ETH zürich

cEvo

Integrating genomic and time series data in phylodynamic analyses with TimTam

Alexander E. Zarebski[1,2], Antoine Zwaans[3,4], Bernardo Gutierrez[1], **Louis du Plessis**[3,4], Oliver G. Pybus[1,5]

[1]Department of Biology, University of Oxford; [2]School of Mathematics and Statistics, University of Melbourne; [3]Department of Biosystems Science and Engineering, ETH Zürich; [4]Swiss Institute of Bioinformatics; [5]Department of Pathobiology and Population Sciences, Royal Veterinary College London
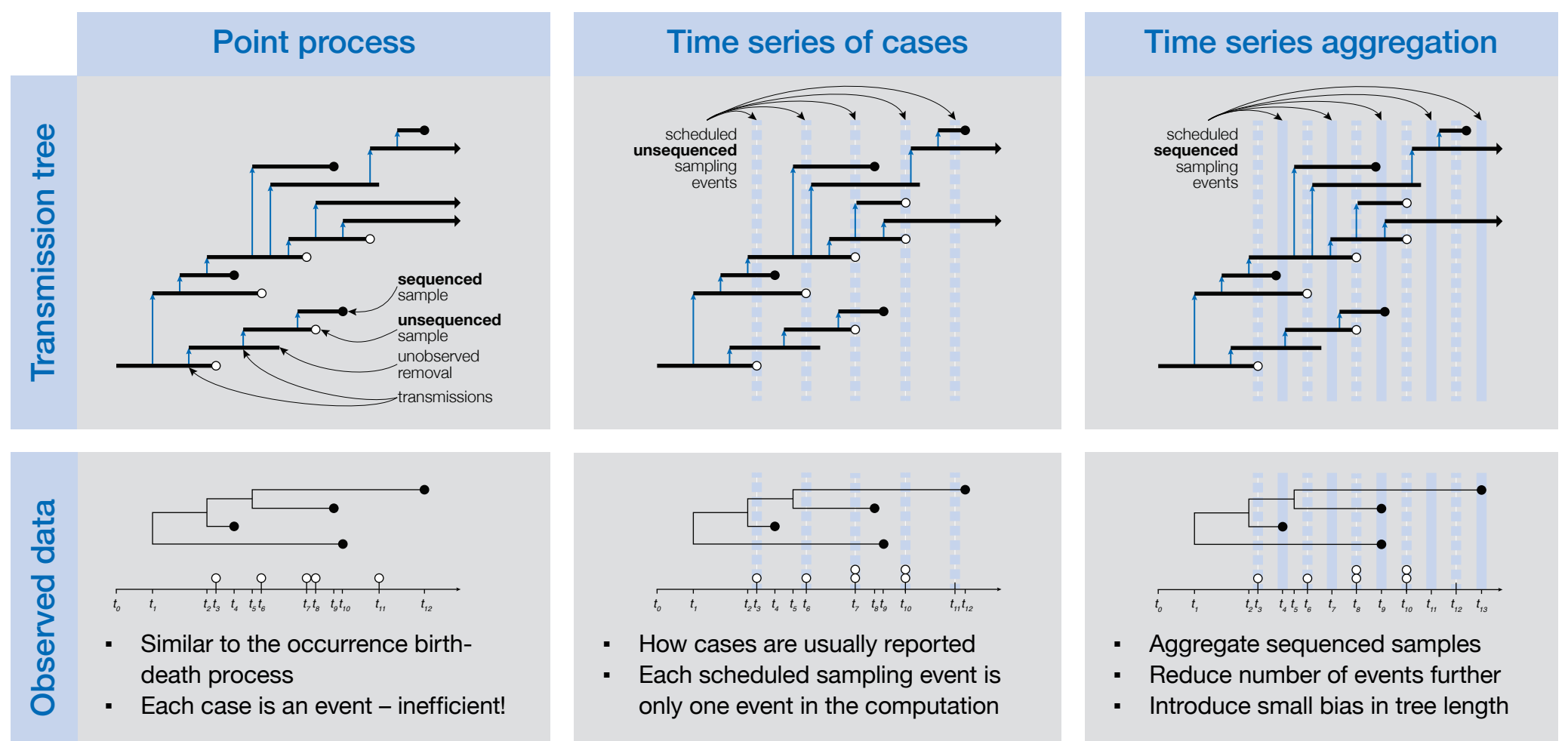
Introduction

How can we efficiently combine sequencing data with a time series of cases in an integrated analysis? The likelihood of the combined model is well-known, but computationally intractable. The time series integration method through approximation of moments (**TimTam**) uses a negative binomial approximation to the number of unobserved lineages to overcome this limitation.
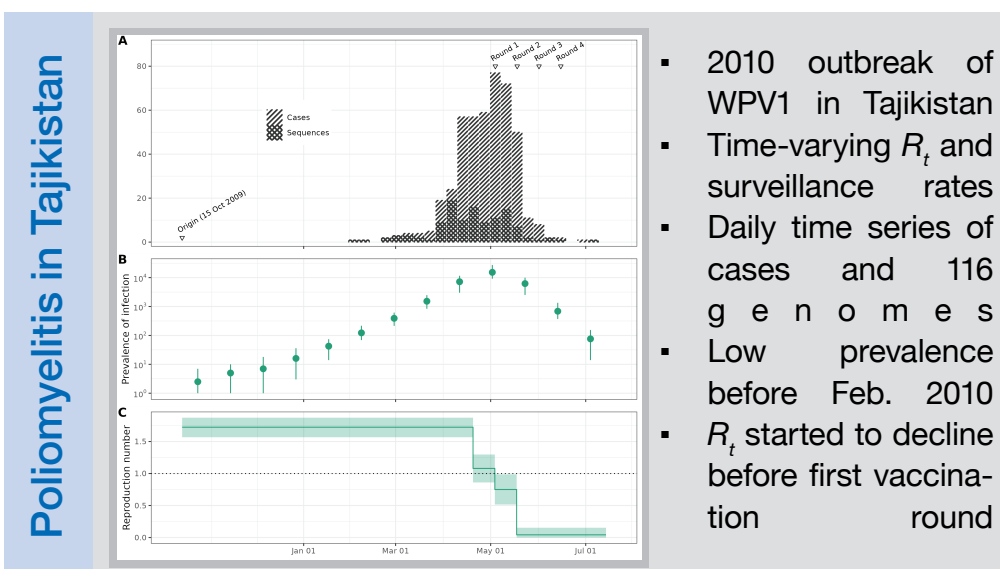
Why use TimTam?

- Available in BEAST2 via BEAUti (tutorials on GitHub)
- Linear time complexity (nr. of genomes and time series length)
- Combine time-series of cases with genomes in one analysis
- Estimate rate changes through time ("skyline")
- Estimate prevalence through time
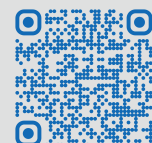- Estimate R_t in the presence of scheduled sampling events

Model specifications

Point process

Time series of cases

Time series aggregation

Transmission tree

sequenced sample
unsequenced sample
unobserved removal
transmissions

scheduled **unsequenced** sampling events

scheduled **sequenced** sampling events

Observed data

- Similar to the occurrence birth-death process
- Each case is an event – inefficient!

- How cases are usually reported
- Each scheduled sampling event is only one event in the computation

- Aggregate sequenced samples
- Reduce number of events further
- Introduce small bias in tree length

Real-world example

Poliomyelitis in Tajikistan

- 2010 outbreak of WPV1 in Tajikistan
- Time-varying R_t and surveillance rates
- Daily time series of cases and 116 genomes
- Low prevalence before Feb. 2010
- R_t started to decline before first vaccination round
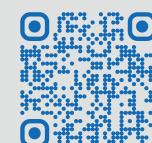
Pandemic-scale phylodynamics?

TimTam is efficient at integrating realistic numbers of cases and the time series aggregation speeds up the computation for large genomic datasets. By combining TimTam with methods for building large trees, like Thorney BEAST or MAPLE, it opens the door to doing pandemic-scale phylodynamics!

Paper
Zarebski et al. PLOS Comp. Biol. 2022

Preprint
Zarebski et al. medRxiv 2024

Package
https://github.com/aezarebski/timtam2

UNIVERSITY OF OXFORD

SIB Swiss Institute of Bioinformatics

THE UNIVERSITY OF MELBOURNE

D BSSE