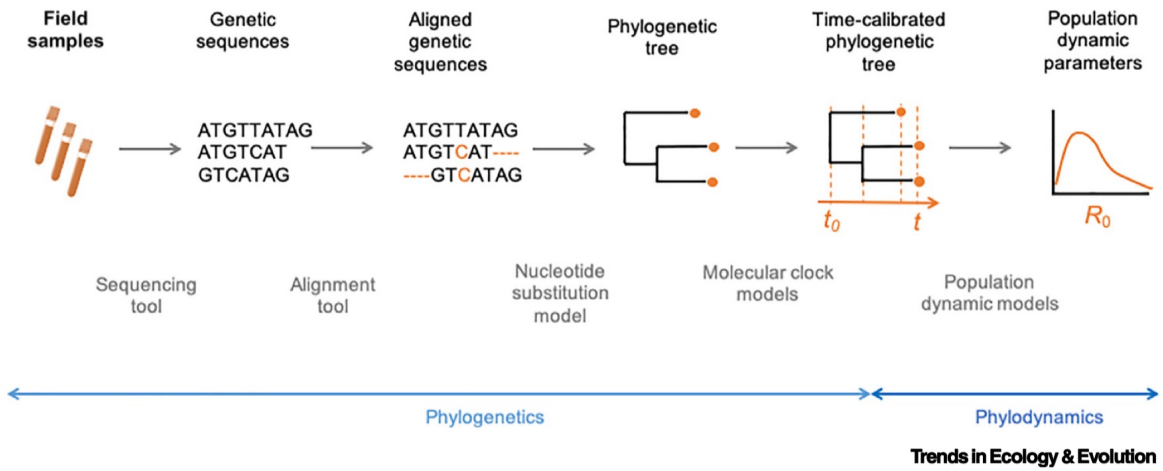
Inferring epidemiological dynamics using temporal patterns of genetic variation

with application to SARS-CoV-2

Yeongseon Park

June 19th, 2023

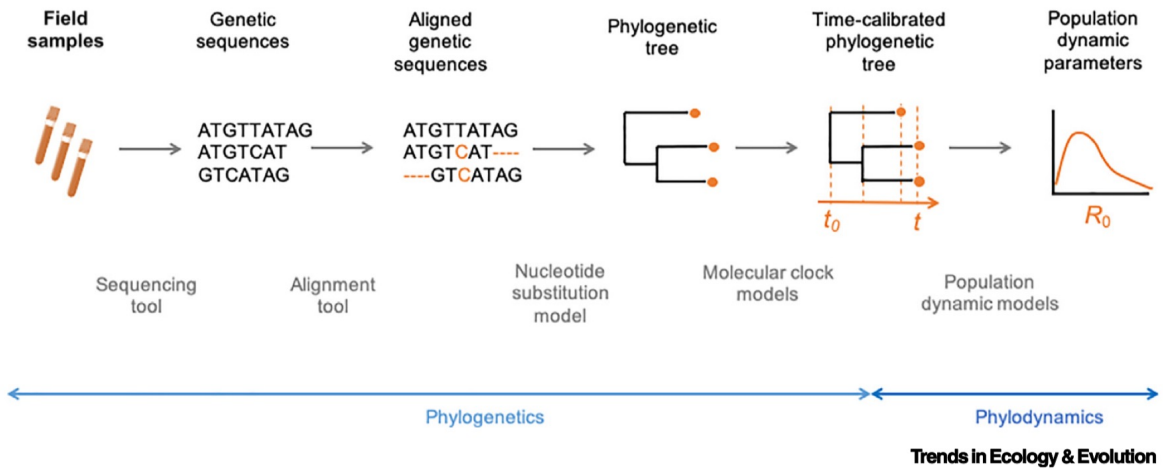
Sequence-based appraoch relying on phylogeny



Guinat et al. (Trends in Ecology & Evolution 2021)

- Significant amount of phylogenetic uncertainty
- More parameters to be estimated with a limited amount of information
- Cannot directly estimate the timing of index case

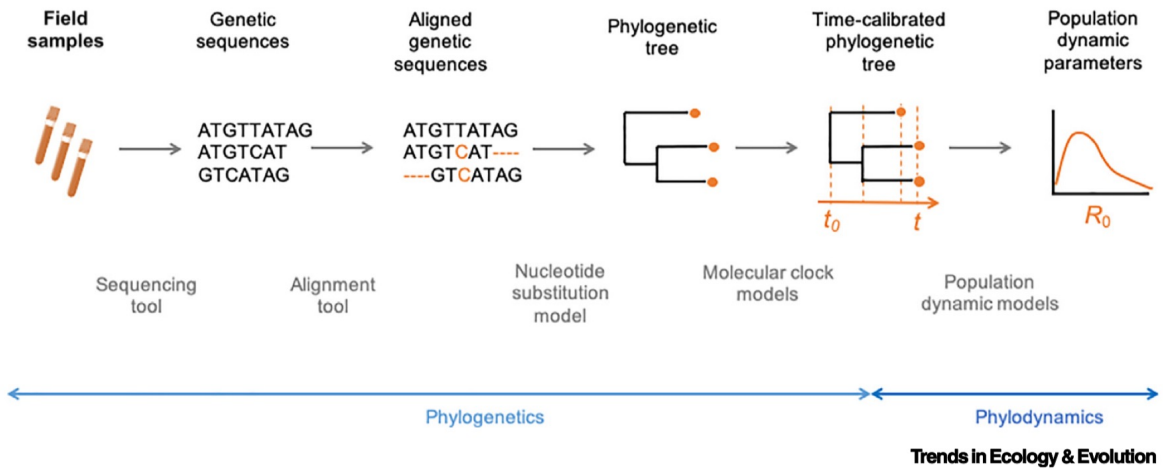
Phylogeny-based approaches could be limited under low viral genetic diversity



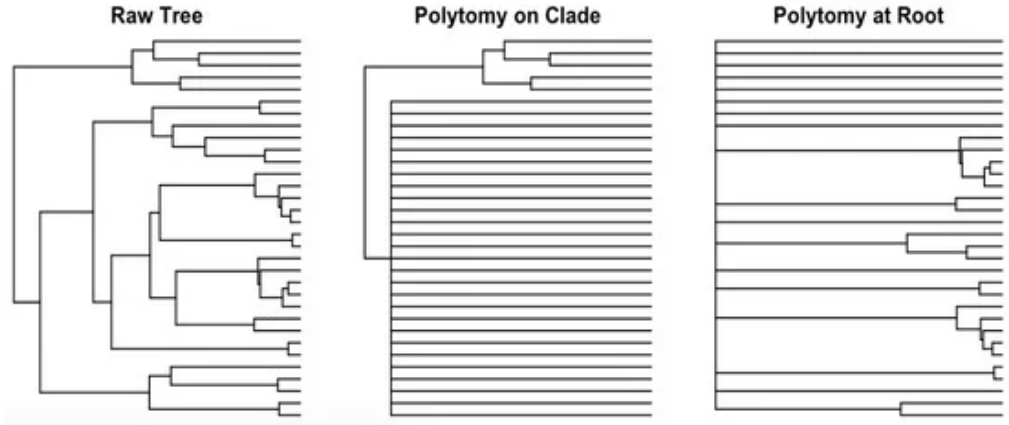
Guinat et al. (Trends in Ecology & Evolution 2021)

- Significant amount of phylogenetic uncertainty
- More parameters to be estimated with a limited amount of information
- Cannot directly estimate the timing of index case

Phylogeny-based approaches could be limited under low viral genetic diversity



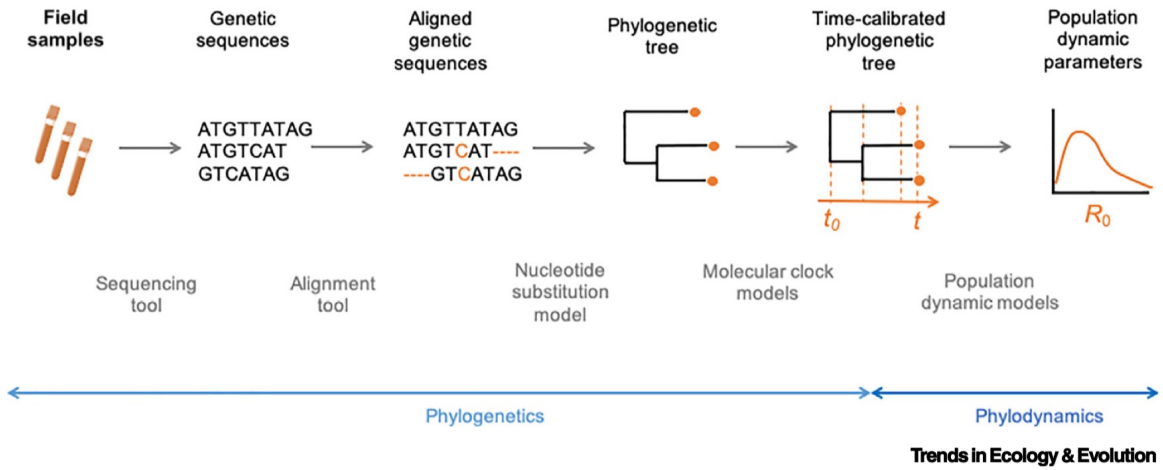
Guinat et al. (Trends in Ecology & Evolution 2021)



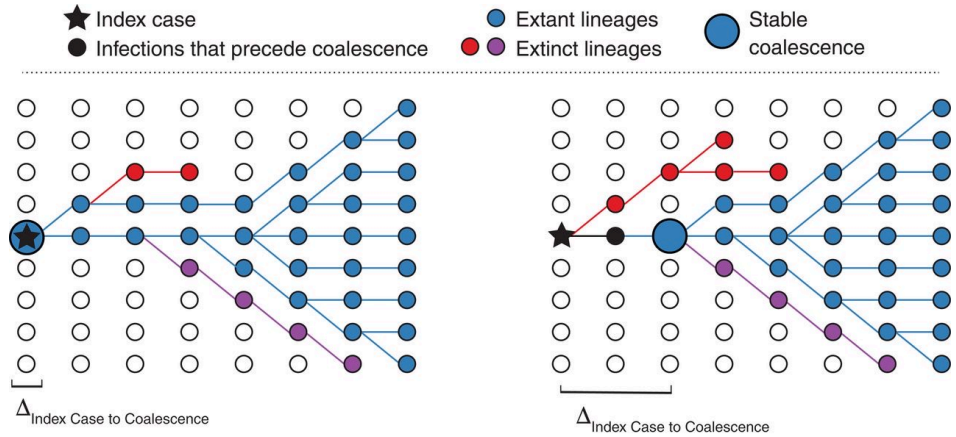
Jhwueng et al. (Diversity, 2022)

- Significant amount of phylogenetic uncertainty
- More parameters to be estimated with a limited amount of information
- Cannot directly estimate the timing of index case

Phylogeny-based approaches could be limited under low viral genetic diversity



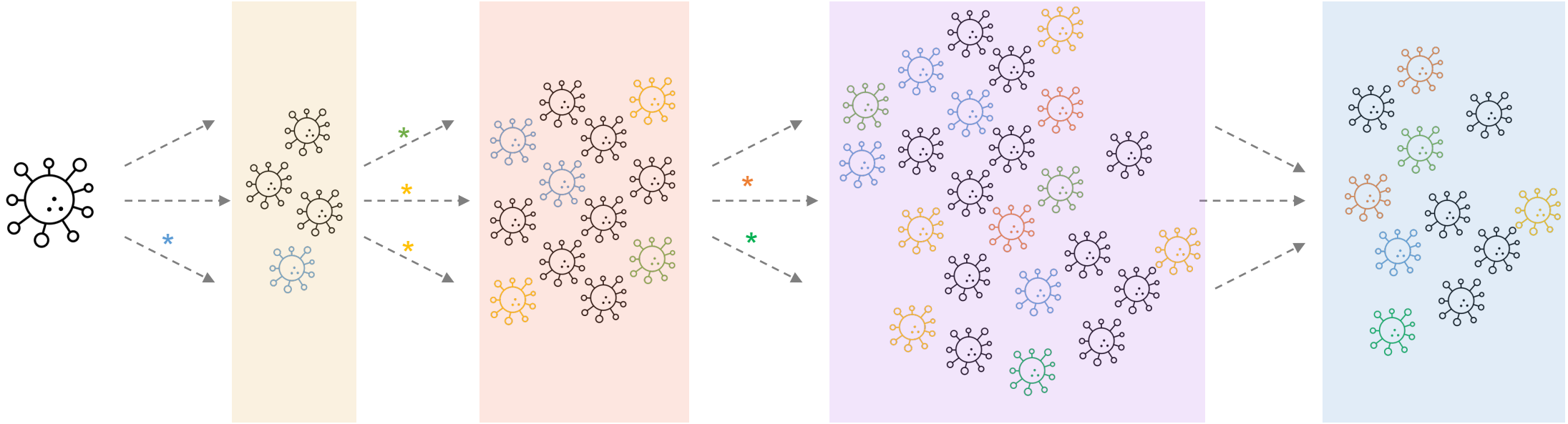
Guinat et al. (Trends in Ecology & Evolution 2021)

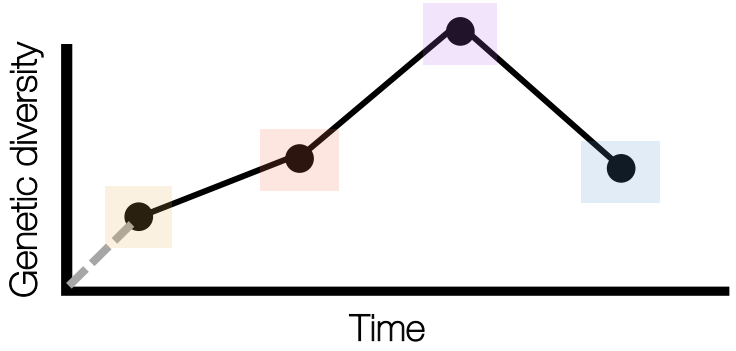


Pekar et al. (Science, 2021)

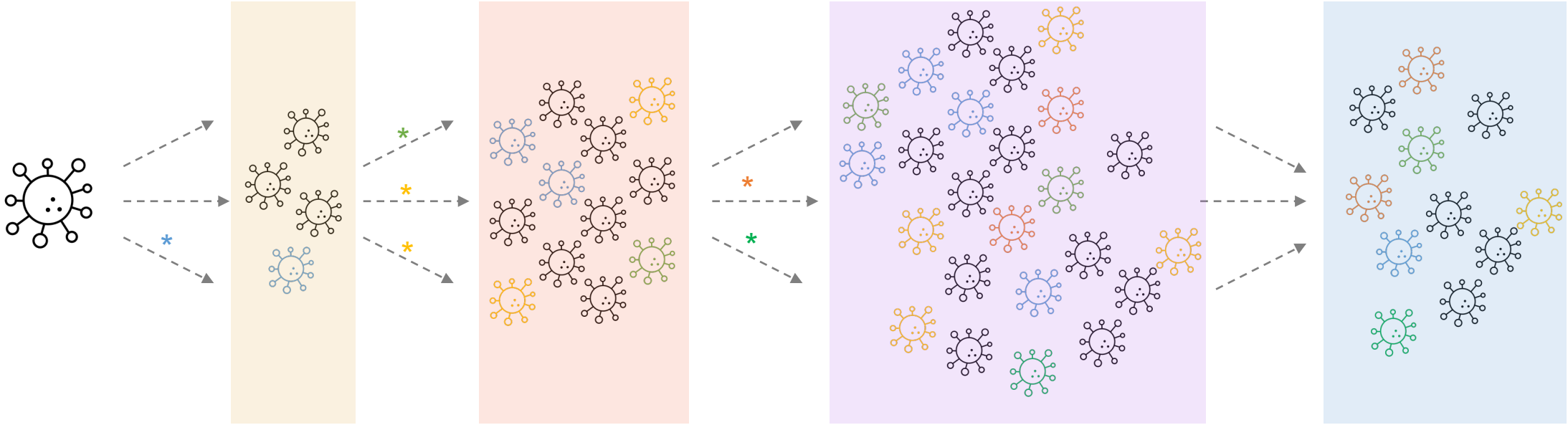
- Significant amount of phylogenetic uncertainty
- More parameters to be estimated with a limited amount of information
- Cannot directly estimate the timing of index case

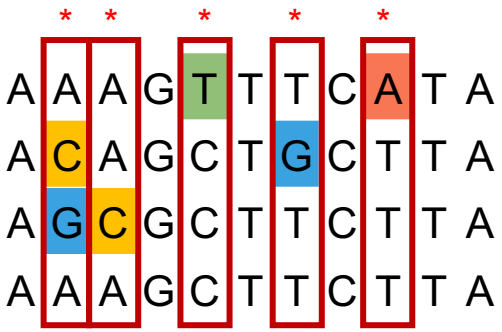
Phylogeny-free approach:
Using genetic variation as time series





Phylogeny-free approach:
Using genetic variation as time series

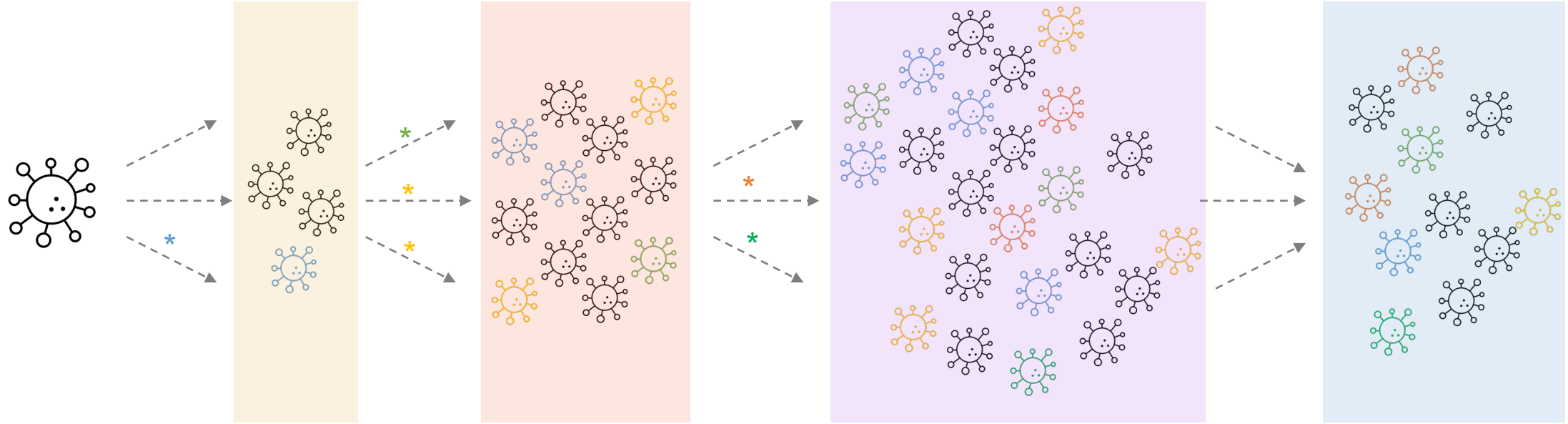


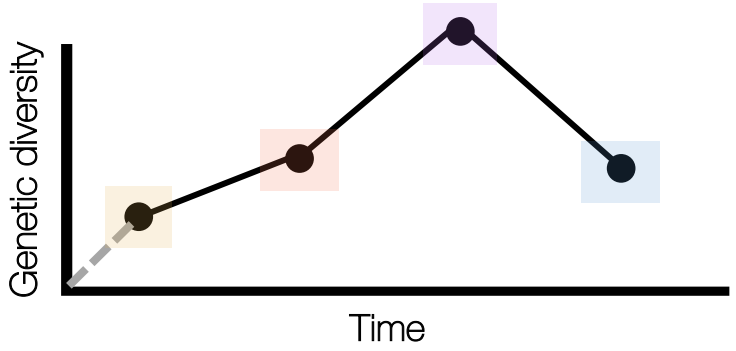


The number of segregating sites

- Classic population genetic statistic summarizing genetic diversity
- The number of sites with more than one allele
- Directly obtained from sequence data
- Depends on the sample size

Phylogeny-free approach: Using genetic variation as time series

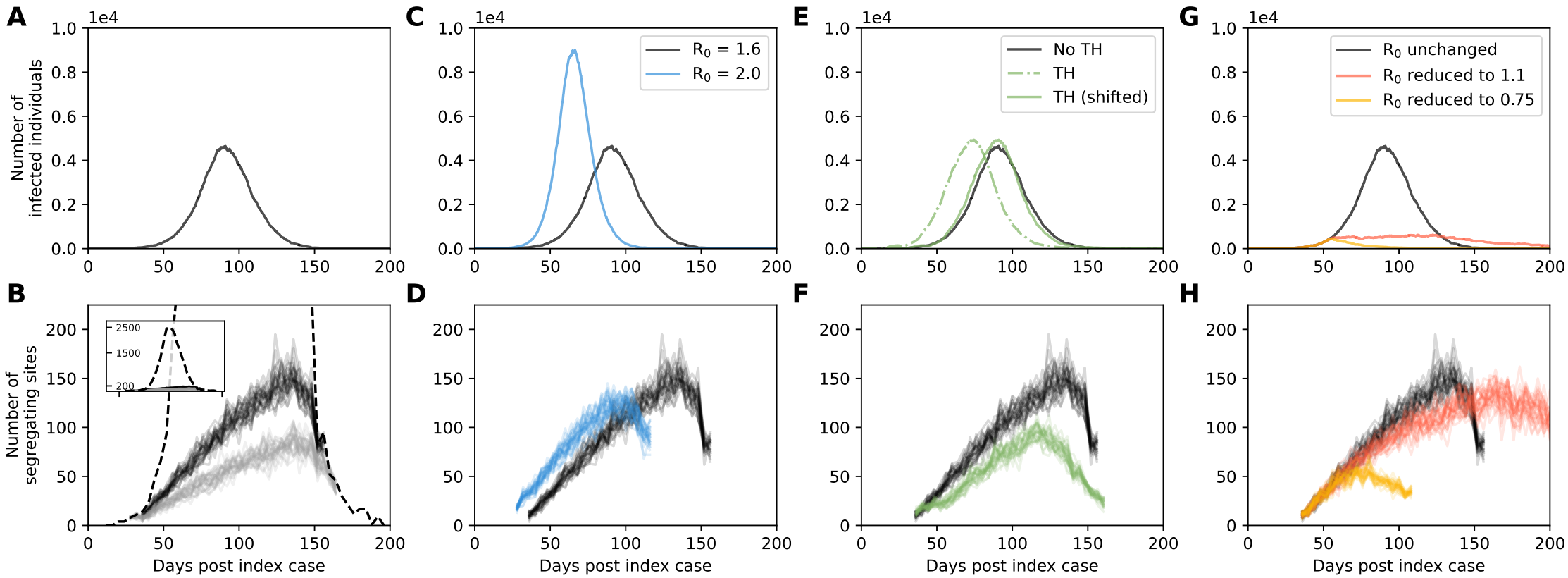




Segregating site trajectory is obtained by

- Determine the window size for time series
- Bin sequences according to the sampling date
- Count the number of segregating sites for each window

Segregating site trajectories are informative of the underlying dynamics

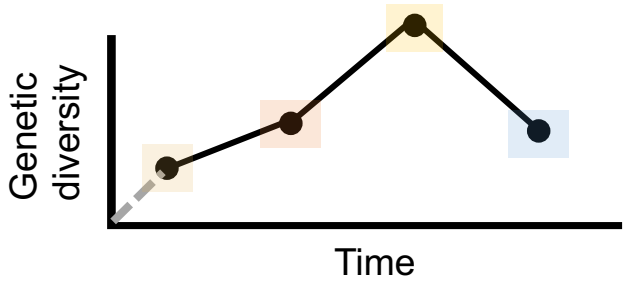


Estimation of R_0 and timing of the index case

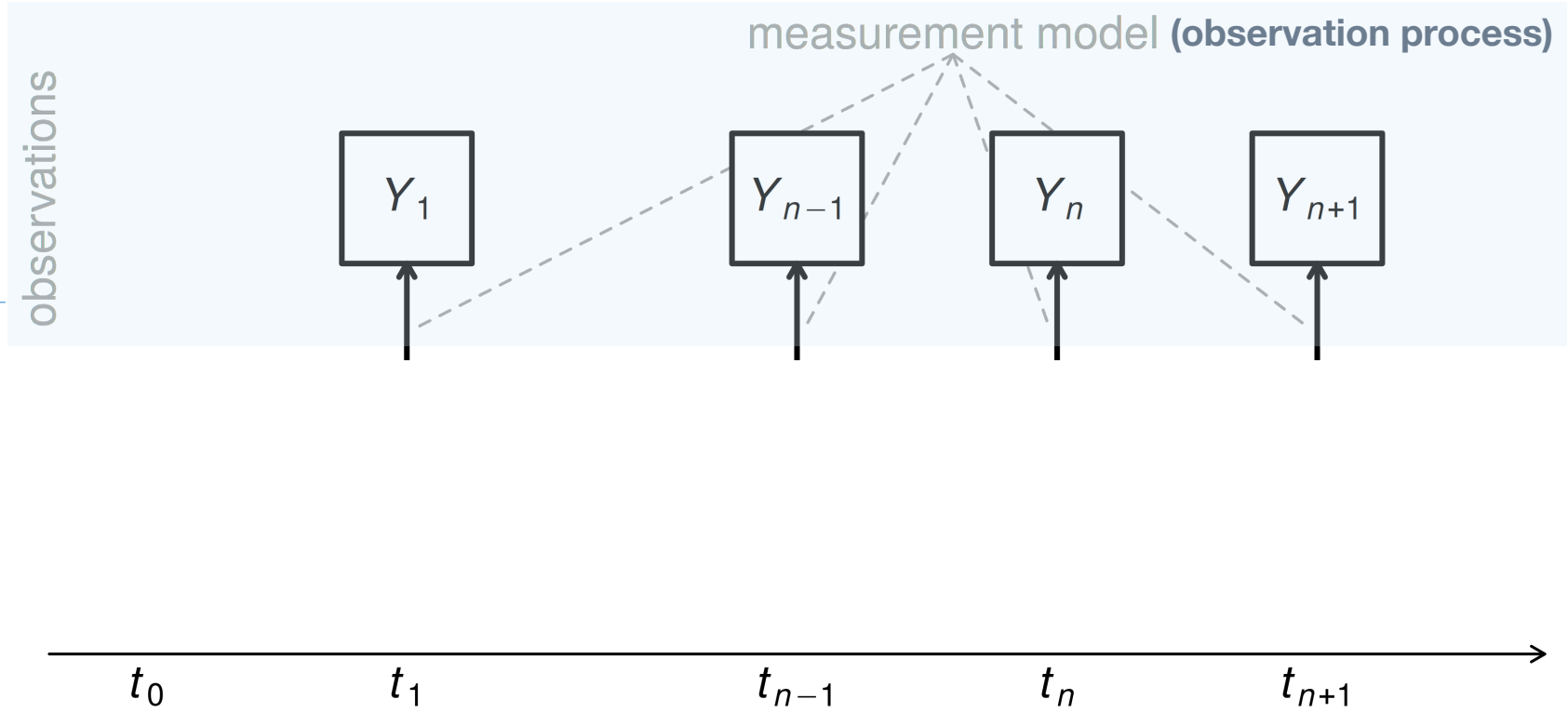
with application to early France

Model and inference framework:

State-space model and particle filtering



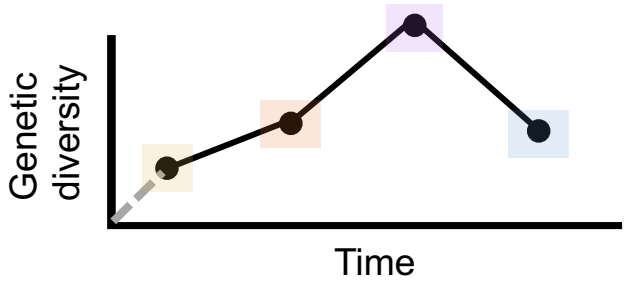
The number of segregating sites



https://kingaa.github.io/pomp/vignettes/getting_started.html

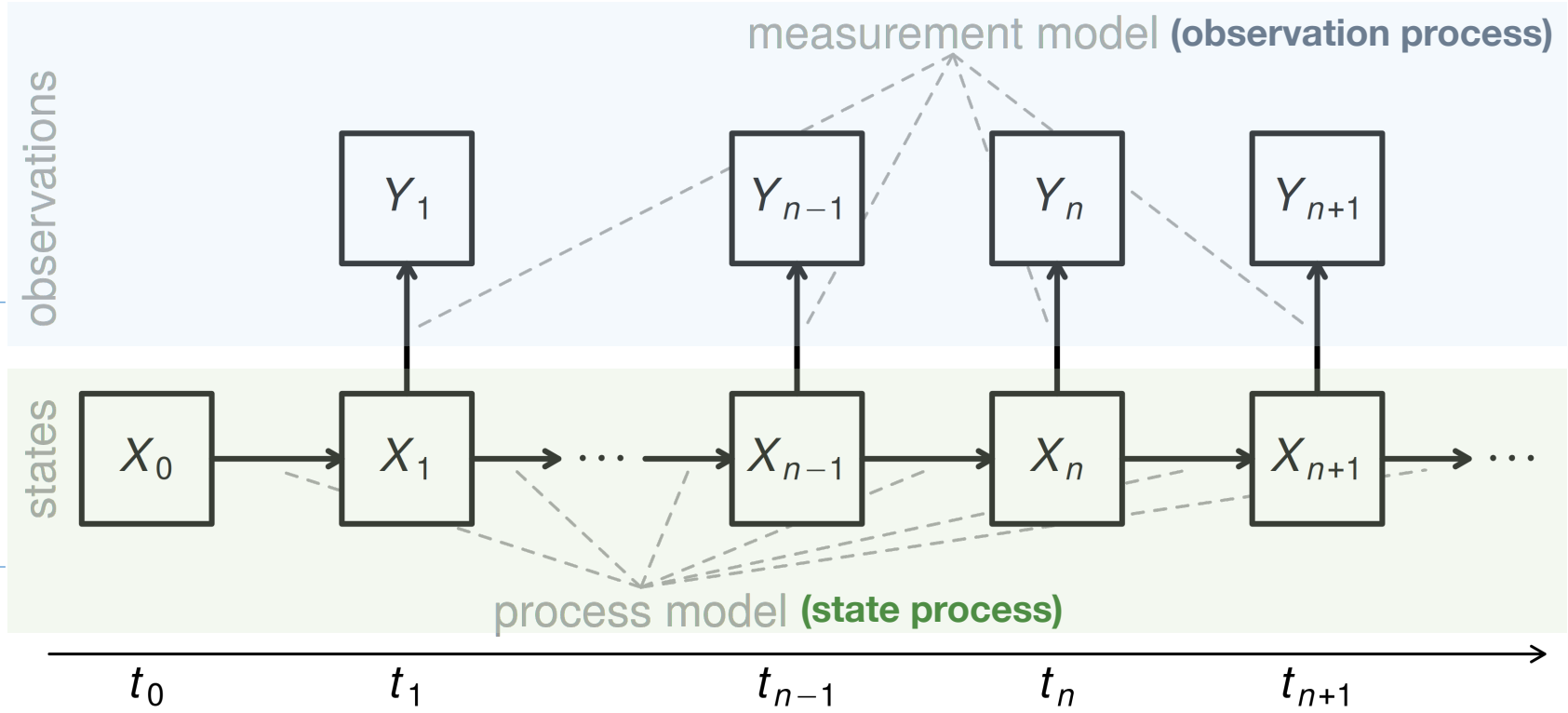
Model and inference framework:

State-space model and particle filtering



The number of segregating sites

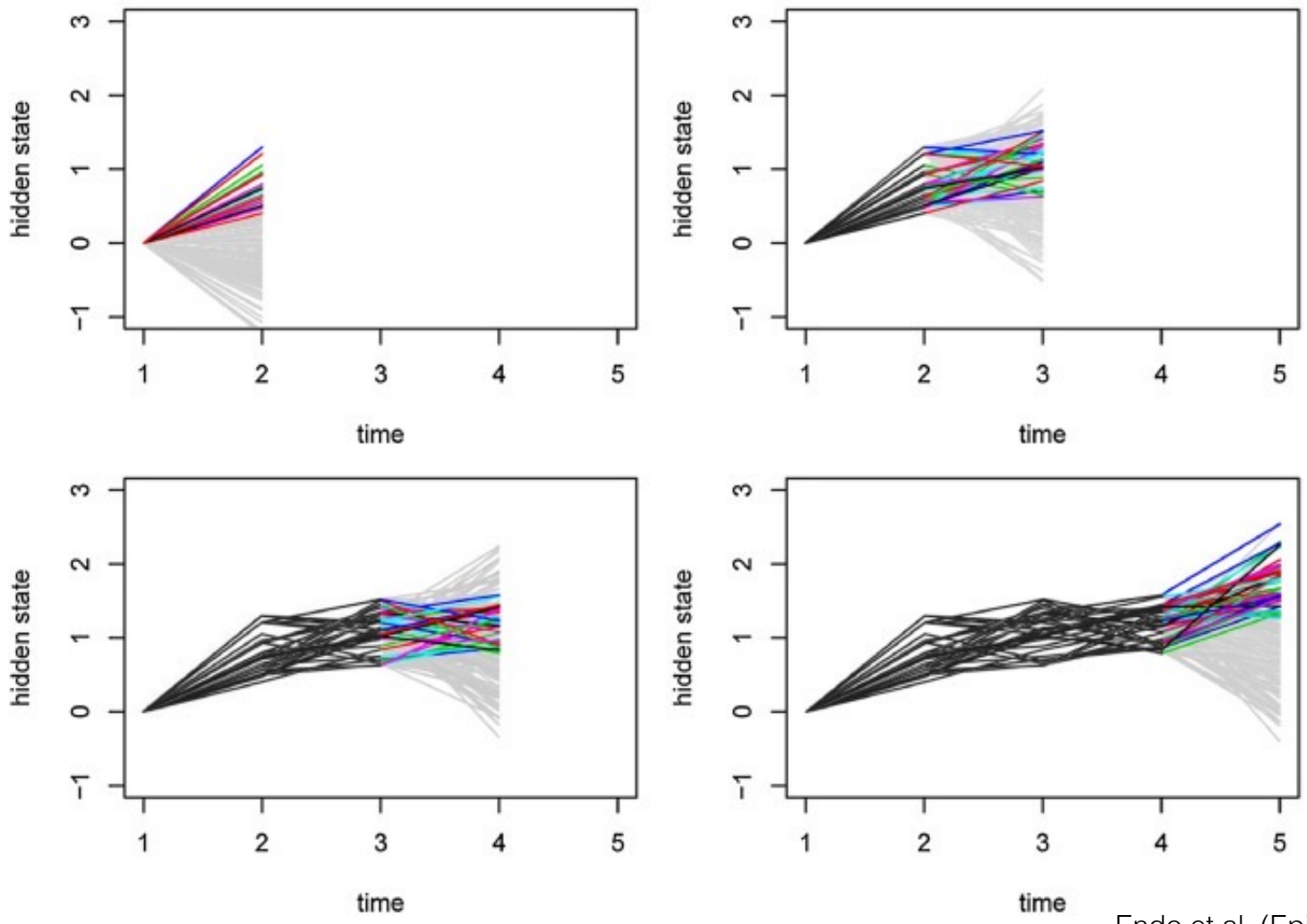
Epidemiological & evolutionary dynamics



https://kingaa.github.io/pomp/vignettes/getting_started.html

Model and inference framework:

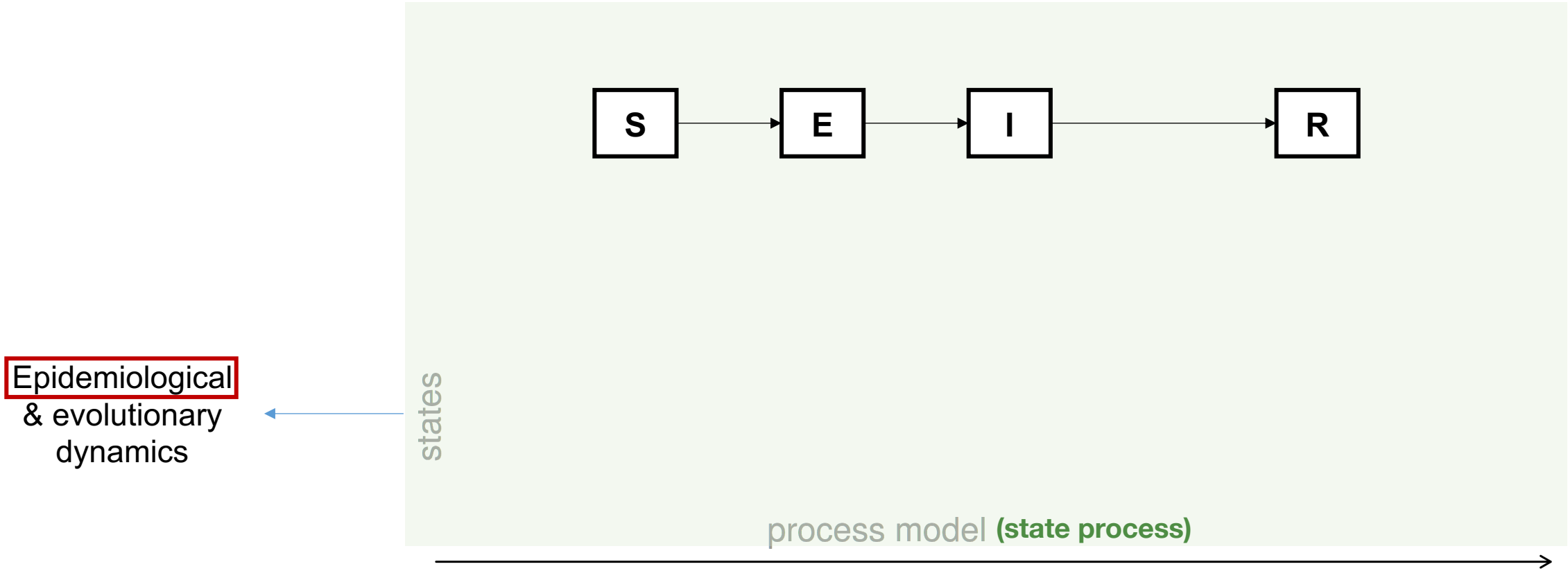
State-space model and particle filtering



Endo et al. (Epidemics. 2020)

Model and inference framework:

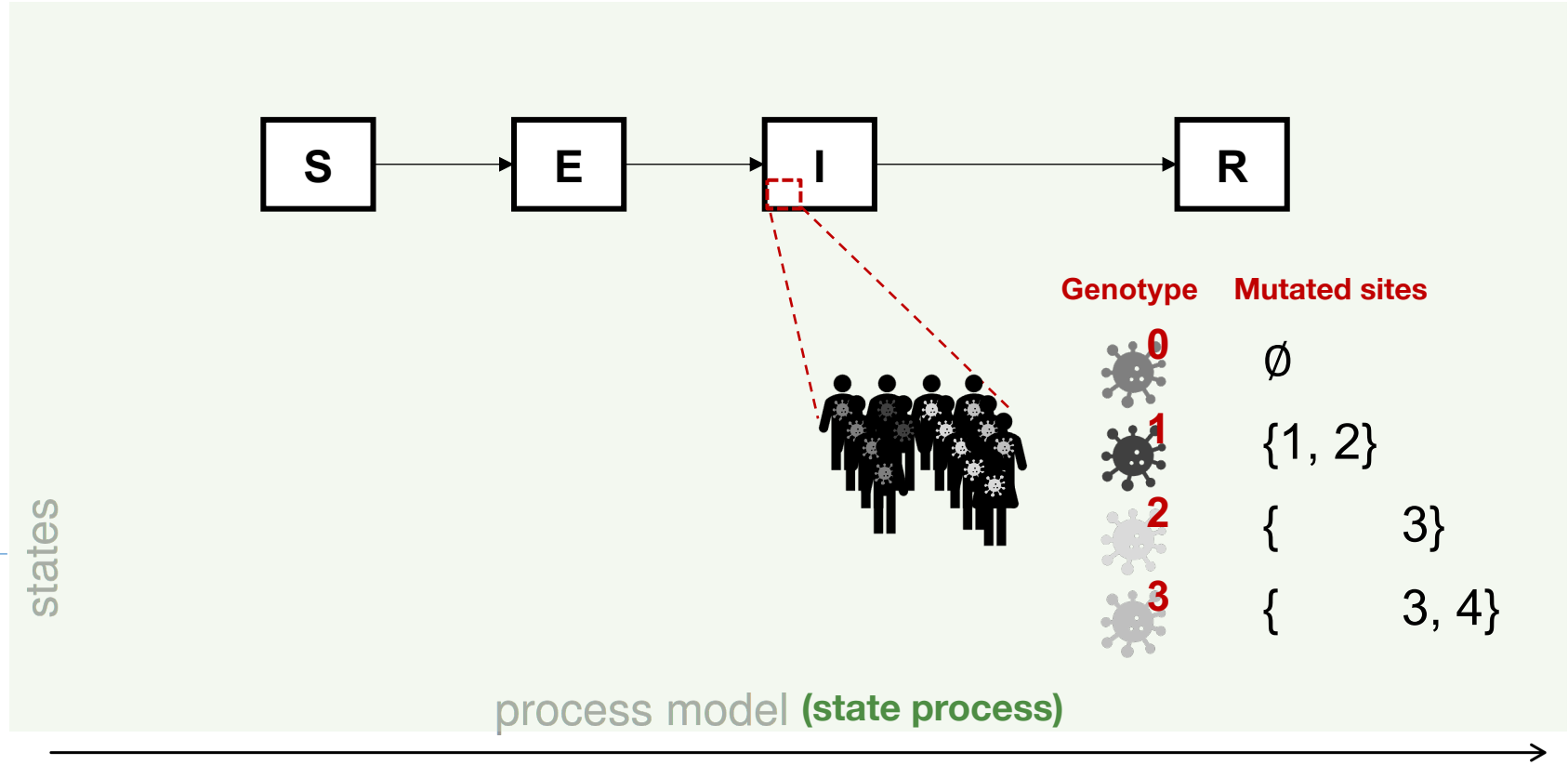
State-space model and particle filtering



https://kingaa.github.io/pomp/vignettes/getting_started.html

Model and inference framework:

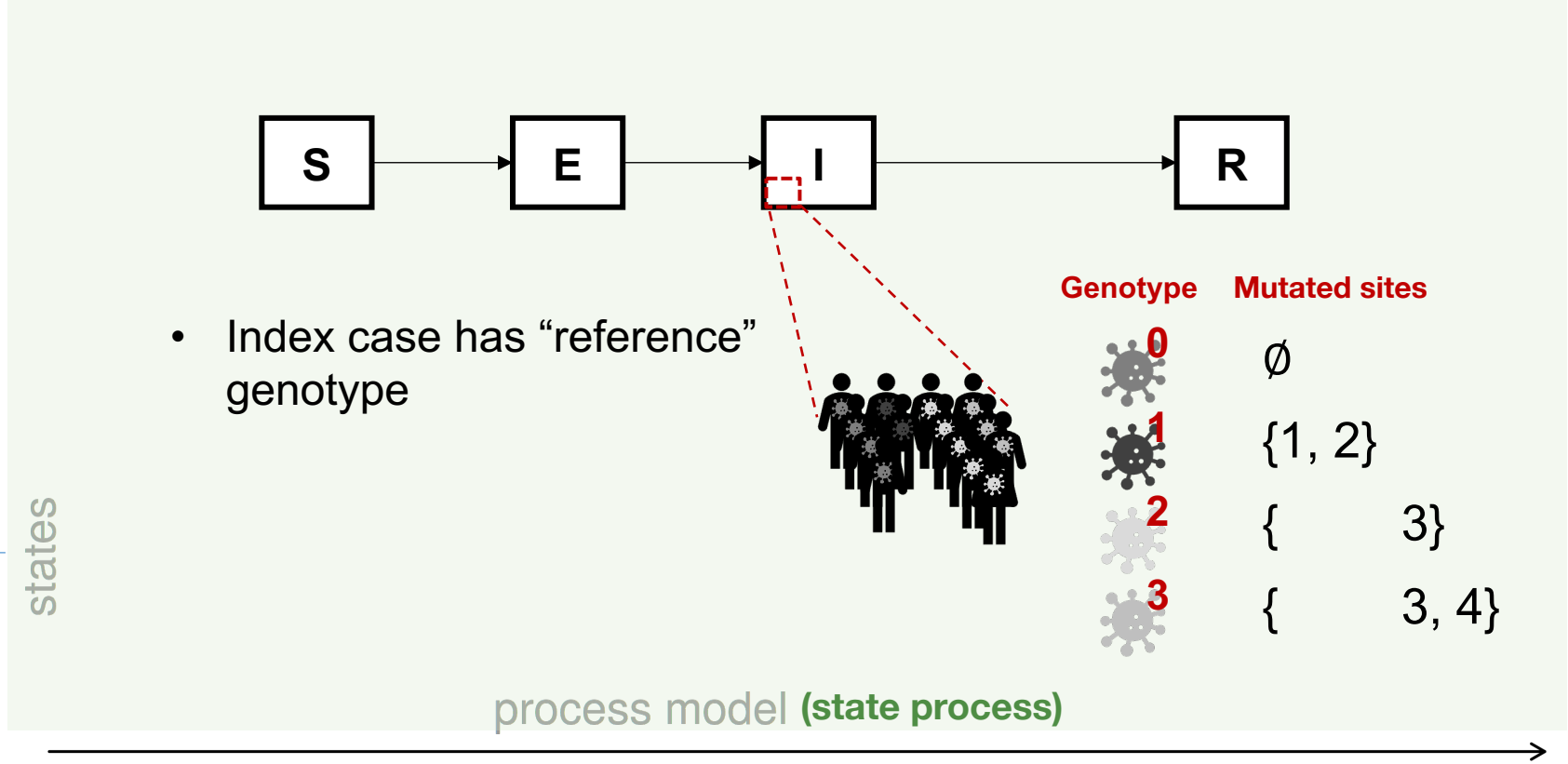
State-space model and particle filtering



https://kingaa.github.io/pomp/vignettes/getting_started.html

Model and inference framework:

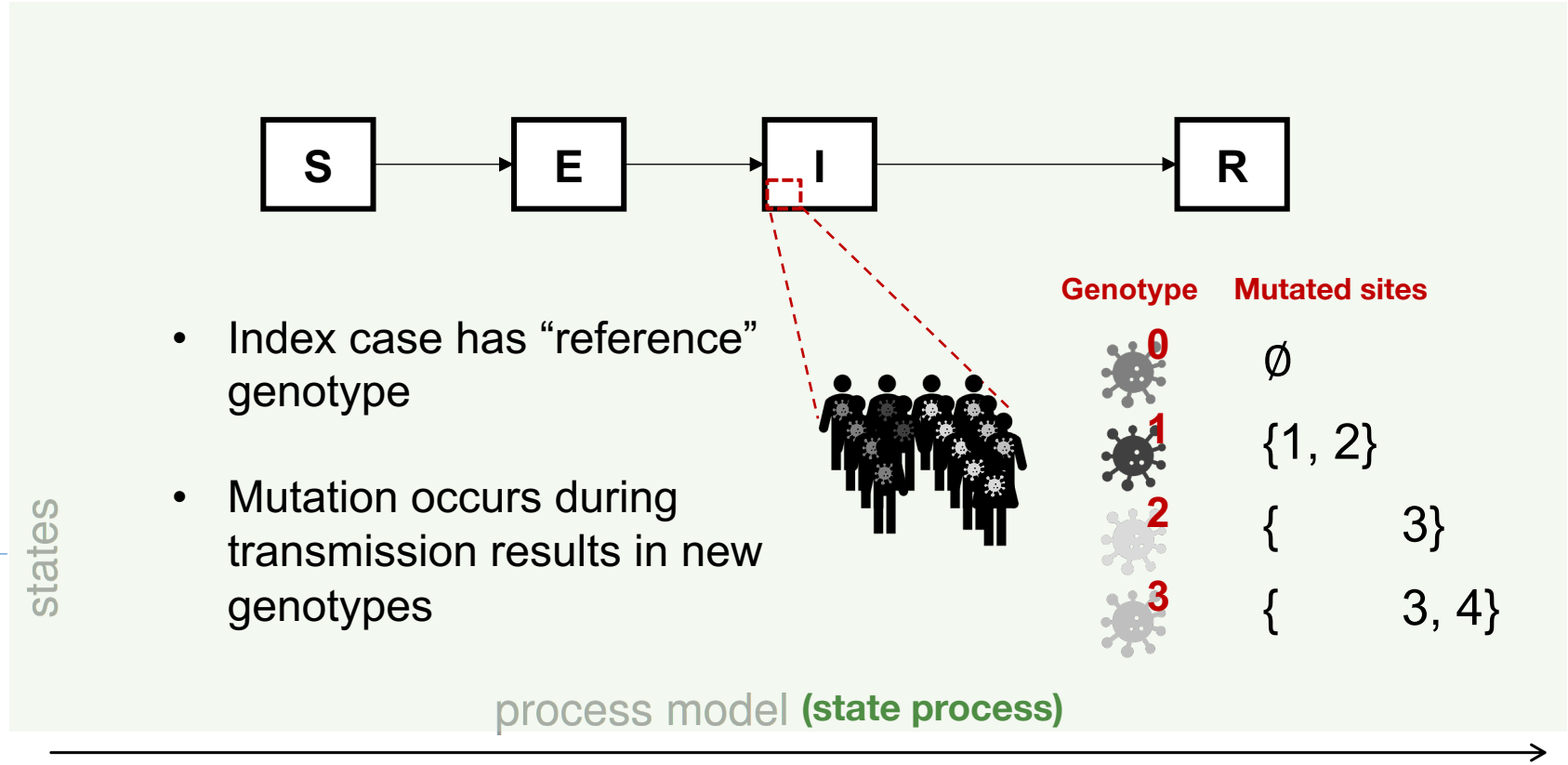
State-space model and particle filtering



https://kingaa.github.io/pomp/vignettes/getting_started.html

Model and inference framework:

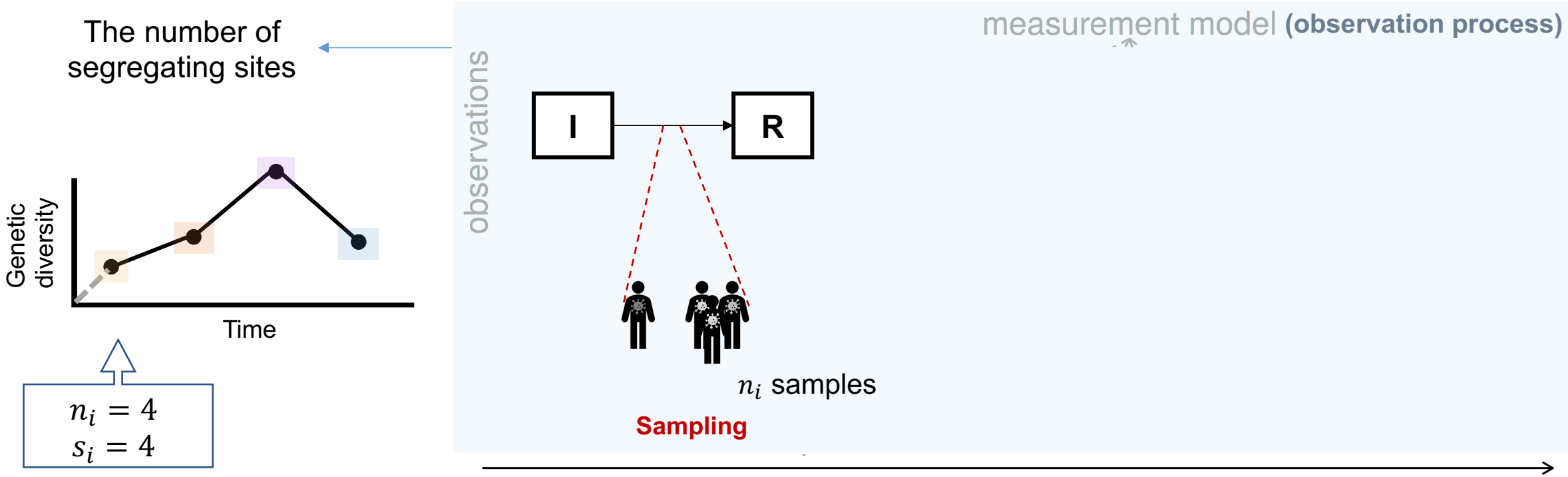
State-space model and particle filtering



https://kingaa.github.io/pomp/vignettes/getting_started.html

Model and inference framework:

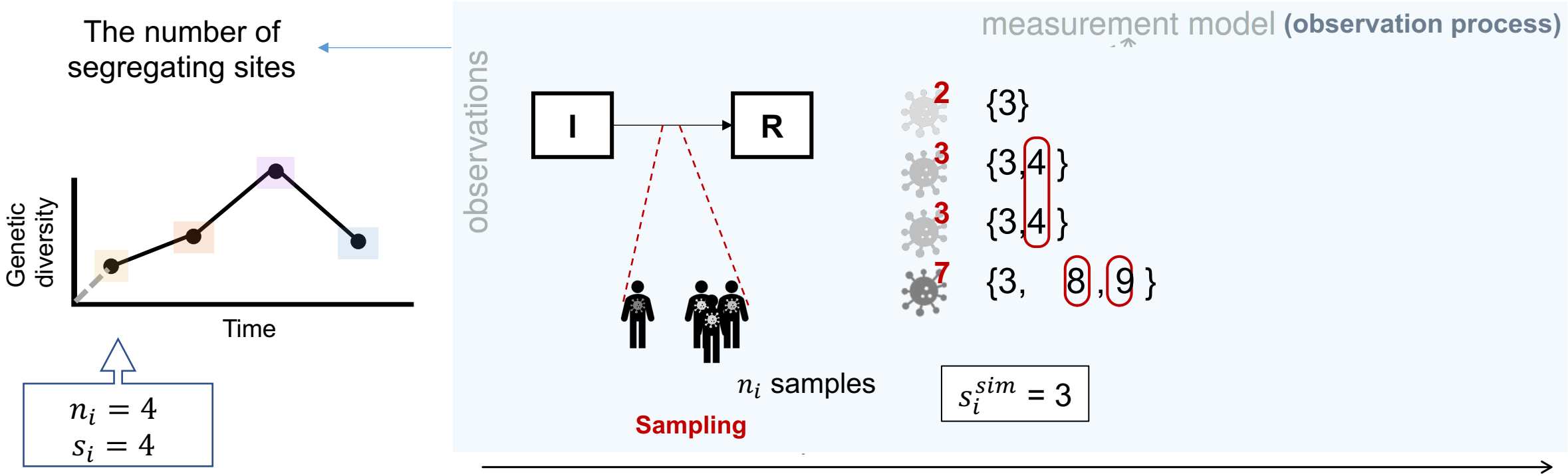
Observation process obtains particle weight



https://kingaa.github.io/pomp/vignettes/getting_started.html

Model and inference framework:

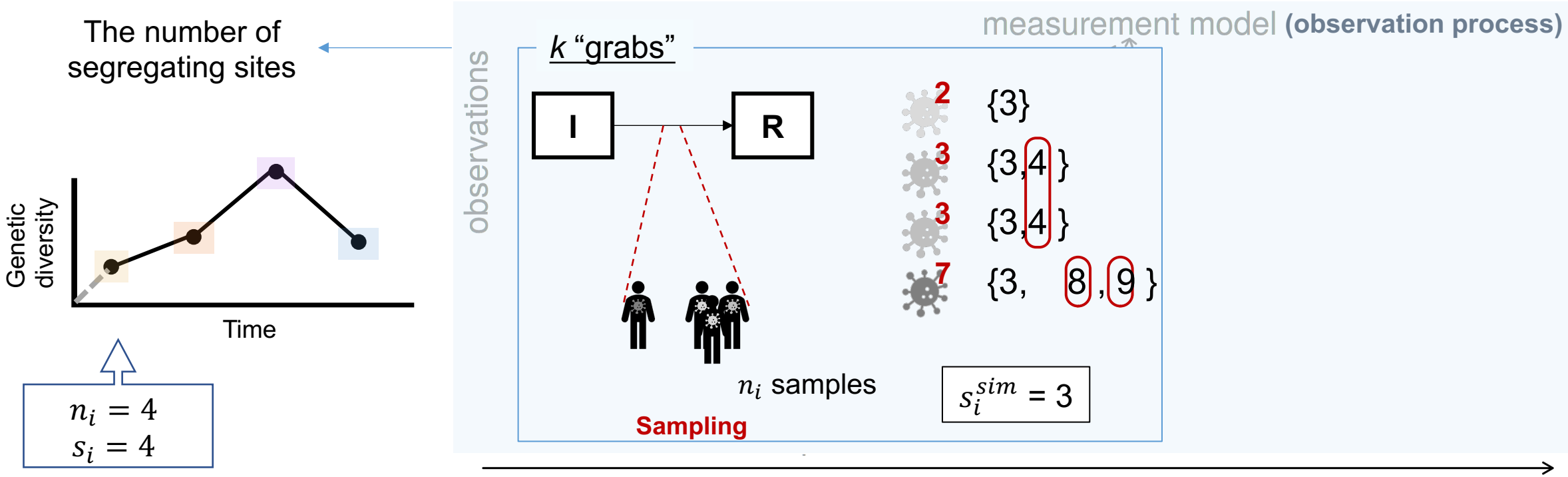
Observation process obtains particle weight



https://kingaa.github.io/pomp/vignettes/getting_started.html

Model and inference framework:

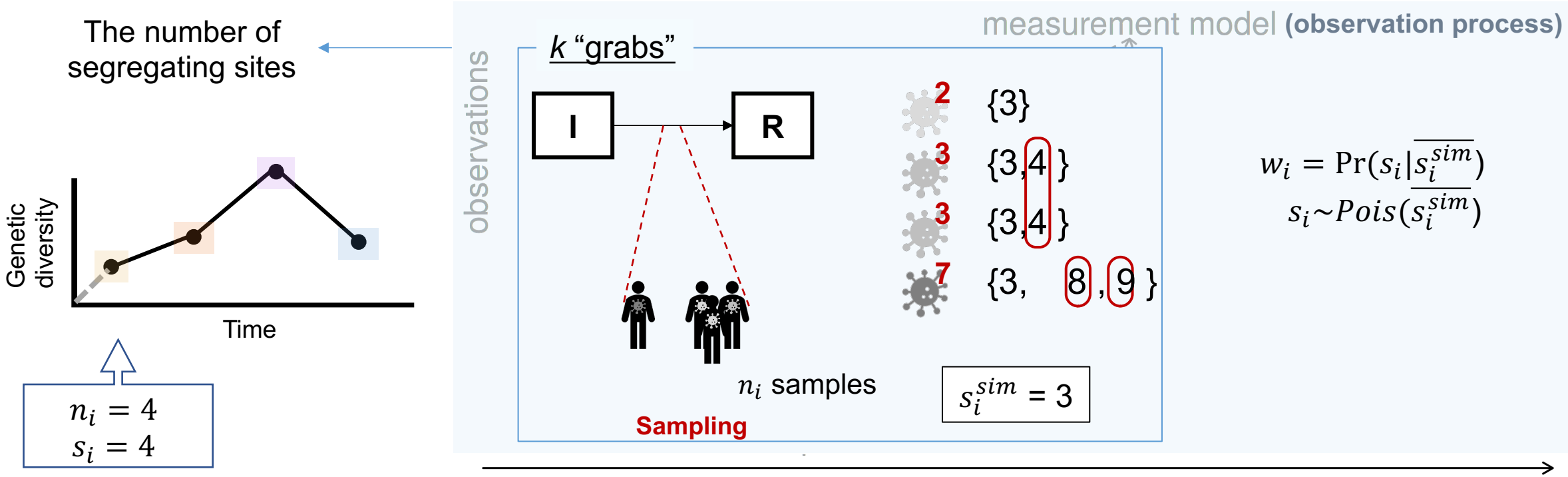
Observation process obtains particle weight



https://kingaa.github.io/pomp/vignettes/getting_started.html

Model and inference framework:

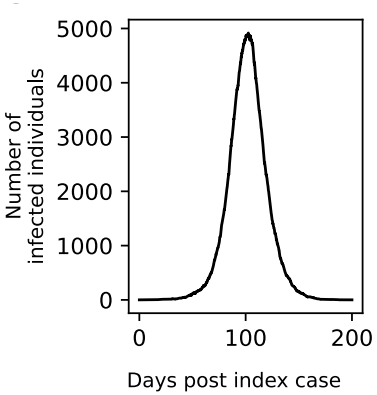
Observation process obtains particle weight



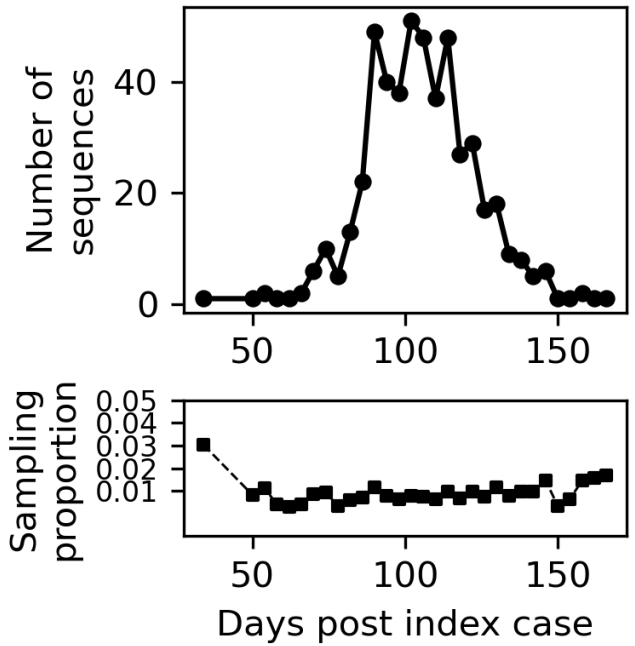
https://kingaa.github.io/pomp/vignettes/getting_started.html

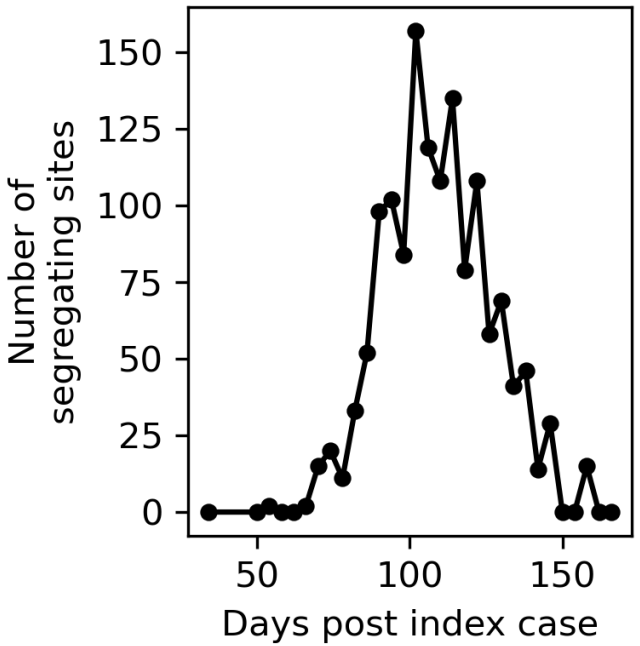
Validation using simulated data:

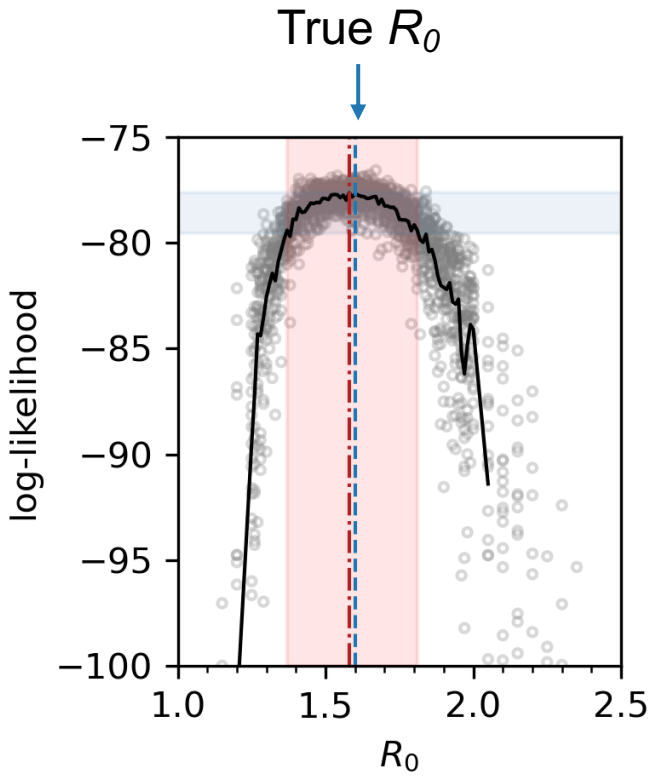
Estimation of R_0



Mock data with proportional sampling is used
True R_0 is recovered



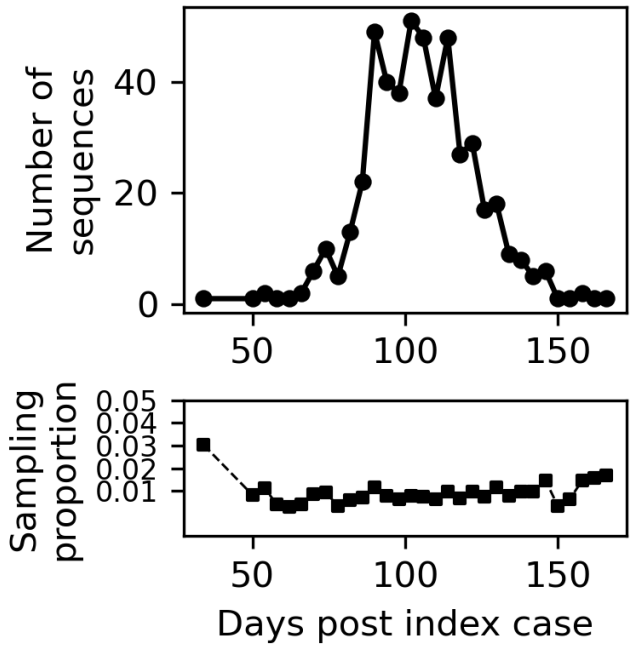


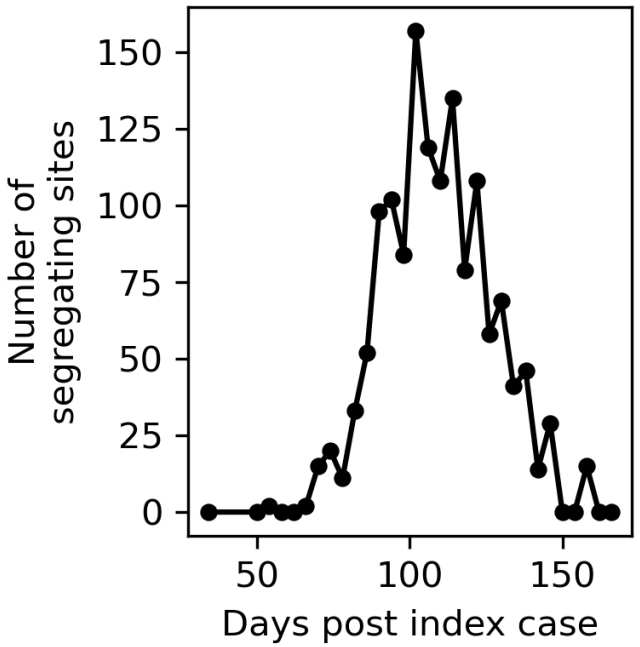


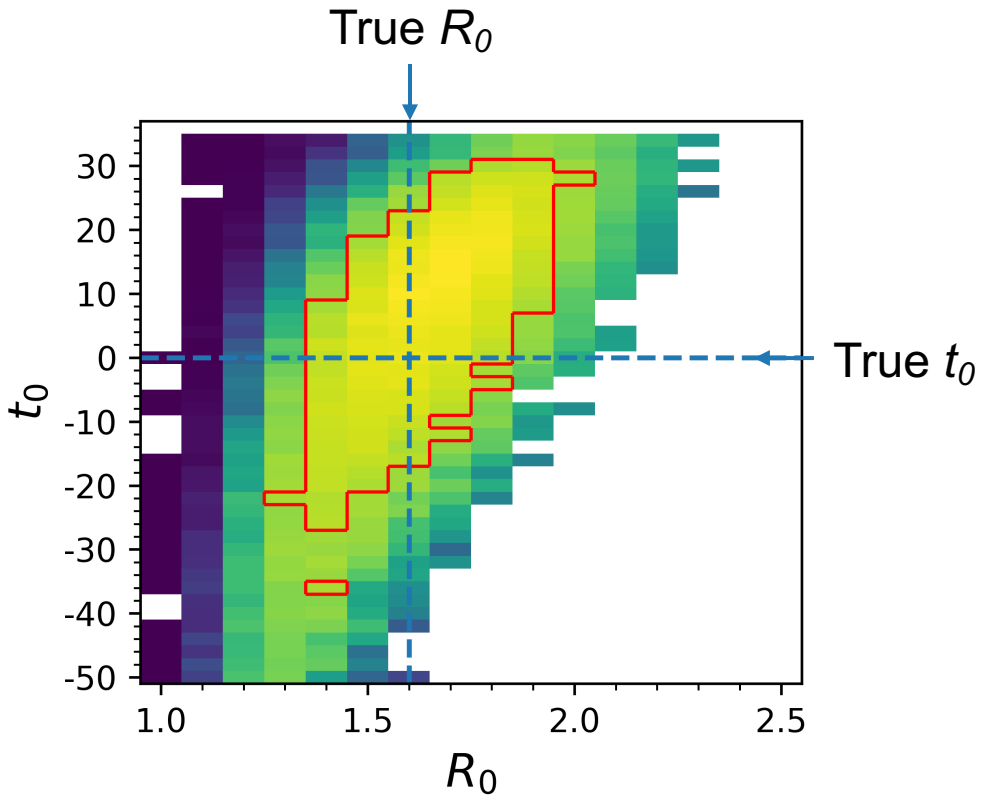
Validation using simulated data:

Joint estimation of R_0 and timing of index case

Mock data with proportional sampling is used
True R_0 and t_0 is recovered





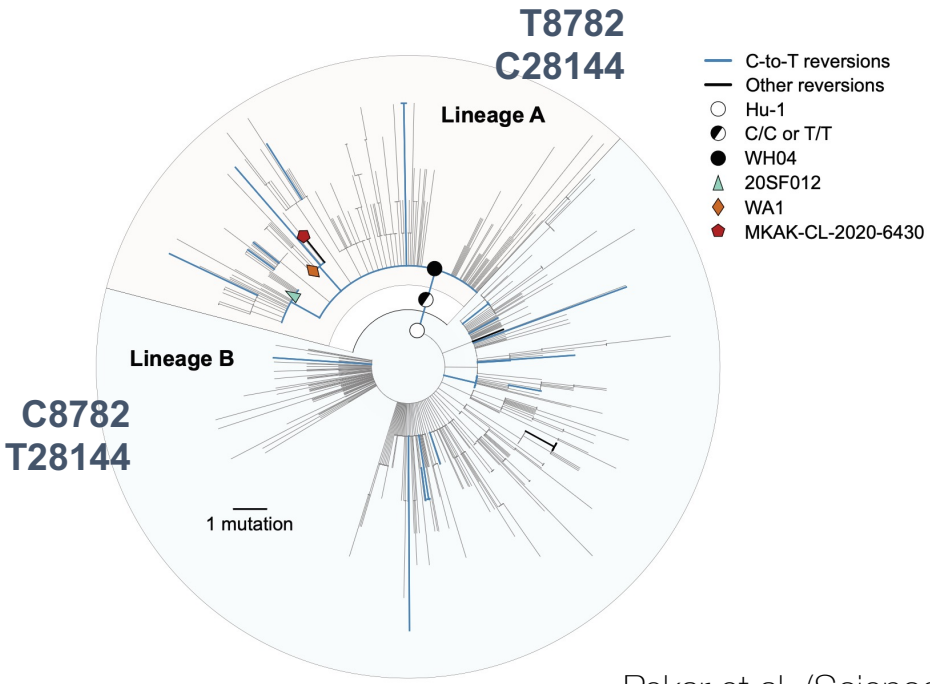


In-progress work:

Statistical evaluation of hypotheses regarding transmission dynamics

with application to early Wuhan

Two SARS-CoV-2 lineage in early Wuhan



Pekar et al. (Science 2022)

CORONAVIRUS

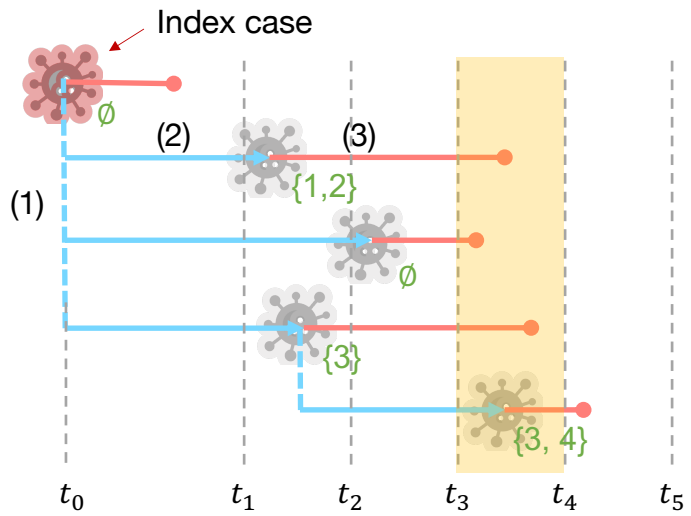
The molecular epidemiology of multiple zoonotic origins of SARS-CoV-2

Jonathan E. Pekar[1,2]*, Andrew Magee[3], Edyth Parker[4], Niema Moshiri[5], Katherine Izhikevich[5,6], Jennifer L. Havens[1], Karthik Gangavarapu[3], Lorena Mariana Malpica Serrano[7], Alexander Crits-Christoph[8], Nathaniel L. Matteson[4], Mark Zeller[4], Joshua I. Levy[4], Jade C. Wang[9], Scott Hughes[9], Jungmin Lee[10], Heedo Park[10,11], Man-Seong Park[10,11], Katherine Ching Zi Yan[12], Raymond Tzer Pin Lin[12], Mohd Noor Mat Isa[13], Yusuf Muhammad Noor[13], Tetyana I. Vasylyeva[14], Robert F. Garry[15,16,17], Edward C. Holmes[18], Andrew Rambaut[19], Marc A. Suchard[3,20,21]*, Kristian G. Andersen[4,22]*, Michael Worobey[7]*, Joel O. Wertheim[14]*

Model selection:

Single vs. multiple introduction hypotheses

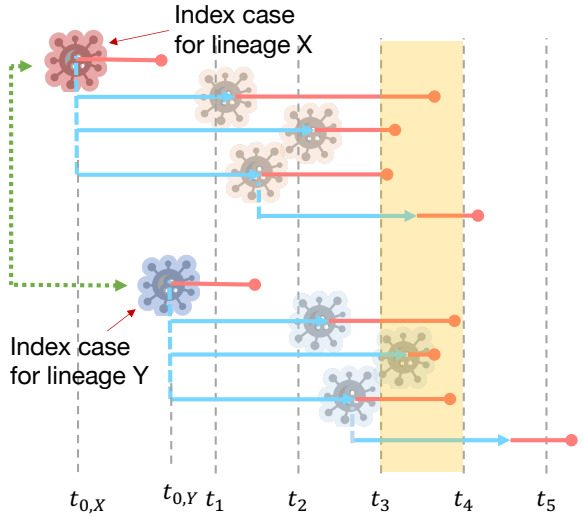
Single-introduction



Compare
AIC

$$\mathrm{AIC} = 2k - 2\ln(\hat{L})$$

Multiple-introduction

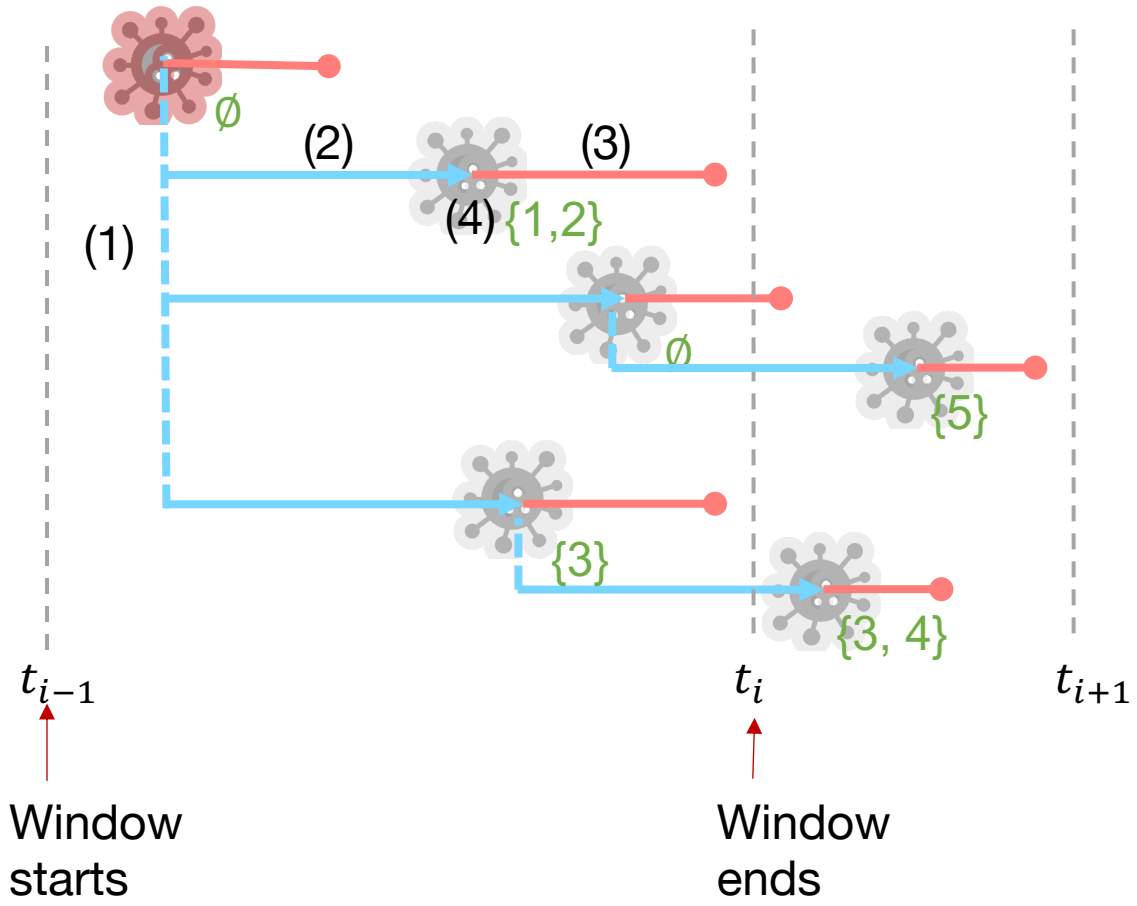


Estimation of t_0 and R_0 for the ancestor lineage

Joint estimation of $t_{0,A}$, $t_{0,B}$ and R_0 for the lineages A and B and n_{diff}

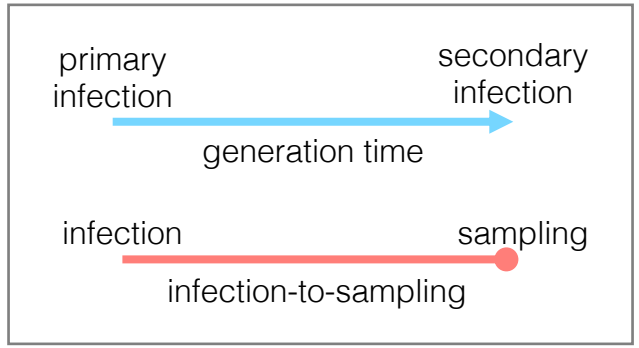
Single-introduction model:

Simulating dynamics using generation time



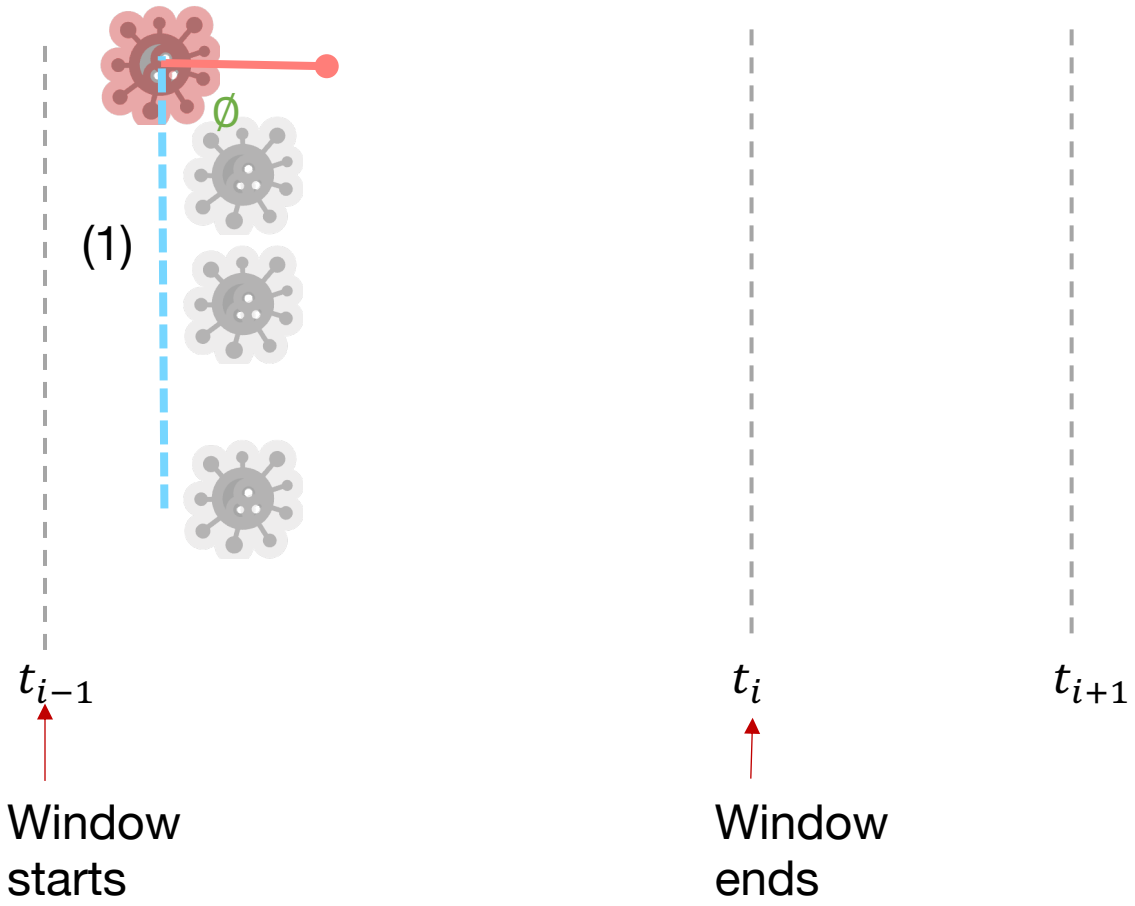
1) Number of secondary infections
2)
3)
4)

Repeated until every individual infected in a window reproduces



Single-introduction model:

Simulating dynamics using generation time

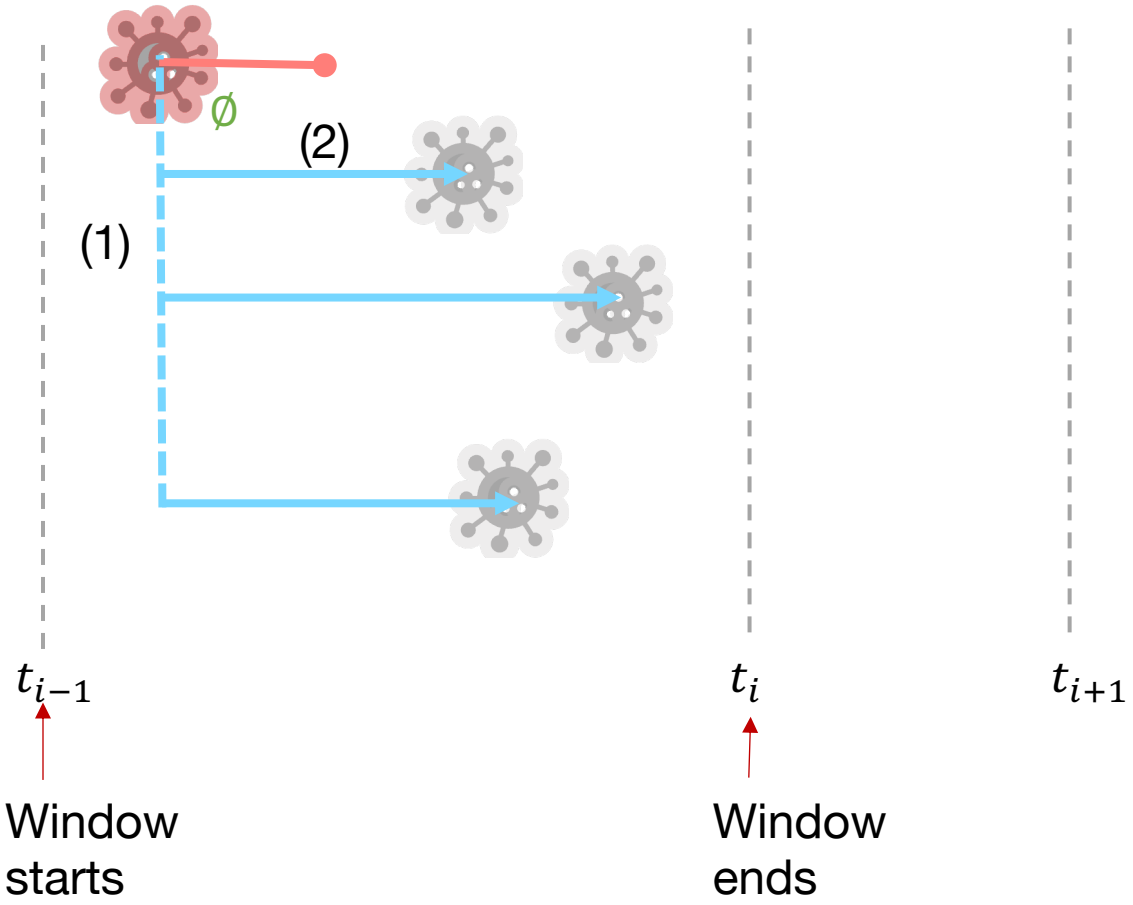


1) Number of secondary infections

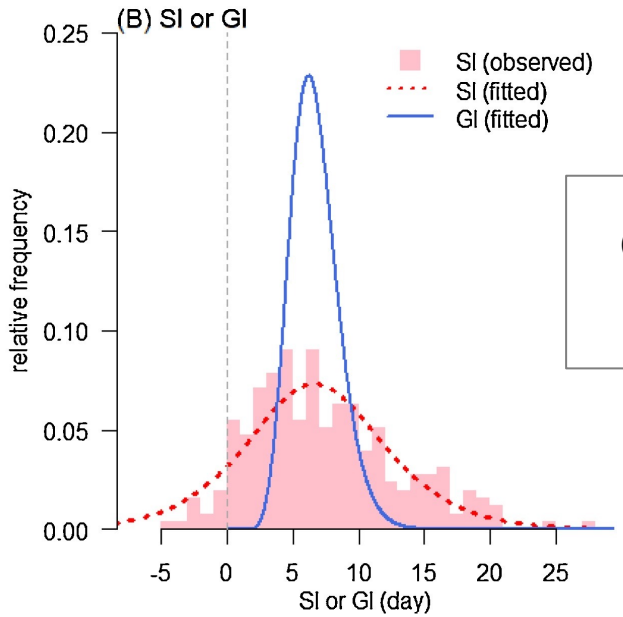
Negative binomial distribution parameterized based on R_e, k

Single-introduction model:

Simulating dynamics using generation time



1) Number of secondary infections
2) Timing of secondary infections

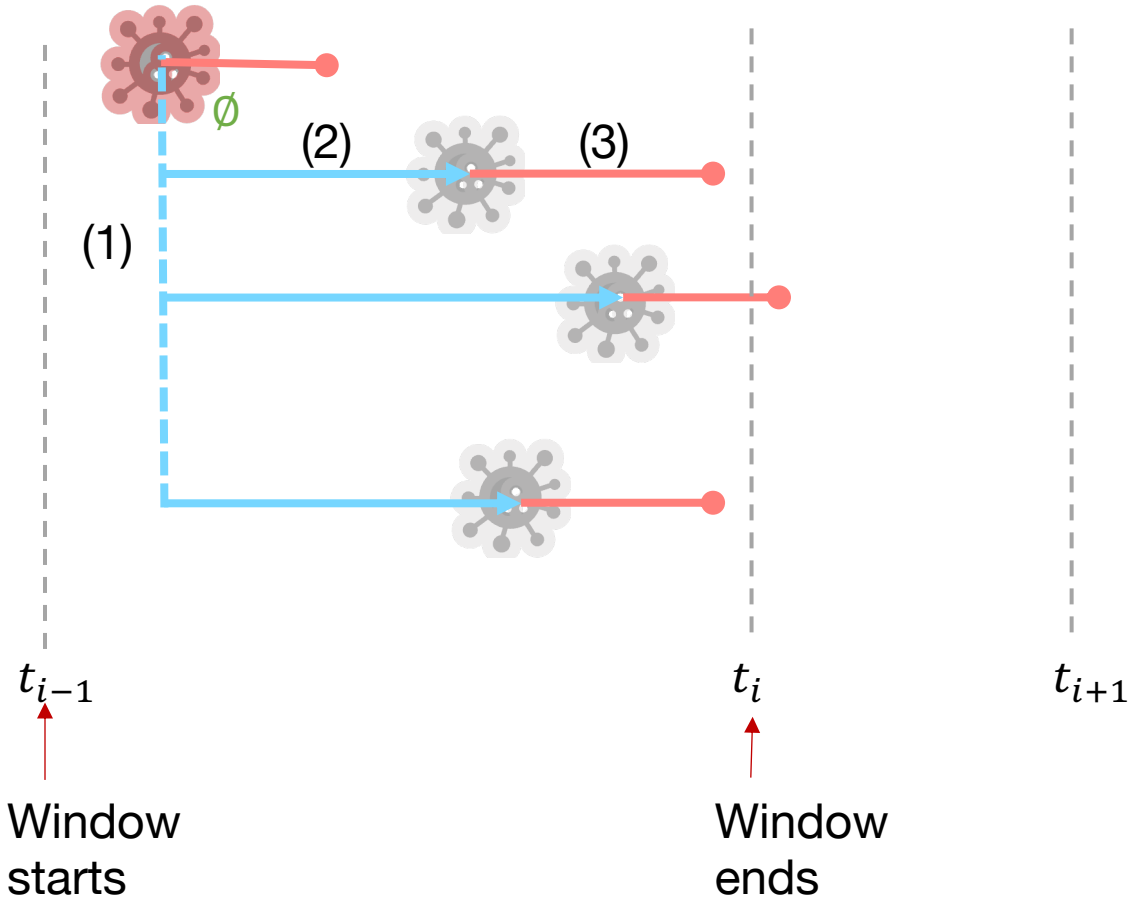


Gamma distribution
($\mu = 6.7, \sigma = 1.8$)

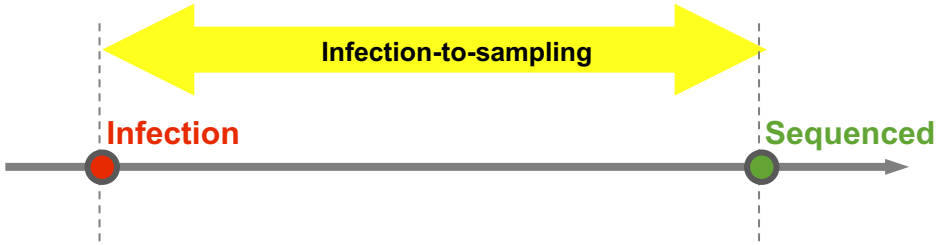
Zhao et al. (Epidemics. 2021)

Single-introduction model:

Simulating dynamics using generation time

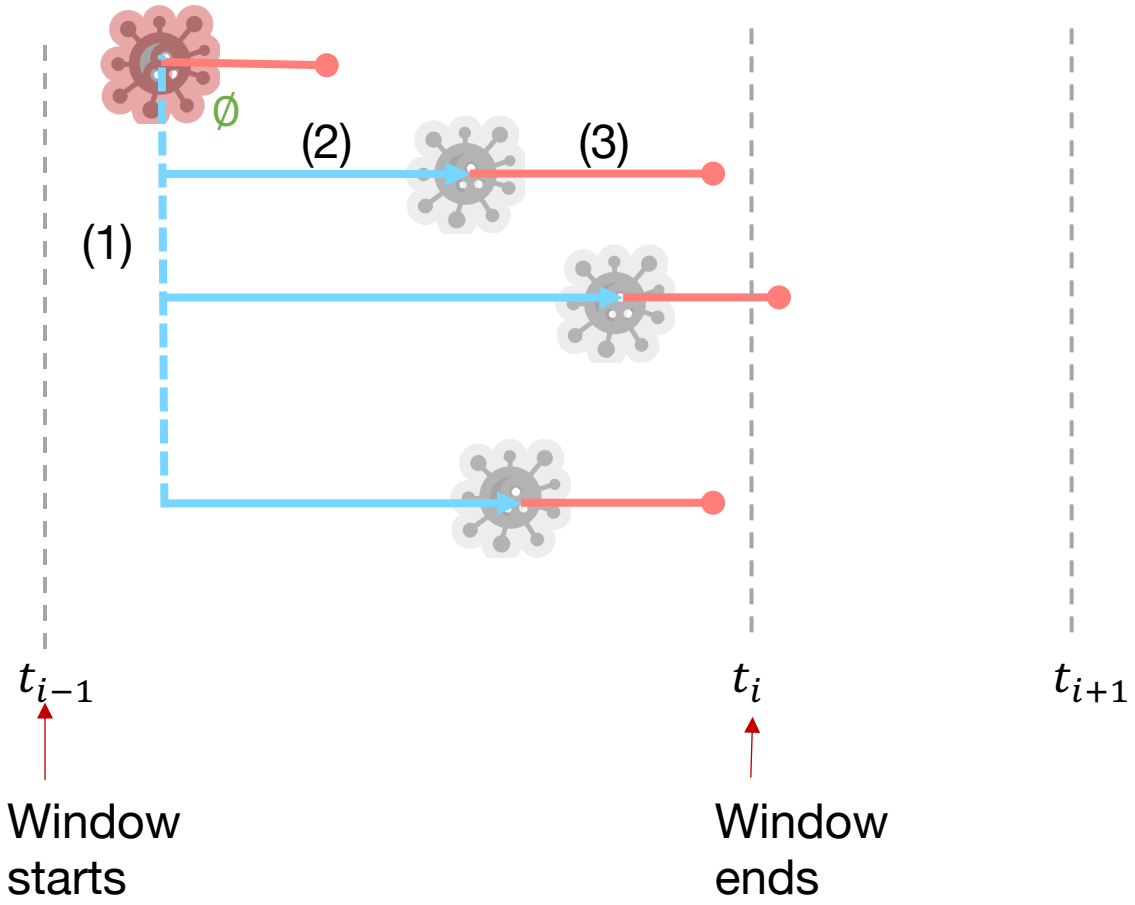


1) Number of secondary infections
2) Timing of secondary infections
3) Sampling time of secondary infections

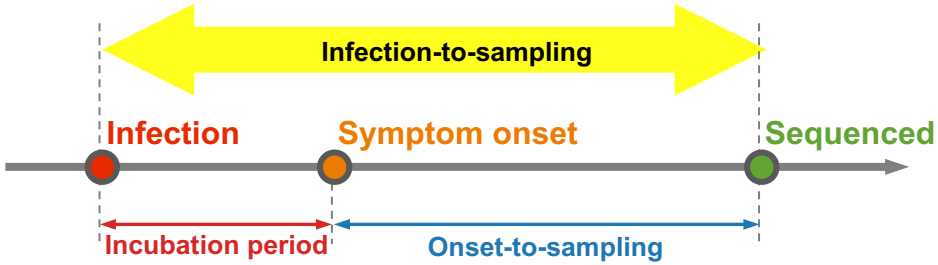


Single-introduction model:

Simulating dynamics using generation time

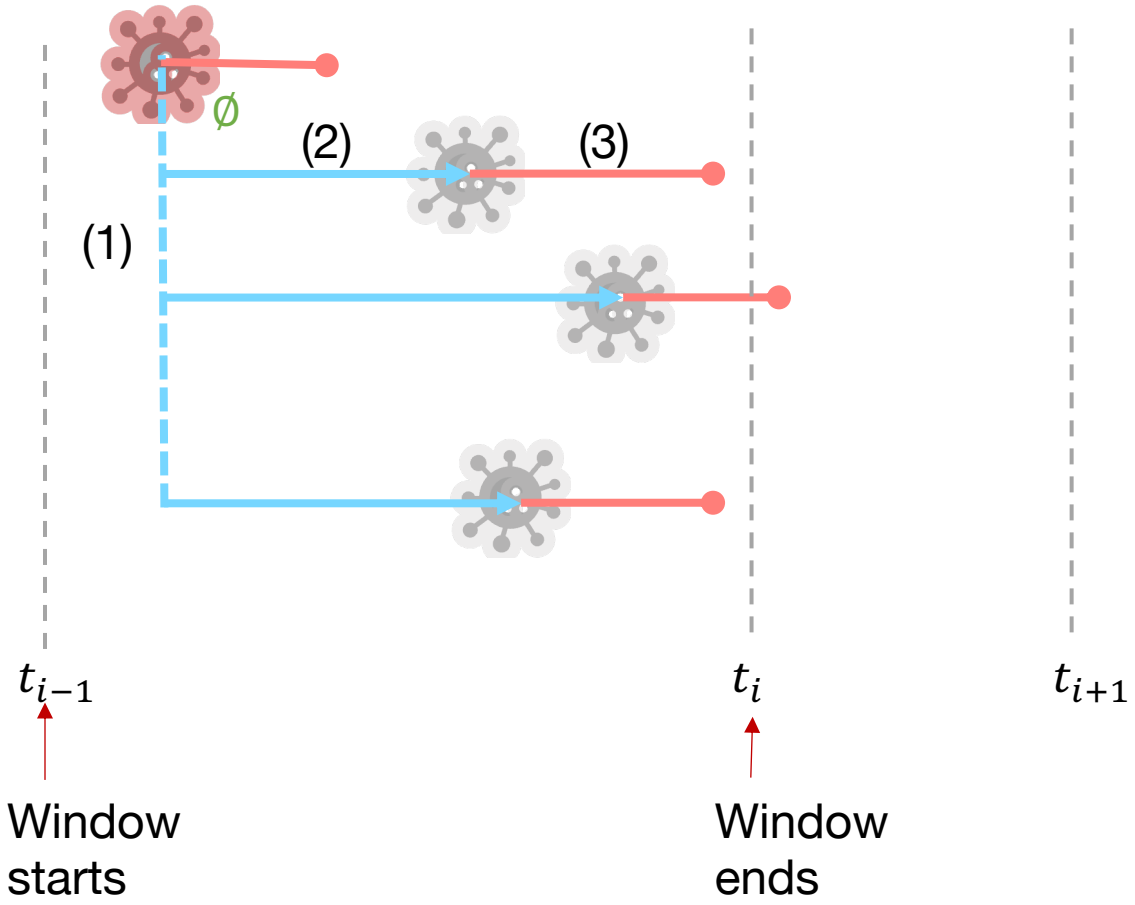


1) Number of secondary infections
2) Timing of secondary infections
3) Sampling time of secondary infections

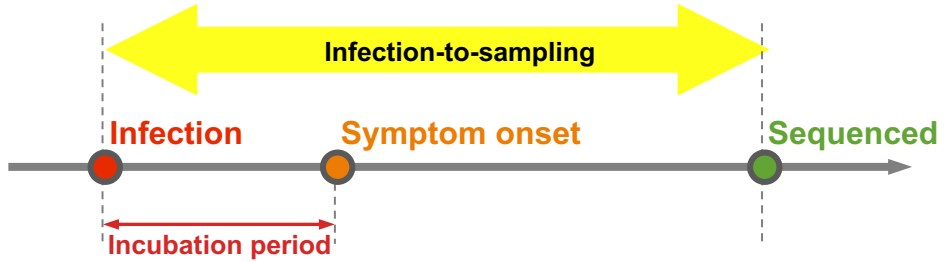


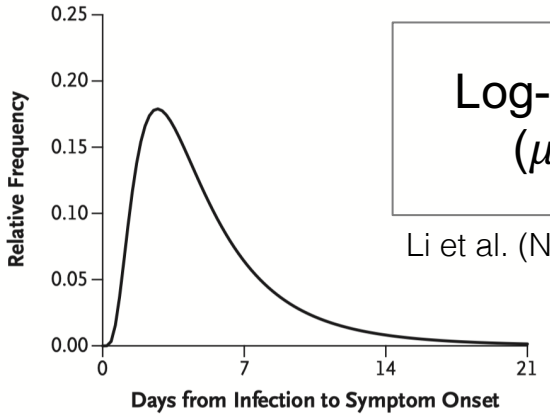
Single-introduction model:

Simulating dynamics using generation time



1) Number of secondary infections
2) Timing of secondary infections
3) Sampling time of secondary infections



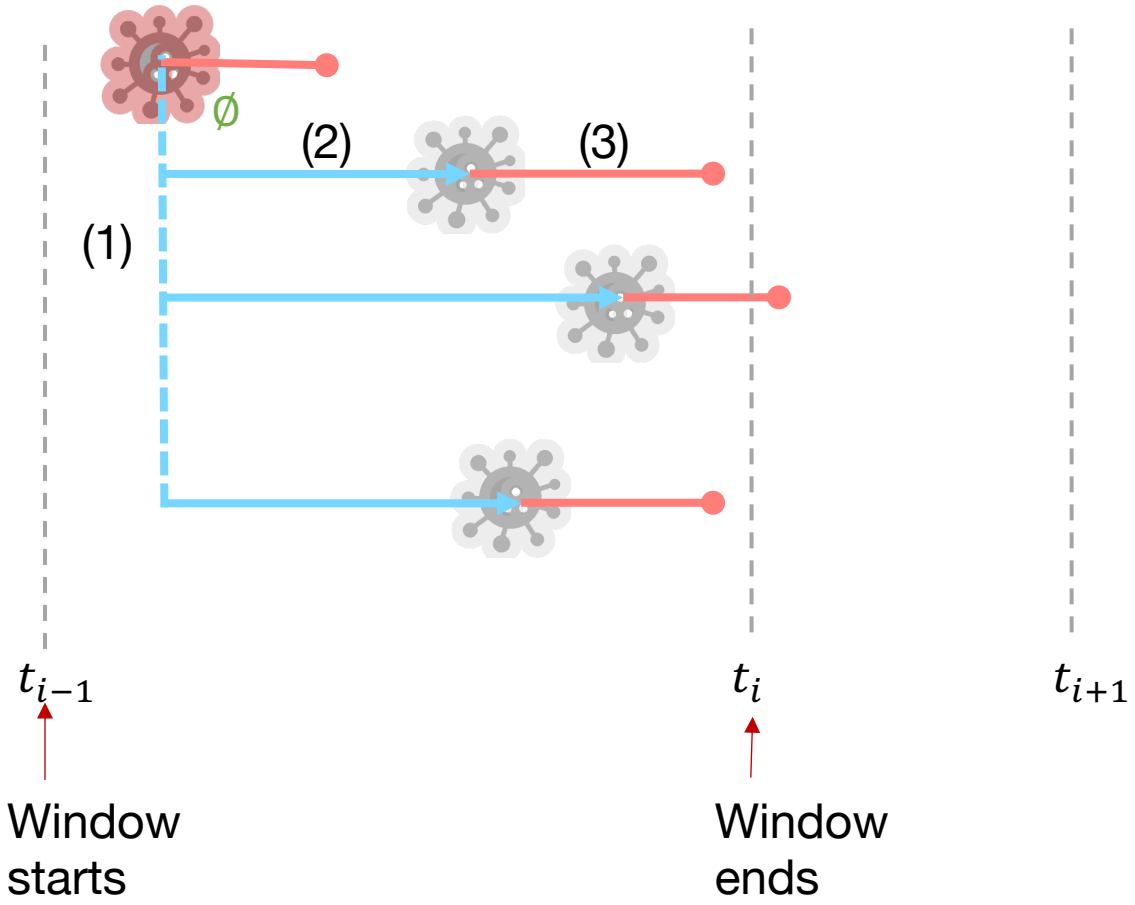


Log-normal distribution
$(\mu = 5.2, \sigma = 3.9)$

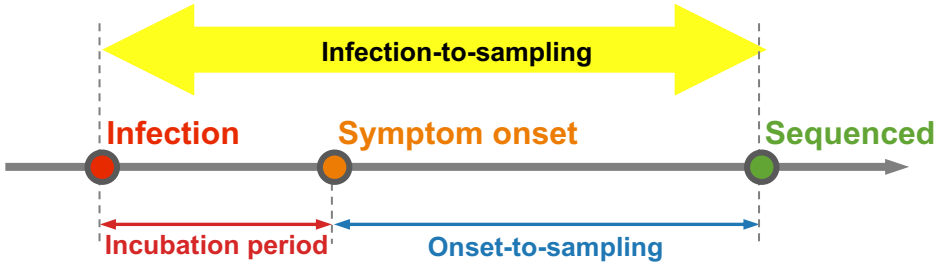
Li et al. (N Engl J Med. 2020)

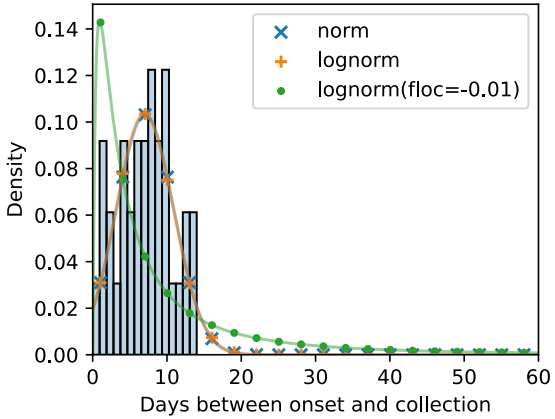
Single-introduction model:

Simulating dynamics using generation time



1) Number of secondary infections
2) Timing of secondary infections
3) Sampling time of secondary infections



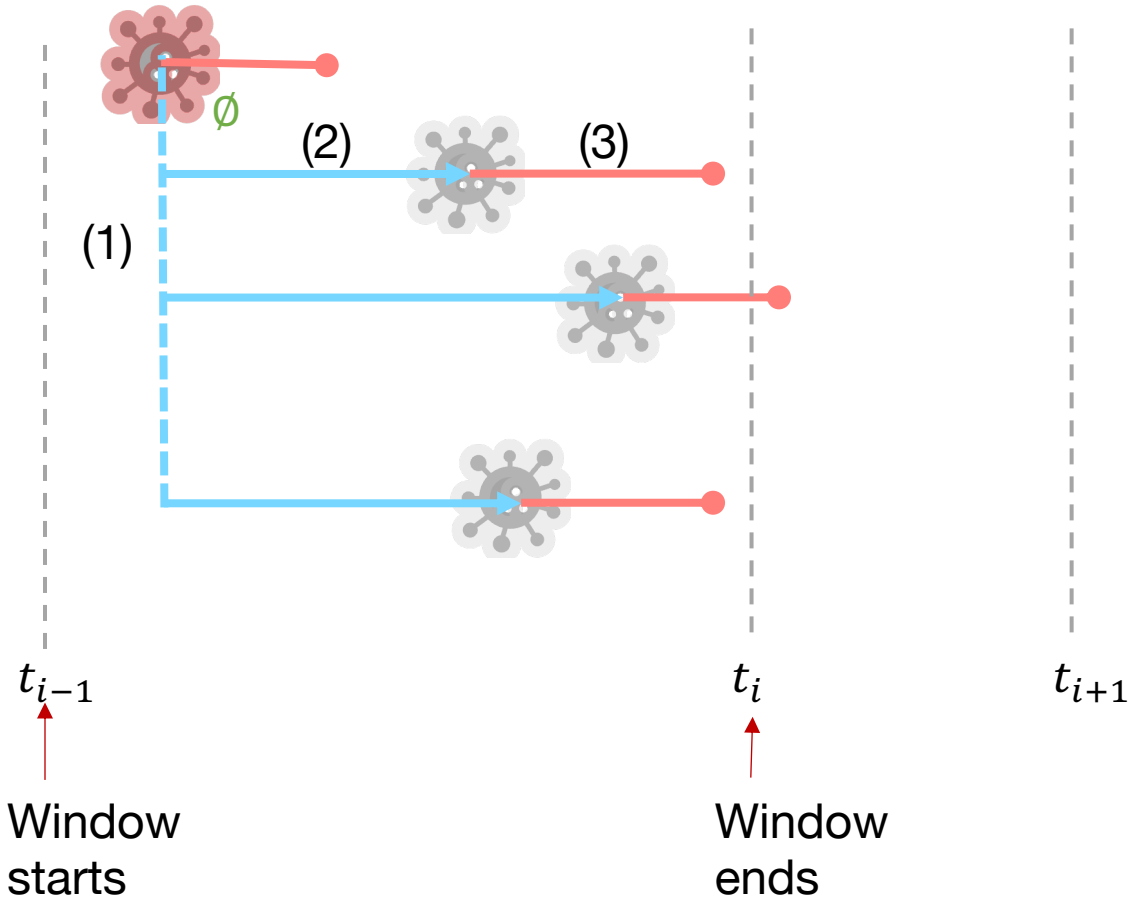


Obtained by matching

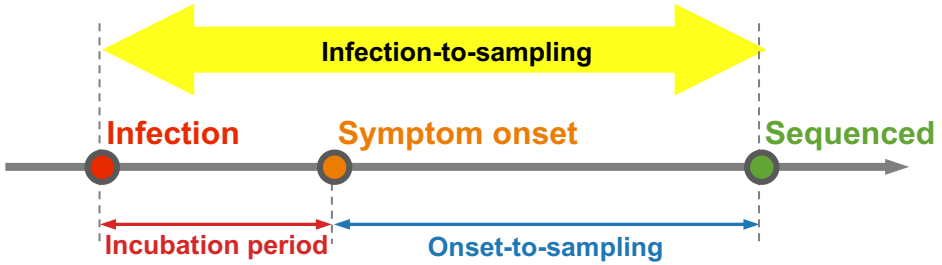
- symptom onset date from the literature
- sampling dates of sequences

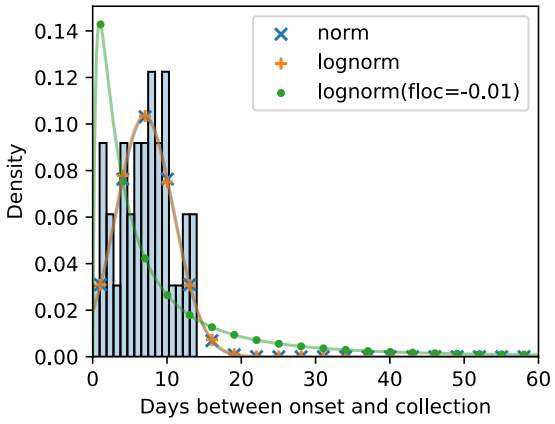
Single-introduction model:

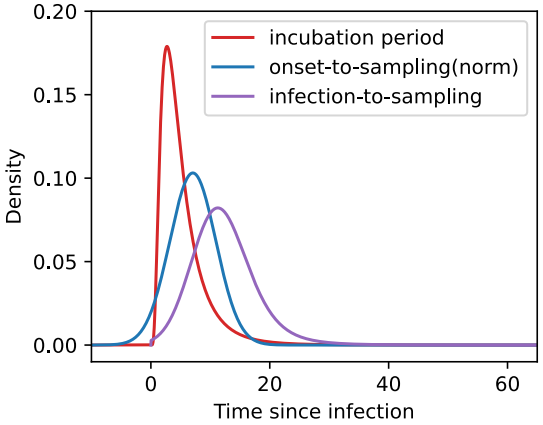
Simulating dynamics using generation time



1) Number of secondary infections
2) Timing of secondary infections
3) Sampling time of secondary infections

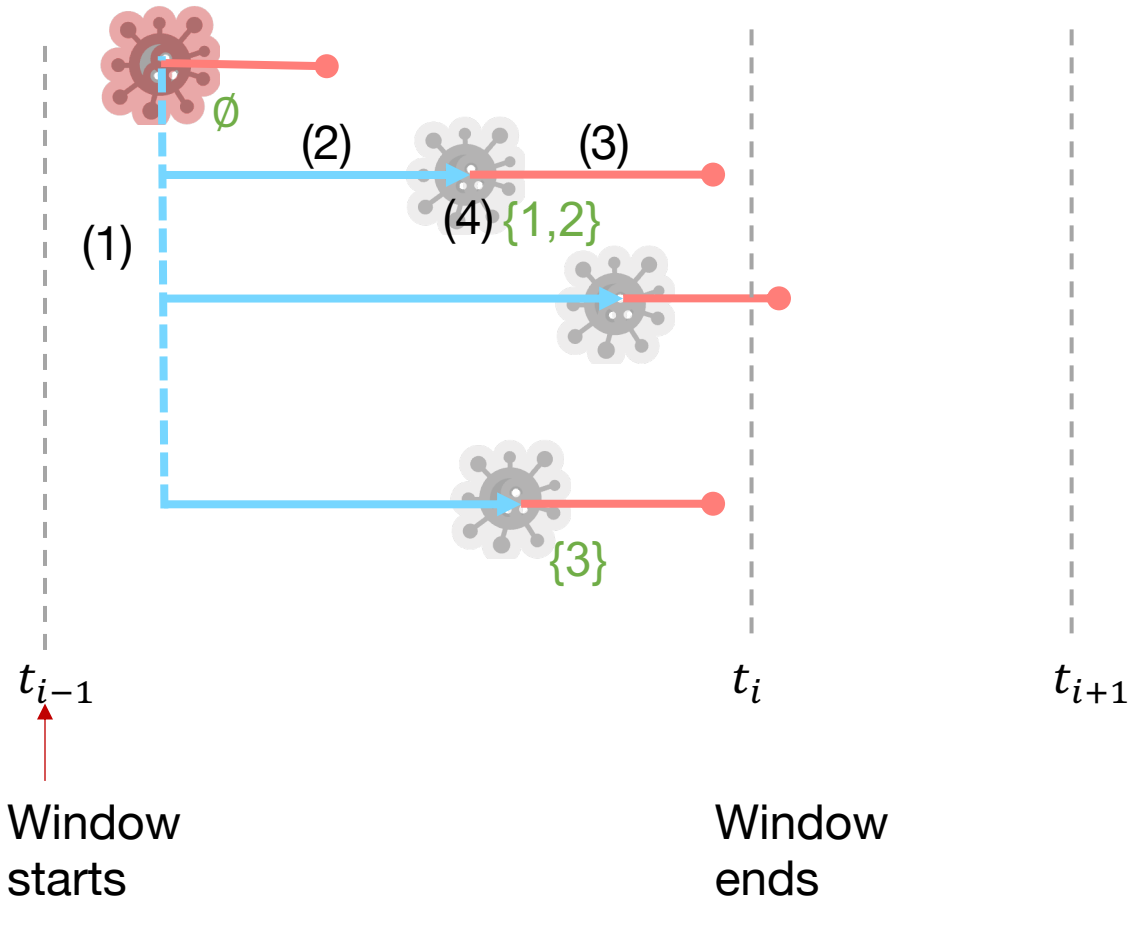






Single-introduction model:

Simulating dynamics using generation time

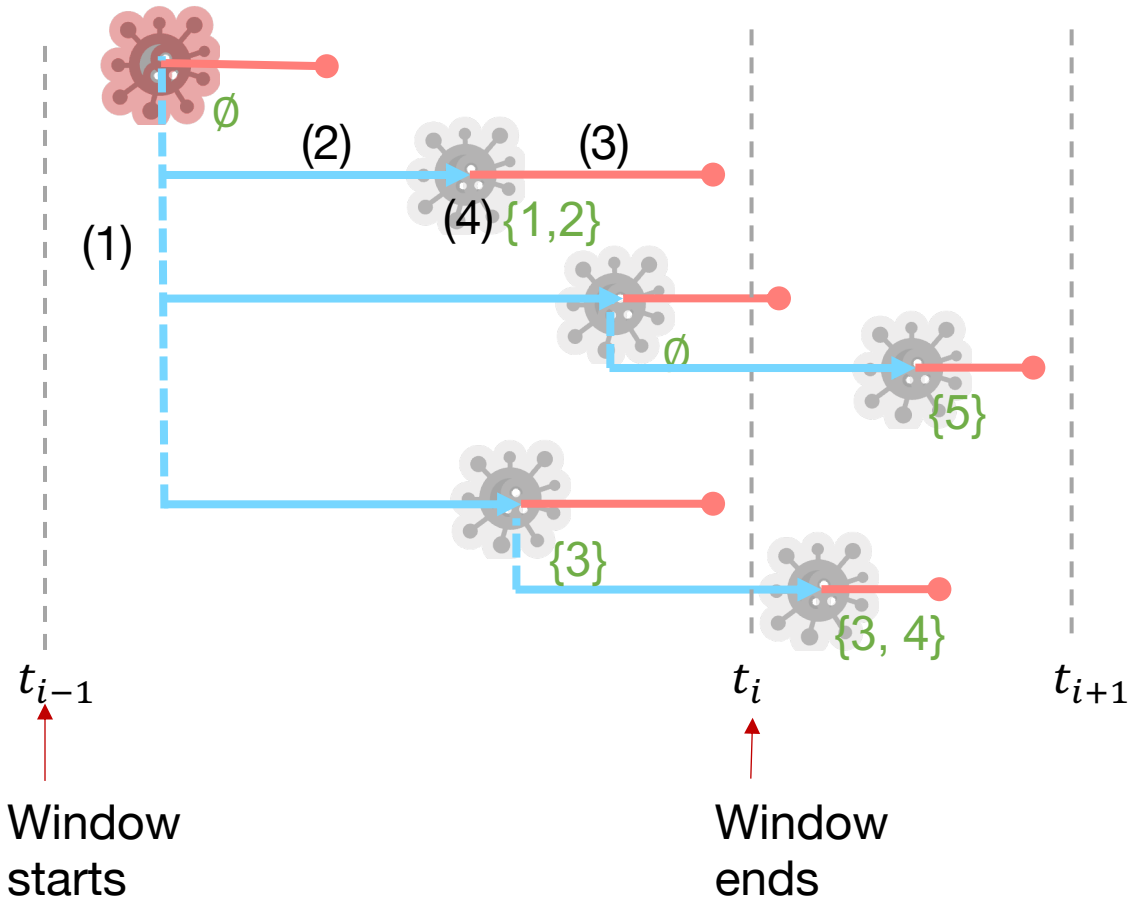


1) Number of secondary infections
2) Timing of secondary infections
3) Sampling time of secondary infections
4) Number of mutations

Poisson distribution with mutation probability per transmission μ

Single-introduction model:

Simulating dynamics using generation time

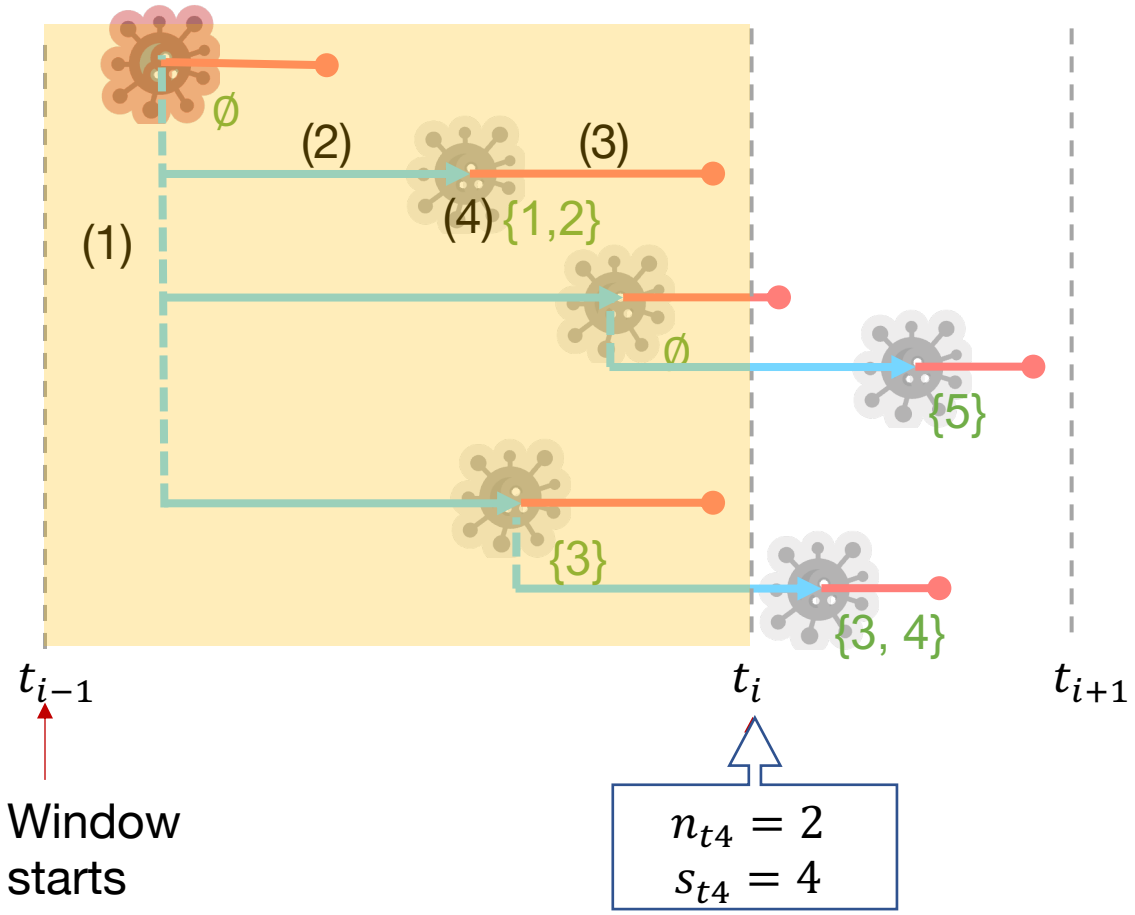


1) Number of secondary infections
2) Timing of secondary infections
3) Sampling time of secondary infections
4) Number of mutations

Repeated until every individual infected in a window reproduces

Single-introduction model:

Observing dynamics as segregating sites

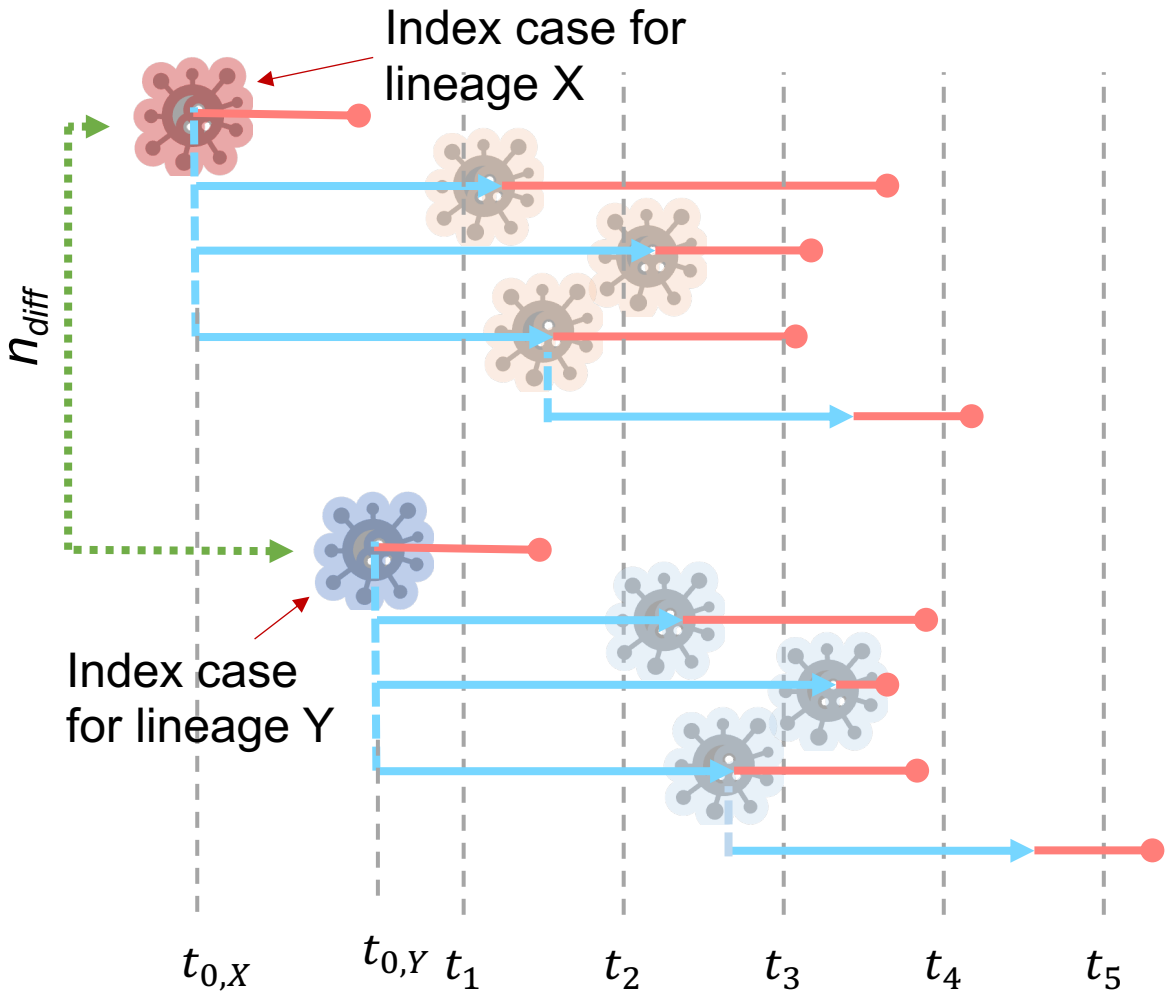


For k grabs

- Sample n_i individuals from candidates
- Count the number of segregating sites

Multiple-introduction model:

Multiple-introduction model with two lineages

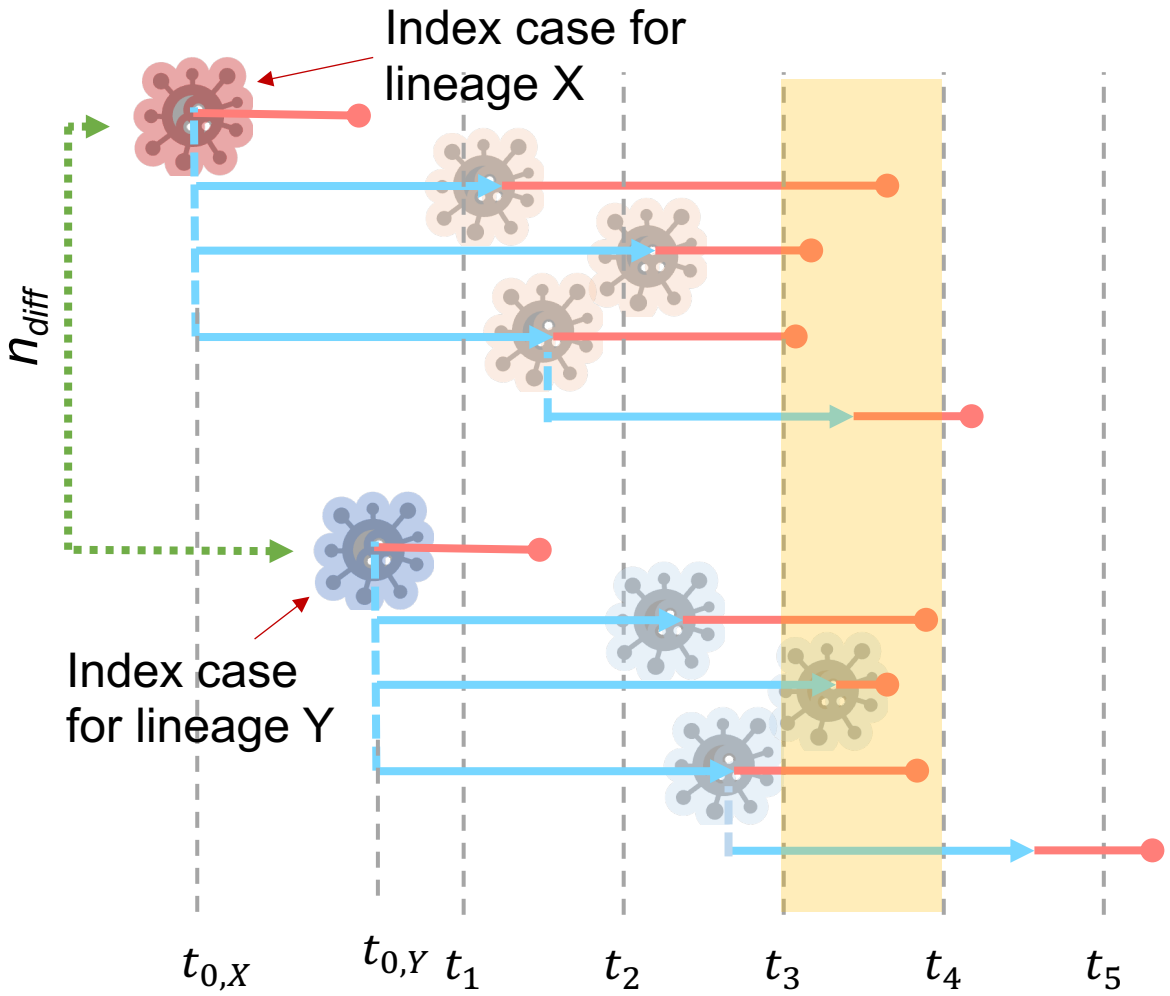


Assumptions for lineages

- Two lineages start with their own index case
- Two lineages are not interacting with each other
- Two index cases have nucleotide difference of n_{diff}, which is a new parameter
- Mutation in each lineages occur in different sites

Multiple-introduction model:

Multiple-introduction model with two lineages



Counting the segregating sites

- Individuals are sampled from each lineage
- Number of segregating sites is obtained and the nucleotide difference between index cases are summed

{1}

{1, 3}

{1'}

{2'}

For t_4,

$n_{t4,X} = 2$

$n_{t4,Y} = 2$

$s_{t4} = 4$

$$s_{t4}^{sim} = 4 + n_{diff}$$

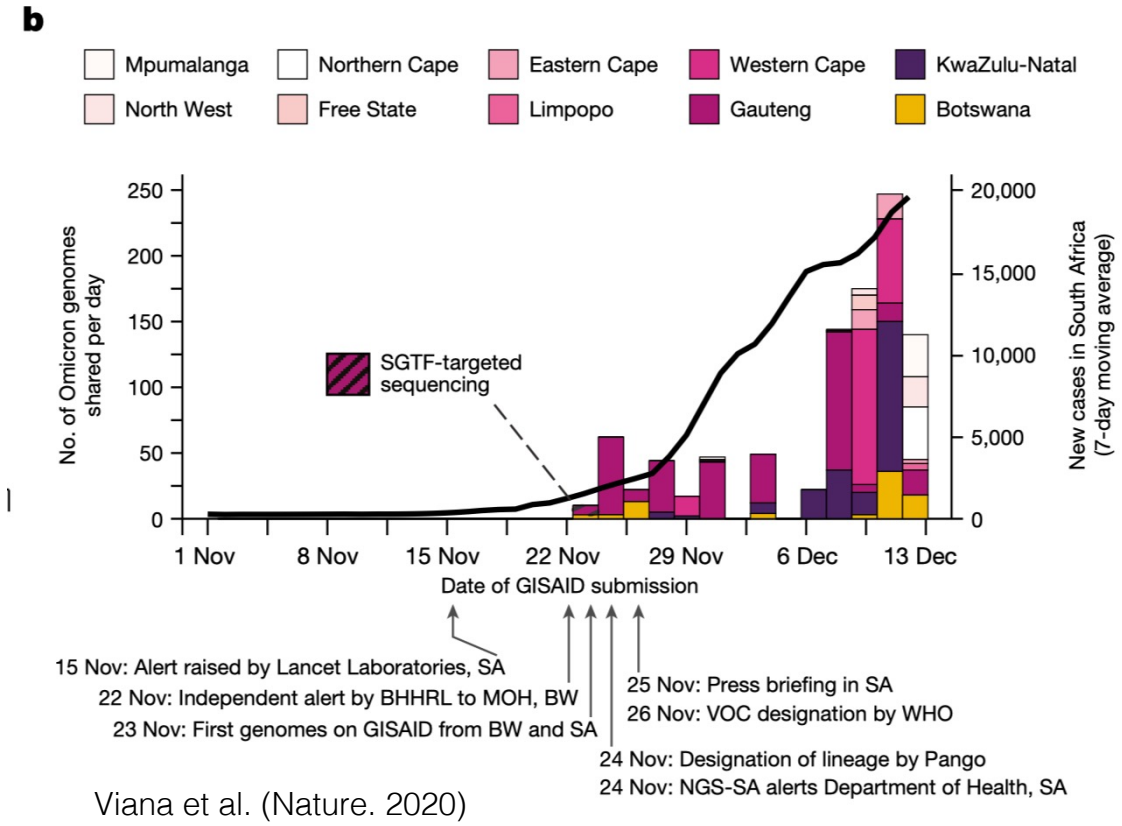
In-progress work:

Data availability during the early phase and inference reliability

with application to Omicron variant in South Africa

Omicron BA.1 variants in South Africa

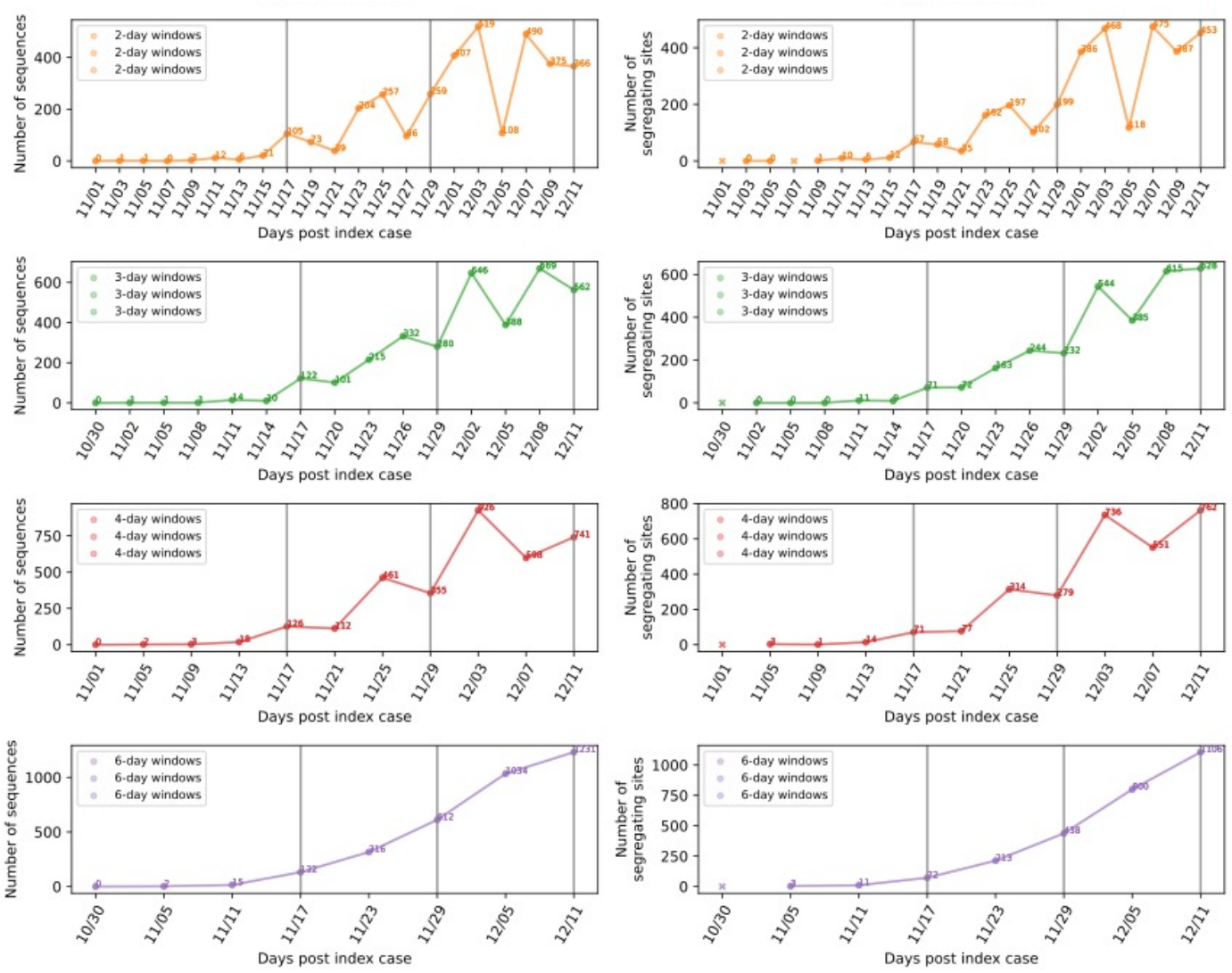
- For more recently emerged VOCs, available sequences have accumulated rapidly.
- However, the temporal signal might not be sufficient



Viana et al. (Nature. 2020)

Snapshots of segregating sites

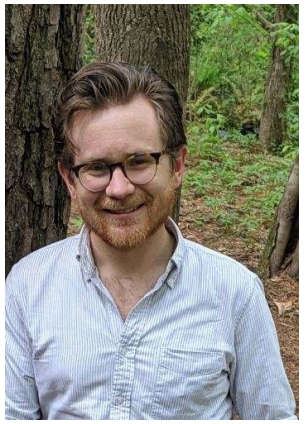
- Sequences sampled until different time points
- Using each snapshot for parameter estimation
- Compare the point estimates and interval estimates from each snapshot



Conclusion

- Viral genome sequences contains information regarding epidemiological and evolutionary dynamics and can be used to infer the epidemiological dynamics
- When genetic diversity is low, inference using segregating site trajectory could be a good complement for tree-based inference.
- Segregating site-based approach can be also used for statistical evaluation of hypotheses or model selection.

EMORY | LANEY GRADUATE SCHOOL

Mike Martin 

Katia Koelle