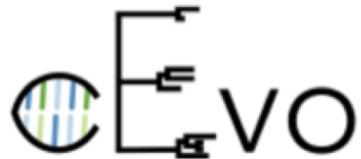
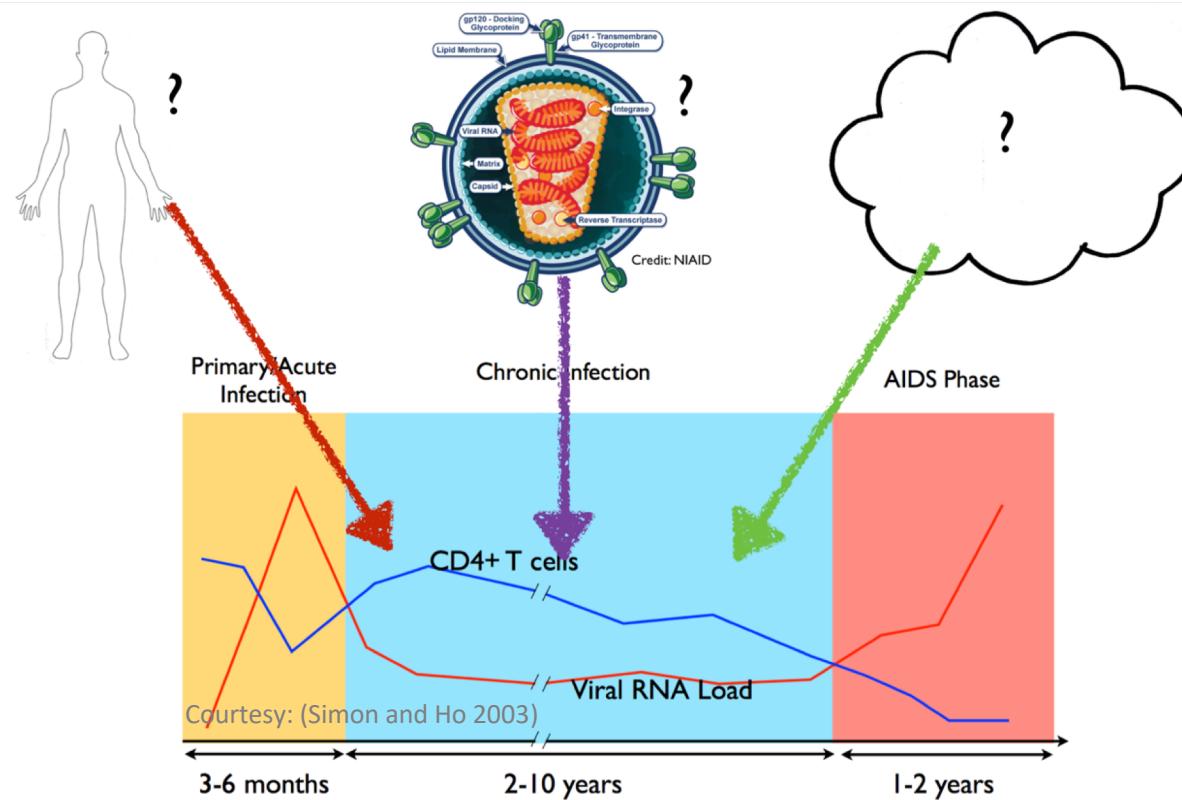


Quantifying the Viral Contribution to Virulence of an HIV Infection

Venelin Mitov, 31 August 2017



Virulence of an HIV Infection



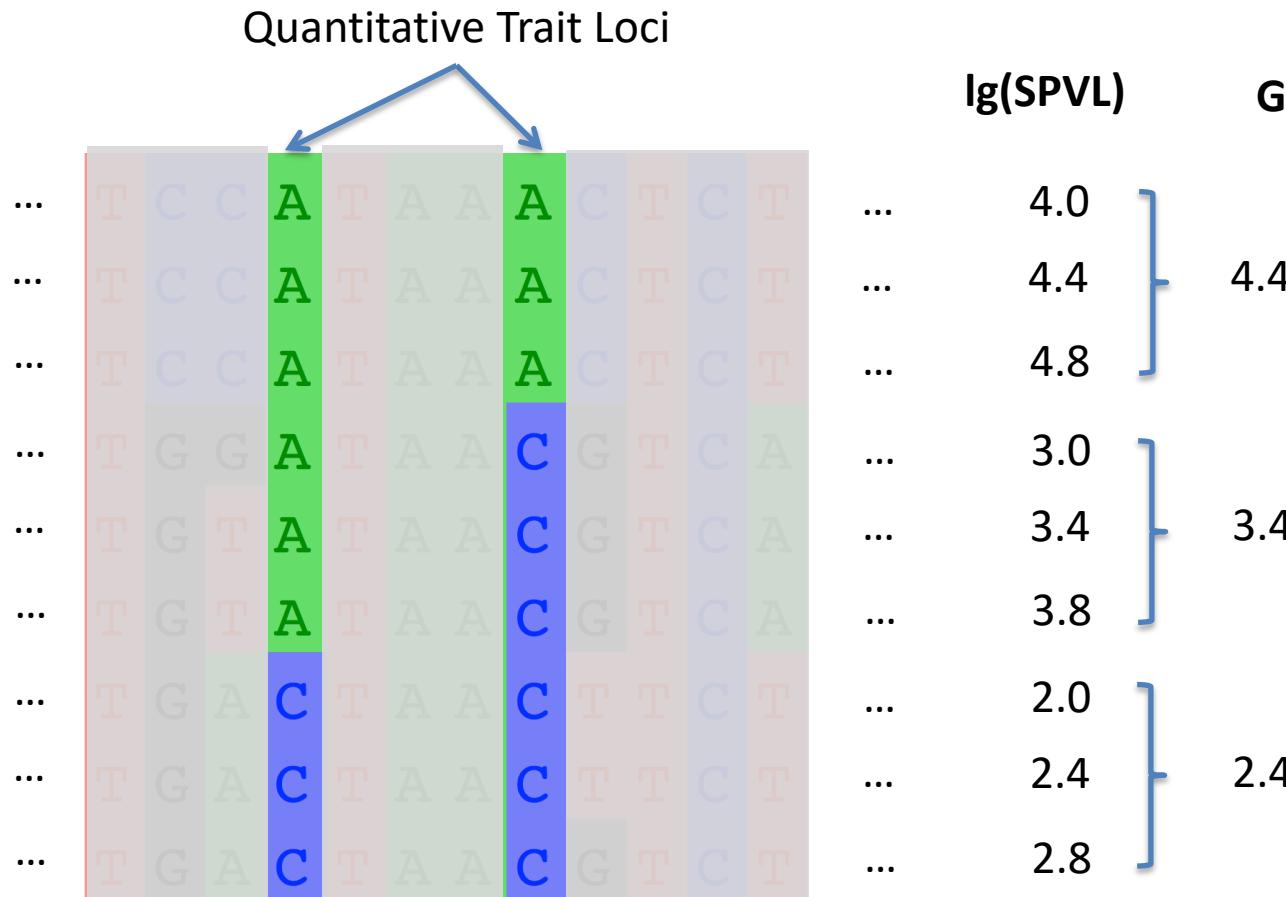
- “Virulence” is the inverse time to AIDS of a patient.
- The viral load during the chronic phase, known as set-point viral load (SPVL) provides a proxy measure for virulence.

Agenda

- Broad-sense heritability of a pathogen trait
- Toy-model simulations for comparing different heritability estimators
- Results in real HIV data
- Implementation

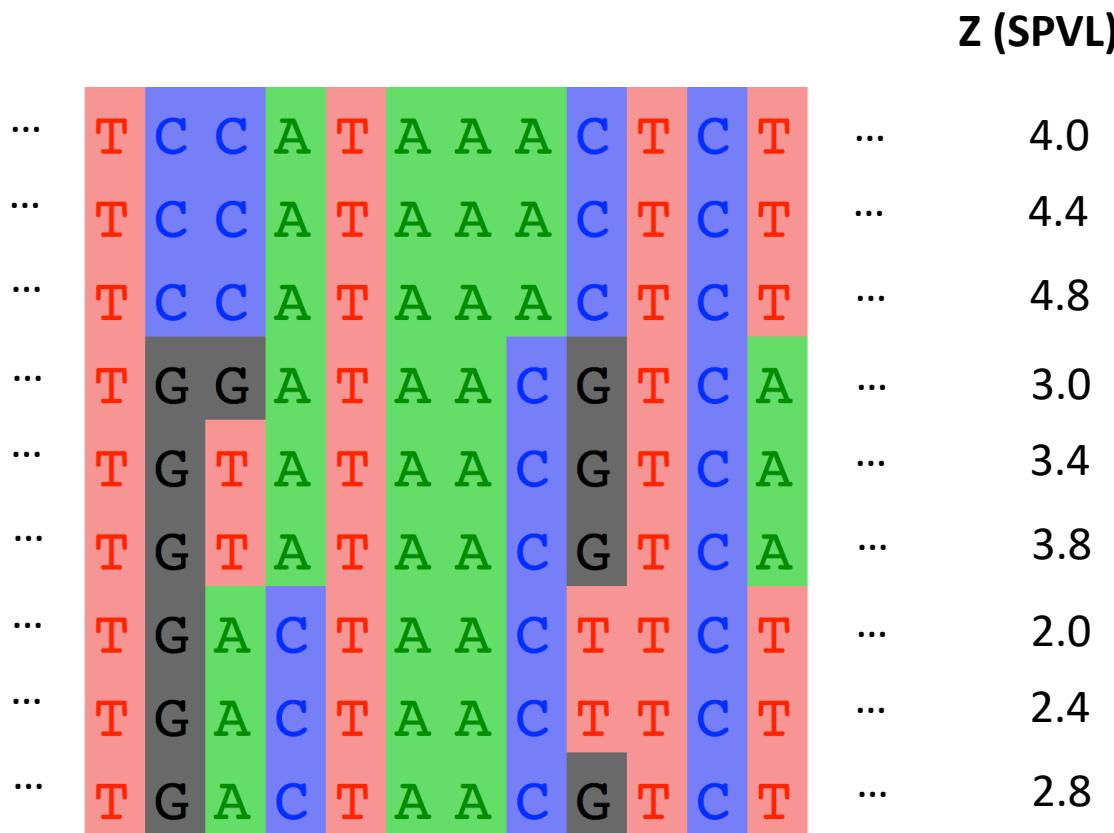
Broad-sense heritability of a pathogen trait

Broad-sense Heritability



$$H^2 = \frac{Var[G]}{Var[\lg(\text{SPVL})]} = \frac{0.75}{0.87} = 86\%$$

Broad-sense Heritability



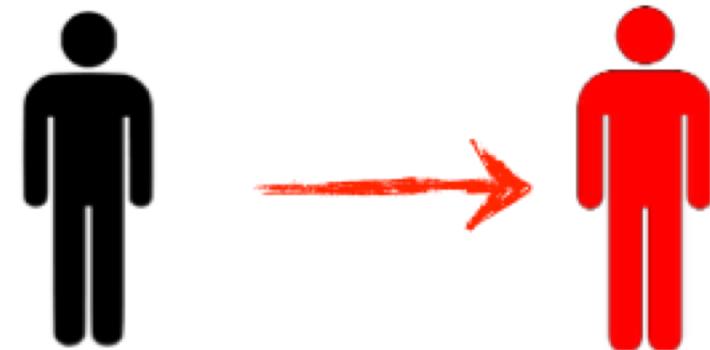
Are there Quantitative Trait Loci for SPVL in the HIV genome?

Yet, it is possible to estimate H^2 !

Resemblance-based estimators:

measuring the relative trait-similarity
within groups of transmission-related
patients

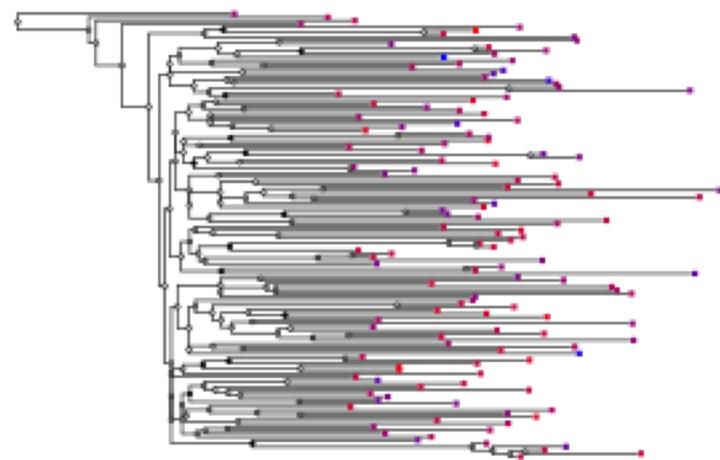
Tool: Donor-recipient regression (DR).



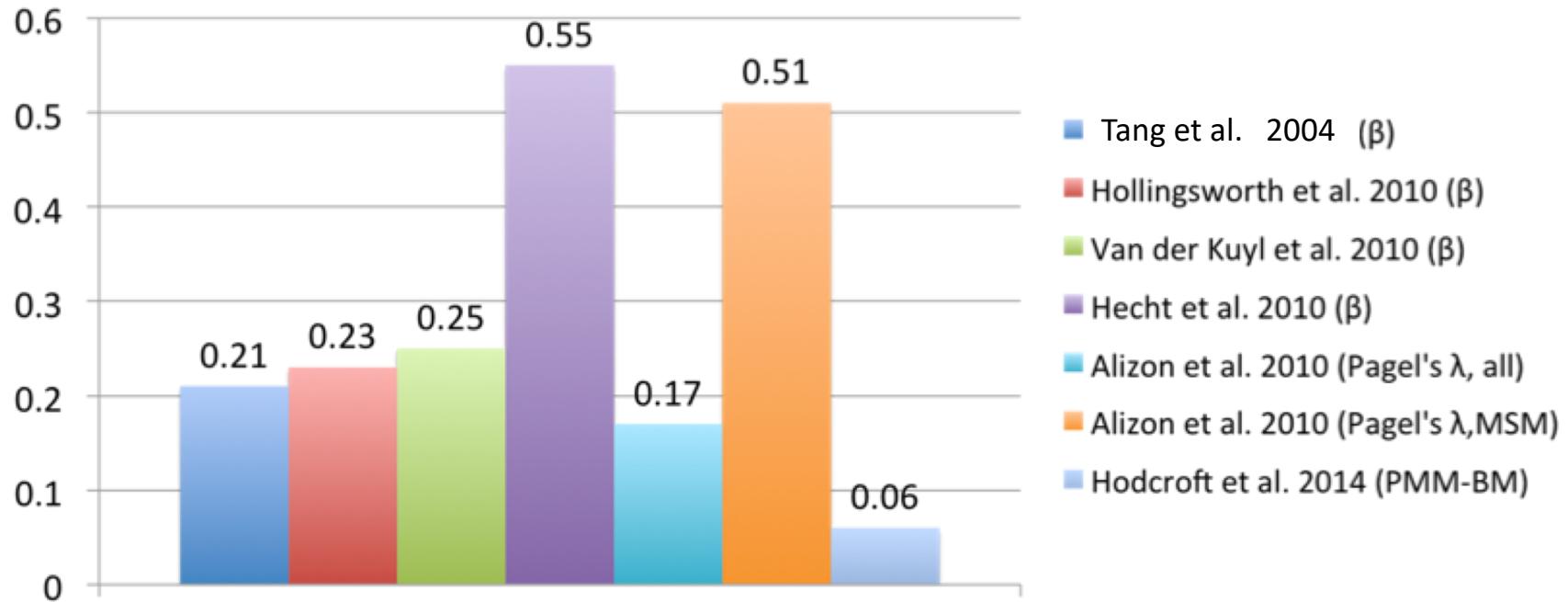
Phylogenetic comparative methods:

measuring the association between
observed trait values from patients and
their (approximate) transmission tree

Tool: Phylogenetic mixed model (PMM).



Controversial Estimates of SPVL-Heritability



We claim that the problem is in the methods, not in the data!

DR and PMM Originate from Quantitative Genetics of Sexually Reproducing Species

Transfer to pathogens

Within-host evolution
Partial quasi-species transmission

DR

Within-host evolution is ignored if trait is measured late in infection:
negative bias (as difference due to evolution is observed as noise)

filter out measurements with long delays:
ANOVA-CPP

PMM

Assumes a Brownian motion process of evolution of the trait along the transmission tree.

Does replacing Brownian motion with Ornstein-Uhlenbeck process:
POUMM

Toy-model simulations for comparing different heritability estimators

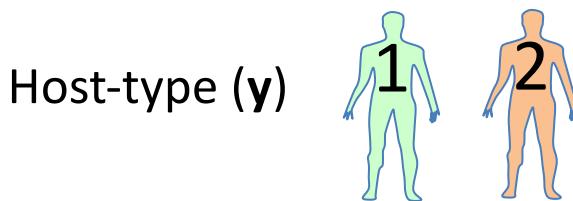
A Toy Model of an epidemic

trait = general effect + host-specific effect

$$z_i(t)$$

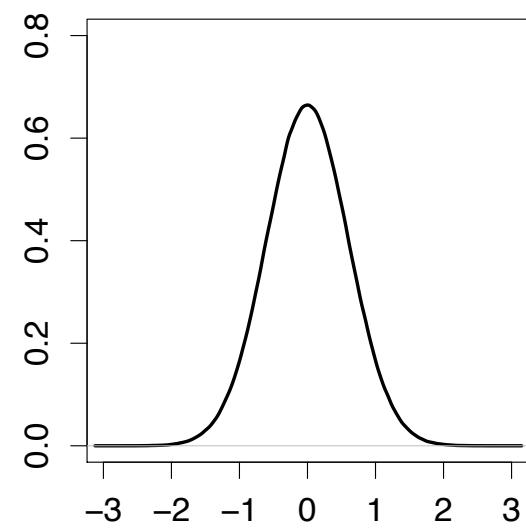
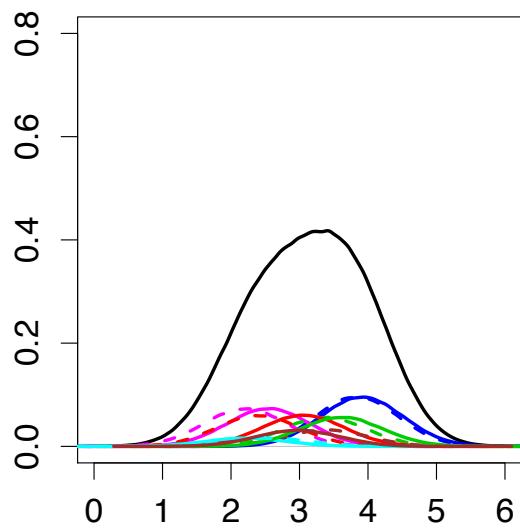
$$GE[\mathbf{y}, \mathbf{x}_i(t)]$$

$$e_i[\mathbf{x}_i(t)]$$



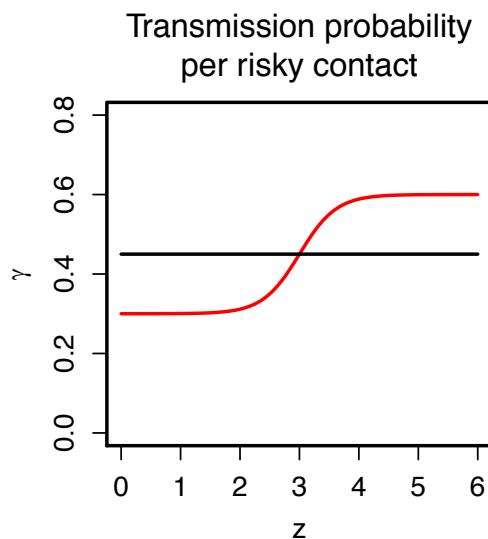
Strain (\mathbf{x})

1:11	3.06	2.40
2:12	3.62	3.37
3:21	3.91	3.83
4:22	2.22	2.57
5:31	2.55	2.21
6:32	2.98	3.40

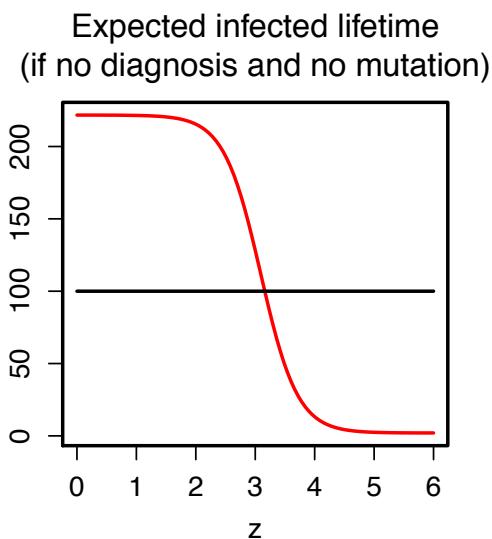
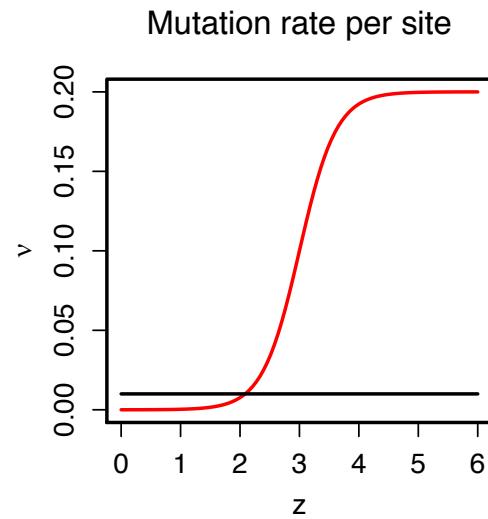


Susceptible Infected Recovered Dynamics

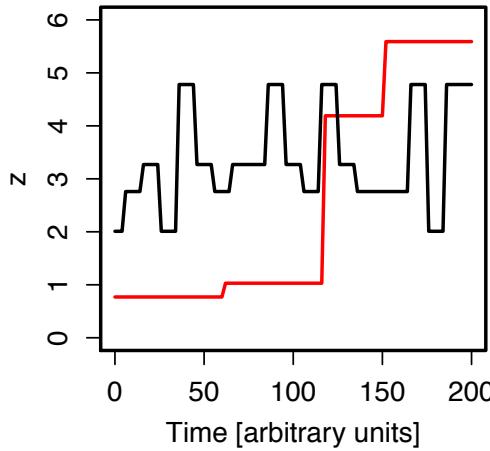
Between hosts:



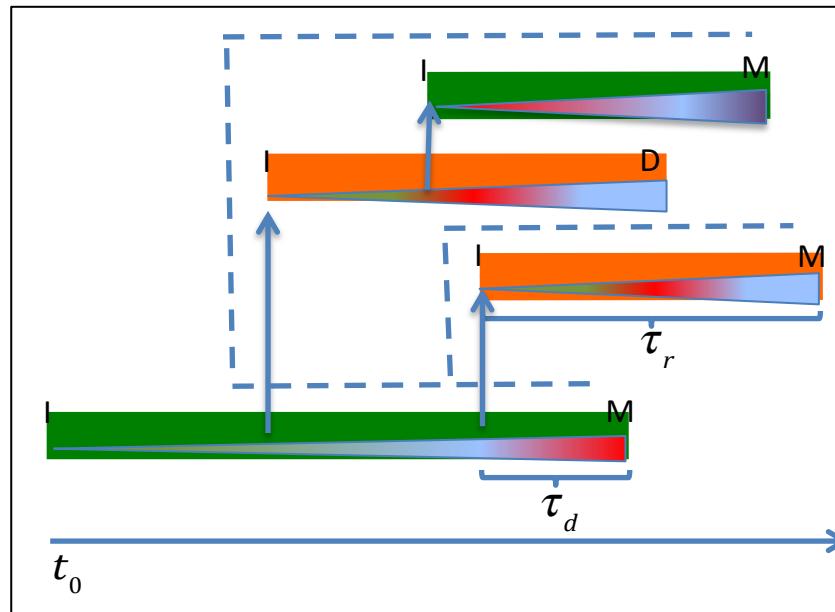
Within host:



Example trait–evolution

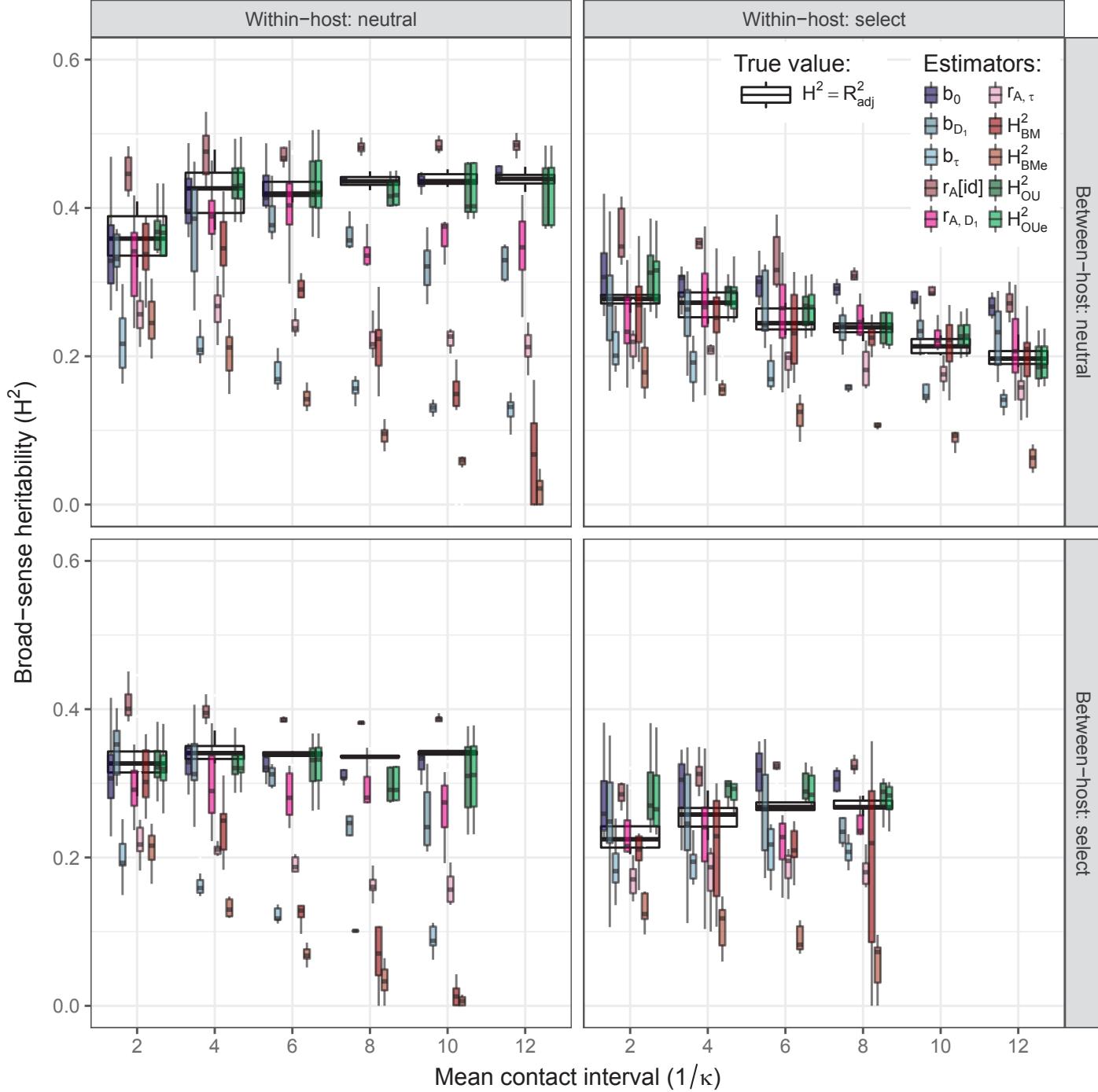


A Toy Model of an Epidemic



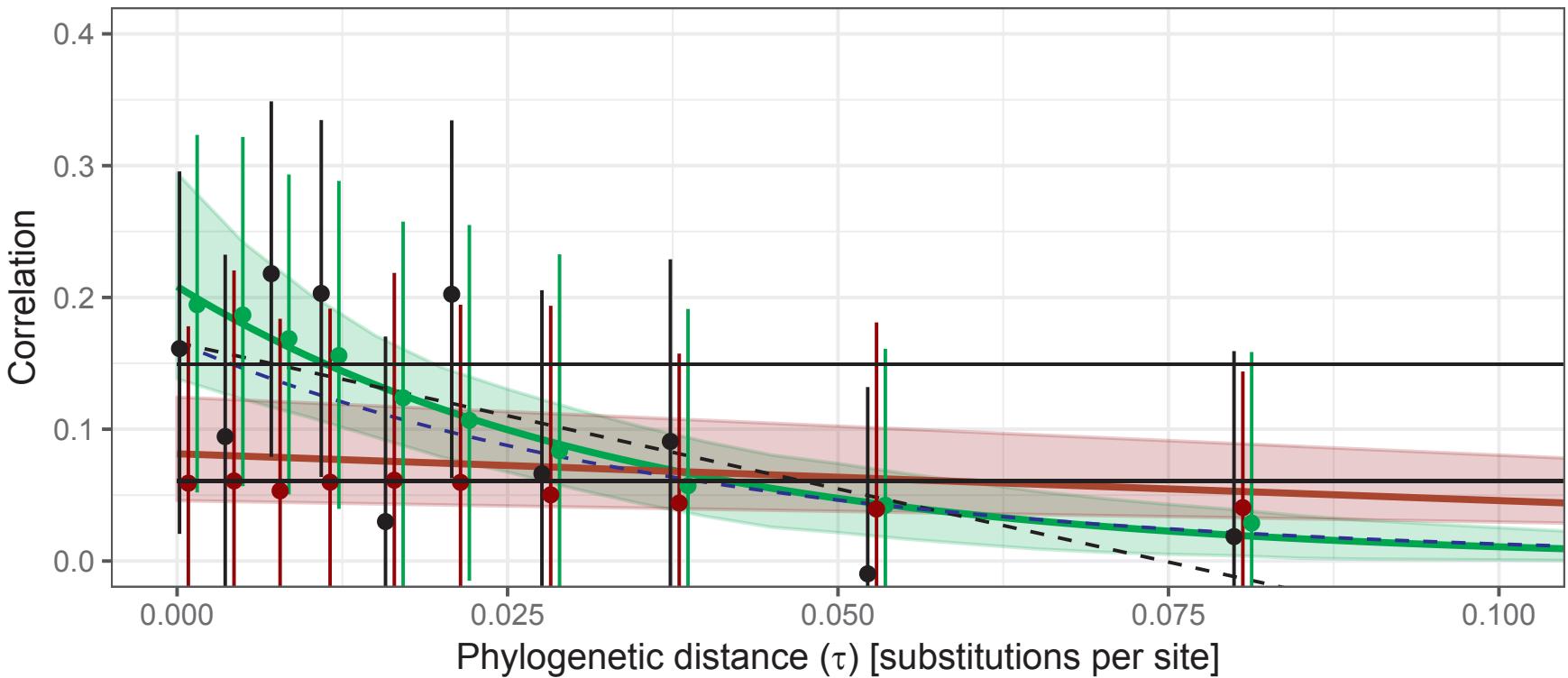
Simulate 240 epidemics of 10'000 diagnosed patients:

- 4 scenarios of within-host between-host dynamics;
- Varying the contact rate of infected hosts

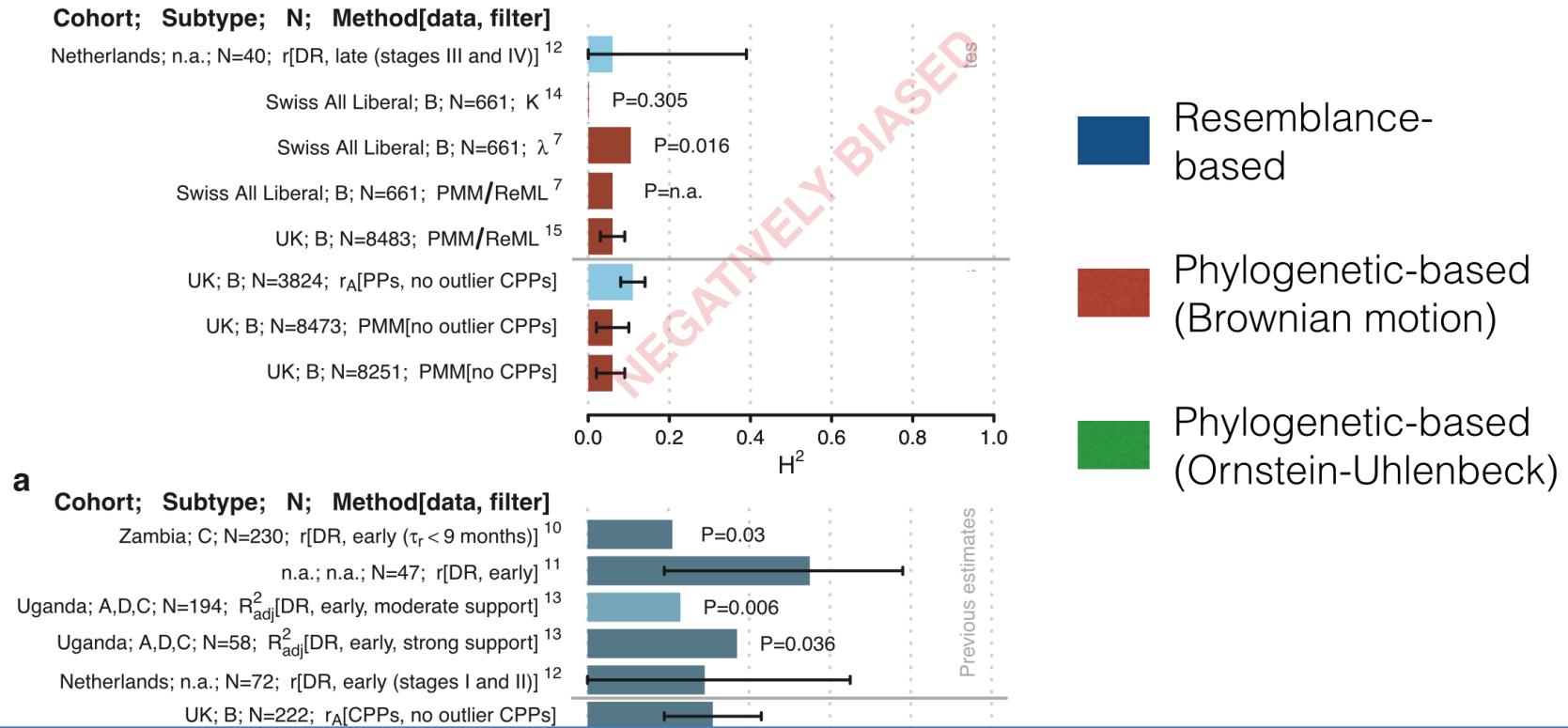


Empirical results

UK HIV Data (8483 patients)



Comparison with previous estimates



The heritability of set-point viral load is estimated at 20%-30%!



New Results

The heritability of pathogen traits - definitions and estimators

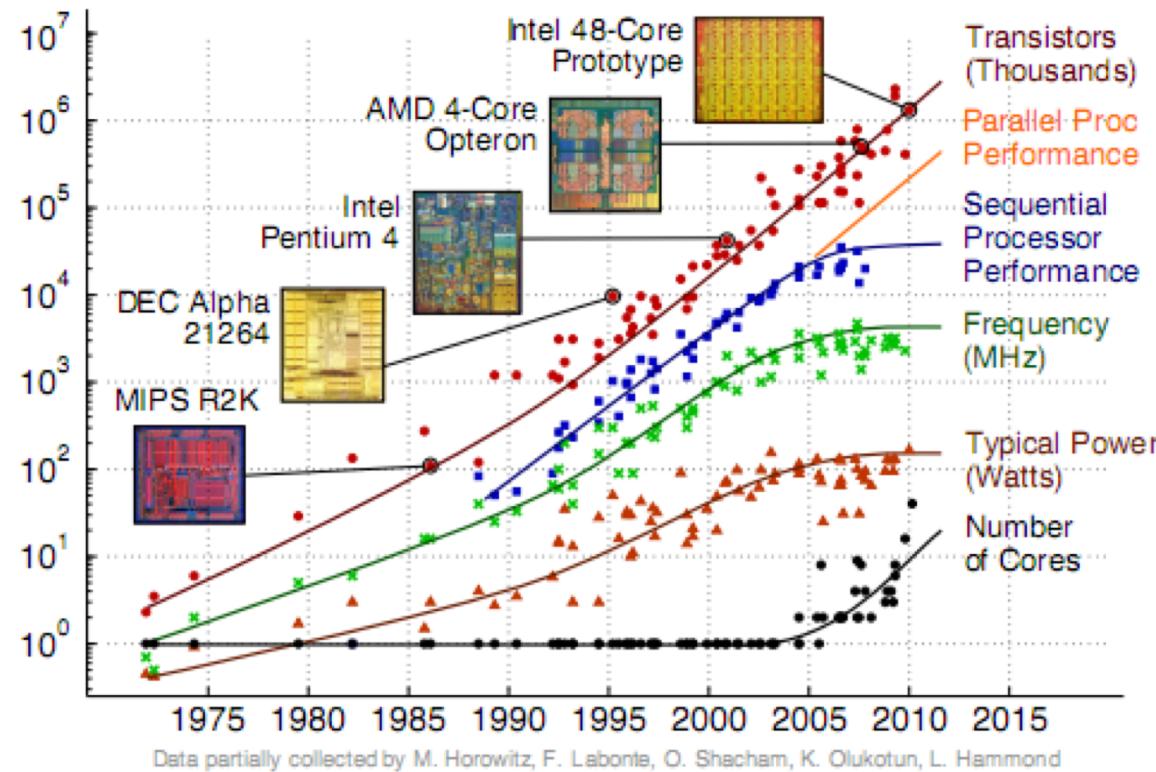
Venelin Mitov, Tanja Stadler

doi: <http://dx.doi.org/10.1101/058503>

This article is a preprint and has not been peer-reviewed [what does this mean?].

Implementation

CPU Clock Rate vs Number of Cores

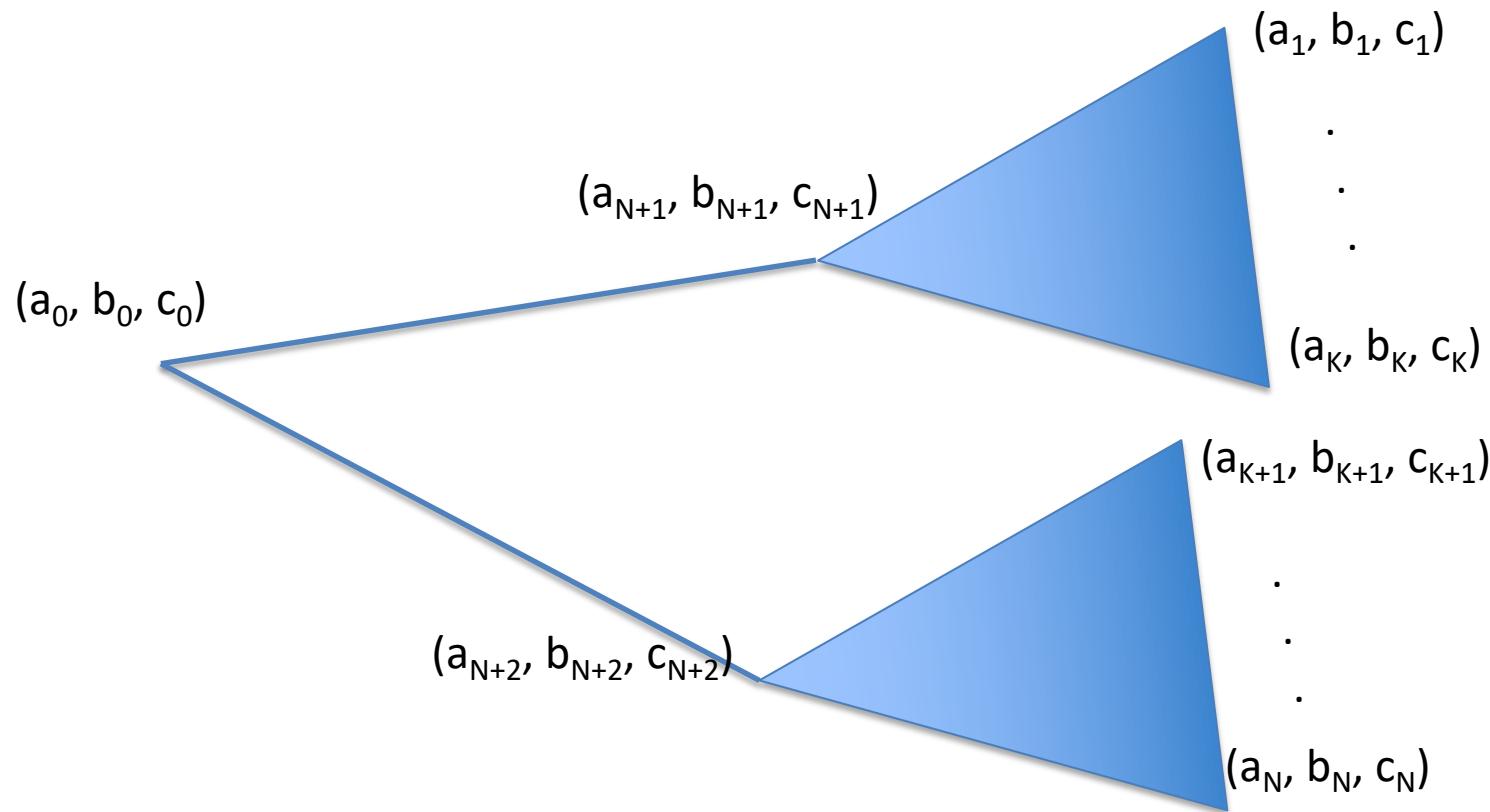


Prepared by C. Batten - School of Electrical and Computer Engineering - Cornell
University - 2005 - retrieved Dec 12 2012 -
<http://www.csl.cornell.edu/courses/ece5950/handouts/ece5950-overview.pdf>

P(OU)MM defines a normal distribution at the tips

	POUMM	PMM ($\alpha \rightarrow 0$)
$\Theta:$	$\langle g_0, \alpha, \theta, \sigma, \sigma_e \rangle$	$\langle g_0, \sigma, \sigma_e \rangle$
$\mu_i(\Theta, \mathcal{T}):$	$e^{-\alpha t_{0i}} g_0 + (1 - e^{-\alpha t_{0i}}) \theta$	g_0
$\Sigma_{ii}(\Theta, \mathcal{T}):$	$\sigma^2 \frac{(1 - e^{-2\alpha t_{0i}})}{2\alpha} + \sigma_e^2$	$\sigma^2 t_{0i} + \sigma_e^2$
$\Sigma_{ij}(\Theta, \mathcal{T}):$	$\sigma^2 \frac{e^{-\alpha \tau_{ij}} (1 - e^{-2\alpha t_{0(ij)}})}{2\alpha}$	$\sigma^2 t_{0(ij)}$
$H_{\bar{t}}^2:$	$\frac{\sigma^2 (1 - e^{-2\alpha \bar{t}})}{\sigma^2 (1 - e^{-2\alpha \bar{t}}) + 2\alpha \sigma_e^2}$	$\bar{t}\sigma^2 / (\bar{t}\sigma^2 + \sigma_e^2)$
$H_{\infty}^2:$	$\sigma^2 / (\sigma^2 + 2\alpha \sigma_e^2)$	1
$H_e^2:$	$1 - \sigma_e^2 / s^2(\mathbf{z})$	$1 - \sigma_e^2 / s^2(\mathbf{z})$

Felsenstein pruning: Recursion along the tree



Proof: Recurrent formulas for the coefficients a, b, c

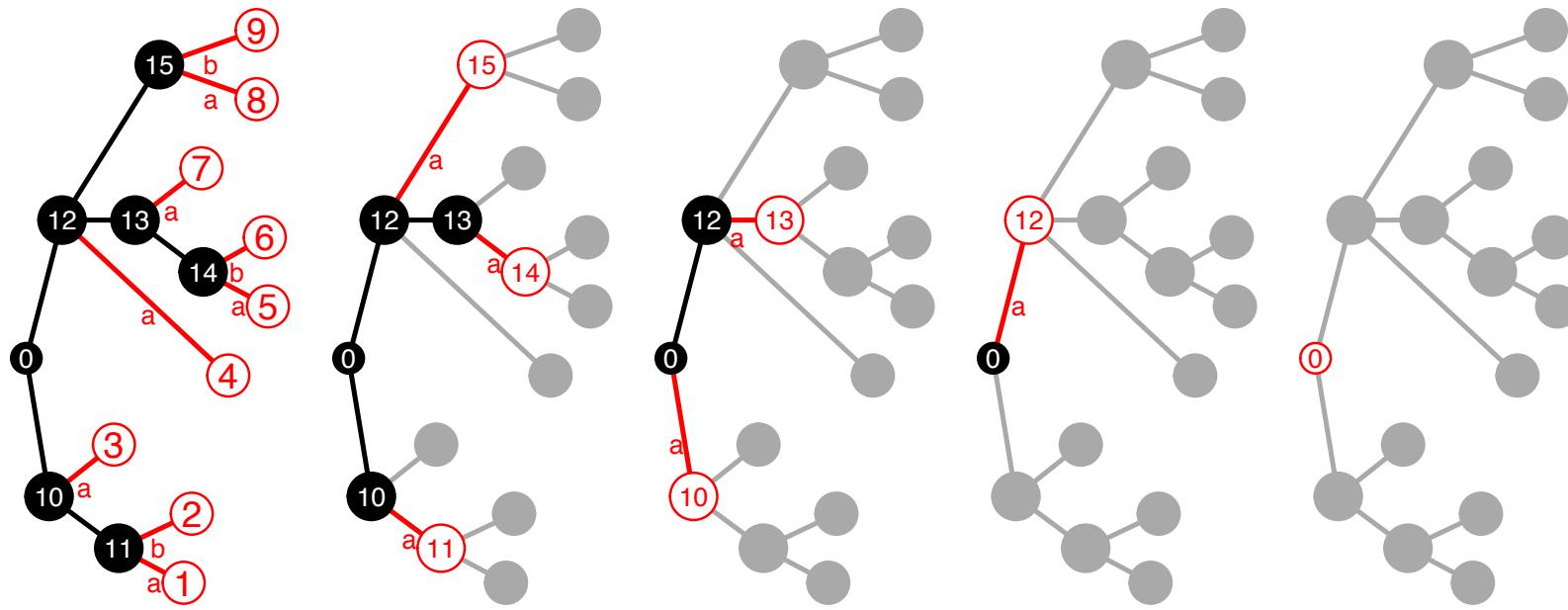
1. For $j \in \{1, \dots, N\}$ (tips):

$$a_j = -\frac{1}{2\sigma_e^2}; \quad b_j = \frac{z_j}{\sigma_e^2}; \quad c_j = -\frac{z_j^2}{2\sigma_e^2} - \ln \sqrt{2\pi\sigma_e^2} \quad (6)$$

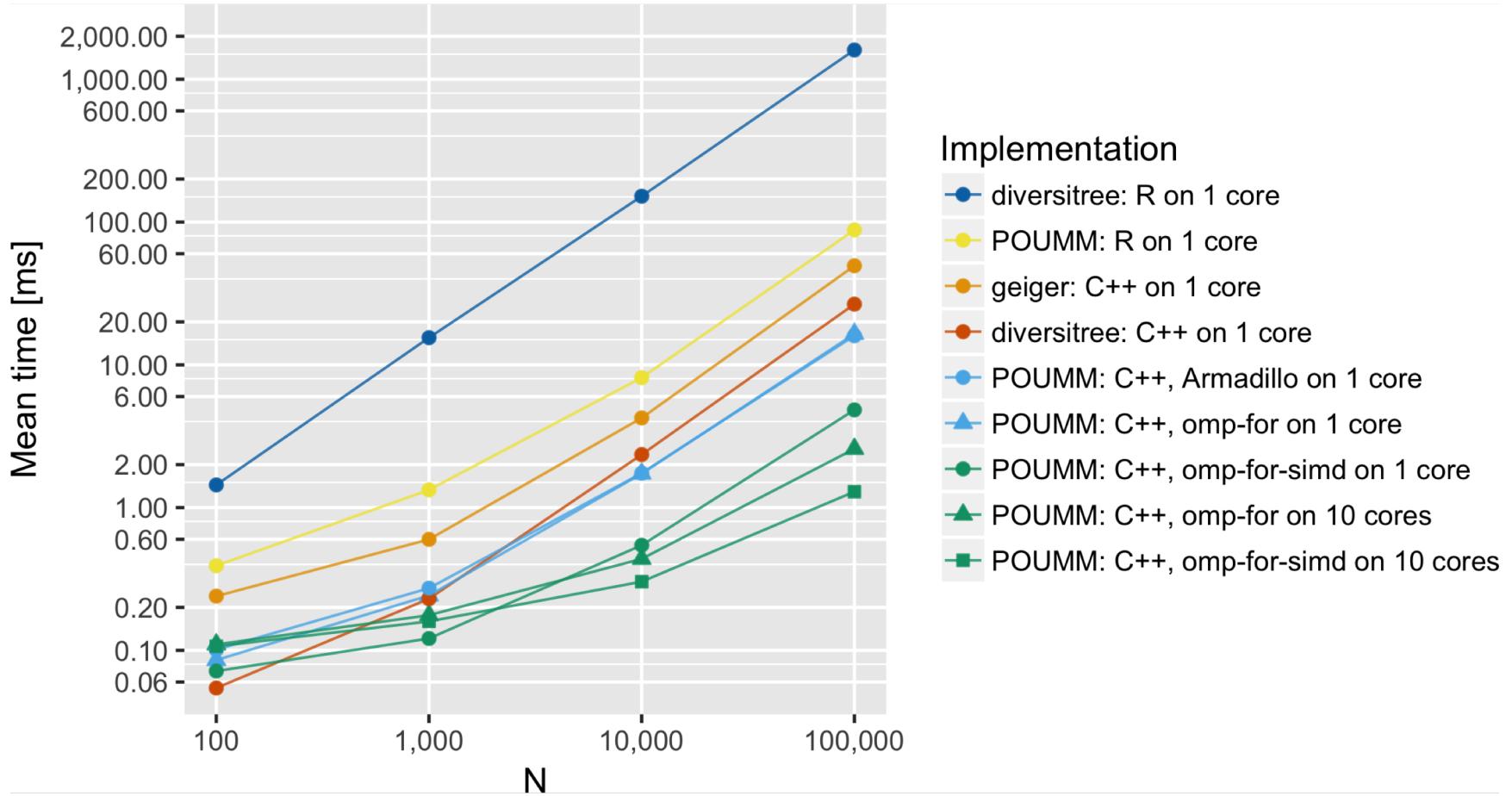
2. For $j > N$ (internal nodes) or $j = 0$ (root):

$$\begin{aligned} a_j &= \sum_{i \in Desc(j)} \frac{a_i u(\alpha, 2t_{ji})}{u(\alpha, 2t_{ji}) - \alpha + \sigma^2 a_i} \\ b_j &= \sum_{i \in Desc(j)} \frac{u(\alpha, 2t_{ji}) [2\theta a_i (e^{\alpha t_{ji}} - 1) + b_i e^{\alpha t_{ji}}]}{u(\alpha, 2t_{ji}) - \alpha + \sigma^2 a_i} \\ c_j &= \sum_{i \in Desc(j)} \left\{ c_i + \alpha t_{ji} - \frac{0.25 b_i^2 \sigma^2}{-\alpha + a_i \sigma^2 + u(\alpha, 2t_{ji})} - \right. \\ &\quad \left. 0.5 \ln \left(\frac{-\alpha + a_i \sigma^2 + u(\alpha, 2t_{ji})}{u(\alpha, 2t_{ji})} \right) + \right. \\ &\quad \left. \frac{\alpha \theta [a_i \theta - (b_i + a_i \theta) e^{\alpha t_{ji}}]}{u(\alpha, 2t_{ji}) + (-\alpha + a_i \sigma^2) (1 + e^{\alpha t_{ji}})} \right\}. \end{aligned} \quad (7)$$

Parallel Pruning



Performance





New Results

Fast and robust inference of phylogenetic Ornstein-Uhlenbeck models using parallel likelihood calculation

This article is a preprint and has not been peer-reviewed [what does this mean?].

Conclusions

- Viral genetic factors account for 20%-30% of the variance of spVL
- Previous discrepancy in the estimates is due to differences in the methods
- Parallelization of phylogenetic algorithms can take advantage of multi-core technology.

Phylogeny of Acknowledgments

