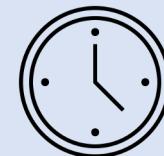
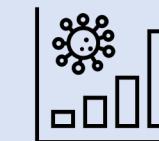
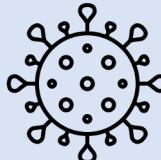
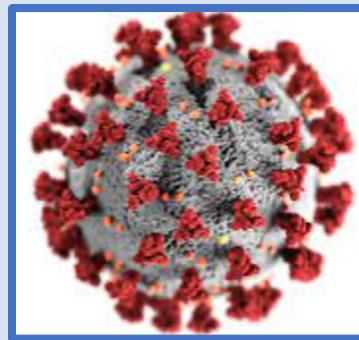
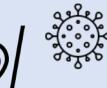
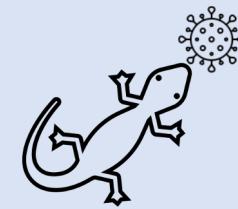
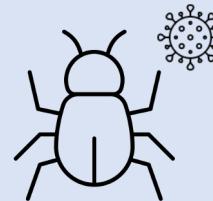
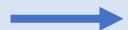
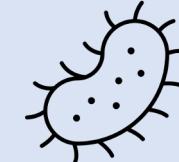
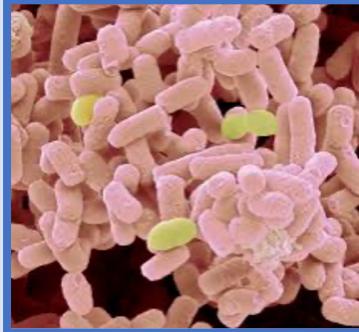




Temporal signal

Ash Porter

Phylogenetics workshop
University of Melbourne, 2021



Evolution and origins of zoonotic viral pathogens

Clock-like evolution

- Nucleic acids (DNA, RNA) gradually change as they are inherited from the previous generation.
- This process can occur linearly with time.
- Applying the neutral theory of molecular evolution;

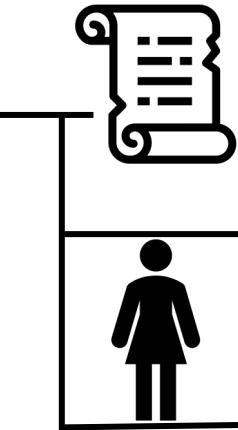


Zuckerkandl, Pauling

Kimura

Most changes are neutral, therefore mutations will be fixed in the population at a clock-like rate equal to the mutation rate.

"A bird, came down the walk"



"A bird, came down the path"



"A bird, came across the path"



"A small bird, came across my path suddenly"



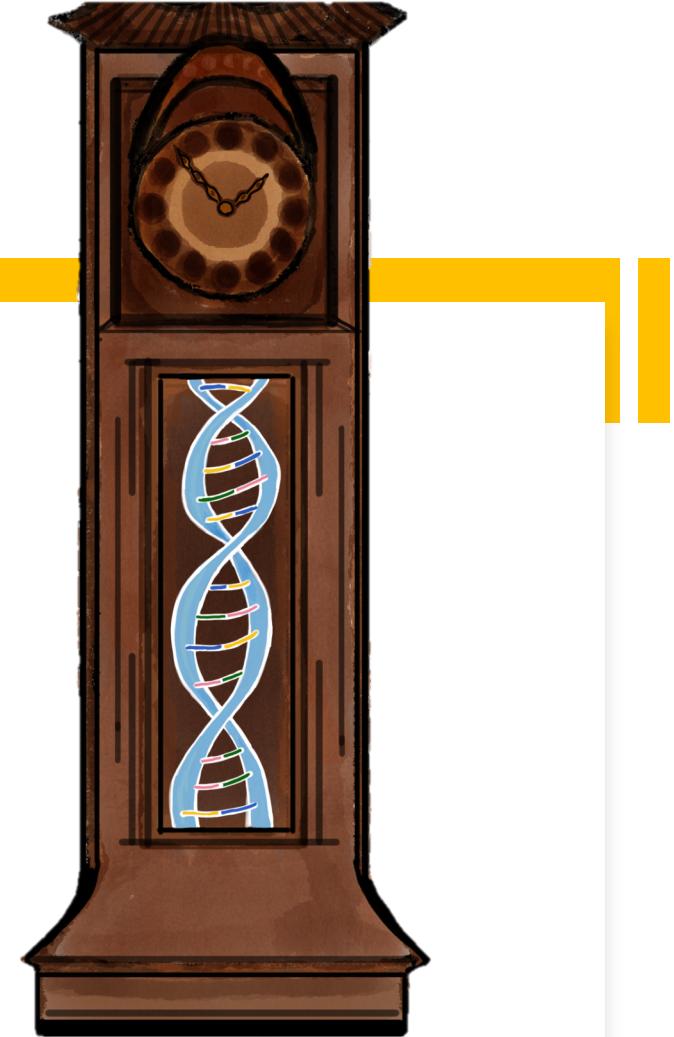
"A small wren, came across my path suddenly"



"A tiny wren, came across my path suddenly"

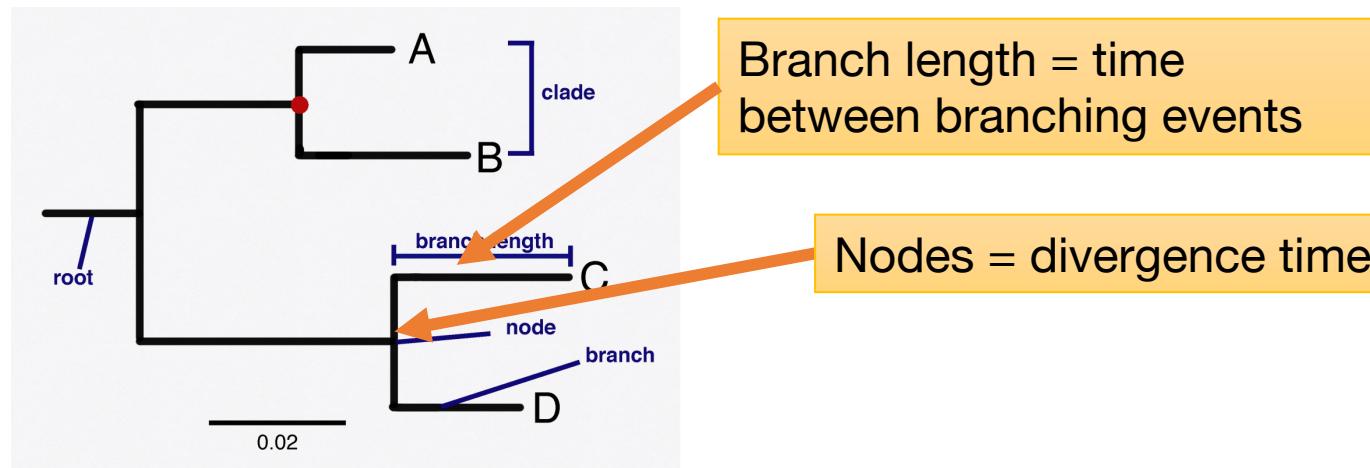
The molecular clock

- Time trees and evolutionary rates
- Strict clocks vs relaxed clocks
- Calibrating the molecular clock



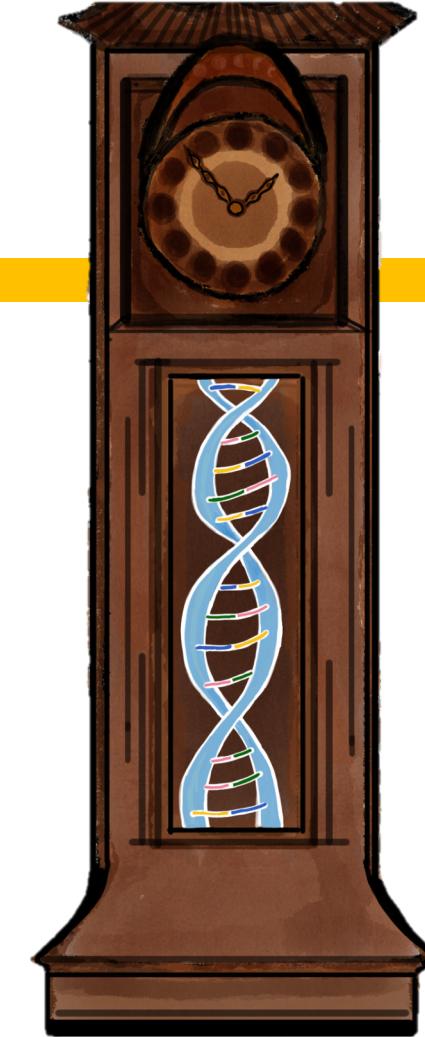
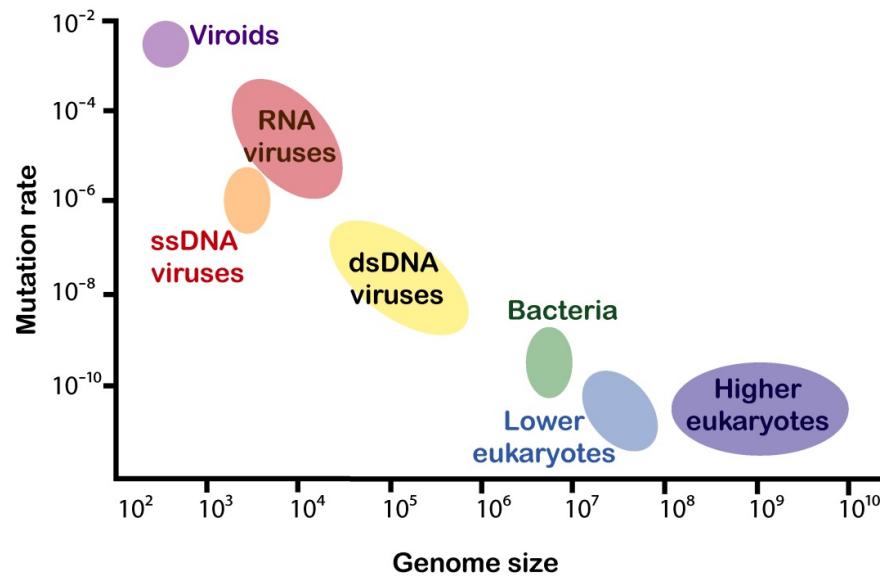
The molecular clock

- **Time-trees and evolutionary rates**
 - Rooted phylogenetic trees scaled to time are known as time-trees.



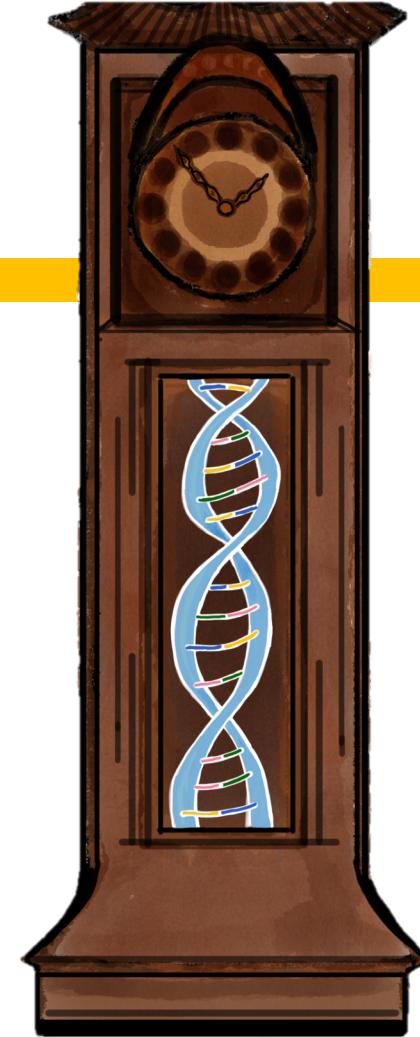
The molecular clock

- Time-trees and evolutionary rates
 - Evolutionary rates vary throughout the tree of life.



The molecular clock

- Strict clocks vs relaxed clocks:
 - Strict clock: assumes that every branch along the tree is evolving at the **same evolutionary rate**
 - Uncorrelated relaxed clock: allows each branch to have a **unique evolutionary rate** (1).
 - Fixed local clock: Allows **one or more clades** in the phylogeny to evolve at a **different rate** to the global rate (2).
 - Random local clock: proposes a series of local clocks over regions of the phylogeny, **allowing branches to change from one local clock to another** (3).



Why is temporal signal useful?



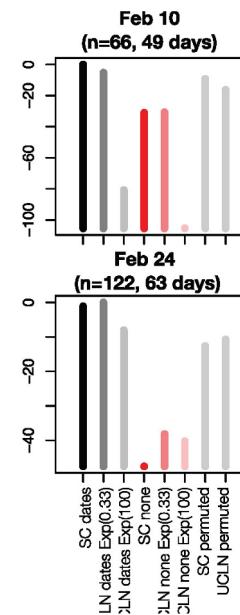
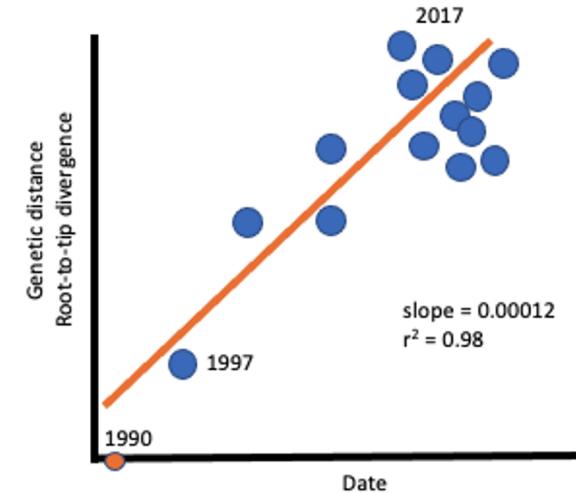
What is temporal signal?

We assume molecular evolution occurs over time at a predictable rate, and that our sampling period is sufficiently large enough to capture measurable evolutionary change

- When using genomic sequences, we can utilize sampling times to calibrate the **molecular clock**:
 - Estimate the evolutionary rate and dates.
- However, the molecular clock assumes that the population is measurably evolving:
 - Sampling window is wide and evolutionary rate is high.
- Sequence data needs to have evidence of temporal signal for the estimates of the molecular clock to be accurate (4,5).

How can we test for temporal signal?

- Informal assessments:
 - **Root-to-tip**
 - Not a statistical test
- Formal assessments:
 - **BETS**
 - Statistical test



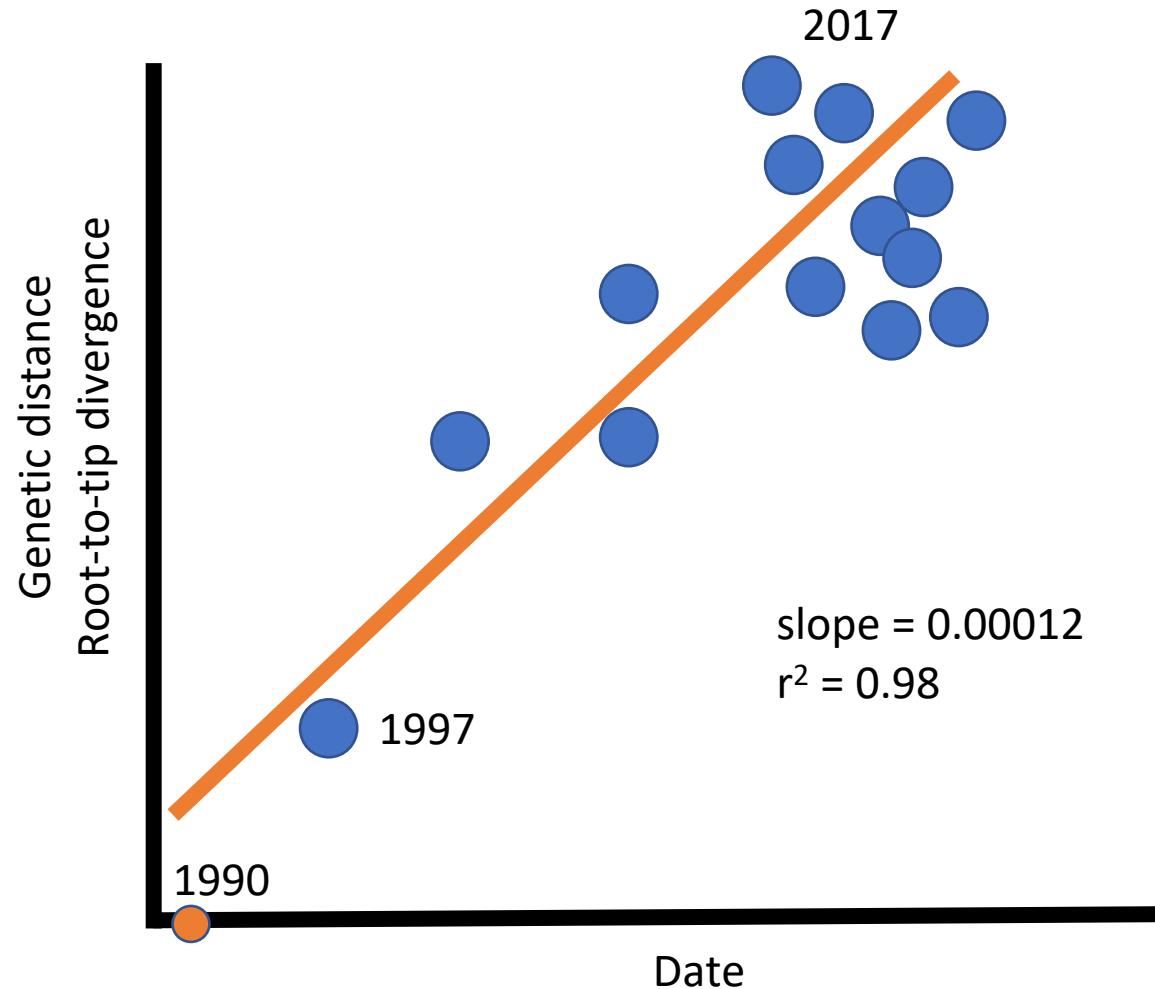
Root-to-tip regression

- We can assess the temporal signal of a dataset via root-to-tip regression, in software such as TempEst (6).
- Uses an unrooted phylogenetic tree (such as ML tree, where branch lengths are in units of substitutions/site).
- Regression of the distance from the root to each of the tips, as a function of sampling times.

An example of root-to-tip regression

- Virus A appears in 1997
- Virus A is sequenced sporadically until 2017
- Under the assumptions that there is measurable evolution:

The **slope** is an estimate of evolutionary rate
 r^2 reflects the degree of clocklike evolution
The **x-intercept** is the time of origin



Bayesian evaluation of temporal signal (BETS)

- BETS compares the fit of models when using:
 1. Correct sampling times
 2. No sampling times
 3. Permuted sampling times
- If the sequence data does have temporal signal, the correct sampling times *should* have the best statistical fit (7).
- The **Bayes factor** (8) is used to determining the most accurate model.
- The model "fitness", in Bayesian terms, is the marginal likelihood. The preference for one model over another is determined by ratio of marginal likelihoods, which is the **Bayes factor**.

An example of BETS

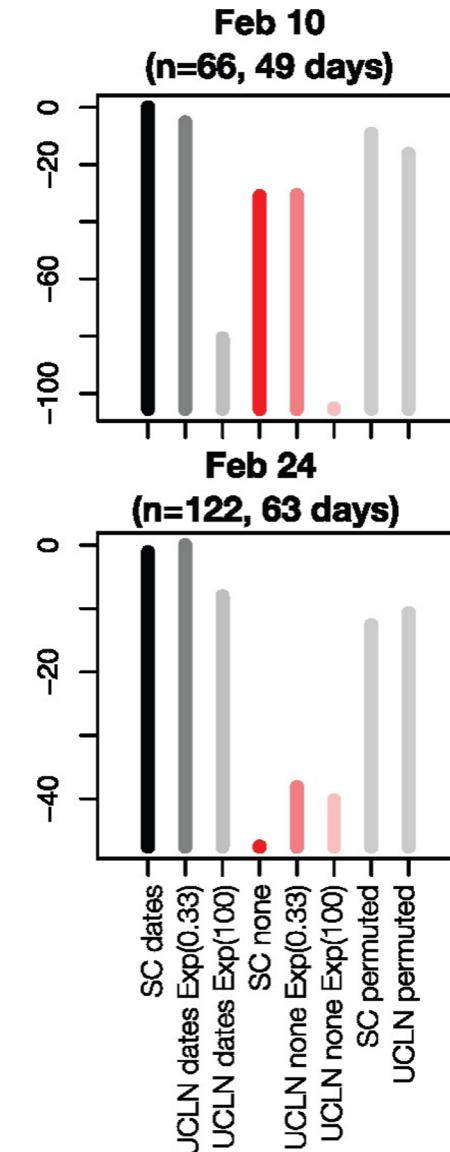
The y-axis represents the log Bayes factors
Compares the strict clock (SC) and relaxed clock
(UCLN).

Black and **grey** bars represent analysis with correct sampling times.

Light red and **red** bars represent no sampling times.

Light grey bars represent permuted sampling times.

The **Bayes factor** is used to determine the preference for one model over another, calculated by the ratio of marginal likelihoods.



Why is temporal signal important?

- Essential for calibrating the molecular clock, and to have accurate inferences in downstream phylodynamics analyses.
- Examining the temporal signal of sequences:
 - To identify **outliers**, and to get a rough estimate of the **TMCRA** and **evolutionary rate**.

What happens if your data doesn't have evidence for temporal signal?

- If there is no temporal signal, you cannot assume that any estimates from the molecular clock are accurate.
- Reporting evolutionary rates from analyses where there is an absence of temporal signal are **invalid**.

References and further reading

References:

1. Drummond AJ, et al. (2006) Relaxed phylogenetics and dating with confidence. *PLoS Biology* 4(5): e88.
2. Yoder A. and Yang Z (2000) Estimation of primate speciation dates using local molecular clocks. *Mol. Biol. Evol.* 17(7):1081-1090.
3. Drummond AJ. and Suchard MA. (2010) Bayesian random local clocks, or one rate to rule them all. *BMC Biology* 2010, 8:114
4. Duchêne S. et al. (2015) 'The Performance of the Date-Randomization Test in Phylogenetic Analyses of Time-Structured Virus Data', *Molecular Biology and Evolution*, 32: 1895–906.
5. Murray G. G. R. et al. (2016) 'The Effect of Genetic Structure on Molecular Dating and Tests for Temporal Signal', *Methods in Ecology and Evolution*, 7: 80–9.
6. Duchene, S., Featherstone, L., et al. (2020). Temporal signal and the phylodynamic threshold of SARS-CoV-2, *Virus Evolution*, 6(2).
7. Kass R. E. Raftery A. E. (1995) 'Bayes Factors', *Journal of the American Statistical Association*, 90: 773–95.
8. Rambaut, L. et al (2016) Exploring the temporal structure of heterochronous sequences using TempEst. *Virus Evolution* 2: vew007

Further reading:

Ho, S. Y., & Duchêne, S. (2014). Molecular-clock methods for estimating evolutionary rates and timescales. *Molecular ecology*, 23(24), 5947–5965.

Duchêne S., Duchêne D.A. (2020) Estimating Evolutionary Rates and Timescales from Time-Stamped Data. In: Ho S.Y.W. (eds) *The Molecular Evolutionary Clock*. Springer, Cham.