A little phylogenetic study on Ebola virus in the 2014 Sierra Leone outbreak

Shofi Andari

Spring 2019 EEOB 563 - Final project

Outline

- EBOV, the 2014 outbreak
- How is molecular phylogeny useful?
- Methods: birth-death Bayesian model
- Results: effective reproduction ratios, origin time estimate, the tree!
- Conclusion
- What can we do more?

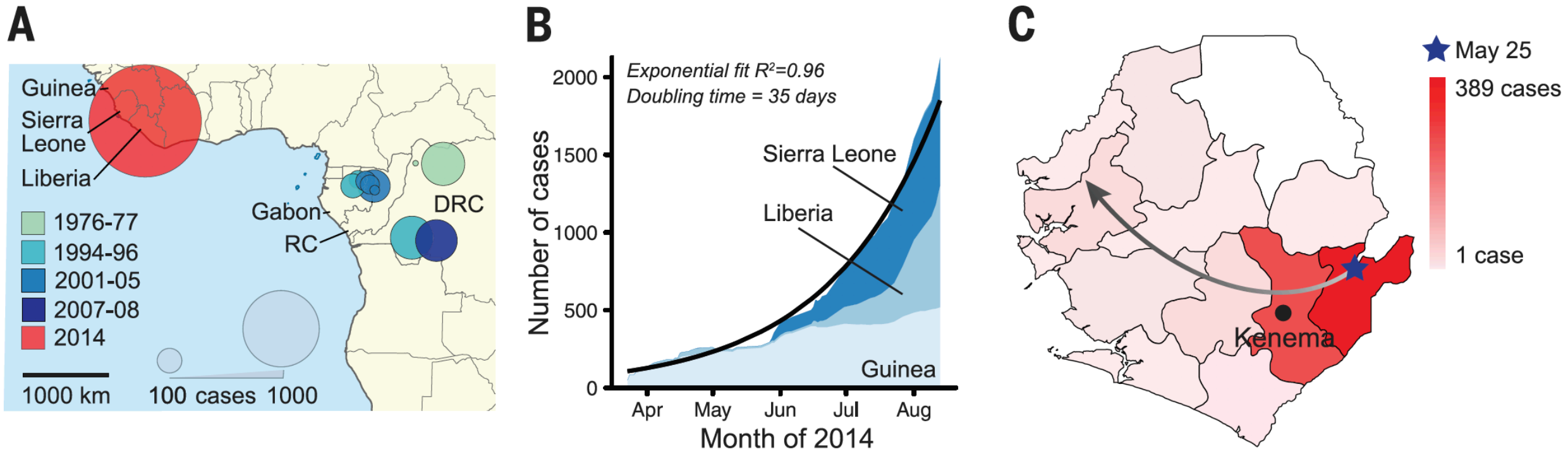


Figure 1. The Ebola virus disease (EVD) outbreak in Sierra Leone during 2014 was one of the most devastating epidemic.
Source: Gire et.al. (2014)

- Historical EVD outbreaks, colored by decade.
- Circle area represents total number of cases.

(**RC** =Republic of the Congo; **DRC** = Democratic Republic of Congo)

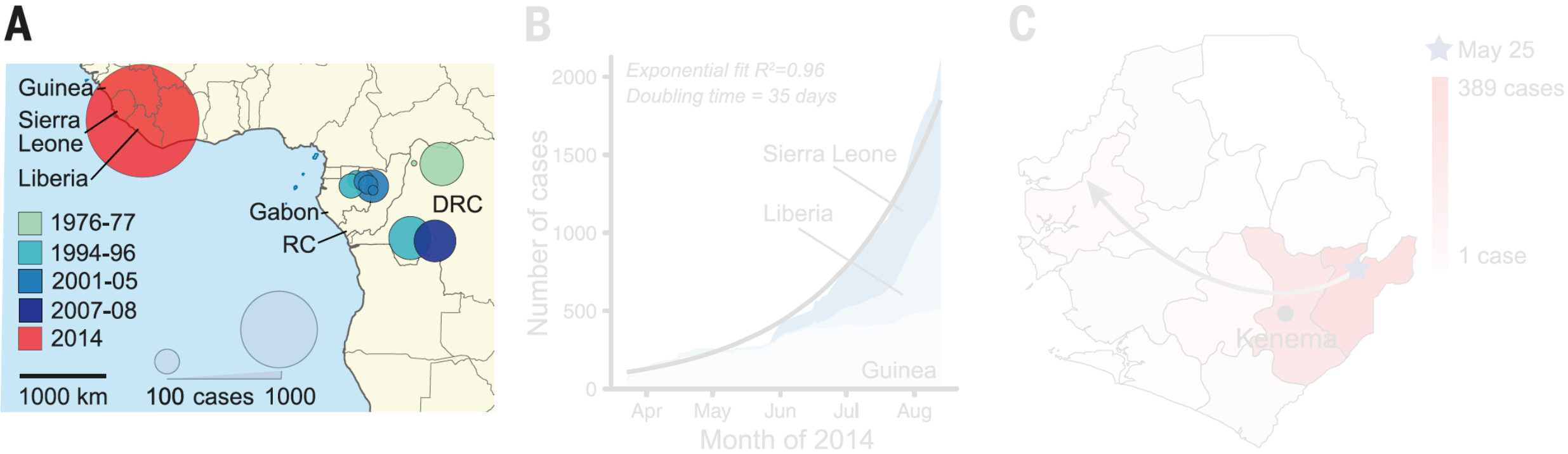


Figure 1. The Ebola virus disease (EVD) outbreak in Sierra Leone during 2014 was one of the most devastating epidemic.
Source: Gire et.al. (2014)

- 2014 outbreak growth was expanding exponentially
- It began to spread in Sierra Leone in May 2014

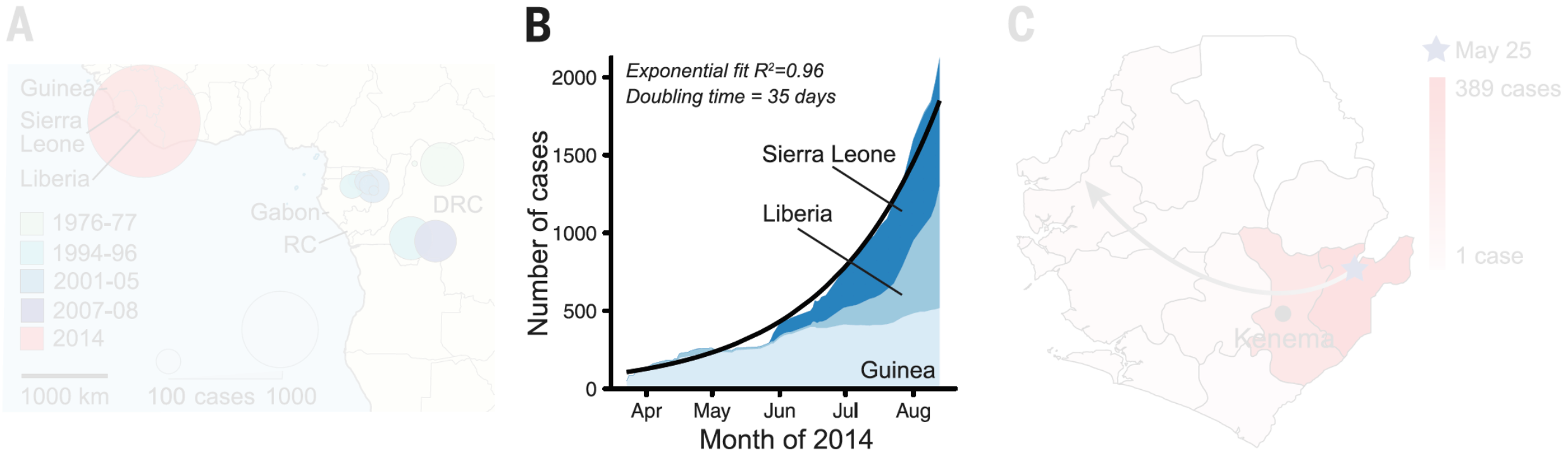


Figure 1. The Ebola virus disease (EVD) outbreak in Sierra Leone during 2014 was one of the most devastating epidemic.
Source: Gire et.al. (2014)

About the outbreak

- In March 2014, Kenema Government Hospital (KGH) established EBOV surveillance in Kenema, Sierra Leone, near the origin of the 2014 outbreak. They performed conventional polymerase chain reaction (PCR)–based EBOV diagnostics.
- On 25 May, KGH scientists confirmed the first case of EVD in Sierra Leone

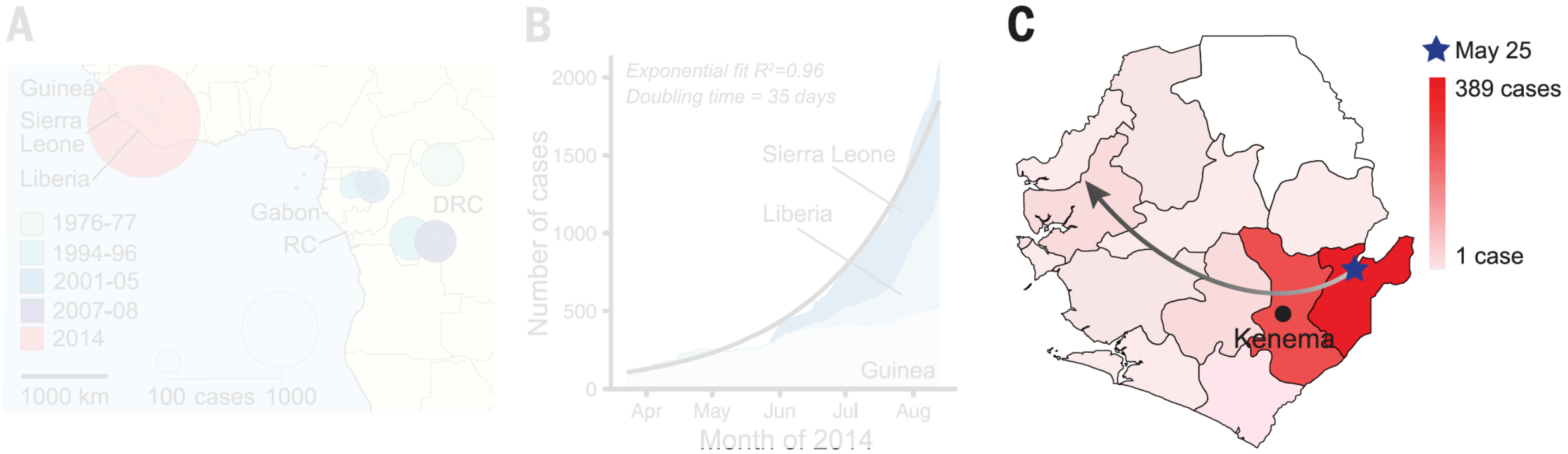


Figure 1. The Ebola virus disease (EVD) outbreak in Sierra Leone during 2014 was one of the most devastating epidemic.
Source: Gire et.al. (2014)

How is molecular phylogeny useful?

- Modeling the phylogeny of EBOV spread is one attempt to **quantify epidemiological dynamics**.

- It is crucial to help us **understanding and forecasting the spread** of an outbreak.

How the data look like?

- Gire et.al. (2014): 78 sequences of EBOV from 78 patients.
- Stadler et.al (2014): only used **72 individuals** (as taxa) dated from **May 26, 2014 to June 18, 2014**.
- The alignment provided in a nexus format, length = 14517 nt.

Birth-death Bayesian model: why?

- Two approaches in phylodynamics Bayesian modeling [*Boskova et.al. (2014)*]:

Coalescent method

Birth-death method

Birth-death Bayesian model: why?

- Two approaches in phylodynamics Bayesian modeling [*Boskova et.al. (2014)*]:

Coalescent method

→ the time is modeled to go **backwards**, from current date (present) towards the origin (past)

→ <u>Assumptions</u>:

1. individuals from one generation give rise to the individuals in the next generation,
2. there exists sufficient genetic diversity within the population to allow reconstruction of the phylogenetic relationships,
3. the population size is large enough (compared to the sample taken), and
4. the population size is small enough to be able to trace back the MRCA.
5. **the population size changes deterministically**.

Birth-death method

Birth-death Bayesian model: why?

- Two approaches in phylodynamics Bayesian modeling [*Boskova et.al. (2014)*]:

Coalescent method

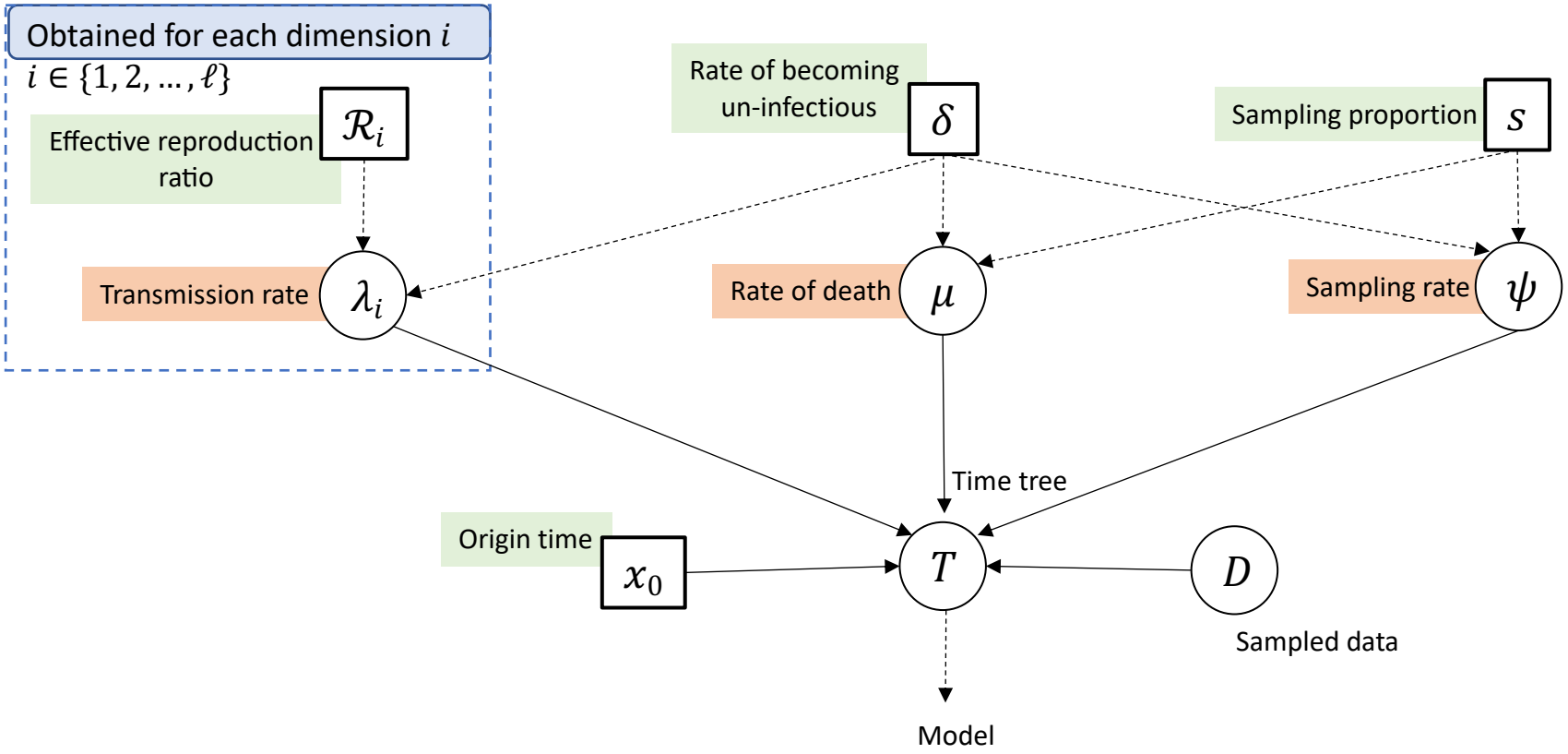
Birth-death method

→ The time is modeled to go **forward**, from the past towards present

→ Assumptions:

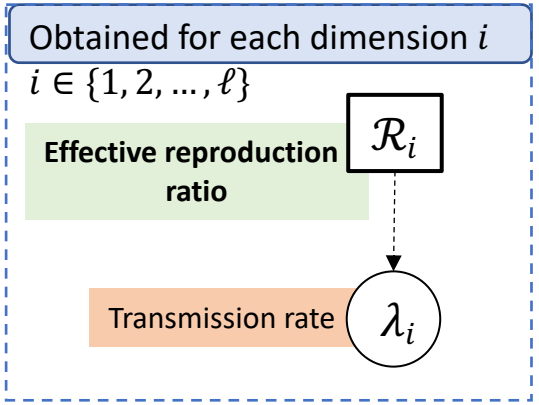
1. ~~individuals from one generation give rise to the individuals in the next generation,~~
2. there exists sufficient genetic diversity within the population to allow reconstruction of the phylogenetic relationships,
3. ~~the population size is large enough (compared to the sample taken), and~~
4. the population size is small enough to be able to trace back the MRCA.
5. **the population size changes ~~deterministically~~ stochastically**

Birth-death Bayesian model: the parameters



Source: Hohna et.al. (2014), Stadler et.al. (2014)

Birth-death Bayesian model



- **Effective reproduction ratio (number)**

→ estimates the average number of secondary cases per infectious case in a population made up of both susceptible and non-susceptible hosts

→ R_0, the basic reproduction number, measures the transmission potential of a disease

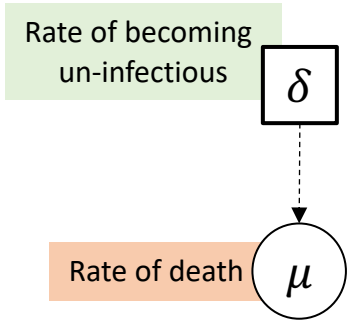
For instance: If the R_0 for EBOV is 5 in a population, then we would expect it to spread quite rapidly because each new case of EVD would produce 5 new secondary cases.

→ The subscript, *i*, is going through ℓ dimensions or intervals

Birth-death Bayesian model

- **Rate of becoming noninfectious**

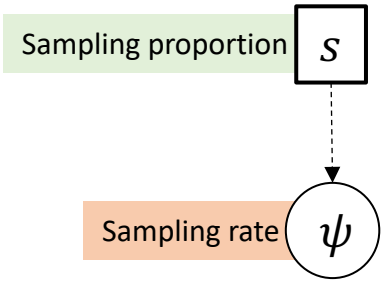
→ the average time a patient can transmit a disease



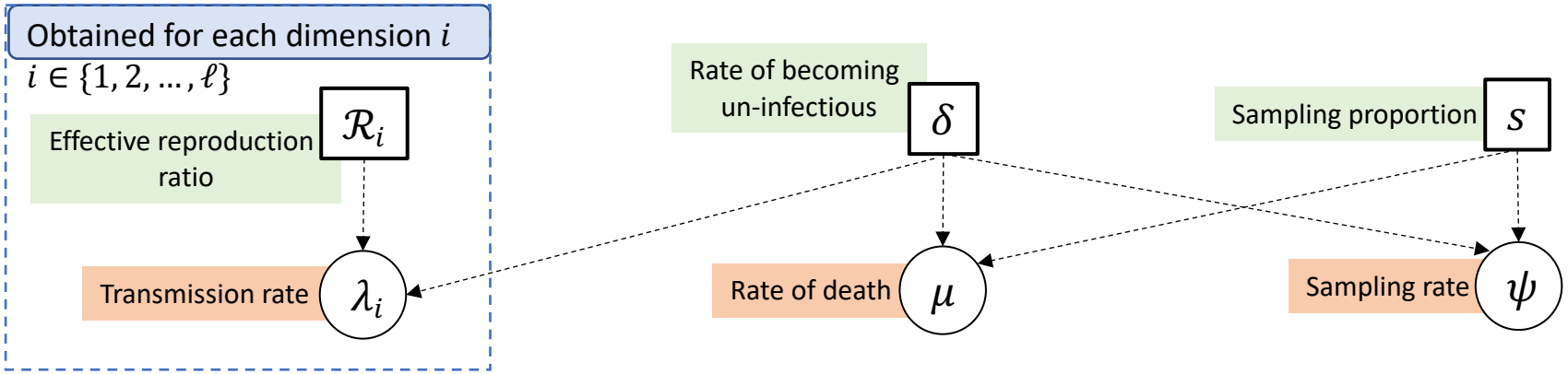
Birth-death Bayesian model

- **Sampling proportion (probability)**

→ probability of sampling an individual upon becoming non-infectious

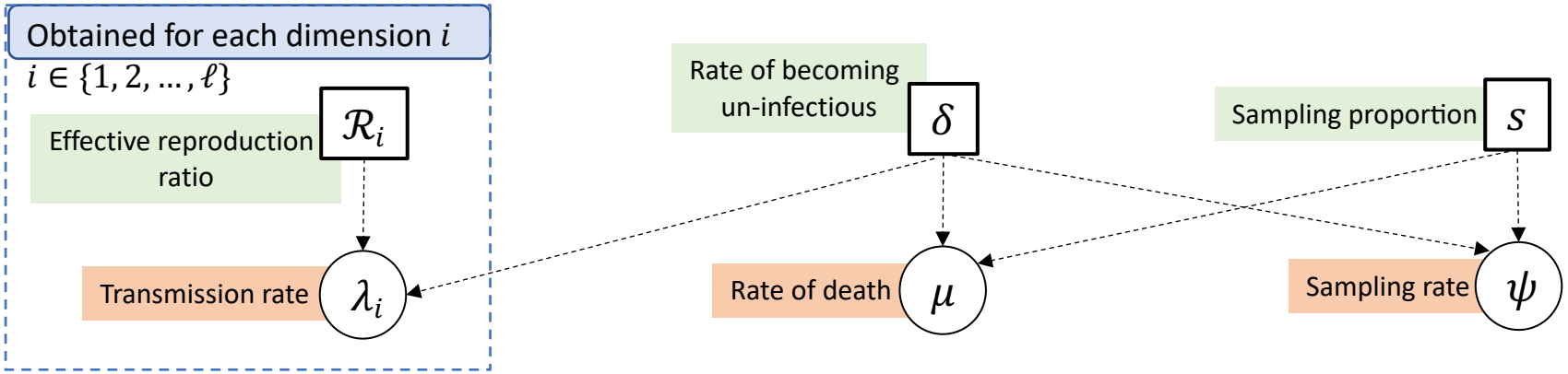


Birth-death Bayesian model: the parameters



$$\delta = \mu + \psi$$

Birth-death Bayesian model: the parameters

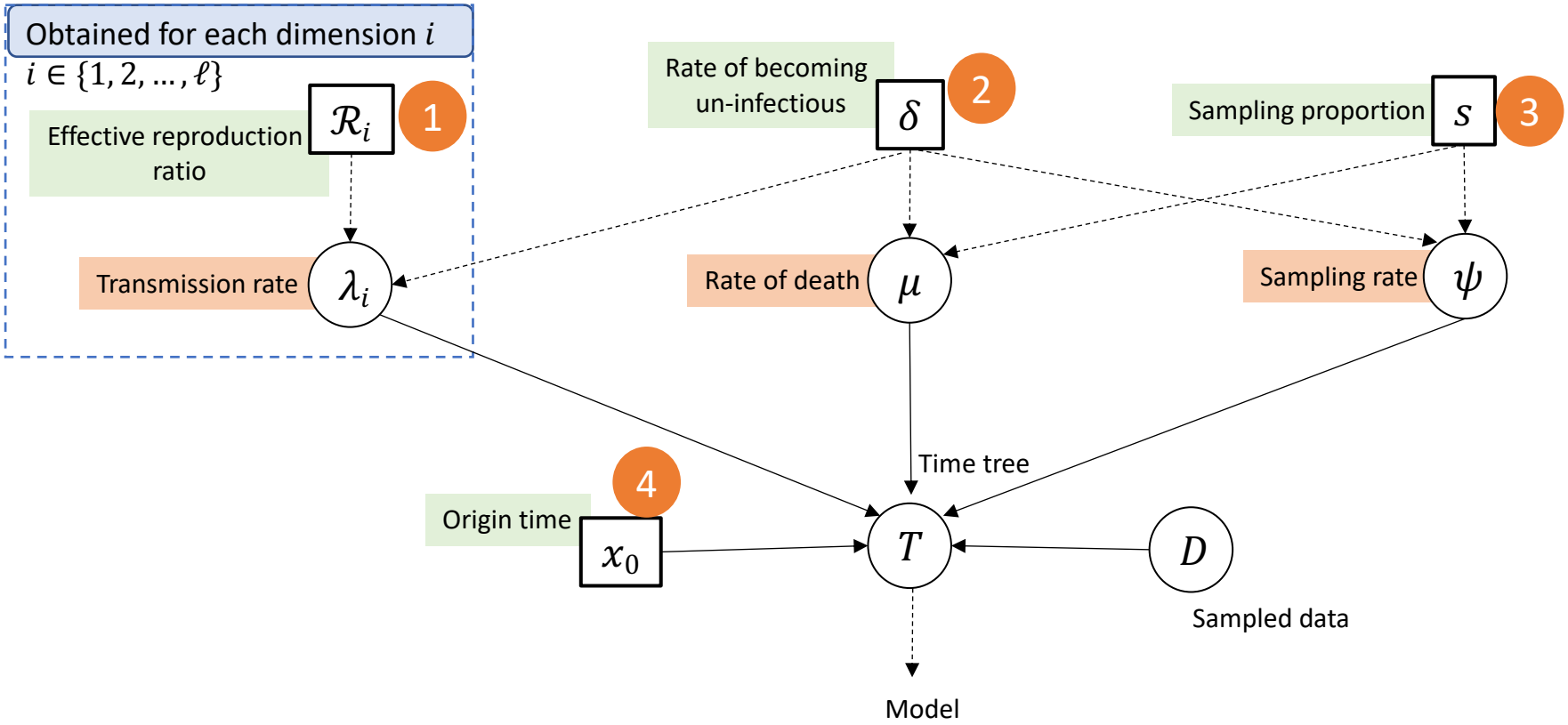


$$R = \frac{\lambda}{\mu + \psi} = \frac{\lambda}{\delta} \Rightarrow \lambda = R\delta$$

$$s = \frac{\psi}{\mu + \psi} = \frac{\psi}{\delta} \Rightarrow \psi = s\delta$$

$$\mu = \delta - \psi = \delta - s\delta = \delta(1 - s)$$

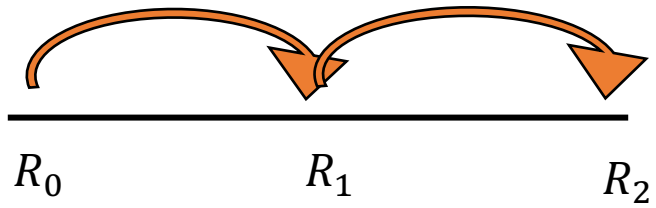
Birth-death Bayesian model: the priors



Birth-death Bayesian model: the priors

1. $R_i \sim LogNormal(0, 0.125)$, with $\ell = 3$

(the **effective reproductive number** changed two times after the start of the epidemic)



Birth-death Bayesian model: the priors

1. $R_i \sim LogNormal(0, 0.125)$, with $\ell = 3$ the effective reproduction ratio
2. $\delta \sim Gamma(\alpha = 0.5, \beta = 61)$ the rate on becoming noninfectious
3. $s \sim Beta(\alpha = 10, \beta = 6)$ the sampling probability

Birth-death Bayesian model

Substitution model: **HKY +** $\boldsymbol{\Gamma_4}$

(transversion rates are equal, transition rates are equal)

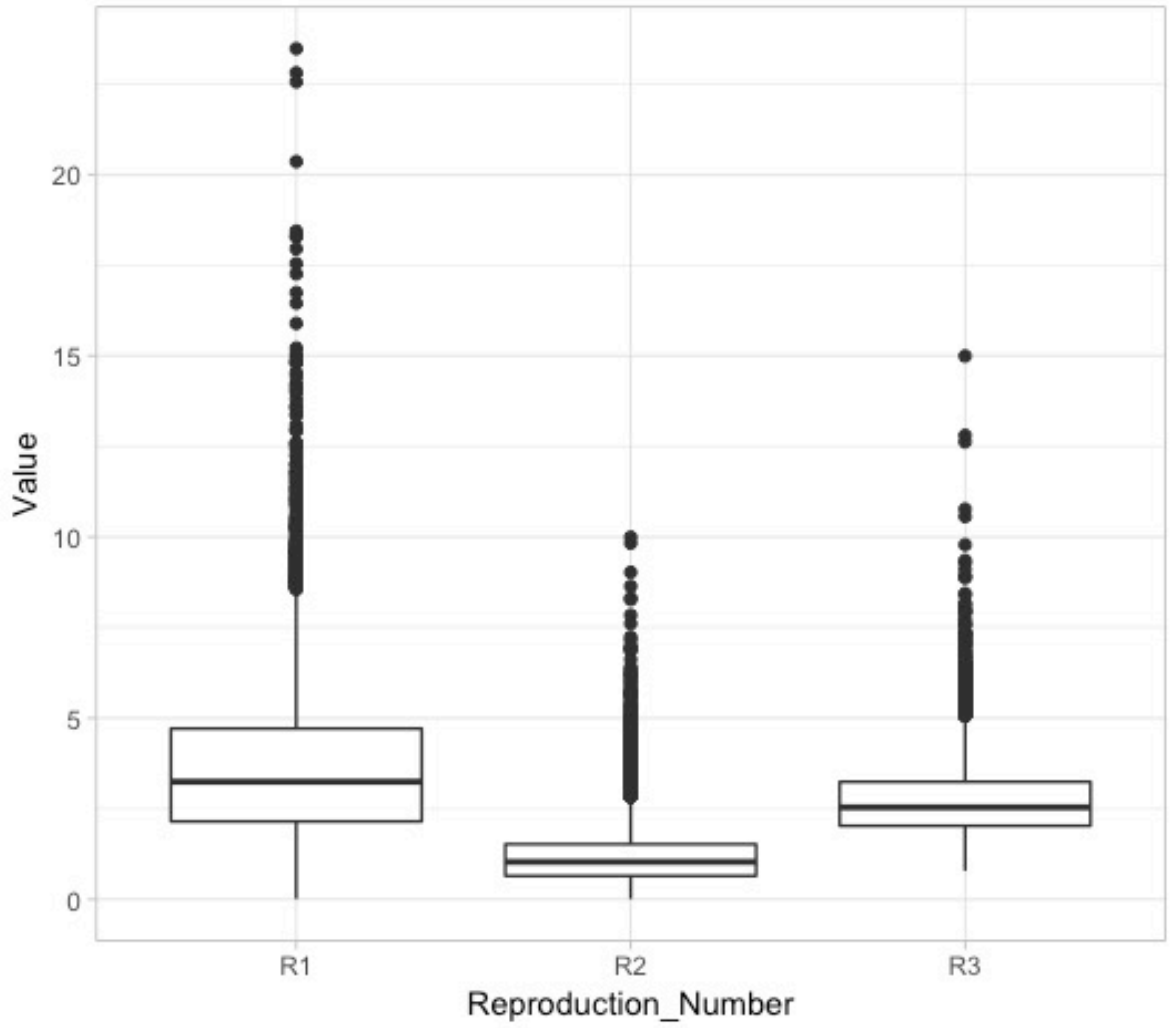
Skyline option: Skyline Bayesian Serial

(since the samples were taken through time)

Software: BEAST v2, FigTree, Tracer, R

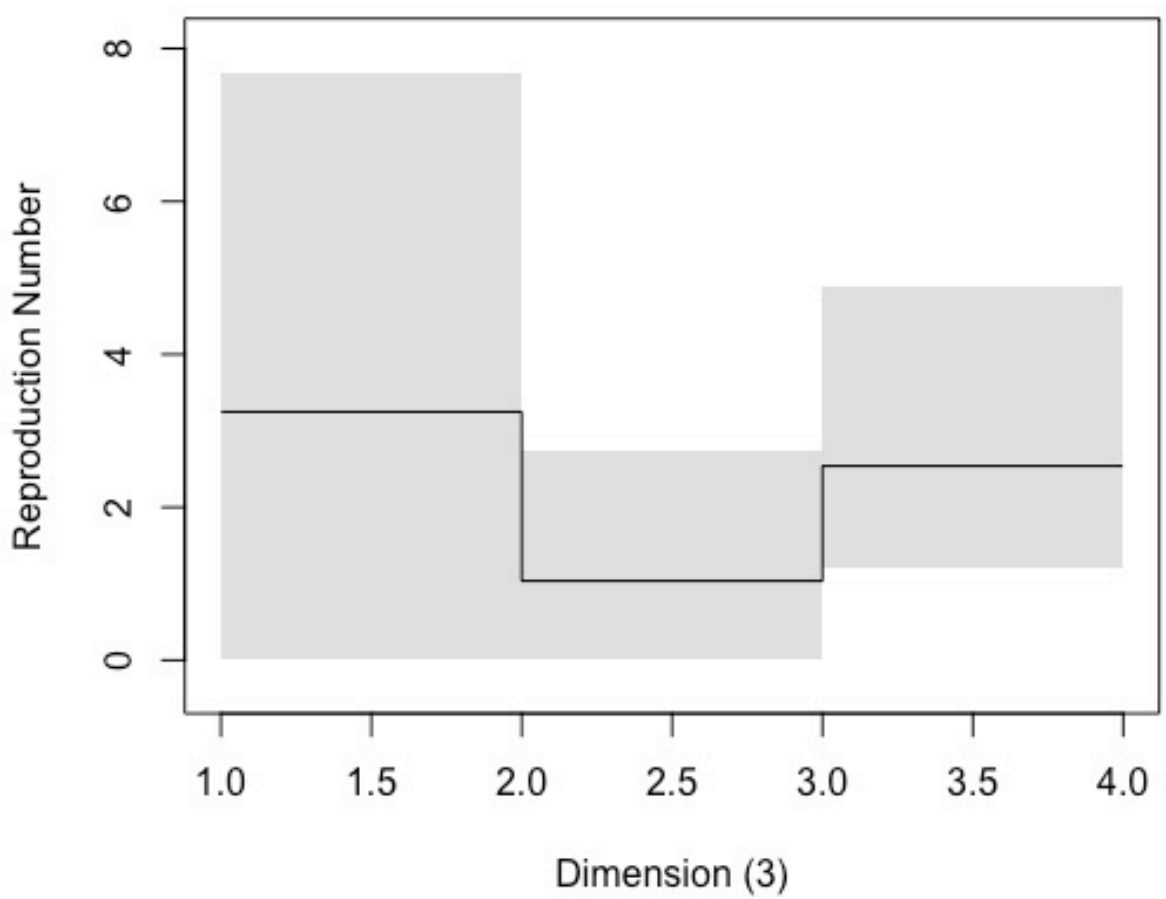
Reproduction ratio and its skyline

- In all 3 dimensions, Ri's are heavy in lower values

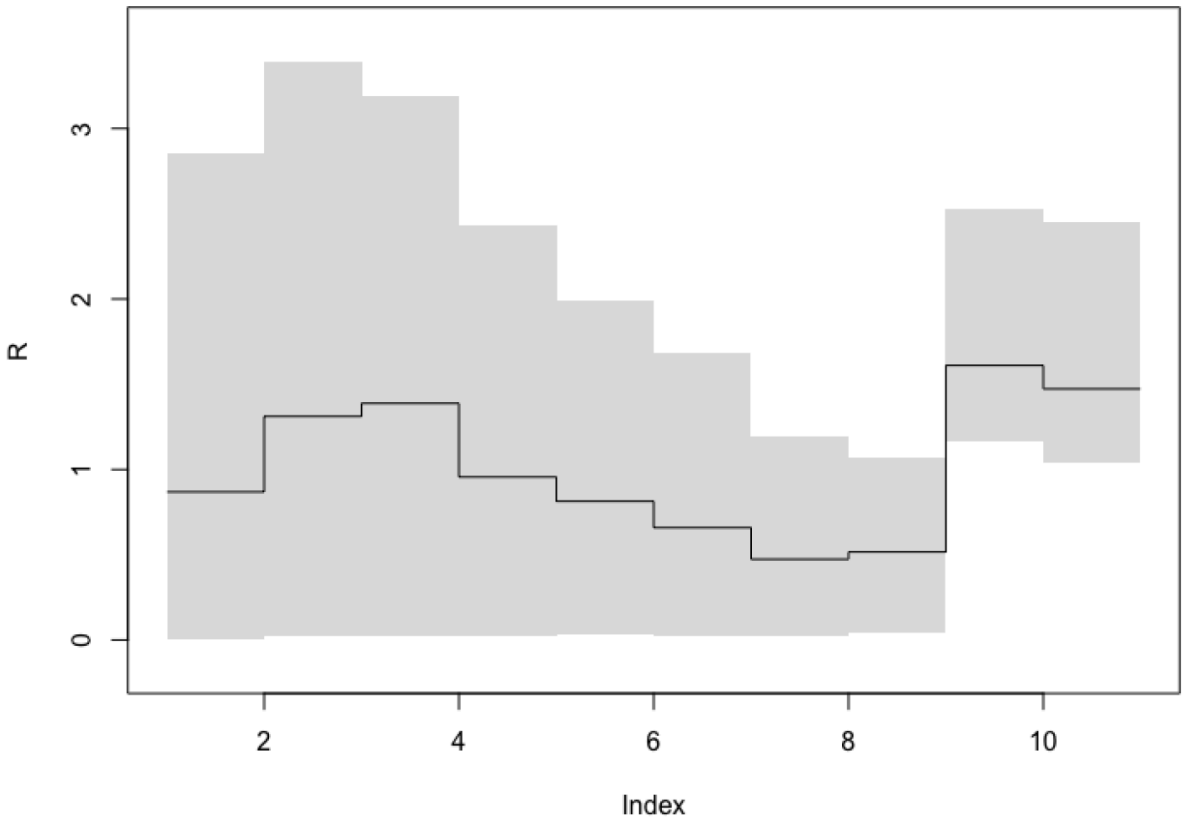


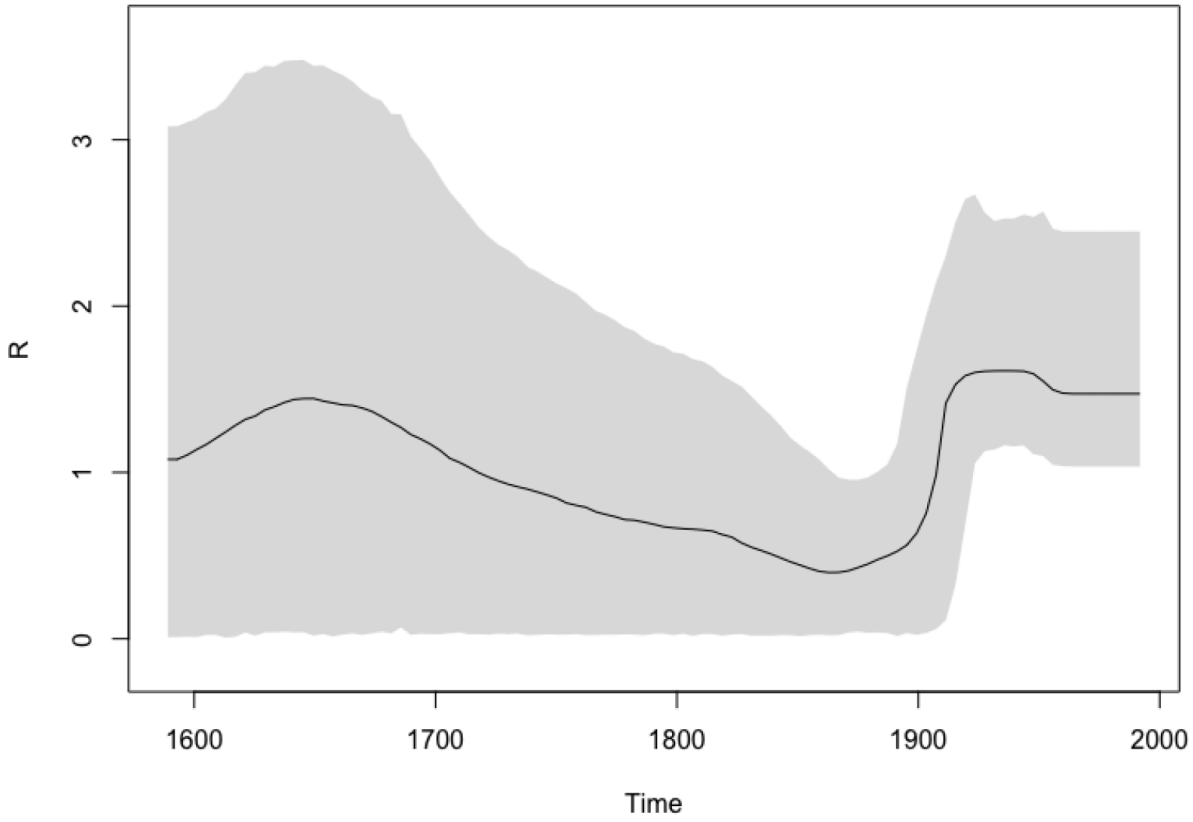
Reproduction ratio and its skyline

- This is the smoothest skyline plot could be obtained
- Attempt to make it smoother only resulted a flat/constant line and does not show the dynamic of R_i



Reproduction ratio and its skyline: illustration

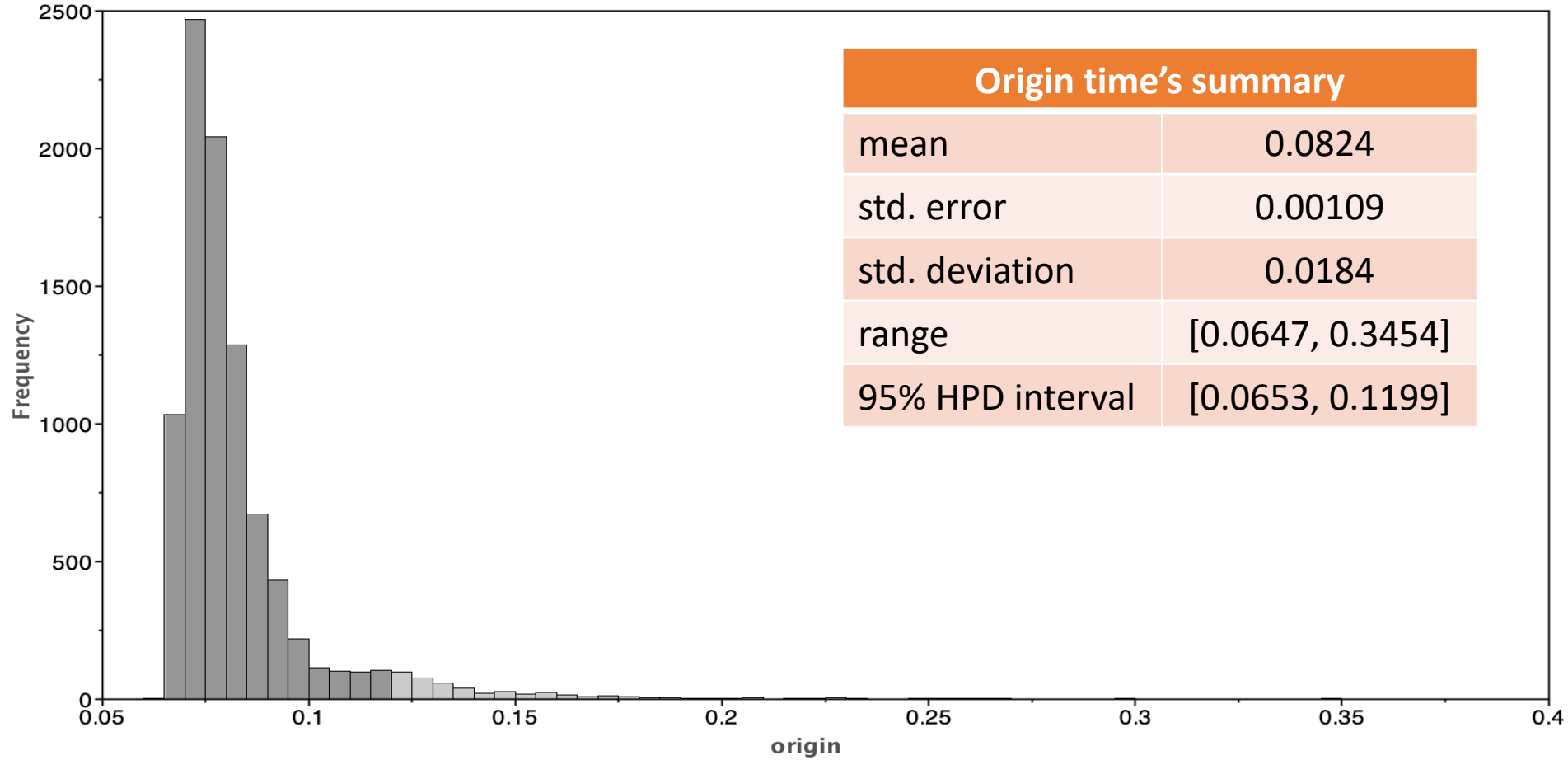




- Skyline of R_i in 10 dimensions

..smoothing!

The estimate of origin time



Origin time's summary	
mean	0.0824
std. error	0.00109
std. deviation	0.0184
range	[0.0647, 0.3454]
95% HPD interval	[0.0653, 0.1199]

The estimate of origin time

- When did the epidemic start?

Origin time's summary	
mean	0.0824
std. error	0.00109
std. deviation	0.0184
range	[0.0647, 0.3454]
95% HPD interval	[0.0653, 0.1199]

The estimate of origin time

- When did the epidemic start?

$$\frac{0.0824}{1/365} = 30.076 \approx 30 \text{ days}$$

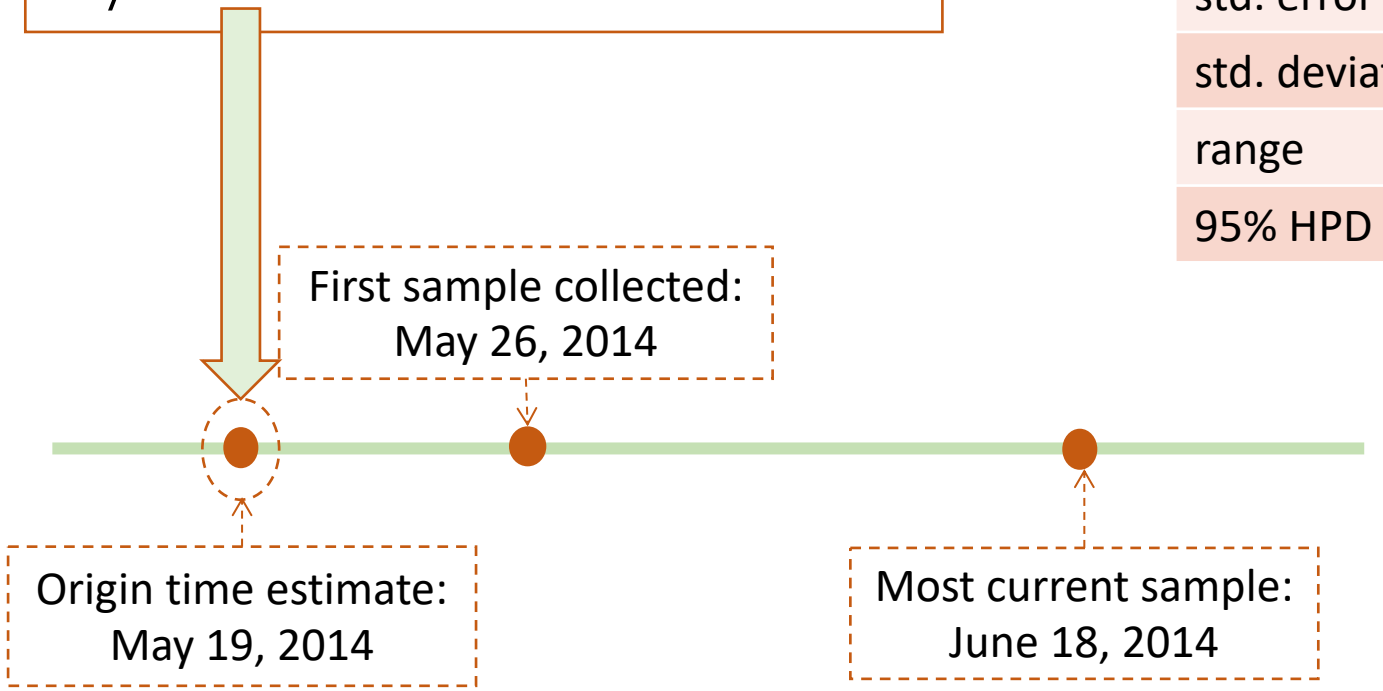
prior to the last sample date

Origin time's summary	
mean	0.0824
std. error	0.00109
std. deviation	0.0184
range	[0.0647, 0.3454]
95% HPD interval	[0.0653, 0.1199]

The estimate of origin time

- When did the epidemic start?

$$\frac{0.0824}{1/365} = 30.076 \approx 30 \text{ days}$$



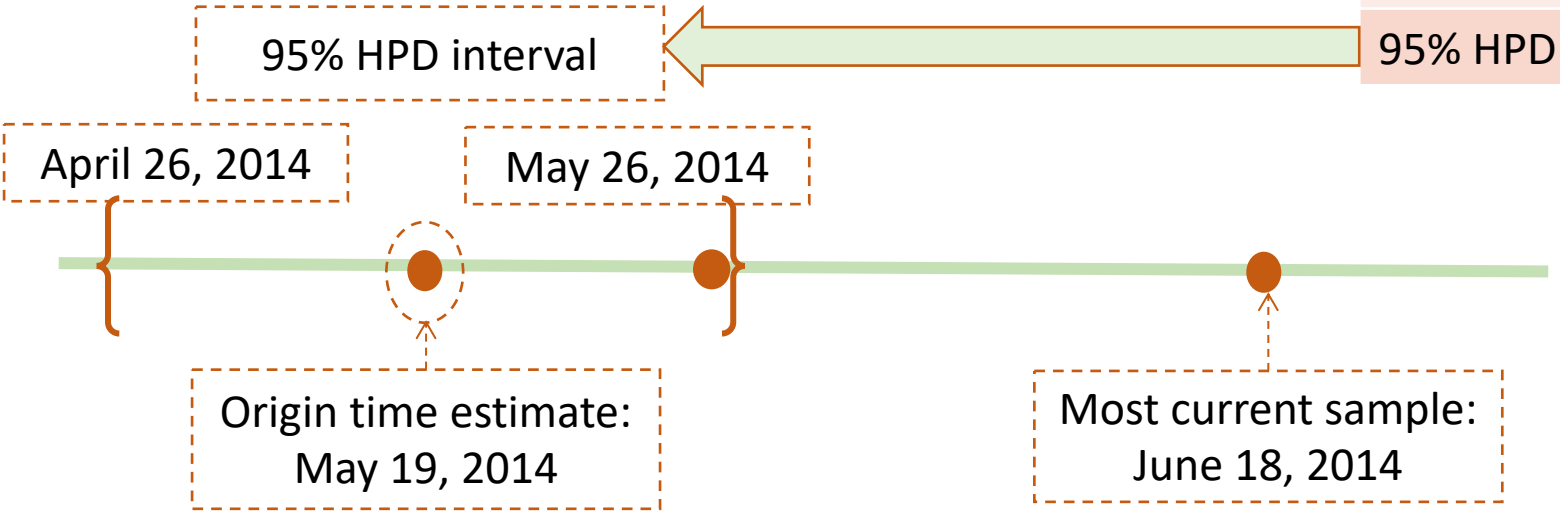
Origin time's summary	
mean	0.0824
std. error	0.00109
std. deviation	0.0184
range	[0.0647, 0.3454]
95% HPD interval	[0.0653, 0.1199]

The estimate of origin time

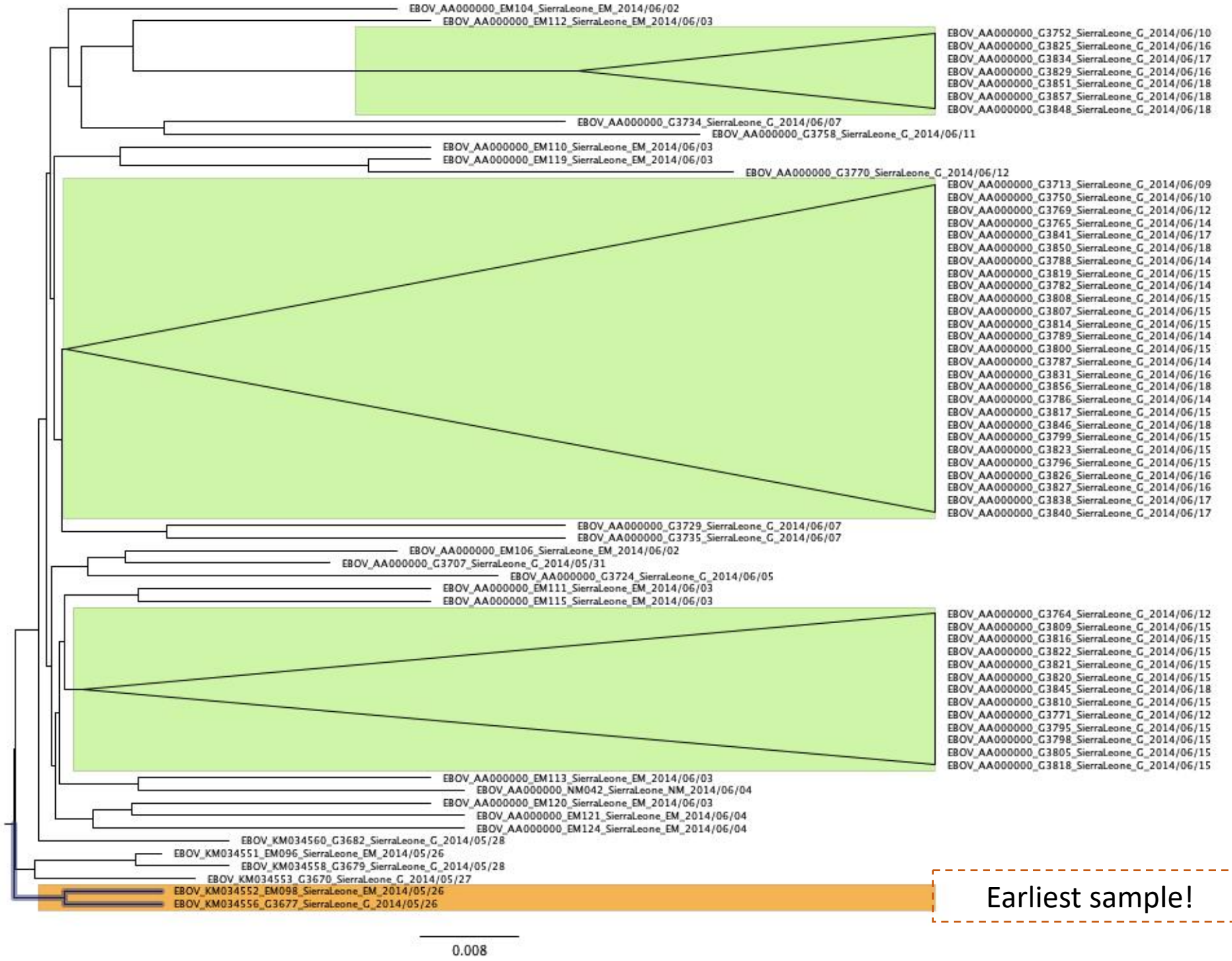
- When did the epidemic start?

$$\frac{0.0824}{1/365} = 30.076 \approx 30 \text{ days}$$

Origin time's summary	
mean	0.0824
std. error	0.00109
std. deviation	0.0184
range	[0.0647, 0.3454]
95% HPD interval	[0.0653, 0.1199]



Results



Earliest sample!

What we have so far….

- 72 taxa from 2014 Sierra Leone EBOV outbreak
- Birth-death Bayesian model + skyline
- The skyline is built over the effective reproduction number (R_i), based on its 95% HPD interval
- Smooth skyline could not be obtained due to flat reproduction numbers
- Estimation of origin of the epidemic: 30 days before last sample

What can be done more?

- Data analysis involving samples from other area of epidemic: spatio-temporal (?)
- Other molecular clock methods
- Compare the results to a simpler tree (Bayesian) to see the improvement

References

1) S. K. Gire et al. Genomic surveillance elucidates Ebola virus origin and transmission during the 2014 outbreak". In: Science 345 (2014).
2) M. Hasegawa, H. Kishino, and T. Yano. Dating of the human-ape splitting by a molecular clock of mitochondrial DNA". In: J Mol Evol 22.2 (1985), pp. 160{74.
3) T. Heath and T. Stadler. Estimating Epidemiological Parameters of an Ebola Outbreak using BEAST2". In: (2014). url: http://phyloworks.org/workshops/Ebola_BEAST2_Exercise.pdf.
4) S. Hohna et al. Probabilistic graphical model representation in phylogenetics. In: Syst. Biol (2014). 63(5):753–771.
5) T. Stadler et al. Birth-death skyline plot reveals temporal changes of epidemic spread in HIV and hepatitis C virus (HCV)". In: PNAS 110 (2013). doi: 10.1073/pnas.1207965110.
6) T. Stadler et al. Insights into the Early Epidemic Spread of Ebola in Sierra Leone Provided by Viral Sequence Data". In: PLOS Currents Outbreaks Edition 1 (2014). doi: 10.1371/currents.outbreaks.02bc6d927ecee7bbd33532ec8ba6a25f.
7) V. Boskova et al. Inference of Epidemiological Dynamics Based on Simulated Phylogenies Using Birth-Death and Coalescent Models. In: PLoS (2014).

Thank you!