

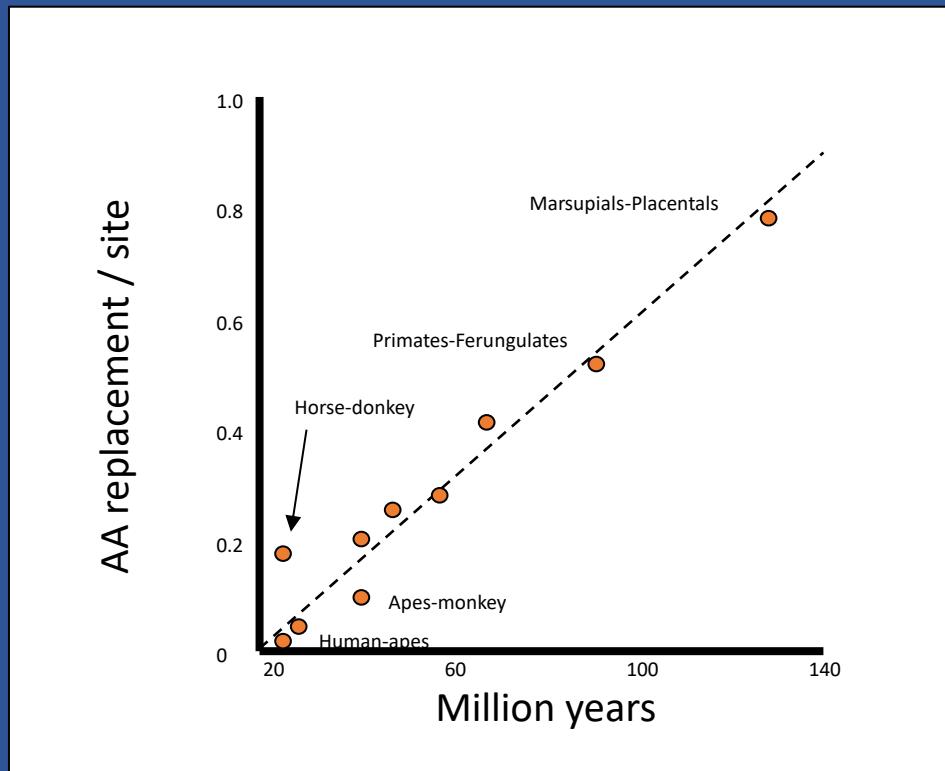
# Molecular Clocks



# Molecular Clocks

Initial observation: rates of amino acid replacement were approximately the same among various mammalian lineages

Hypothesis: for any given protein, the rate of molecular evolution is approximately constant over time in all lineages



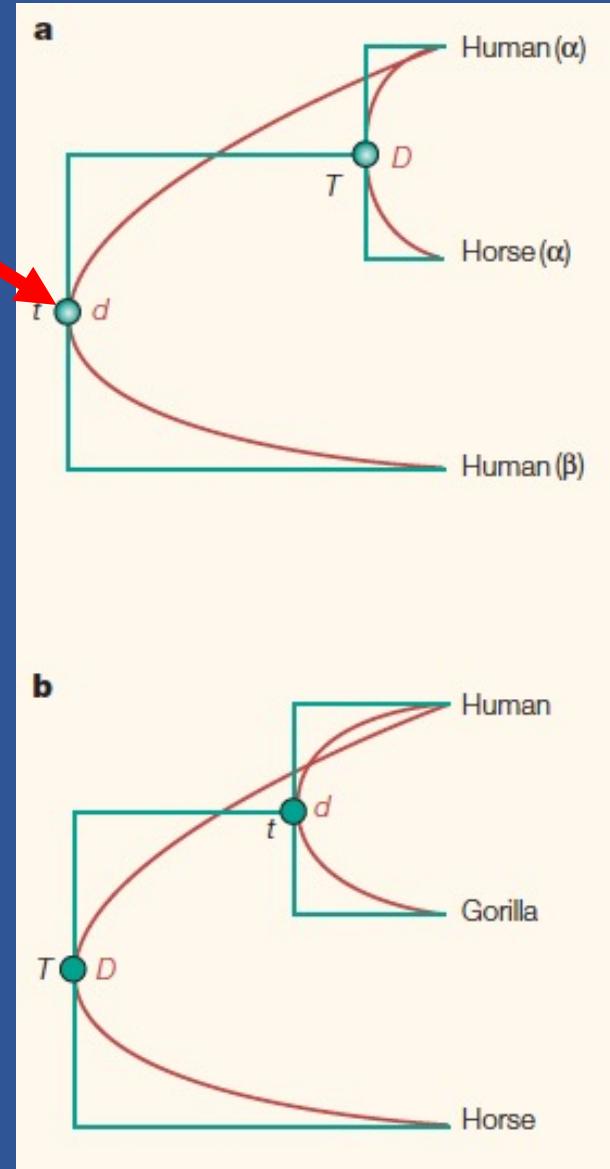
# Molecular Clocks

- Originally described for vertebrates (Zuckerkandl & Pauling 1962, 65):

*“Genetic divergence between species increases linearly since the time these species shared a common ancestor”*

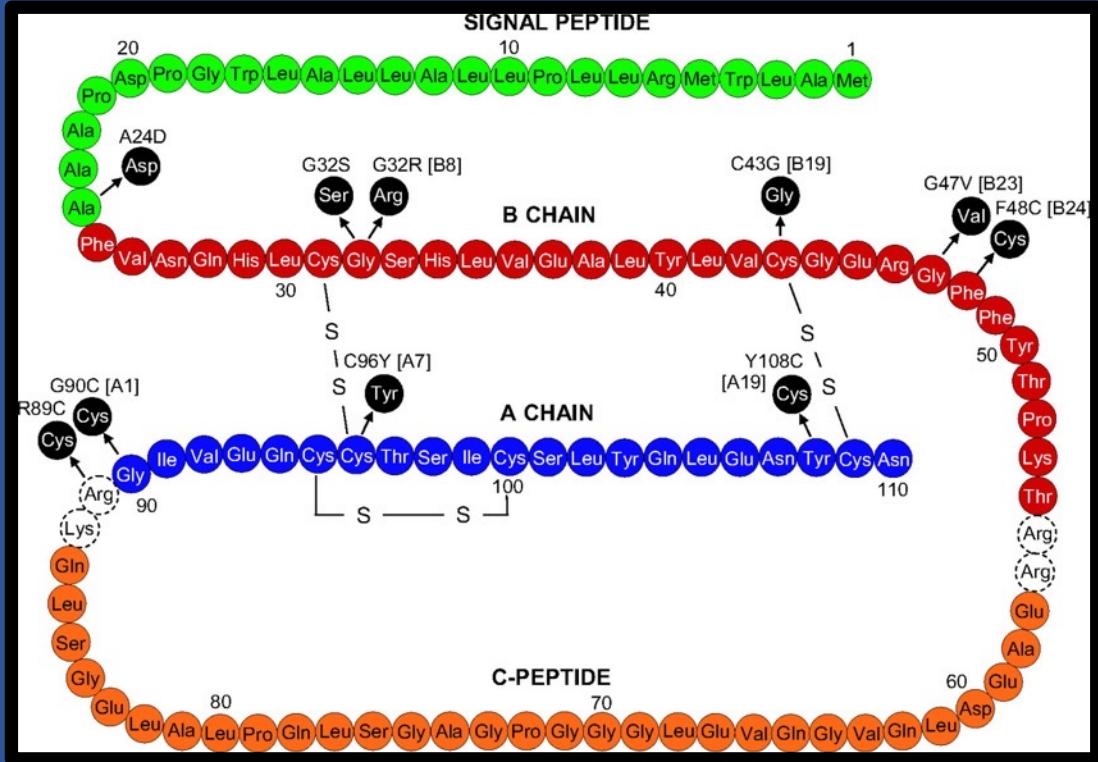
- Estimated time of divergence of four members of the haemoglobin gene family. Calibrated based on time (from the fossil record).

Fossil record



# Molecular Clocks

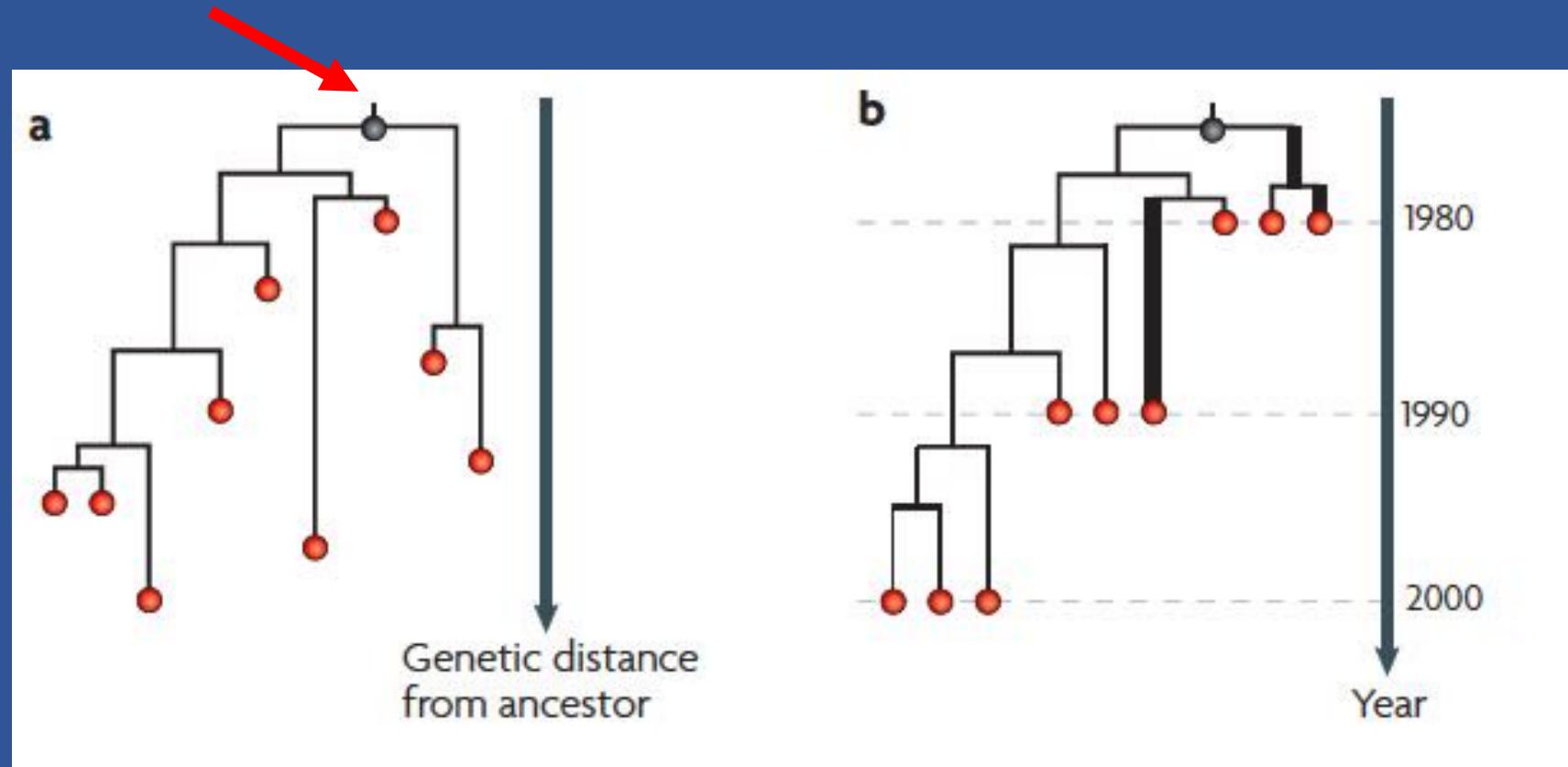
- Assumptions of rate constancy may be questionable
- Controversy of applying molecular clock assumptions to estimate time(s) of divergence



Classic example: Insulin evolves faster in the guinea pig lineage(s) than along other lineages

# Molecular Clocks for Pathogens

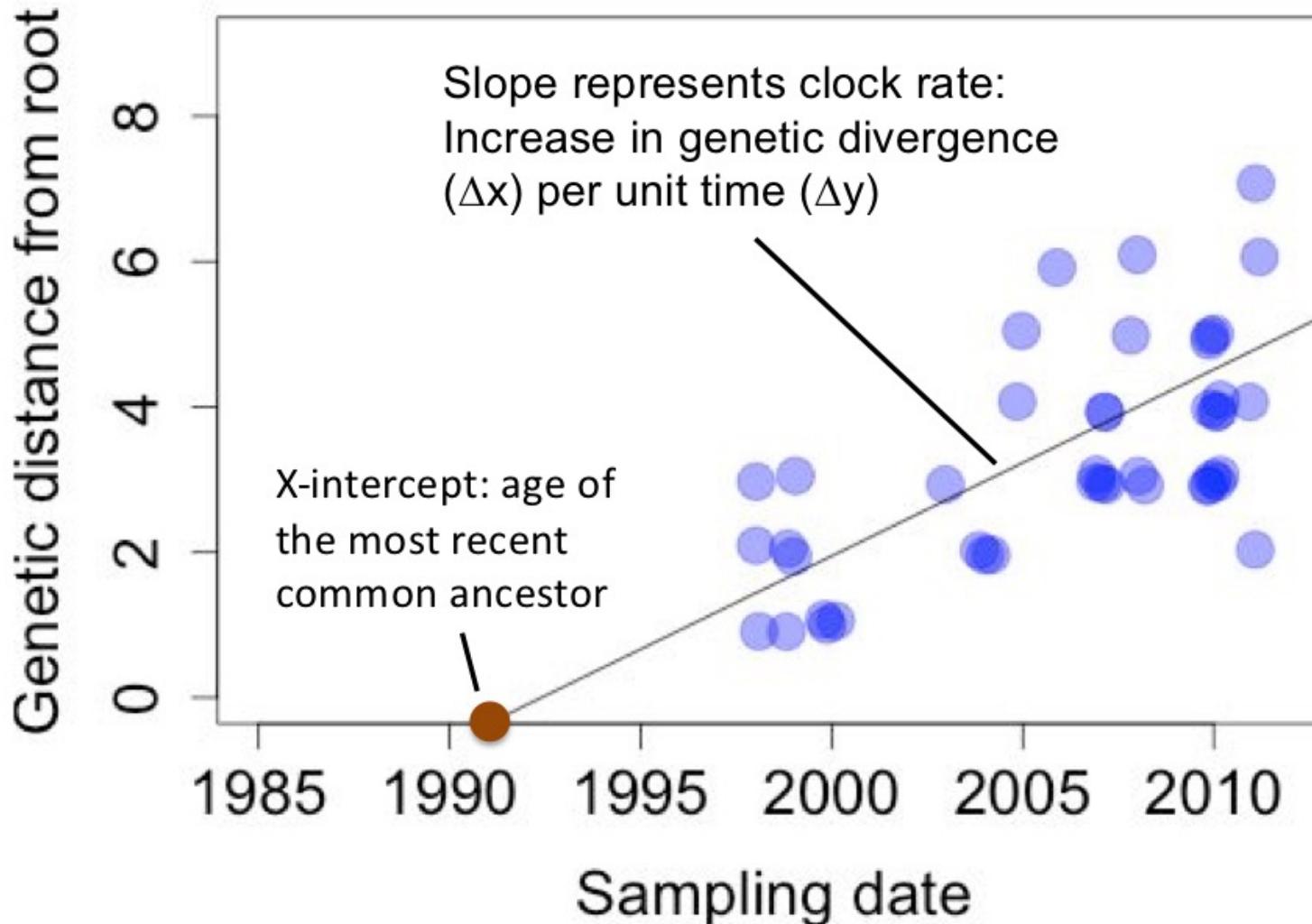
Most recent common ancestor



Are genetic and temporal distances correlated?

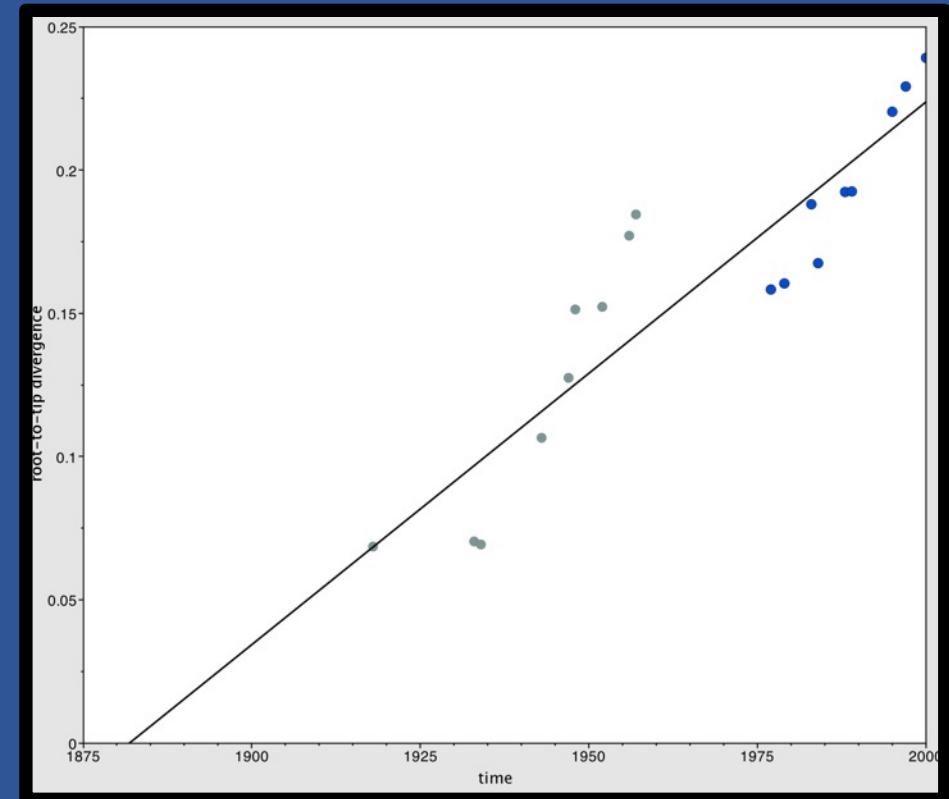
Pybus and Rambaut  
2009, Nature Rev.  
Genetics

# Molecular Clocks for Pathogens

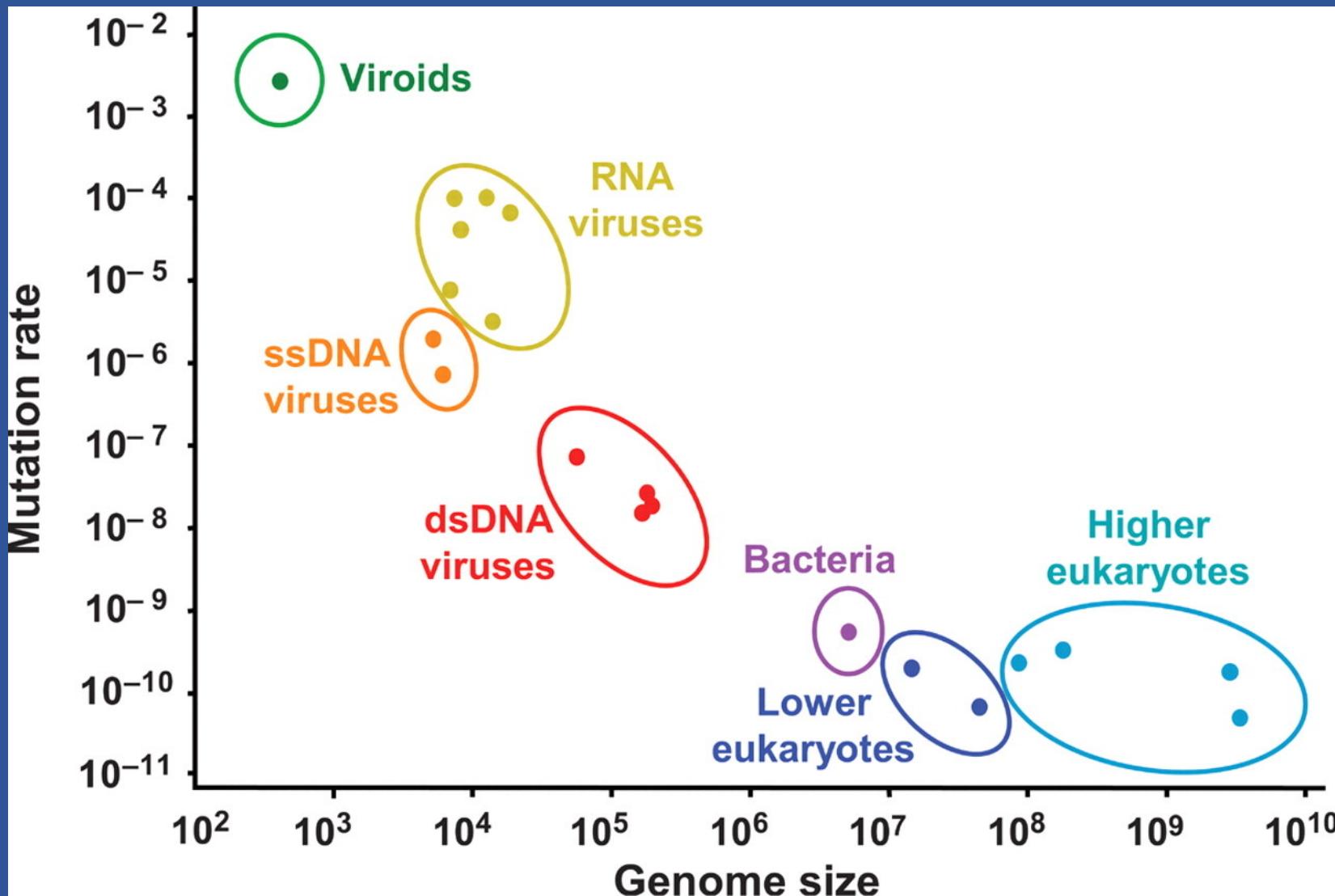


# Molecular Clocks for Pathogens

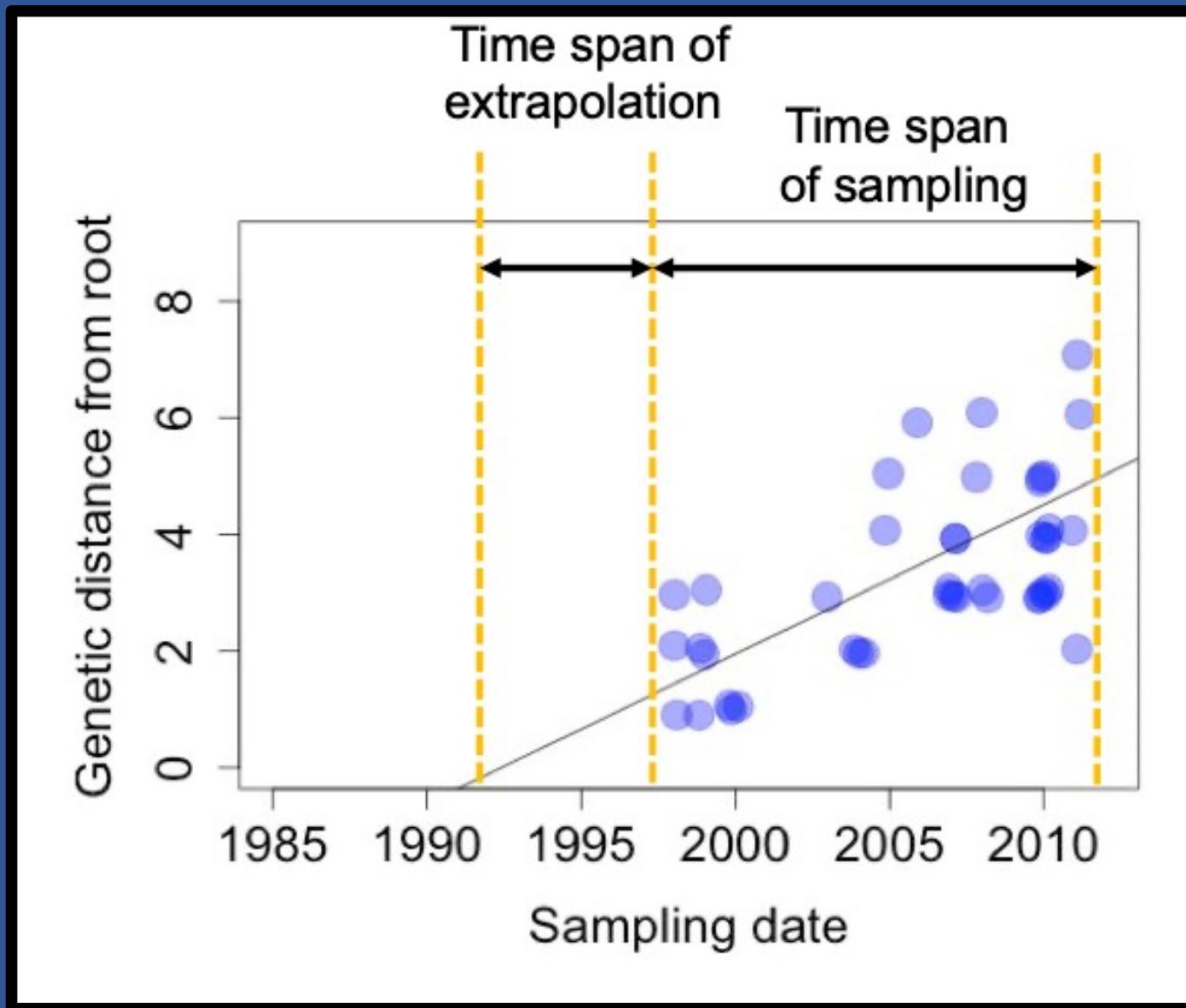
- If positive relationship between genetic divergence and sampling date is found, population is said to be **measurably evolving** (Drummond et al. 2003)
- Increased divergence likely caused by **mutations that are effectively neutral**  
=> expect limited fitness consequences  
(at least in the short term)



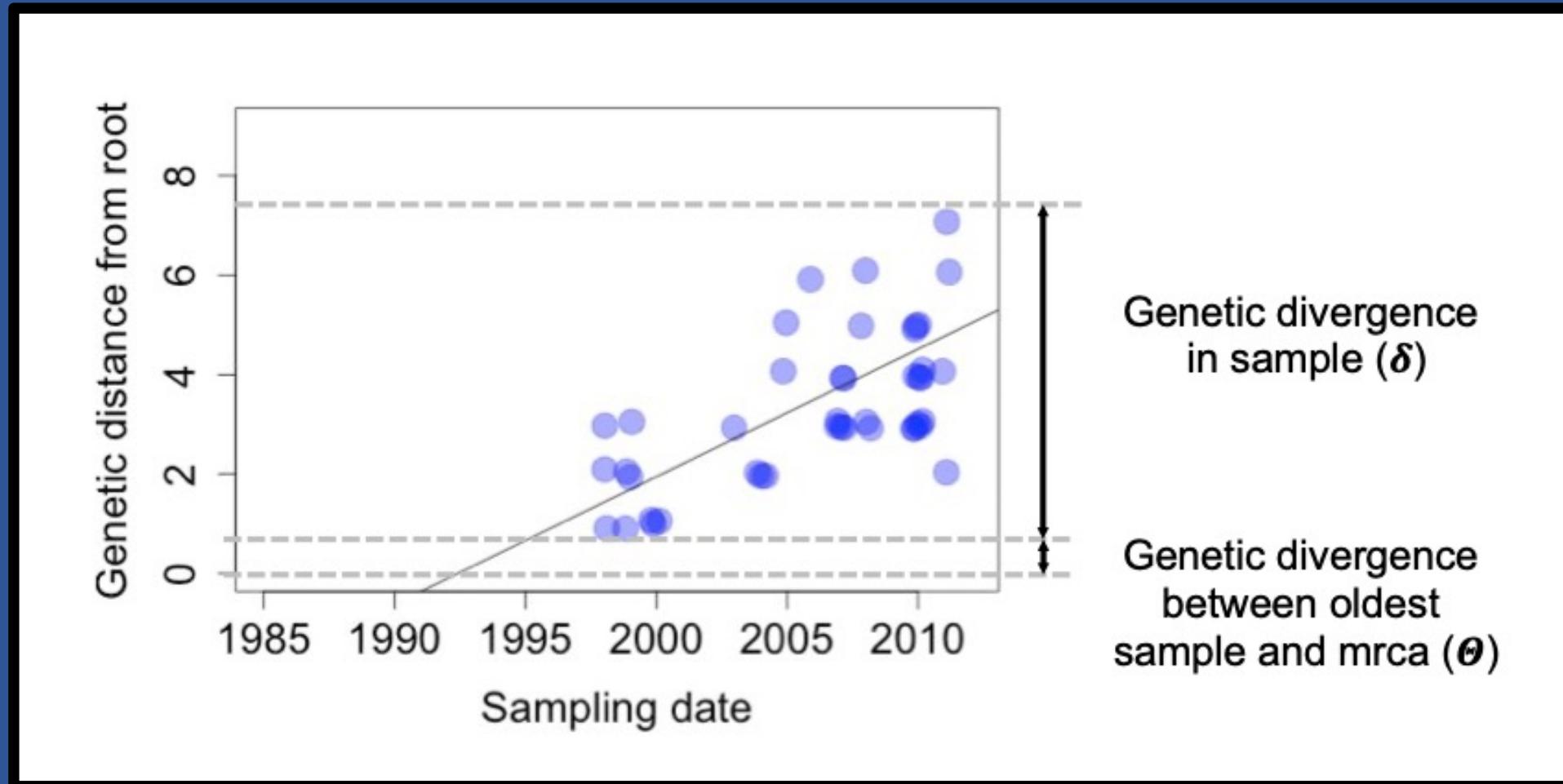
# Per-site mutation rate versus genome size



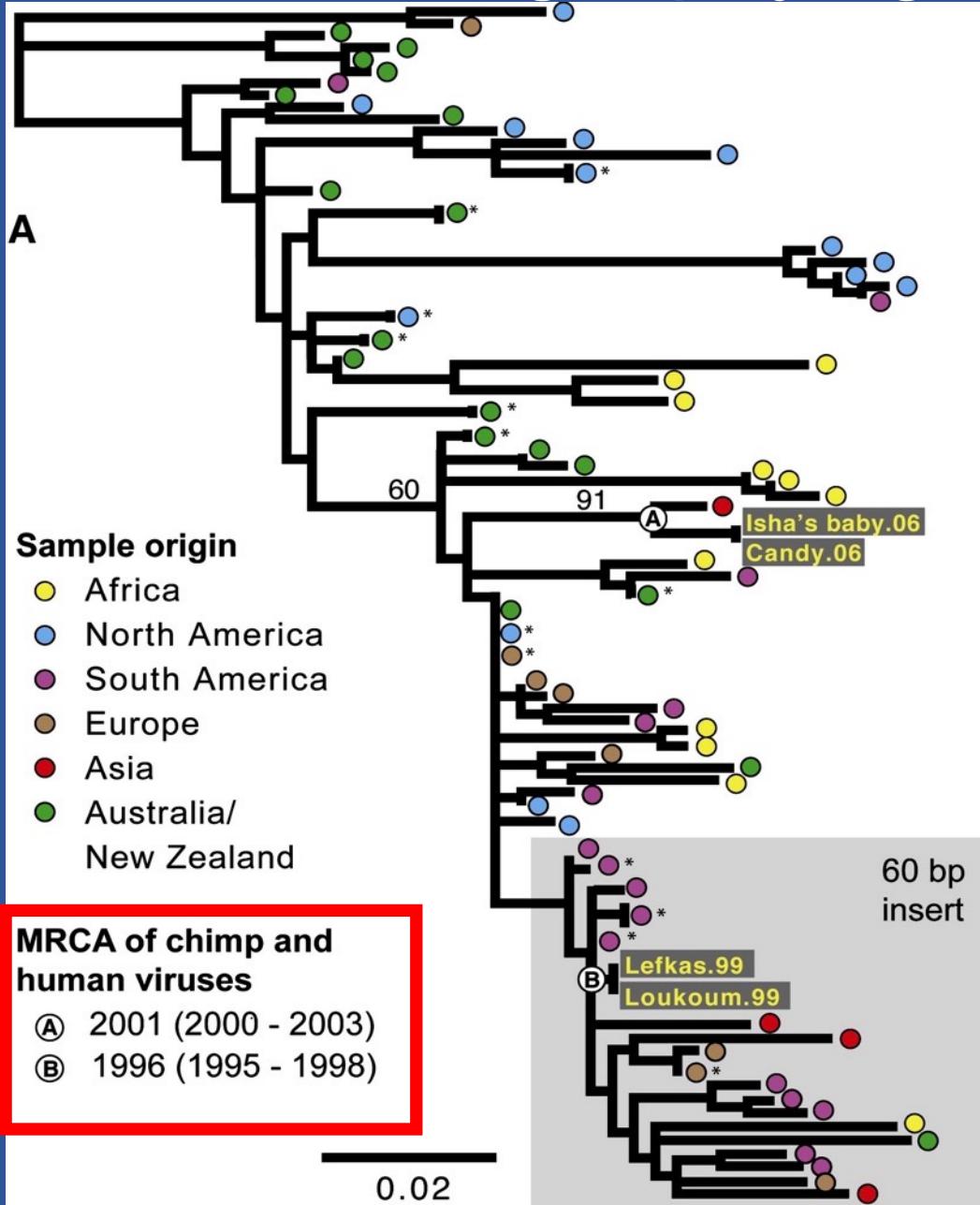
# Testing for clock-like signal using root-to-tip regression



# Testing for clock-like signal using root-to-tip regression



# Using a phylogeny to date internal nodes

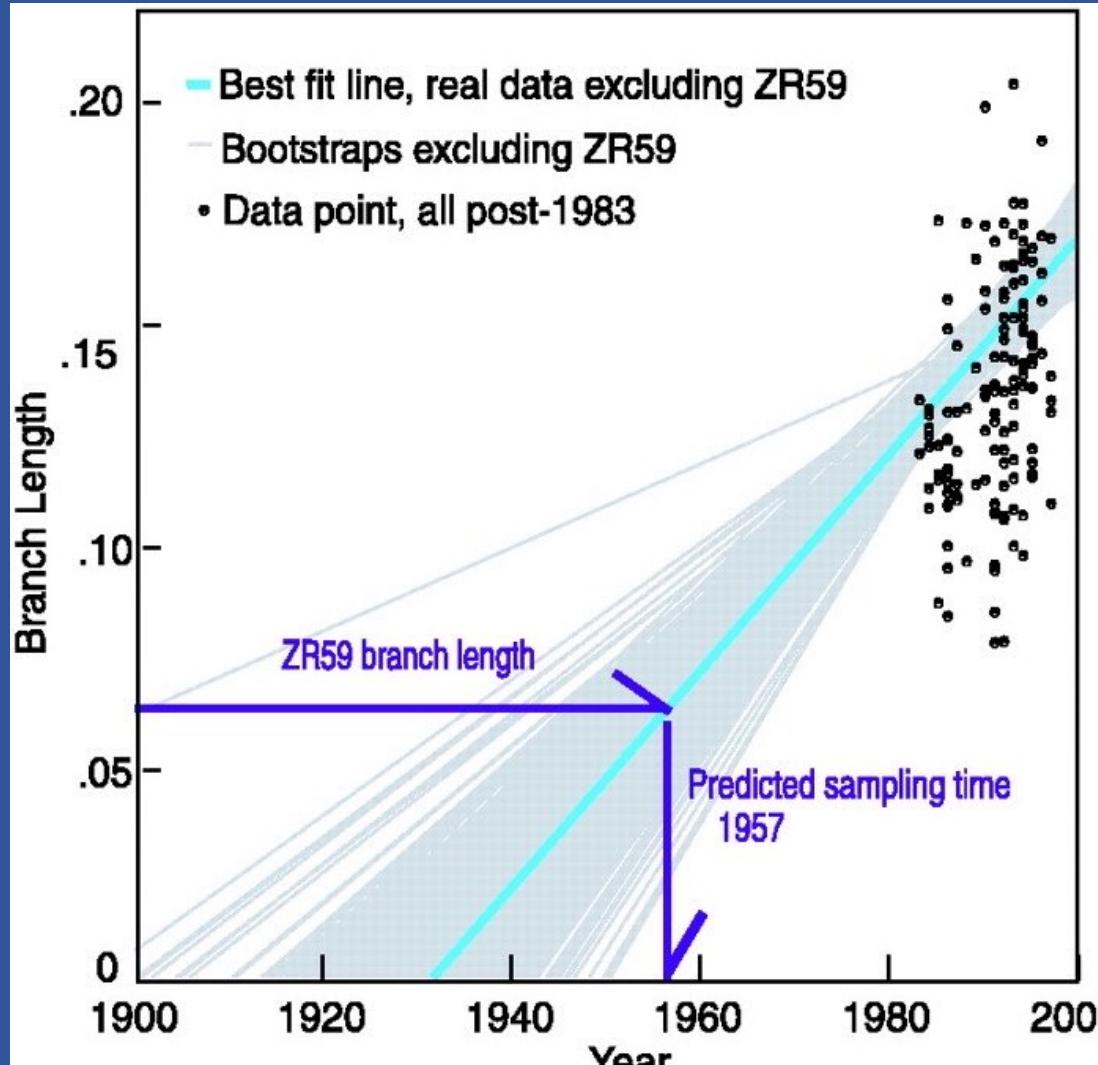


Transmission of a human respiratory virus (HRSV) from humans to chimps:  
Using molecular clock to date the spillover event

Oldest samples 1960-1962 used to root tree.

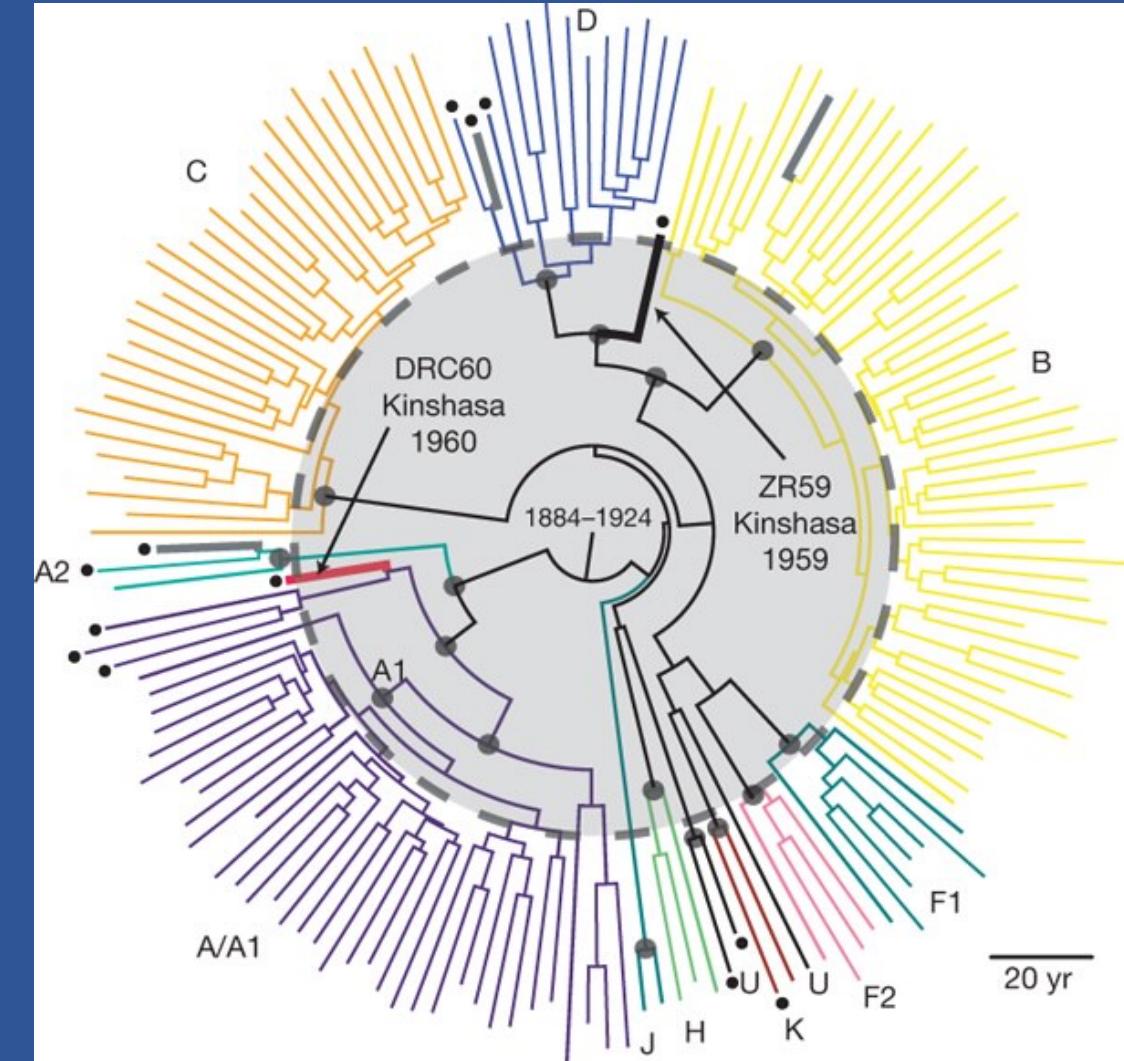
# Dating the origin of the HIV pandemic

Early estimates 1930s



Korber et al. 2000, Science

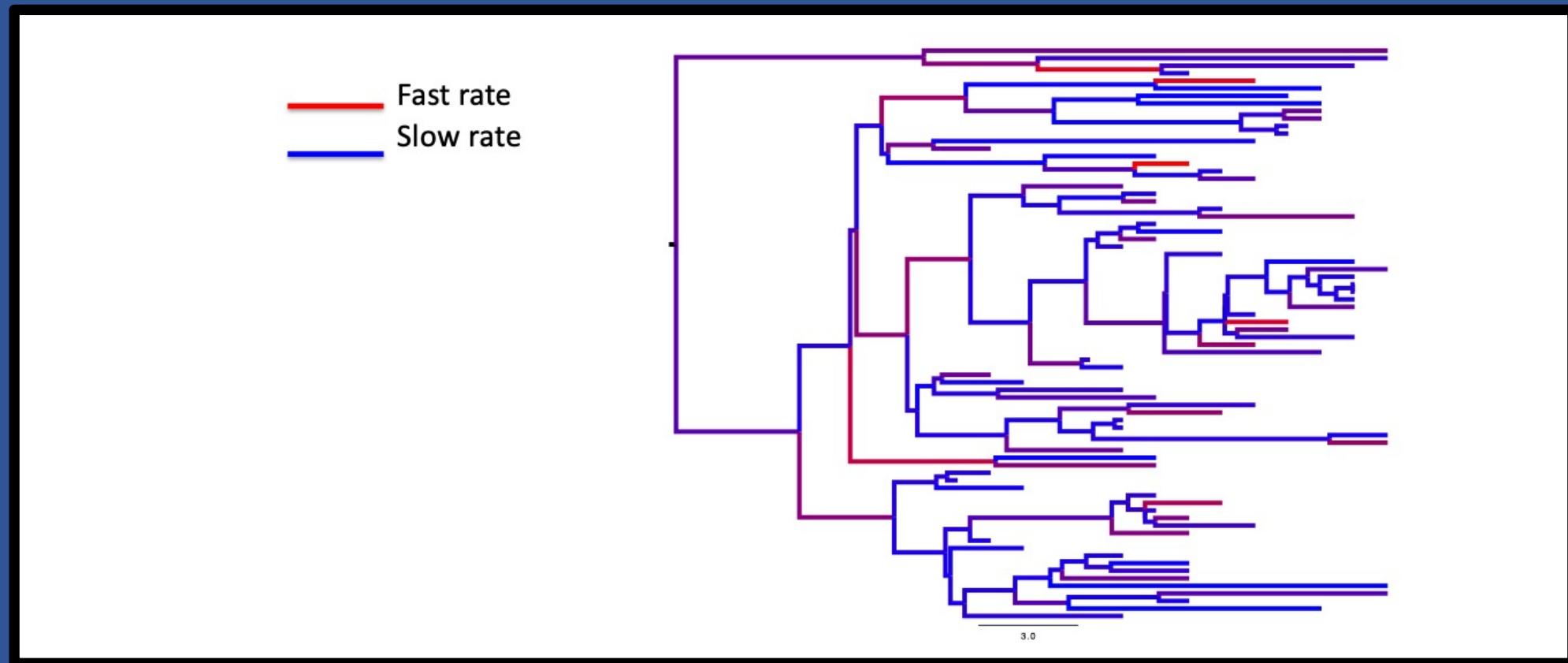
Current estimates prior to 1924



Worobey et al. 2008, Nature

# Modelling the reality?

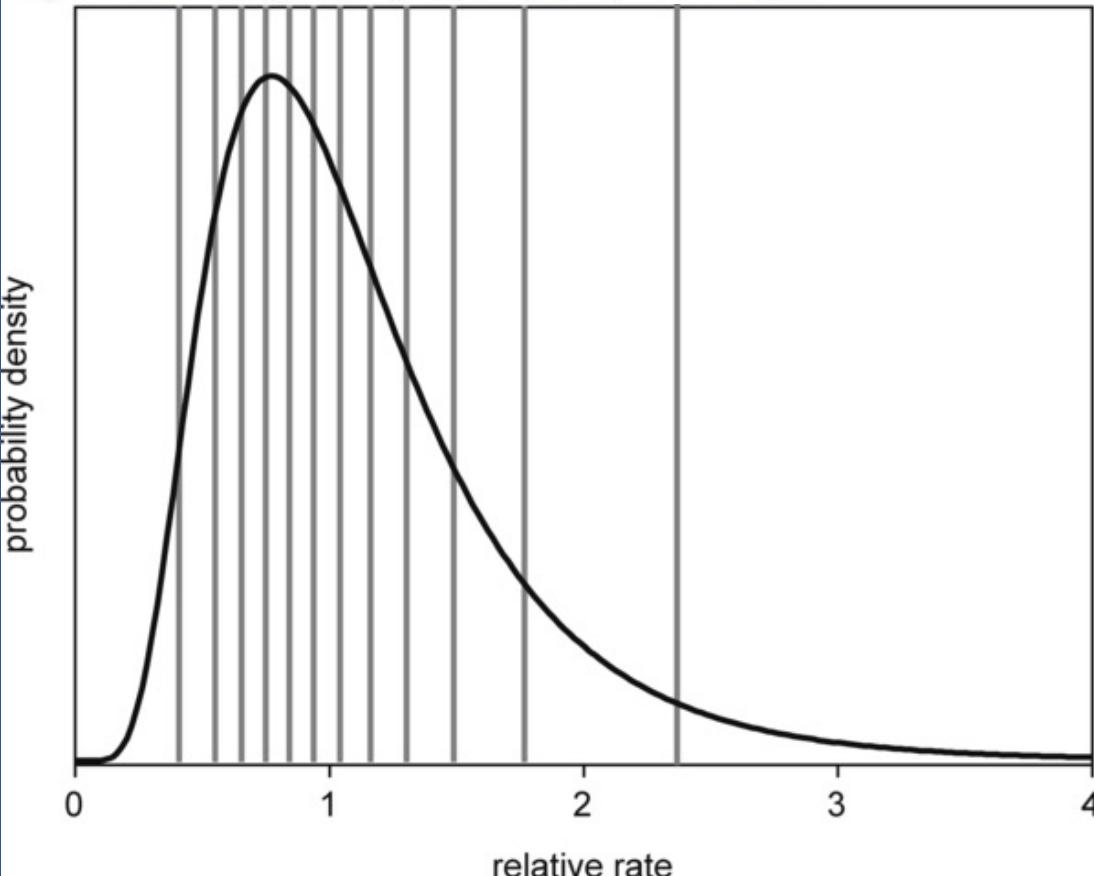
- Evolution is often not strictly clock-like
- Relaxed clocks allow for rates to vary among branches  
(Drummond 2006, PLoS Biol)



# Relaxed Phylogenetics and Dating with Confidence

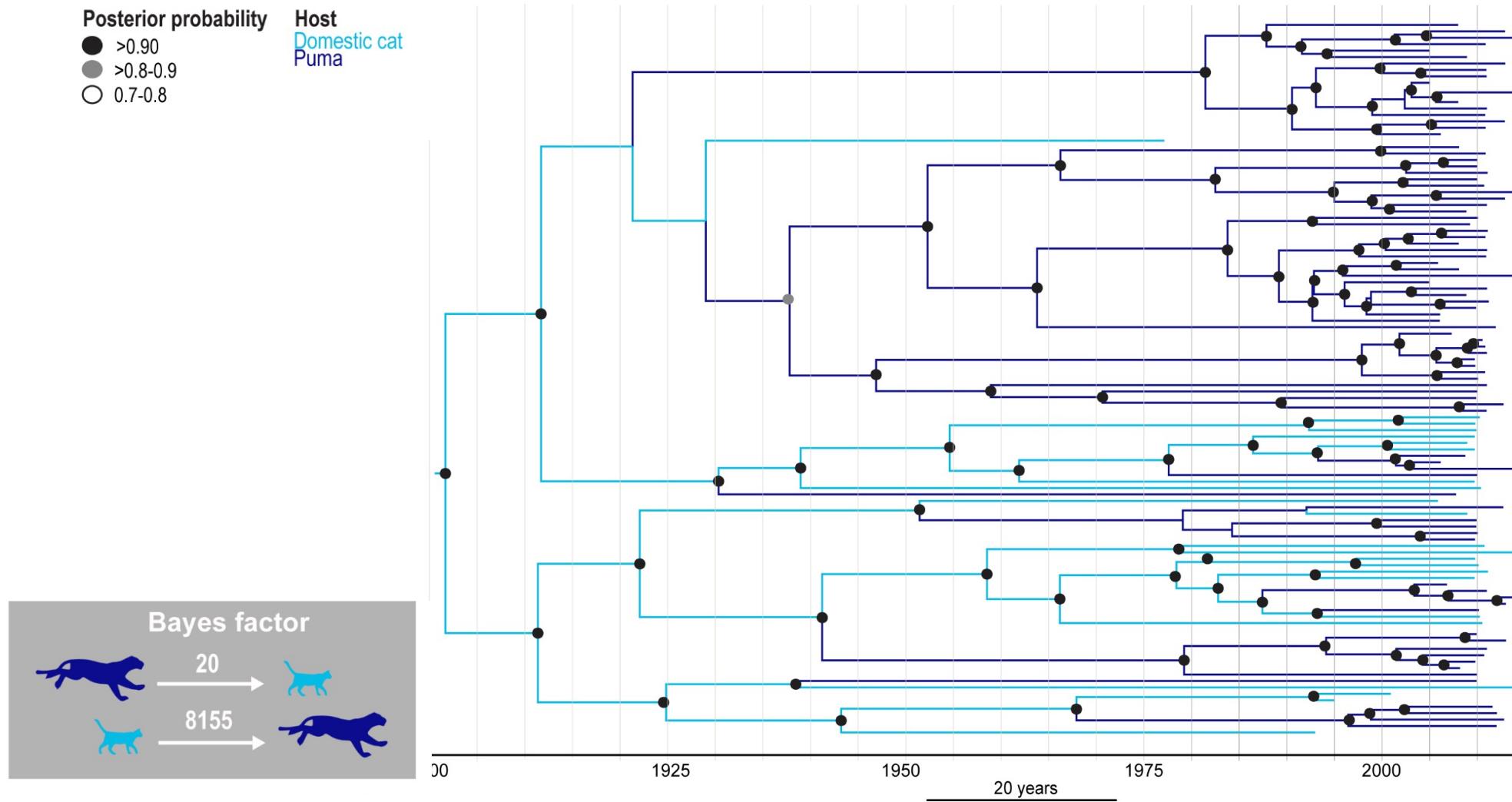
Alexei J. Drummond<sup>✉</sup>\*, Simon Y. W. Ho, Matthew J. Phillips, Andrew Rambaut<sup>✉</sup>\*

Department of Zoology, University of Oxford, Oxford, United Kingdom



Relaxed molecular clock (uncorrelated):  
rate for each branch drawn randomly  
from distribution (lognormal or  
exponential)

# Ancestral host reconstruction of FFV (pol)



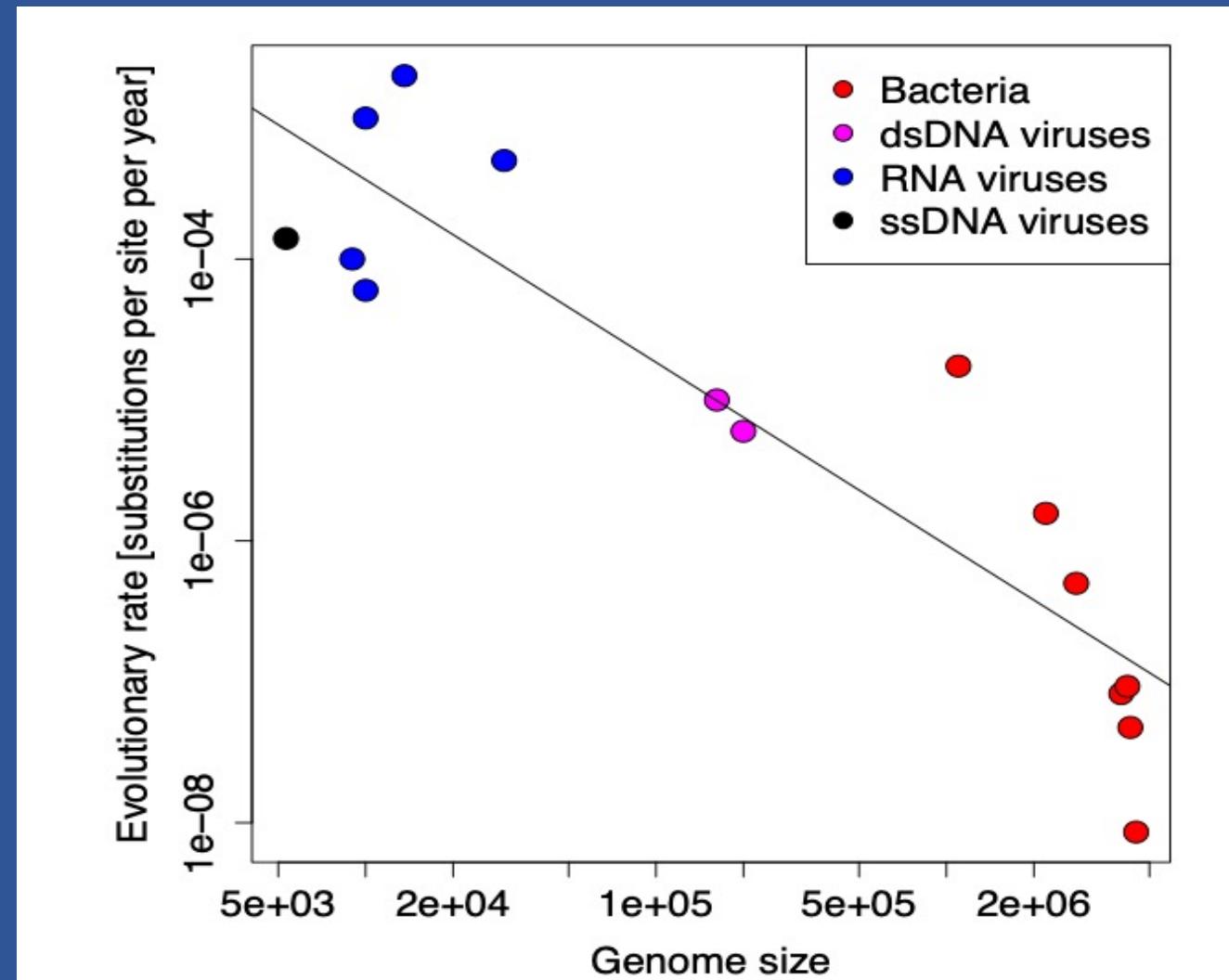
Krabberger et al. 2020

Overwhelming support for DC to puma transmission

# Rate of evolution scales with genome size

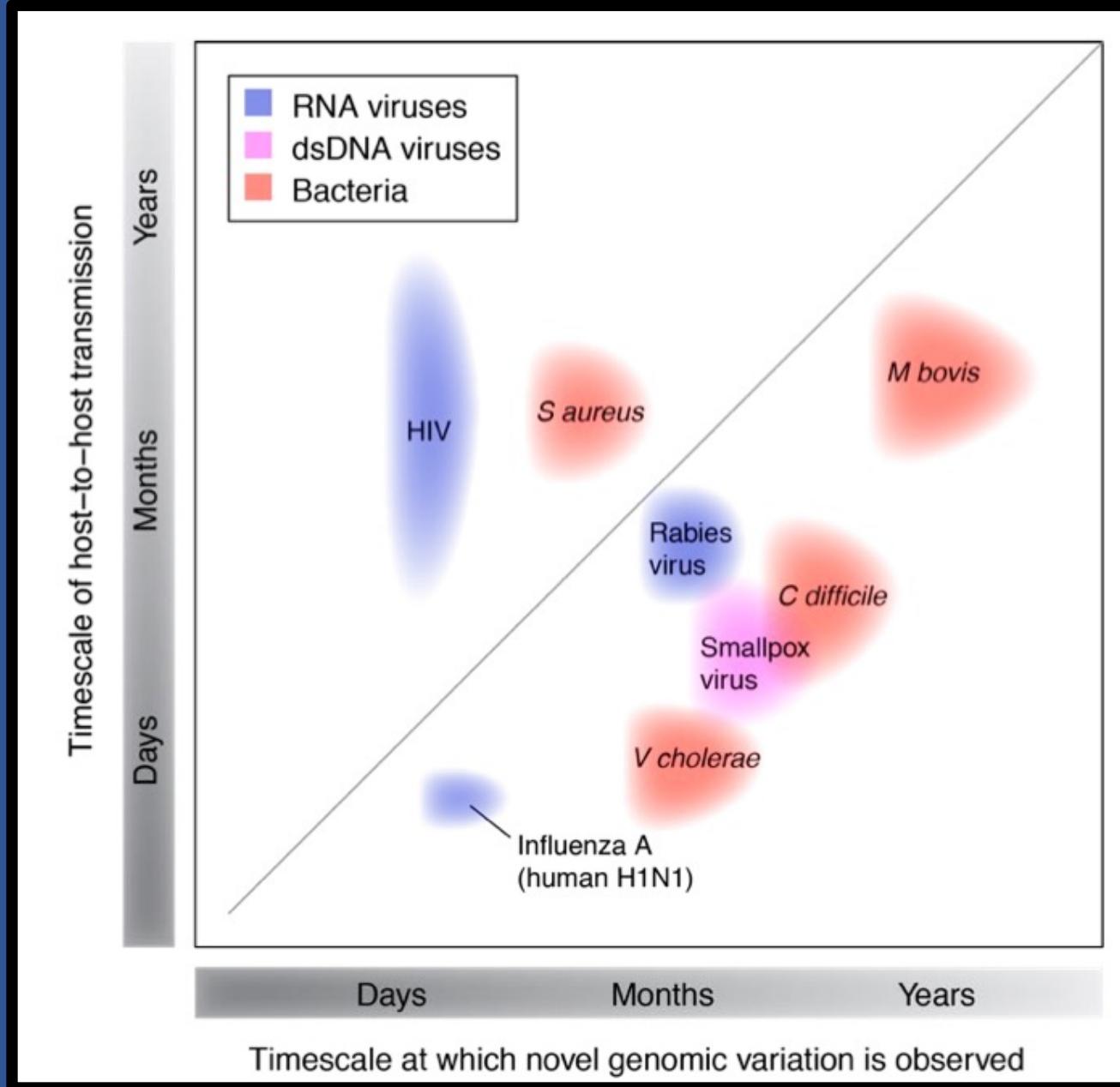
Evolutionary rate per genome  
much more similar than per  
site

DNA viruses and bacteria are  
measurably evolving – can  
use whole genome  
sequencing



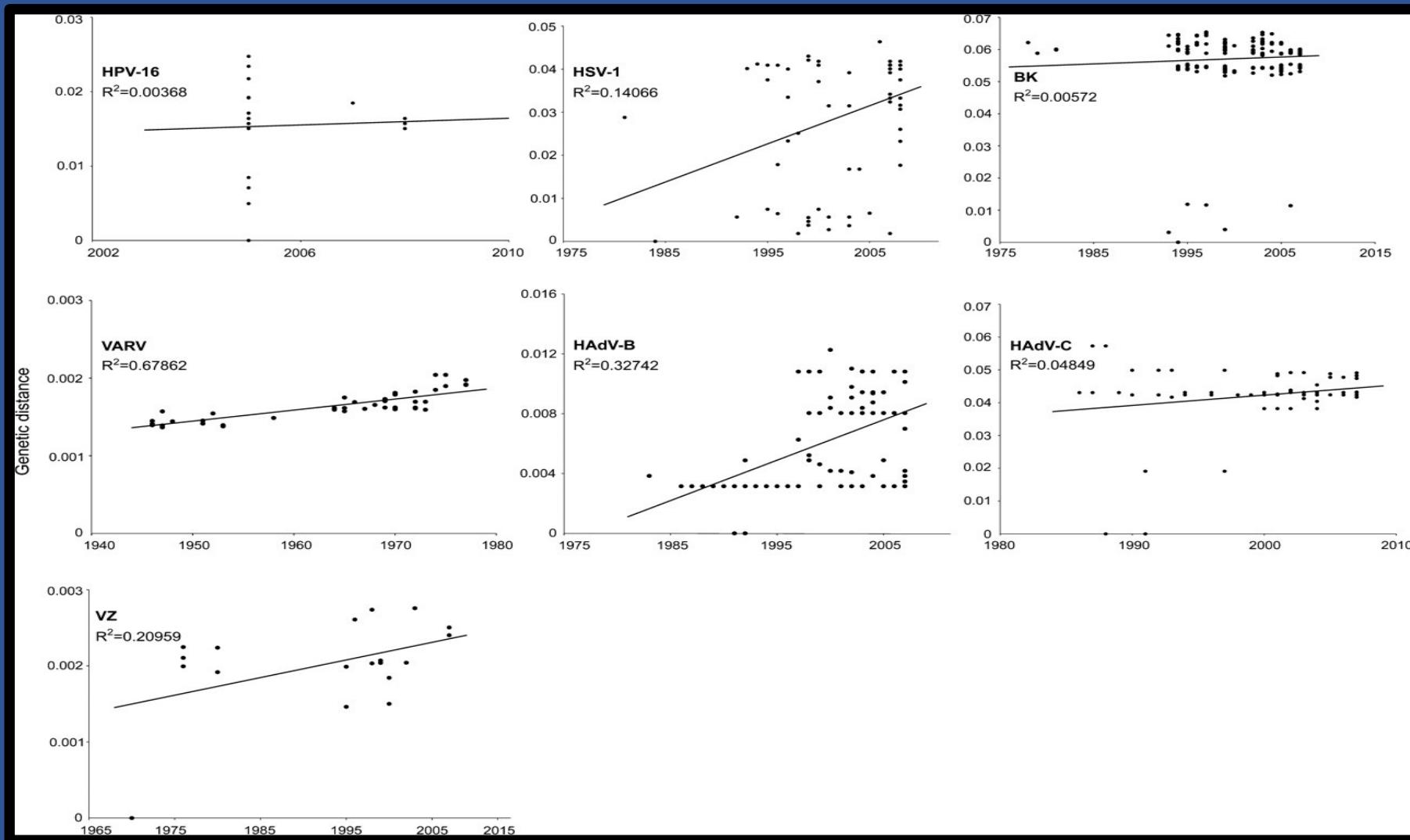
Biek et al. 2015, Trends Ecol Evol

# Relative time scales of mutational and epidemiological events



Biek et al. 2015,  
Trends Ecol Evol

# Genetic distance versus sampling year for the dsDNA viruses



# Virus origins and molecular clocks

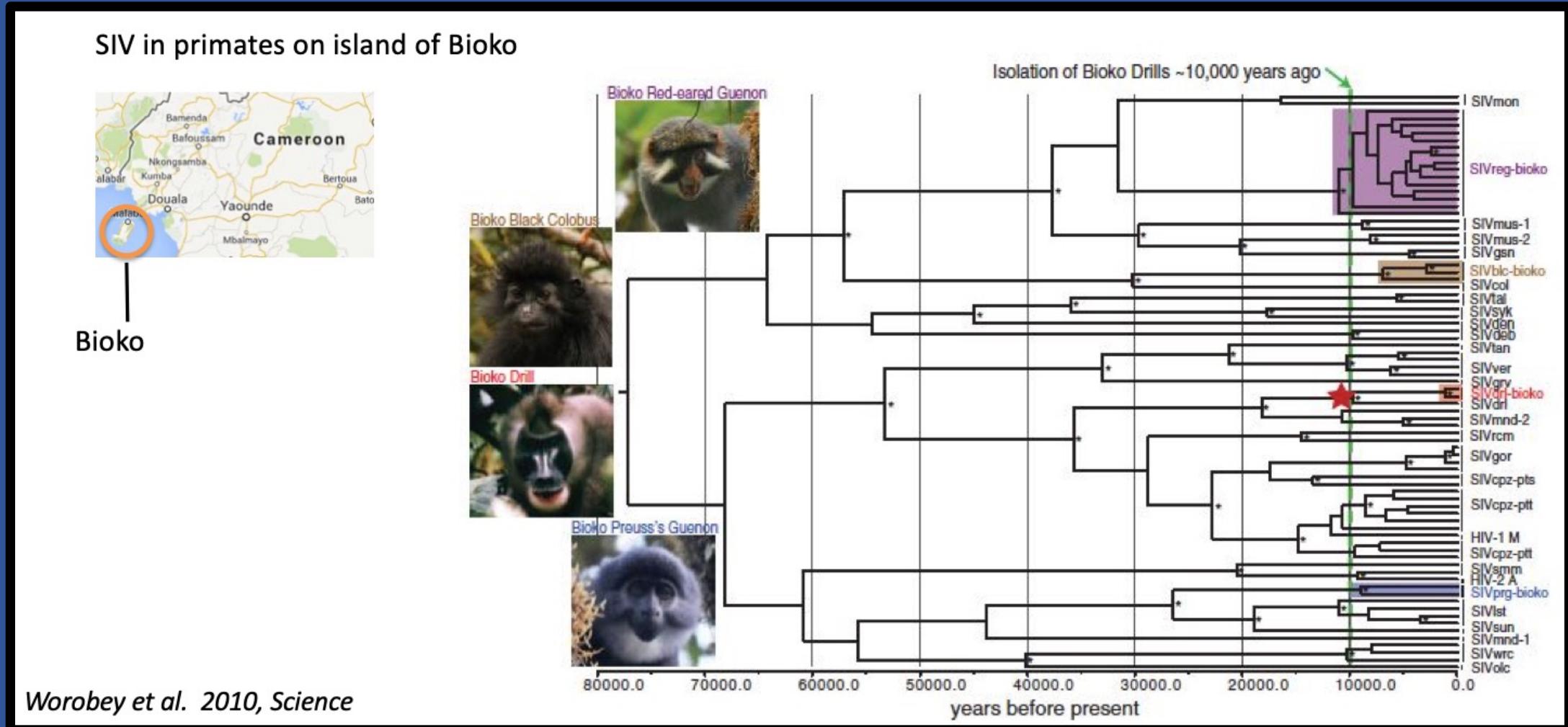
RNA viruses have been infecting vertebrates for probably millions of years

Some clear examples of co-evolution (e.g. foamy viruses, Switzer et al. 2005, Nature)

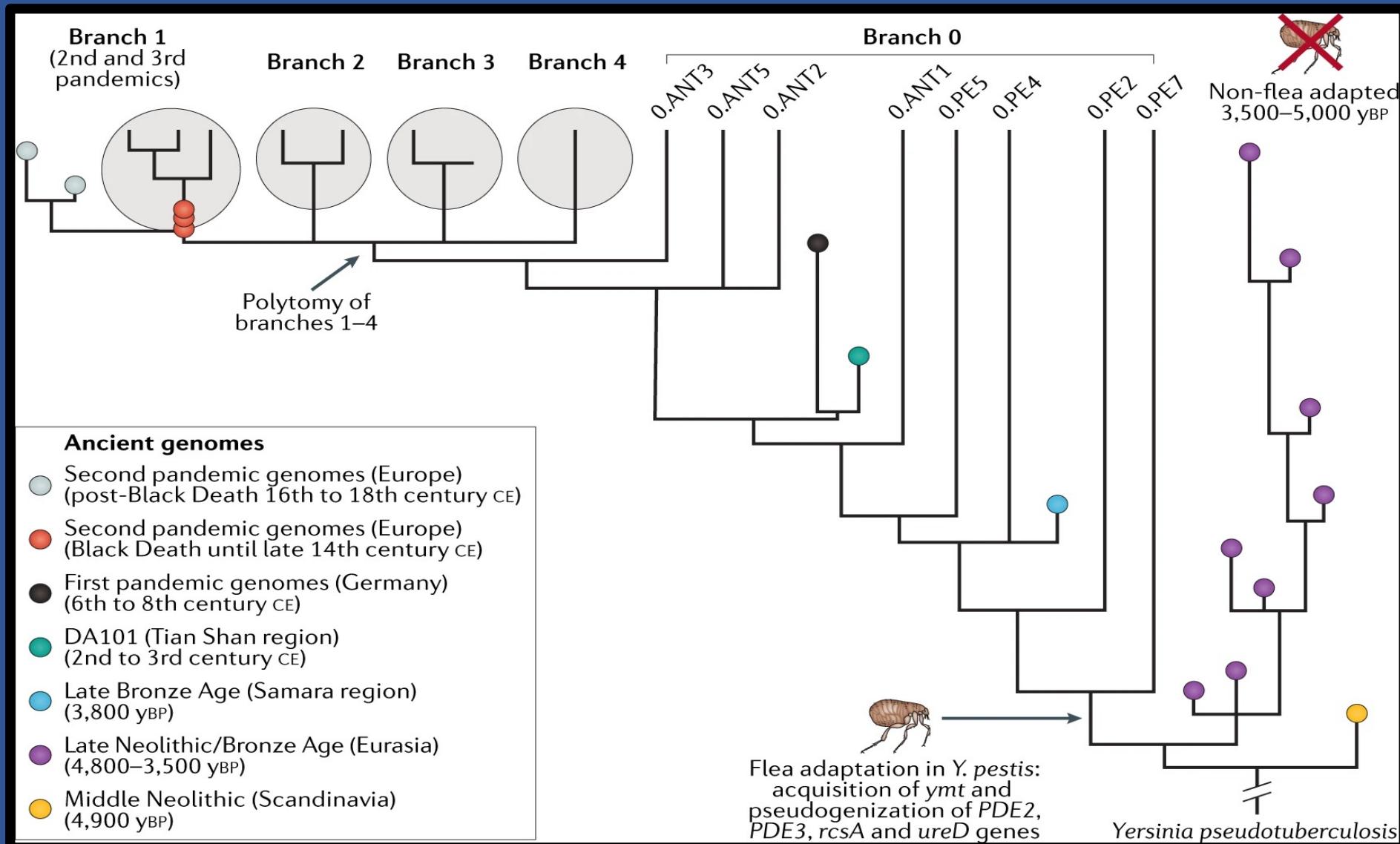
But for most RNA viruses, molecular clocks suggest common ancestors of a few thousand years (Holmes 2003, J Virol)

Clock rates estimated over time-spans of years or decades only valid on those time scales!

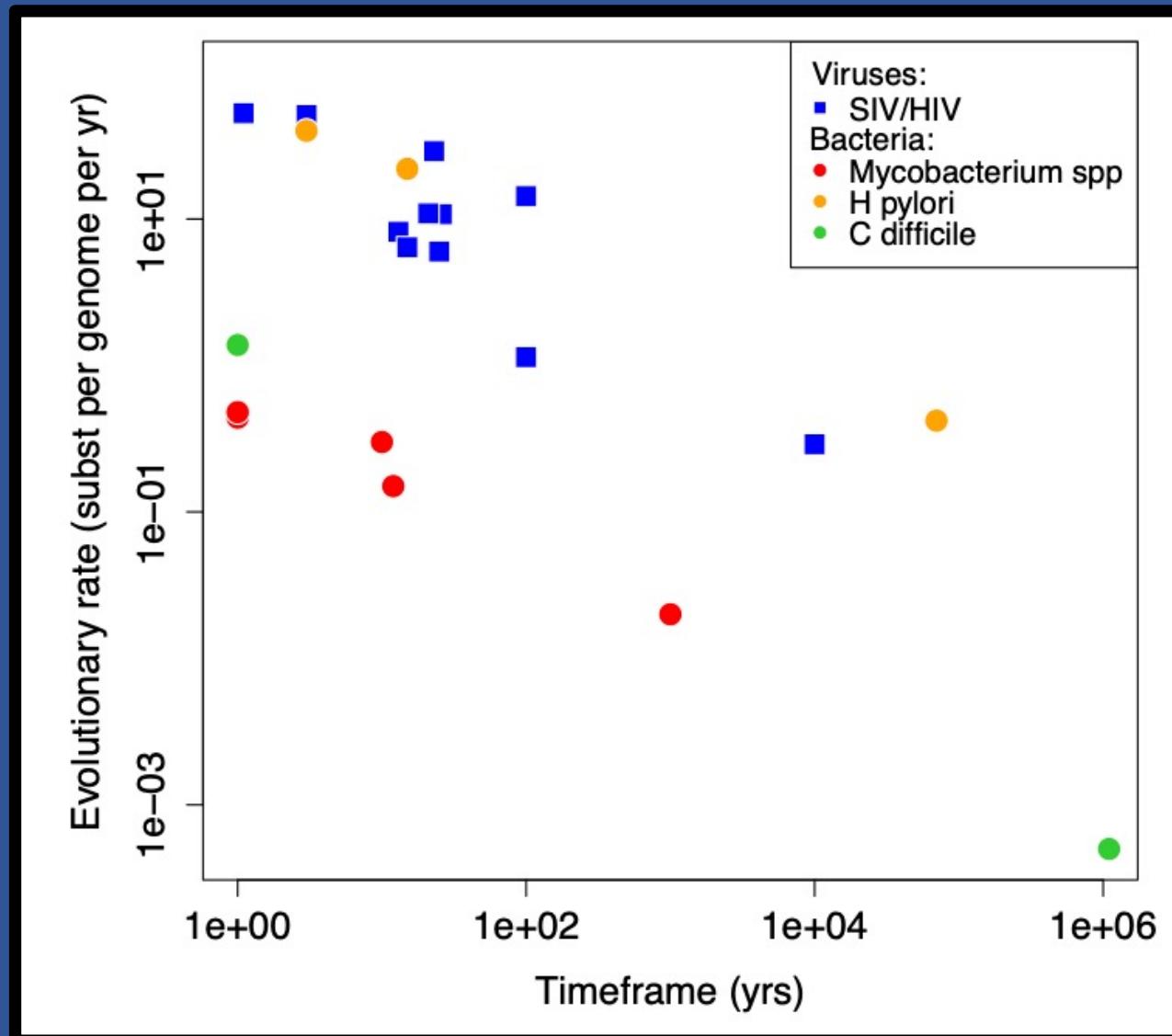
# Improving rate estimates through node calibrations



# Evolutionary history of *Yersinia pestis*



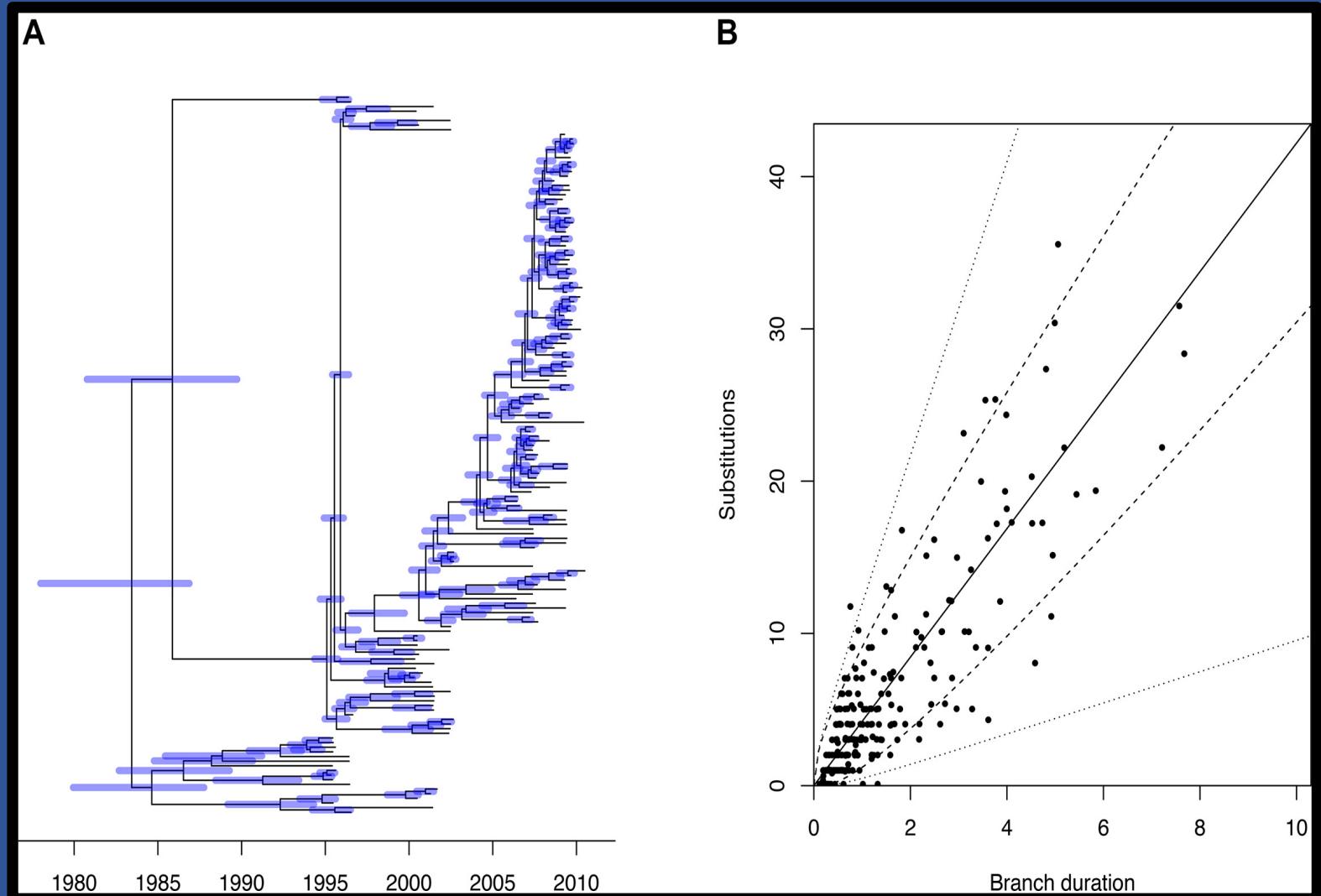
# Ancient genomes reveal time-dependent decline in evolutionary rates



Ancient genomes reveal  
time-dependent decline in  
evolutionary rates

# Fast molecular clock analyses of bacterial genomes

- BacDating – R package that implements Bayesian method to construct dated phylogenies for bacterial genomes (Didelot 2018, Nucleic Acids Research)
- Based on single genome so ignoring phylogenetic uncertainty



# Using molecular clocks to inform transmission trees

Moving from overall properties of an epidemic to specifically how samples relate to each other in a transmission chain.

Multiple approaches that use the concept of molecular clocks and rapidly evolving pathogens

But, unlike coalescent models assume nearly complete sampling

## **Genomic Epidemiology Analysis of Infectious Disease Outbreaks Using TransPhylo**

Xavier Didelot,<sup>1,8</sup> Michelle Kendall,<sup>1</sup> Yuanwei Xu,<sup>2</sup> Peter J. White,<sup>3,4,5,6</sup> and Noel McCarthy<sup>7</sup>

# PHYLOSCANNER: Inferring Transmission from Within- and Between-Host Pathogen Genetic Diversity

Chris Wymant,<sup>\*,†,1,2</sup> Matthew Hall,<sup>†,1,2</sup> Oliver Ratmann,<sup>2,3</sup> David Bonsall,<sup>1,4,5</sup> Tanya Golubchik,<sup>1,5</sup> Mariateresa de Cesare,<sup>5</sup> Astrid Gall,<sup>6</sup> Marion Cornelissen,<sup>7</sup> Christophe Fraser,<sup>\*,1,2</sup> STOP-HCV Consortium, The Maela Pneumococcal Collaboration, and The BEEHIVE Collaboration<sup>‡</sup>

<sup>1</sup>Big Data Institute, Li Ka Shing Centre for Health Information and Discovery, Nuffield Department of Medicine, University of Oxford, United Kingdom

<sup>2</sup>Department of Infectious Disease Epidemiology, Medical Research Council Centre for Outbreak Analysis and Modelling, Imperial College London, London, United Kingdom

<sup>3</sup>Department of Mathematics, Imperial College London, London, United Kingdom

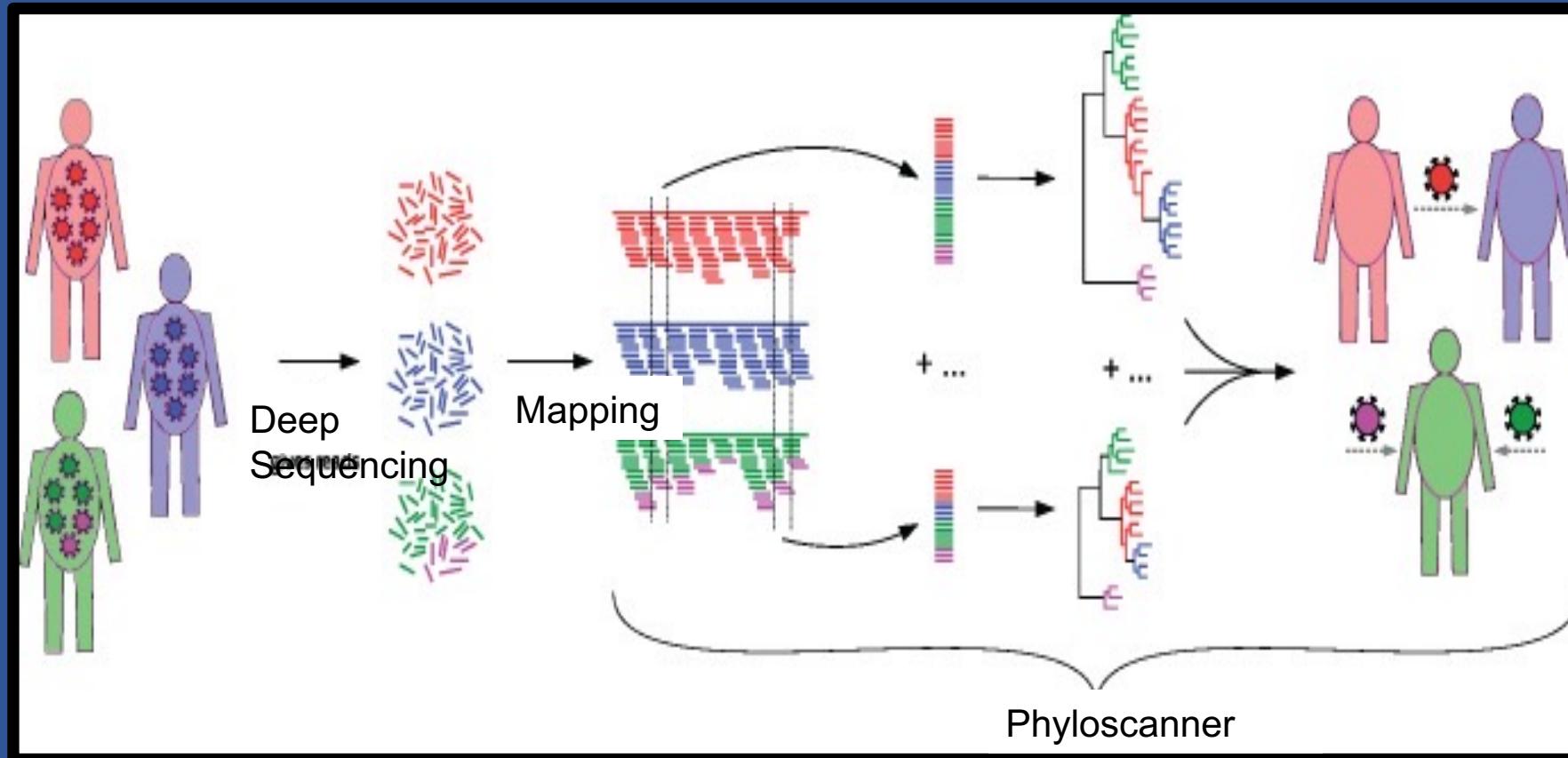
<sup>4</sup>Peter Medawar Building for Pathogen Research, Nuffield Department of Medicine and the NIHR Oxford BRC, University of Oxford, United Kingdom

<sup>5</sup>Wellcome Centre for Human Genetics, Nuffield Department of Medicine, University of Oxford, United Kingdom

<sup>6</sup>Department of Veterinary Medicine, University of Cambridge, Cambridge, United Kingdom

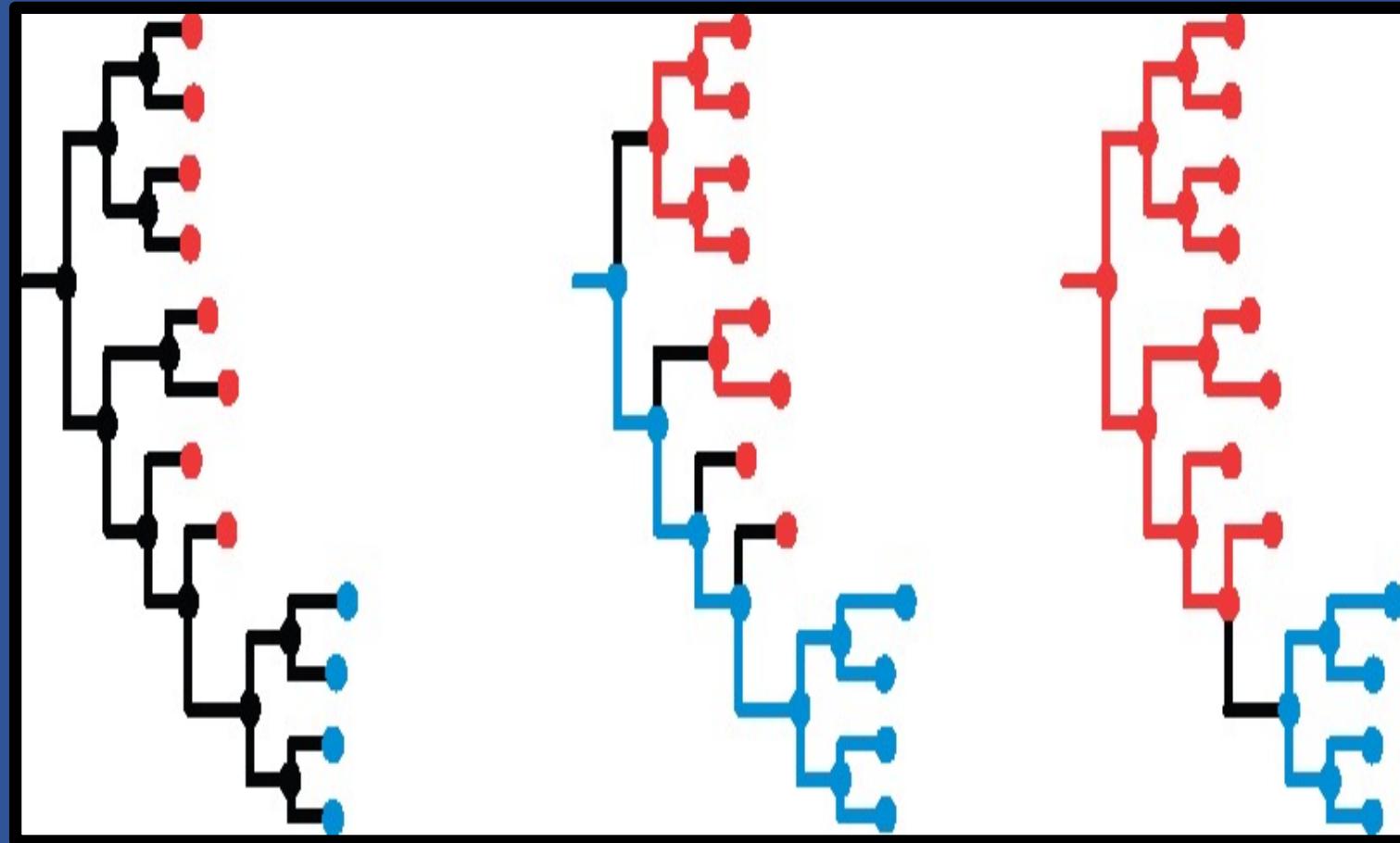
<sup>7</sup>Laboratory of Experimental Virology, Department of Medical Microbiology, Center for Infection and Immunity Amsterdam (CINIMA), Academic Medical Center of the University of Amsterdam, Amsterdam, The Netherlands

# Incorporates within host variation

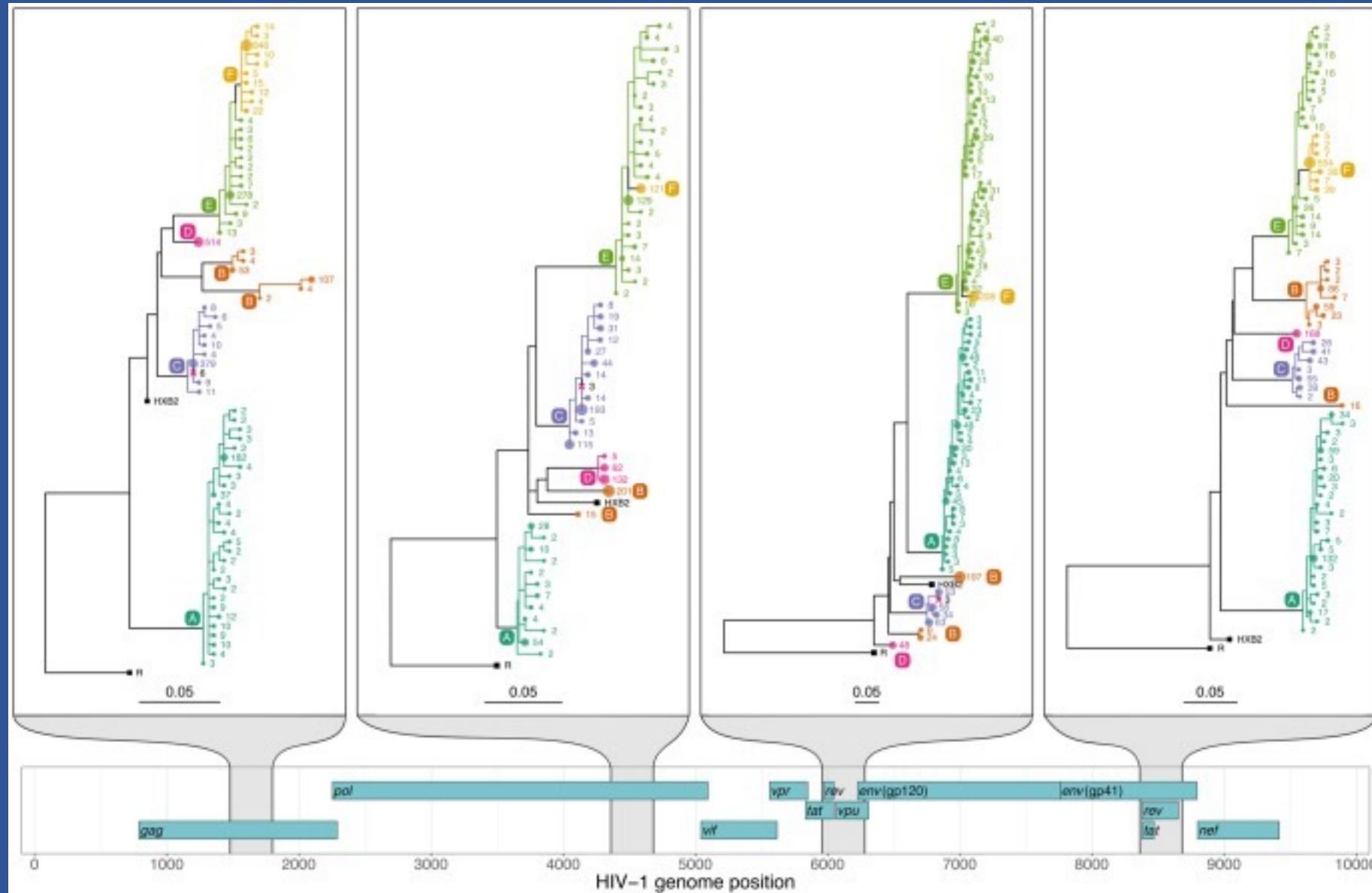


Wymant, et al. *Molecular biology and evolution*, 35(3), 719-733.

# Incorporates within host variation



# Transmission trees informed by within host variation



# Molecular clock summary

- Measurably evolving populations accumulate substitutions over the time span of sampling
- Means that molecular clock can be estimated and used to put time-scale on phylogeny
- Used to be RNA viruses only with genomic data also applicable to bacteria and DNA viruses
- Pitfalls:
  - insufficient temporal signal (should check!)
  - rate depends on time scale of sampling

# Working with...





Program geared towards molecular clock analyses

Estimating phylogenies is not the focus: instead uses MCMC to average over tree space

Original developers:  
Andrew Rambaut ( U Edinburgh) – maintains version 1

Alexei Drummond (U Auckland) – started BEAST2  
Permits more user packages

# Some common questions applied to BEAST

- How does geography influence evolution between samples/populations?
- What is the timeline of evolution across...(samples, species, etc)?
- What is the rate of evolution across my tree?
  - Does the evolutionary rate differ along different branches?
- Approximately when did xx variant emerge?
- What is the relationship between hosts/pathogens?
  - Cross-species transmission vs. host specificity
- How has a particular phenotype evolved across taxa?

# More of a package than a single software...



BEAUTi – Input data, set parameters, create the BEAST input file

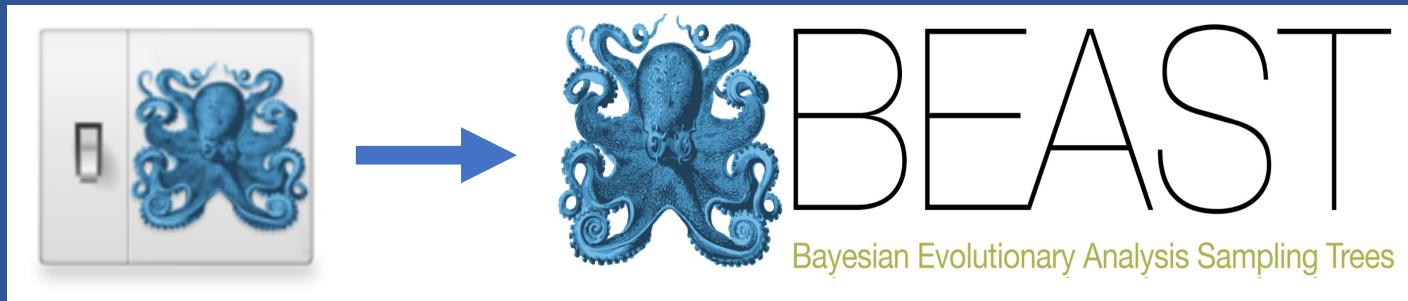
# More of a package than a single software...



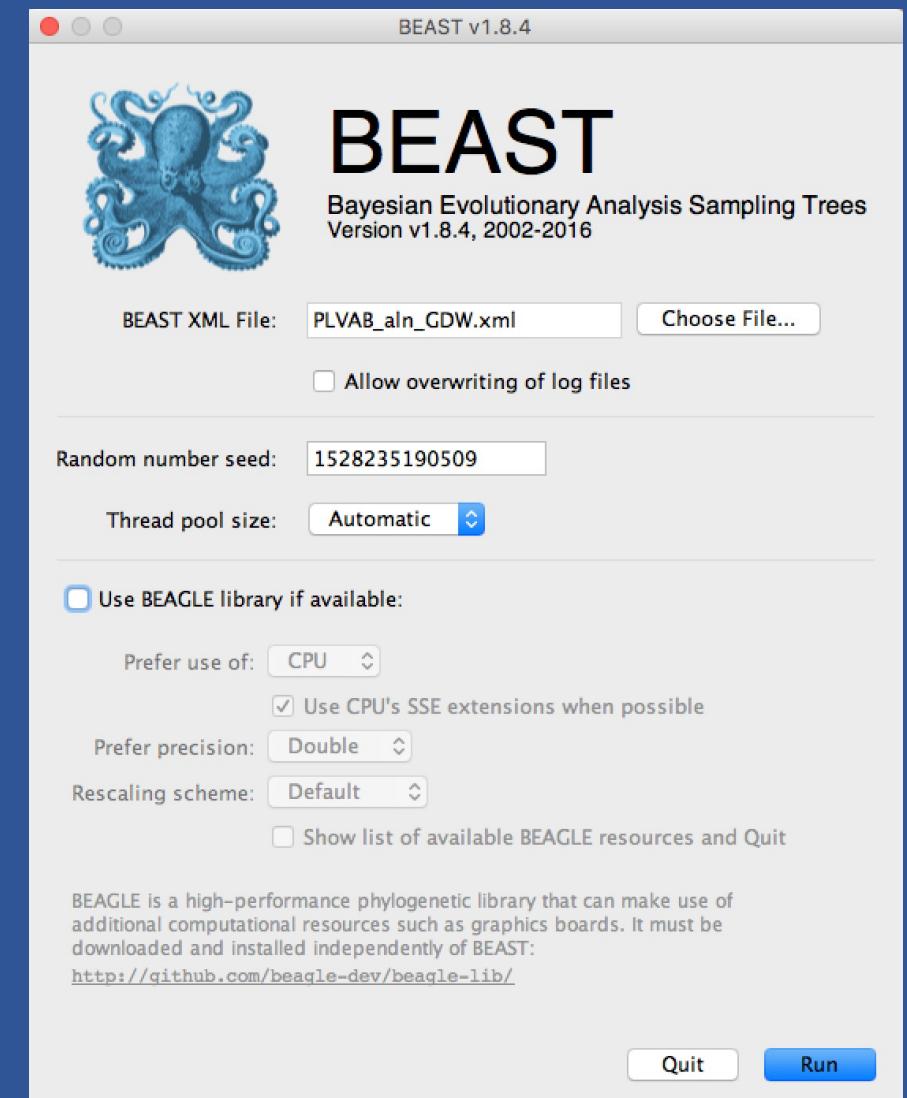
## BEAUTi XML Format

```
<!-- The uncorrelated relaxed clock (Drummond, Ho, Phillips & Rambaut (2006) PLoS
Biology 4, e88 )-->
<discretizedBranchRates id="branchRates">
  <treeModel idref="treeModel"/>
  <distribution>
    <logNormalDistributionModel meanInRealSpace="true">
      <mean>
        <parameter id="ucl.d.mean" value="0.1" lower="0.0"/>
      </mean>
      <stdev>
        <parameter id="ucl.d.stdev" value="0.3333333333" lower="0.0"/>
      </stdev>
    </logNormalDistributionModel>
  </distribution>
  <rateCategories>
    <parameter id="branchRates.categories"/>
  </rateCategories>
</discretizedBranchRates>
<rateStatistic id="meanRate" name="meanRate" mode="mean" internal="true"
external="true">
  <treeModel idref="treeModel"/>
  <discretizedBranchRates idref="branchRates"/>
</rateStatistic>
<rateStatistic id="coefficientOfVariation" name="coefficientOfVariation"
mode="coefficientOfVariation" internal="true" external="true">
  <treeModel idref="treeModel"/>
  <discretizedBranchRates idref="branchRates"/>
</rateStatistic>
<rateCovarianceStatistic id="covariance" name="covariance">
  <treeModel idref="treeModel"/>
  <discretizedBranchRates idref="branchRates"/>
</rateCovarianceStatistic>
```

# More of a package than a single software...



BEAUTi

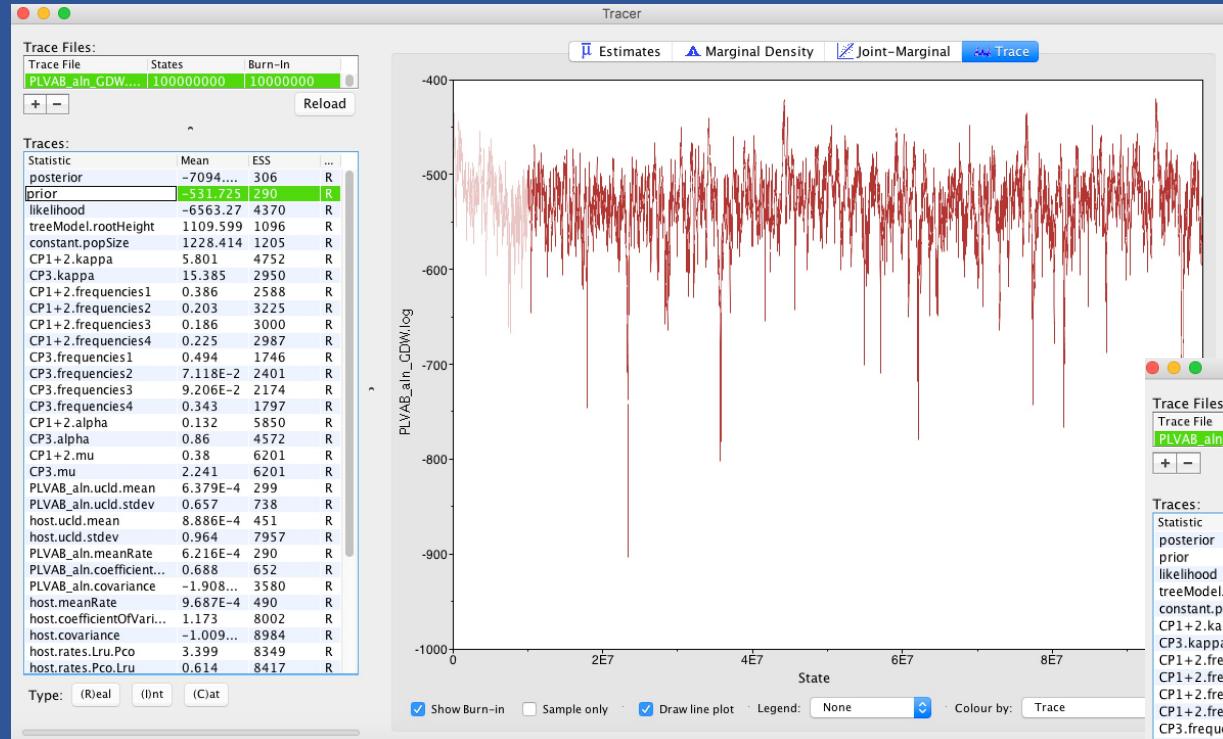


# More of a package than a single software...

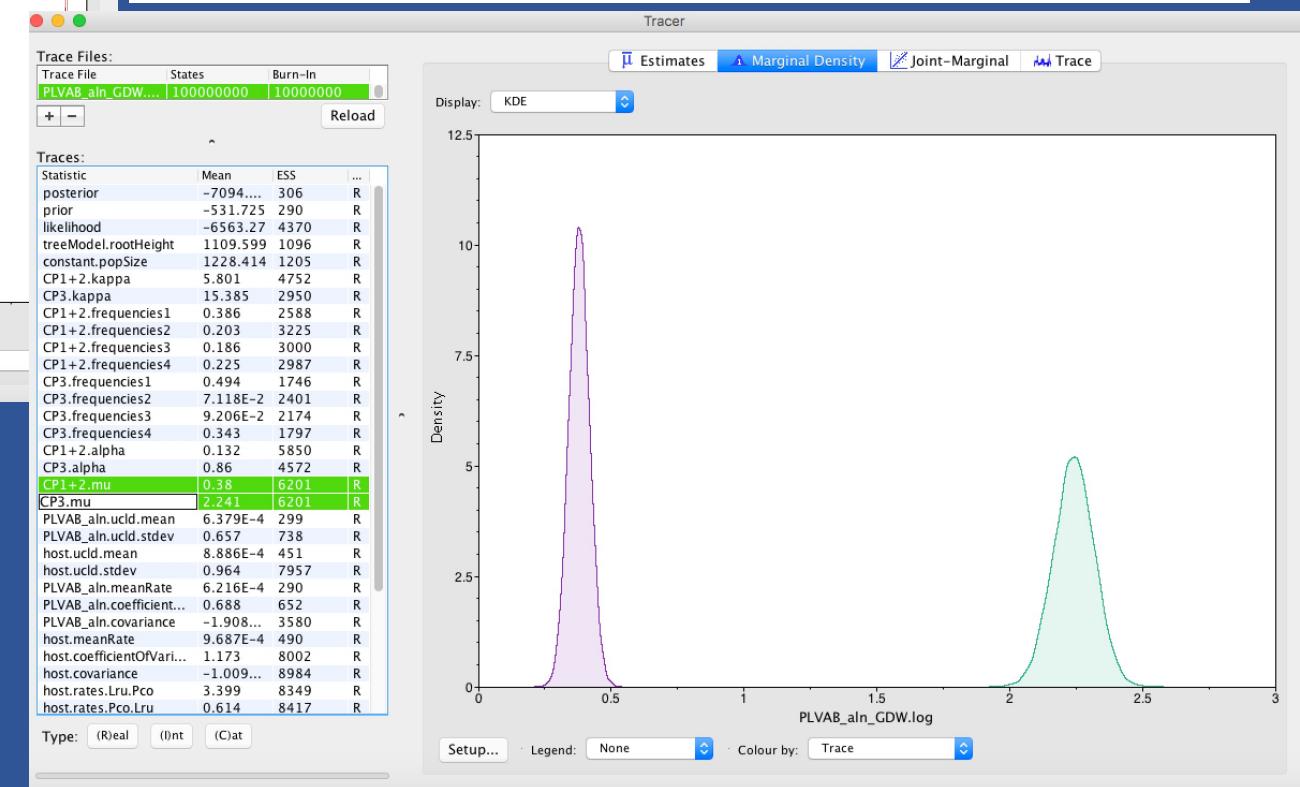


BEAUTi

Tracer – view  
BEAST output and  
summarize results



**Tracer** | This is a graphical program for exploring the output of BEAST, diagnosing problems, and summarizing the results.



# More of a package than a single software...



BEAUTi

Tracer



Tree Annotator

Produces a consensus tree  
from BEAST output

# More of a package than a single software...



BEAUTi

Tracer



Tree Annotator

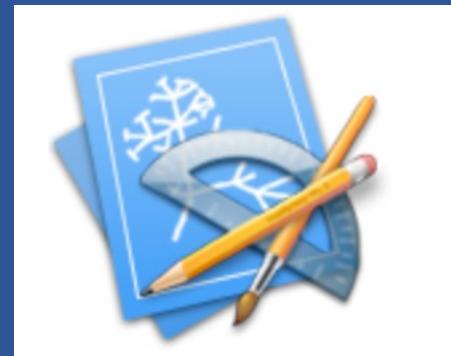
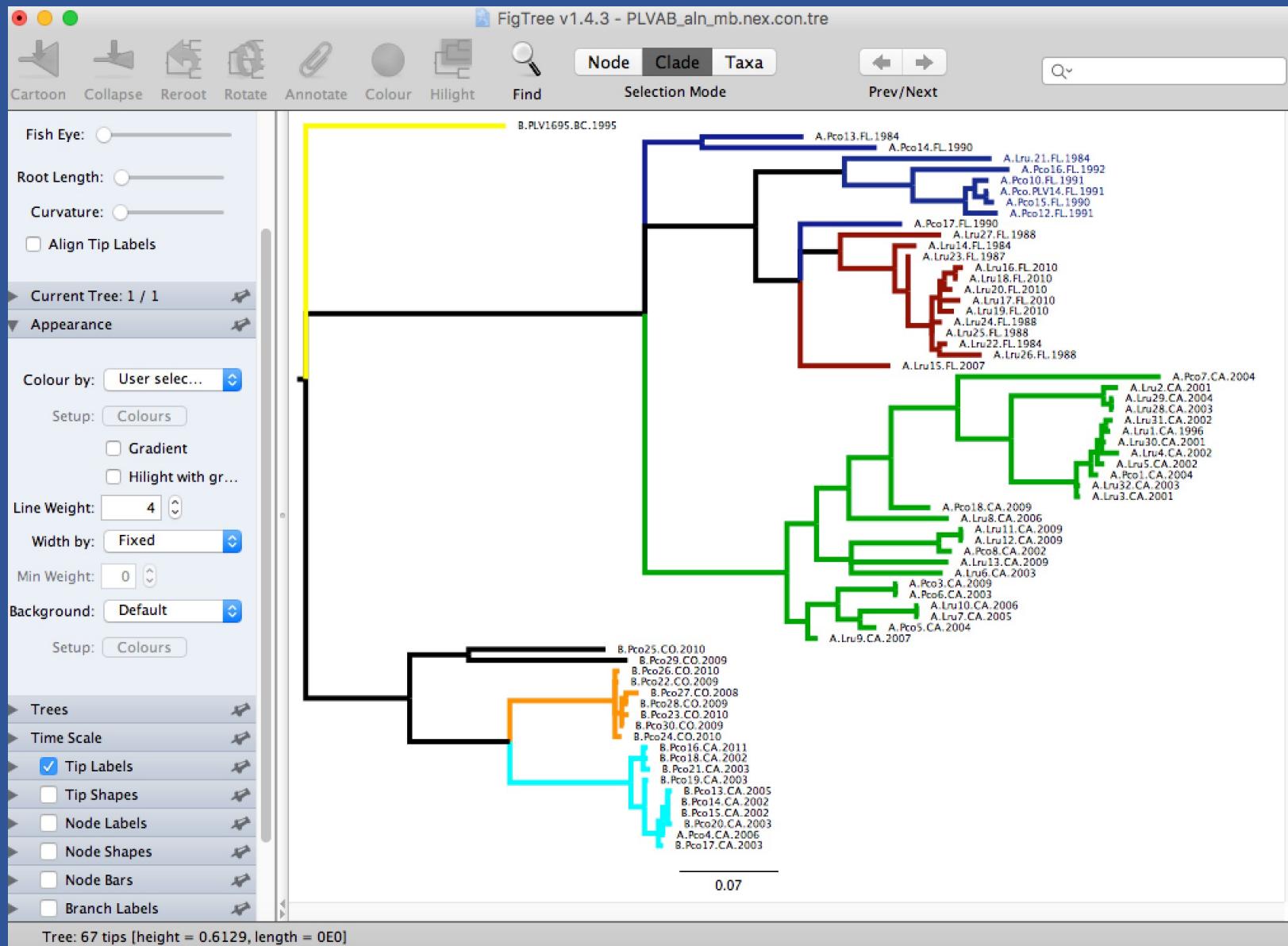


Fig Tree

View consensus tree  
and parameter estimates



# Tracer



# Fig Tree

View/edit consensus tree  
and parameter estimates

# Recap of process leading up to BEAST...

- Question/hypothesis > Grant > Samples > Nucleic acid extraction
- > NGS/Sanger > De novo assembly/mapping to reference
- > Consensus sequences (+/- metadata) > Alignment/trimming
- > Test for recombination > Nucleotide substitution model selection
- > Revisit your question
- > Choose appropriate phylogenetic analysis to answer your question!
- > If above involves BEAST...proceed!
- > Validate BEAST model/run and priors\*