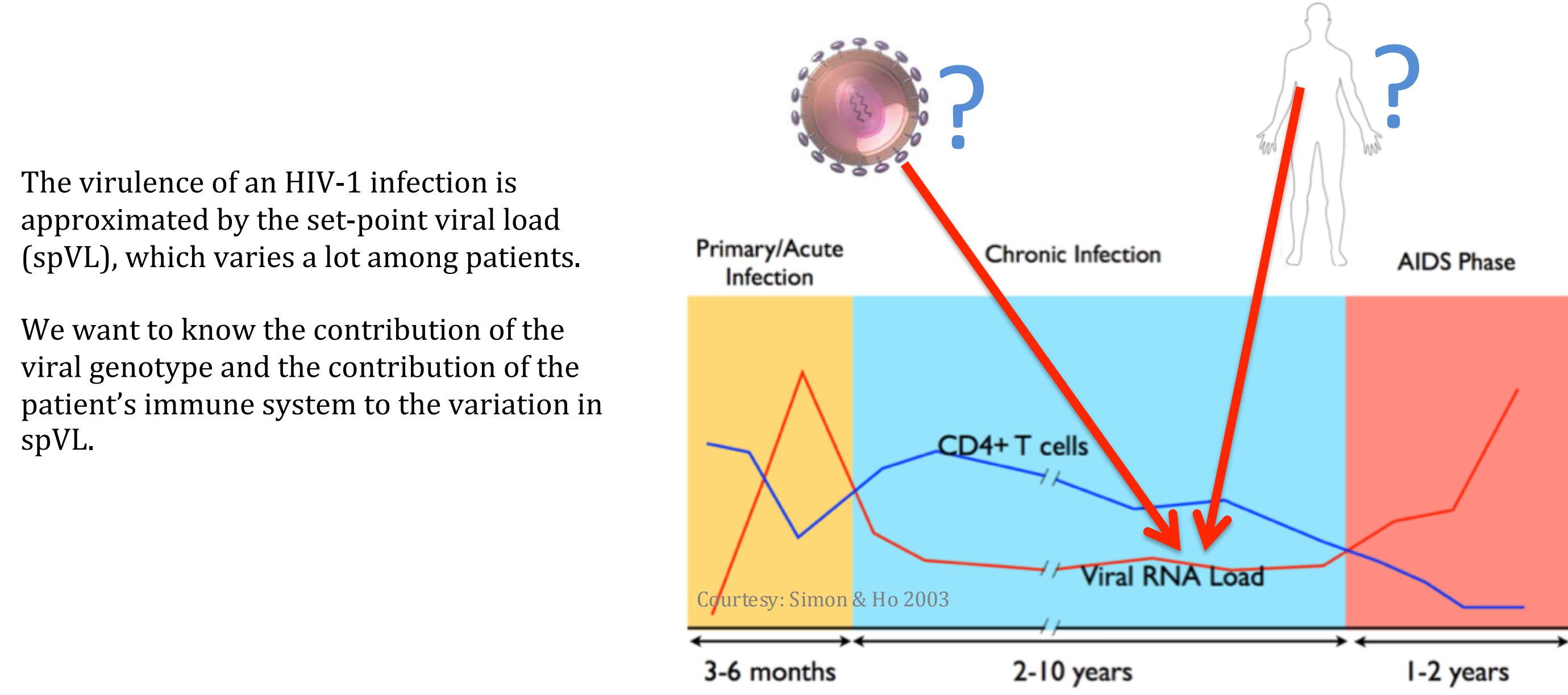


Importance of Pathogen Mutation in Quantifying the Viral Contribution to Virulence of an HIV-1 Infection

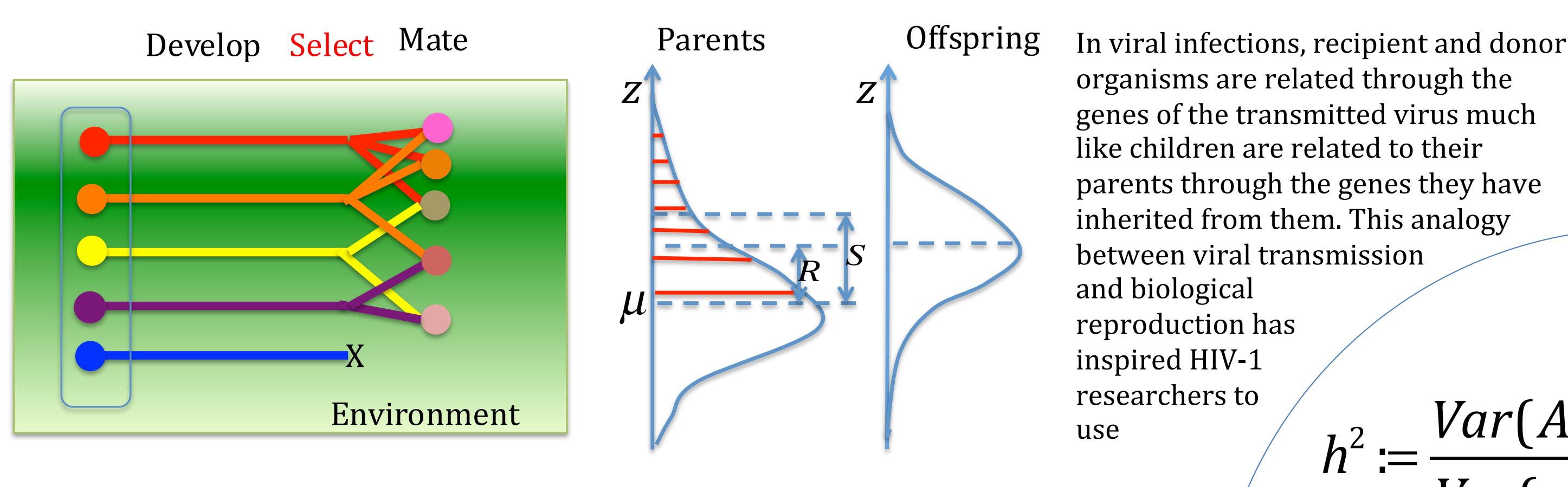
Venelin Mitov^{1,2}, Tanja Stadler^{1,2}

1 Department of Biosystems, Science and Engineering, Swiss Federal Institute of Technology Zurich, Basel, Switzerland
2 Swiss Institute of Bioinformatics

1. The Virulence of an HIV-1 Infection Has a Huge Variation Among Patients



2. Quantitative Genetics is Used to Understand the Variation in Virulence



In Quantitative Genetics, a population of sexually reproducing organisms is modeled in terms of generations, which follow phases of development, selection for reproduction and random mating. For a continuous (quantitative) trait of interest, the goal is to describe the observed trait values in a generation in terms of heritable (genetic) and environmental contributions. This separation of trait values into heritable and non-heritable components allows predicting the response to natural (or artificial) selection in offspring generations and has found a broad application in animal and plant breeding. The equality between the parent-offspring regression slope, β , the narrow-sense heritability, h^2 , and the breeding heritability, h_b^2 , plays a central role, because it enables estimating genetic contributions to traits without knowing which genes affect the trait. → "The A, B, C, G... of Quantitative Genetics".

The A, B, C, G... of Quantitative Genetics

A stays for "Additive genetic value" of an individual

The trait value, z , of an individual in a population can be written as the sum

$$z = \mu + A + I + E,$$

where μ is the population mean, A is the sum of single-loci additive genetic effects, I is the sum of epistatic and dominance interactions and E is the sum of environmental effects and measurement error.

Summarized at the population level, A defines the narrow-sense heritability:

$$h^2 := \frac{Var(A)}{Var(z)}.$$

B stays for "Breeding value" of an individual

For a diploid sexually reproducing organism, B is defined as twice the expected deviation of the trait values of its progeny with respect to the population mean.

By applying ANOVA on B w.r.t. the population variance, we define the intraclass-correlation (ICC):

$$r_A := \frac{Var(B)}{Var(z)}.$$

G stays for "Genotypic value" of a genotype

G is defined for a given genotype as the expected value of individuals having this genotype in the population.

Together, $\mu + A + I$ equal G .

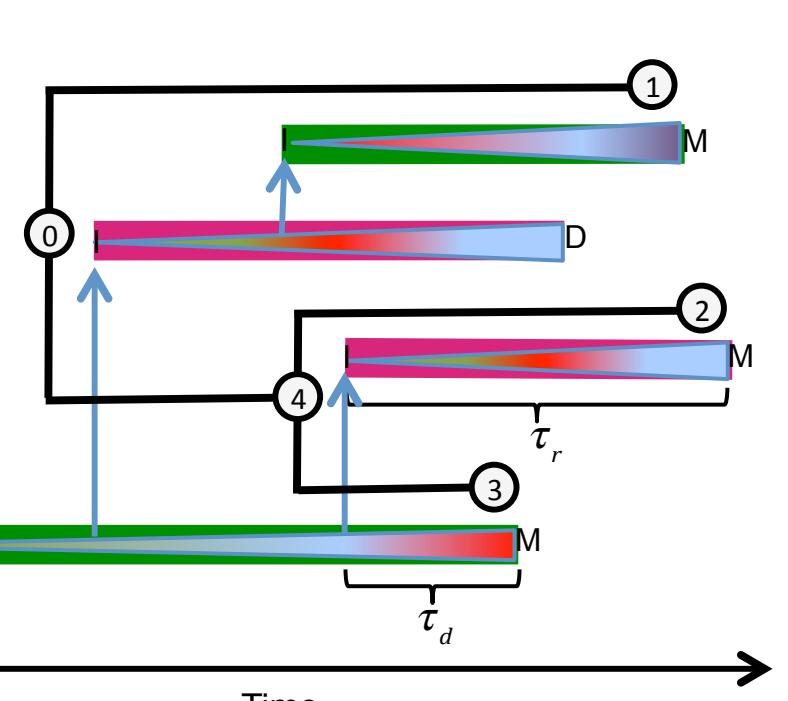
Summarized at the population level, G defines the broad-sense heritability:

$$H^2 := \frac{Var(G)}{Var(z)}$$

MUTATION

$$r_A := \frac{Var(B)}{Var(z)} \quad b := \frac{Cov(C)}{Var(z)}$$

3. Can We Apply Quantitative Genetics to Pathogen Traits?



While genetic mutation usually doesn't occur during the lifetime of animals and plants, it constitutes a hallmark in the lifecycle of infections. The theory of quantitative genetics has not been designed to account for changes in the genotype of individuals during their lifetime. Thus, we question the validity of its principles and methods in the case of pathogen evolution.

A schematic representation of an epidemic. Colored rectangles represent infectious periods of hosts, different colors corresponding to different host-types. Triangles inside hosts represent pathogen quasispecies, change of color indicating substitution of dominant strains. Capital letters denote host-events as follows: I: beginning of infection, M: phenotype measurement and recovery D: host death. Arrows show the time and direction of transmission events.

4. Different Methods to Estimate spVL-Heritability Report Controversial Estimates

Two major types of methods:

- Resemblance based
 - Regression slope (b) from pairs of donor and recipient patients.
 - ANOVA (r_A) on groups of transmission related patients:
 - phylogenetic pairs – PP;
 - closest phylogenetic pairs – CPP;



- Phylogenetic based: fitting a model of evolution of G to a phylogeny approximating the transmission network between patients.

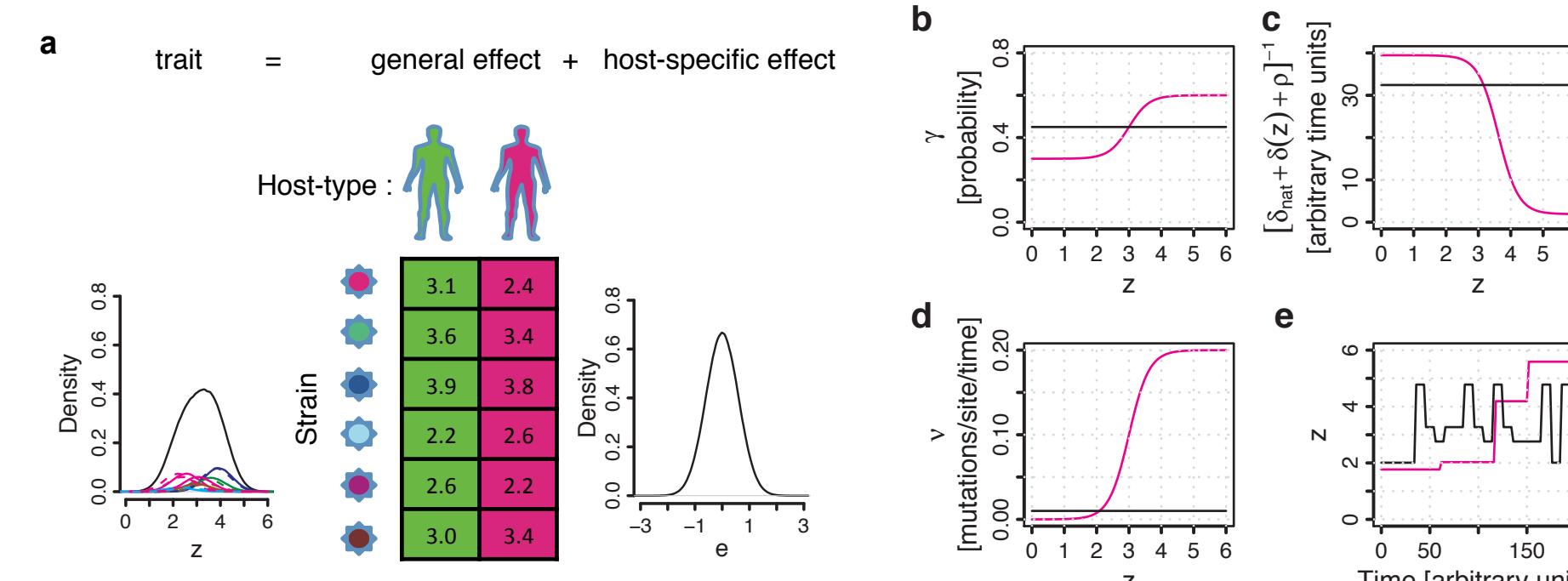
Model parameters:

- PMM: σ, σ_e
- POUMM: $\alpha, \theta, \sigma, \sigma_e$

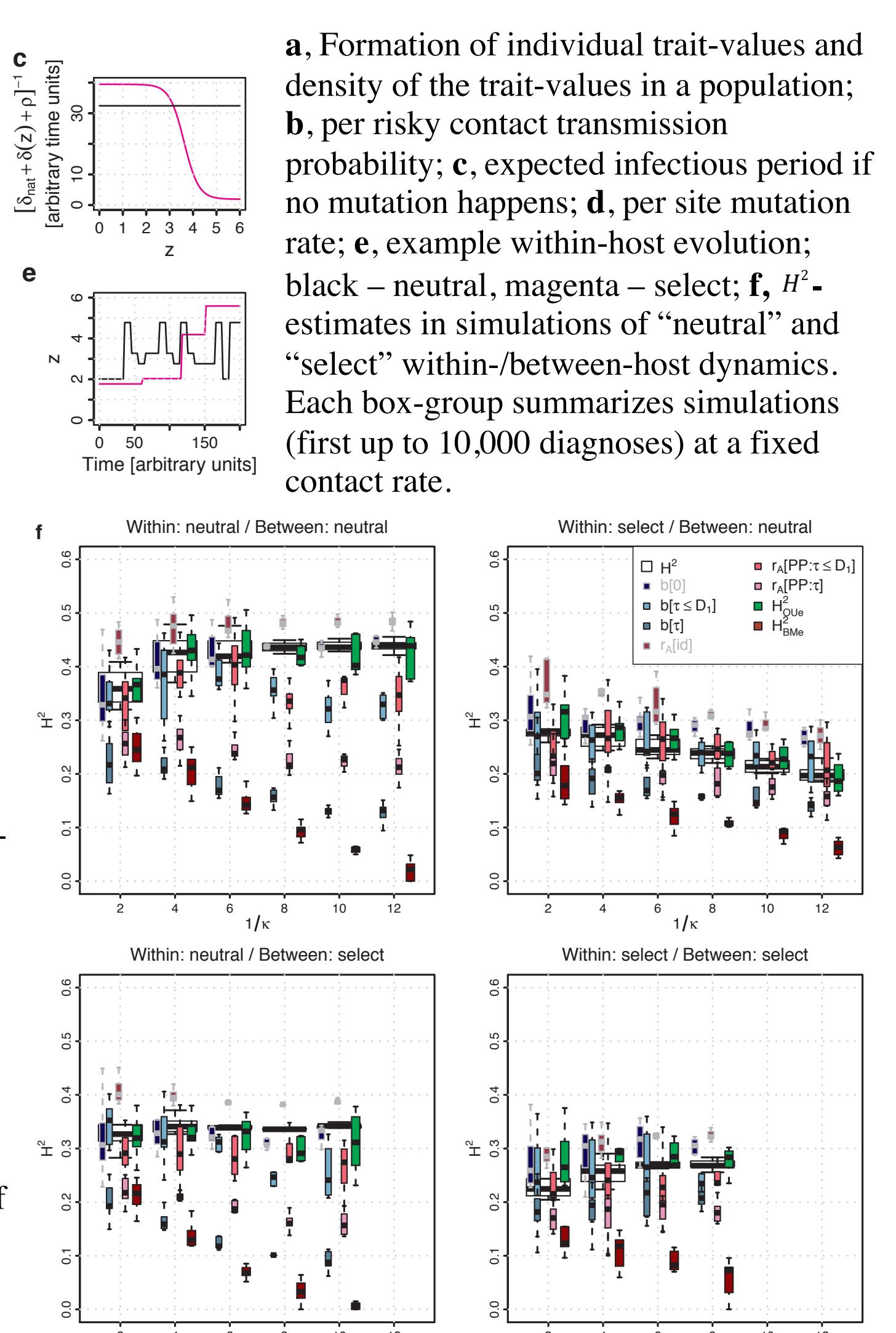
Phylogenetic heritability:

$$H^2_{OUE} = 1 - \sigma_e^2 / \sigma^2(z)$$

5. Results on Toy Model Simulations



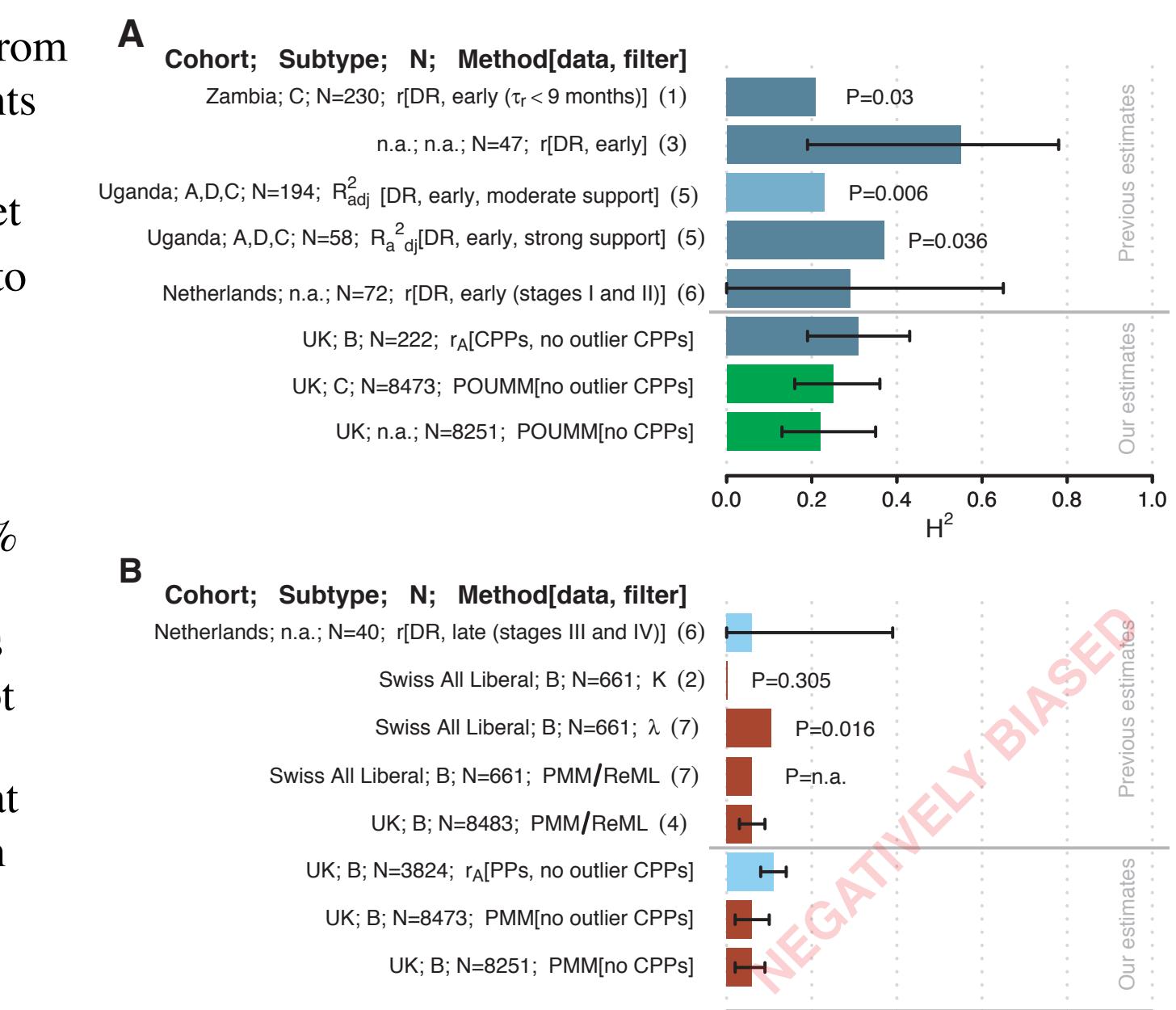
To test the different estimators of heritability, we developed a toy-model of an epidemic, in which an imaginary pathogen trait, Z , was determined by the interaction between the alleles at two loci in the pathogen genotype and one of two immune system types encountered at equal frequencies in the susceptible population (a). This toy-model was embedded into a stochastic Susceptible-Infected-Recovered (SIR) epidemic model with demography and frequency dependent transmission, implementing "neutral" and "selection" modes of within- and between-host dynamics (b-e). Using different contact-rates between individuals, $\kappa \in \{\frac{1}{2}, \frac{1}{4}, \dots, \frac{1}{12}\}$, we performed 240 simulations, of which 175 resulted in epidemic outbreaks of at least 1,000 diagnosed individuals.



6. Results

We performed ANOVA-CPP and POUMM on data from the UK HIV-1 cohort comprising Ig(spVL) measurements and a tree of viral (pol) sequences from 8,483 patients inferred previously in. The goal was to test our conclusions on a real dataset and compare the H^2 -estimates from ANOVA-CPP and POUMM to previous PMM/ReML-estimates on exactly the same data.

Excluding outlier CPPs, ANOVA-CPP (222 patients) reported Ig(spVL)-heritability estimates of 0.31, 95% CI [0.19, 0.43]. POUMM (8,473 patients) reported agreeing estimates of 0.25, 95% CI [0.16, 0.36] and 0.22, CI [0.13, 0.35] upon omitting all 222 patients belonging to CPPs. The slightly lower POUMM estimates could be explained by errors in the transmission tree, which are not present in CPPs. These results show first, that ANOVA-CPP and POUMM agree on disjoint subsets of the UK data and, second, that POUMM provides an alternative to resemblance-based methods in the absence of early-diagnosed cases.



7. Conclusions

- Our results show that when there is mutation during the lifetime of individuals, the heritability of pathogen traits cannot be estimated by donor-recipient regression or a PMM method assuming a Brownian Motion model of evolution of G .
- A modification of the PMM method assuming an Ornstein-Uhlenbeck Model of evolution for G provides an accurate estimate of the broad-sense heritability under different scenarios of within- and between-host selection.
- In summary, POUMM and ANOVA-CPP yield agreeing estimates for H^2 in the UK data and these estimates agree with DR-based estimates in datasets with short measurement delay. Similar to the toy-model simulations, we notice a well-pronounced pattern of negative bias for the other estimators, PMM and ANOVA-PP, as well as for the previous DR-studies on data obtained under long measurement delay.

References

- Tang J, et al. (2004) HLA allele sharing and HIV type 1 viremia in seroconverting Zambians with known transmitting partners. *AIDS Res Hum Retroviruses* 20(1):19-25.
- Alizon S, Wyl, von V, Stadler, T, Kouyou, R. D., Yerly, S., Hirschel, B., et al. (2010) Phylogenetic approach reveals that virus genotype largely determines HIV set-point viral load. *PLoS Pathogens* 6(9), e1001123. <http://doi.org/10.1371/journal.ppat.1001123>
- Hecht, F. M., Hartogson, W., Bragg, L., Bacchetti, P., Atchison, R., Grant, R., et al. (2010) HIV RNA level in early infection is predicted by viral load in the transmission source. *AIDS (London, England)*, 24(7), 941-945. <http://doi.org/10.1097/QAD.0b013e328337b12e>
- Hodcroft, E., Hadfield, J. D., Fearnhill, E., Phillips, A., Dunn, D., O'Shea, S., et al. (2014) The Contribution of Viral Genotype to Plasma Viral Set-Point in HIV Infection. *PLoS Pathogens*, 10(5), e1004112. <http://doi.org/10.1371/journal.ppat.1004112>
- Hollingsworth, T. D., Layden, O., Shirreff, G., Donnelly, C. A., Serwadda, D., Wawer, M. J., et al. (2010) HIV-1 transmitting couples have similar viral load set-points in Rakai, Uganda. *PLoS Pathogens*, 6(5), e1000876. <http://doi.org/10.1371/journal.ppat.1000876>
- van der Kuy, A. C., Juriaans, S., Pollakis, G., Bakker, M., & Cornelissen, M. (2010) HIV RNA levels in transmission sources only weakly predict plasma viral load in recipients. *AIDS (London, England)*, 24(10), 1607-1608. <http://doi.org/10.1097/QAD.0b013e32833b318f>
- Shirreff G, et al. (2013) How effectively can HIV phylogenies be used to measure heritability? *Evolution, Medicine, and Public Health* 2013(1):209-224.