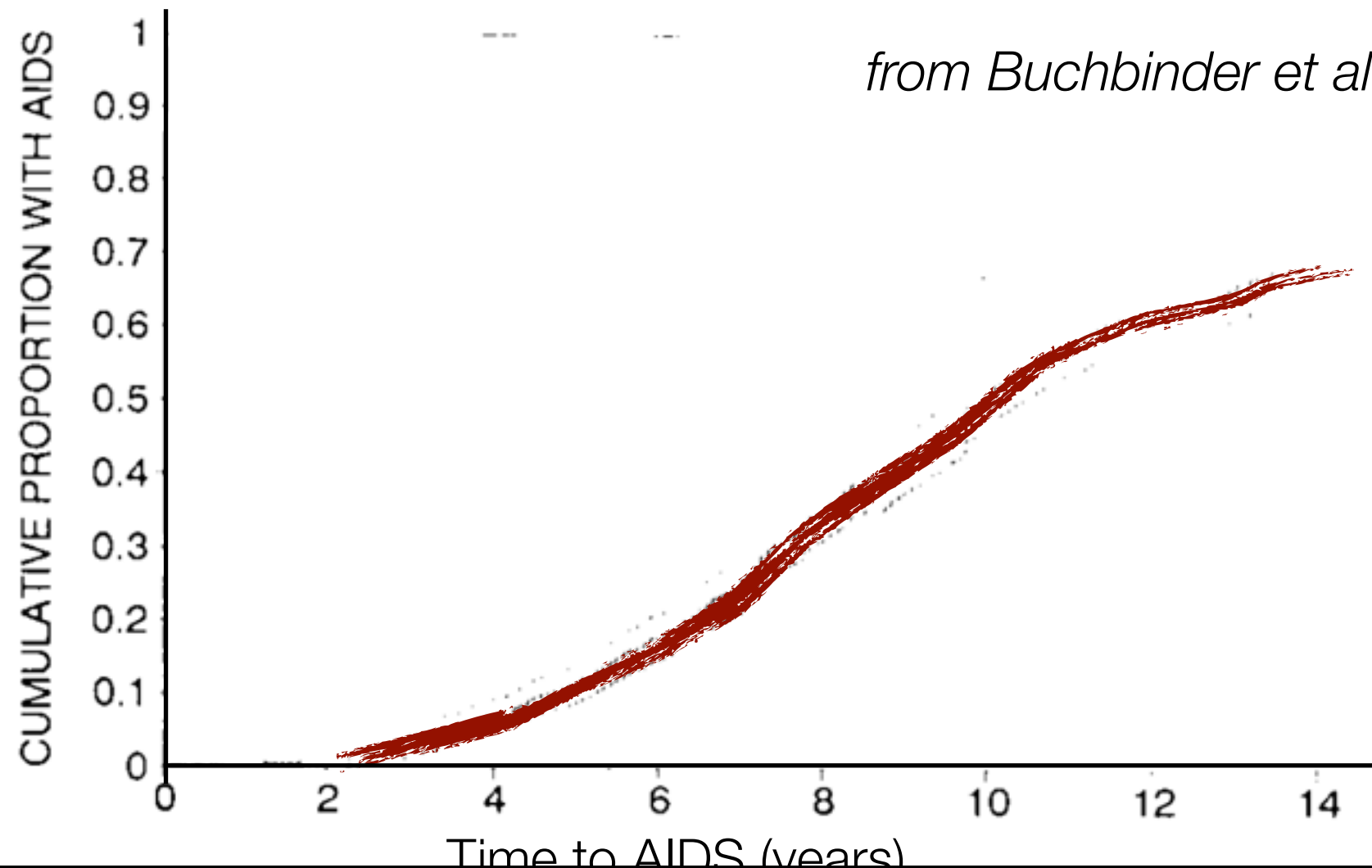# To what extend influences the virus genotype the virulence of an HIV infection?

- estimating heritability of spVL -

Venelin Mitov & Tanja Stadler

Tanja Stadler
Computational Evolution group — Department of Biosystems Science and Engineering
tanja.stadler@bsse.ethz.ch

ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

High variance in virulence of HIV
virulence = 1/ (time to AIDS)
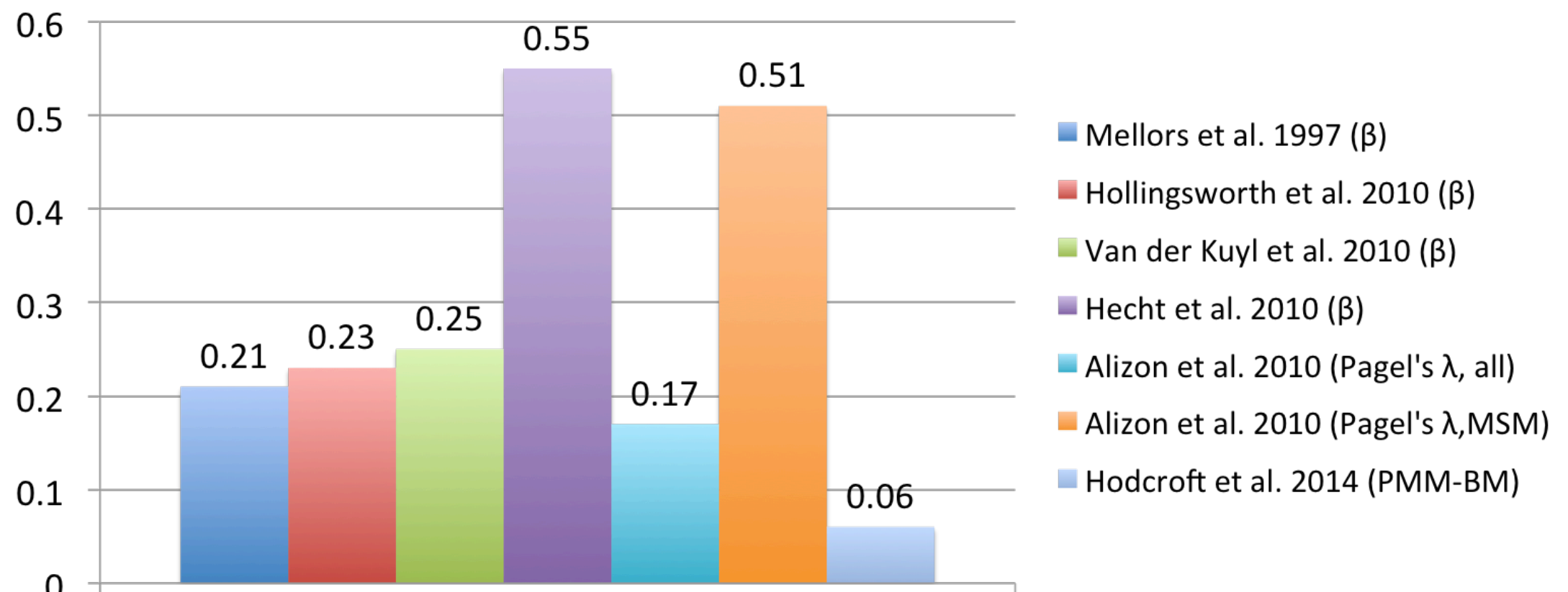

from Buchbinder et al. (1994, AIDS)

What determines the virulence of an infection?
Virus? Host? Environment?

Trait spVL is early predictor of time to AIDS
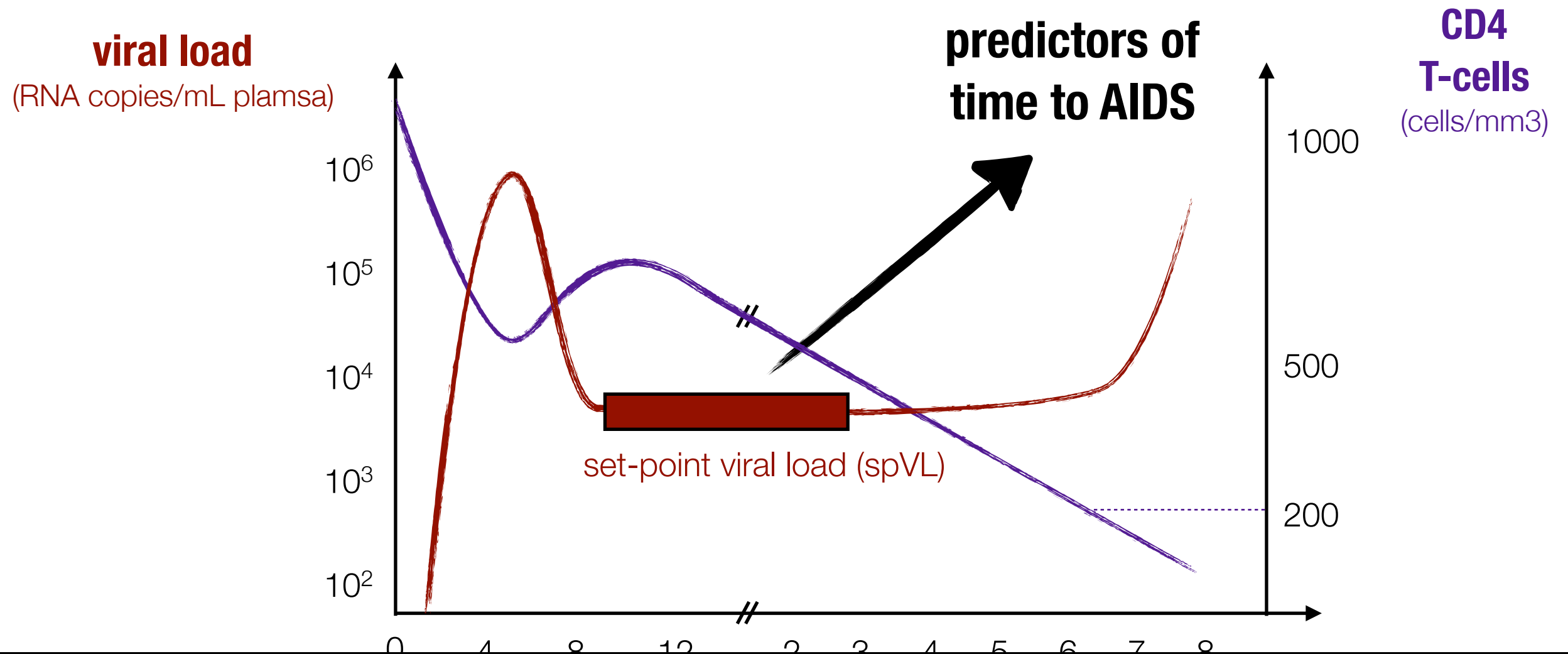
What determines the virulence of an infection?
To what extend does the virus determine spVL?

Virus control is discussed controversially!

We claim that the methods are the problem, not the data!

# Approaches to determine the importance of the virus

Quanti-fication measure	**Heritability H^2**: Amount of variation in a trait explained by the virus genotype

Estimator

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).
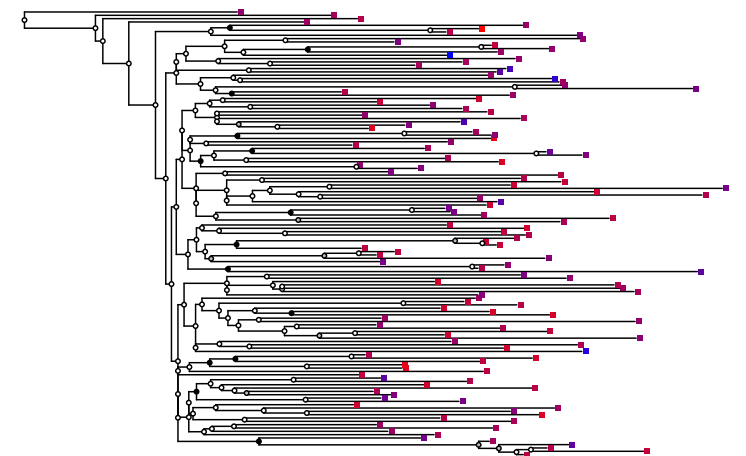
Infection

H^2 comes from quantitative genetics for sexual reproducing species

Transfer to pathogens

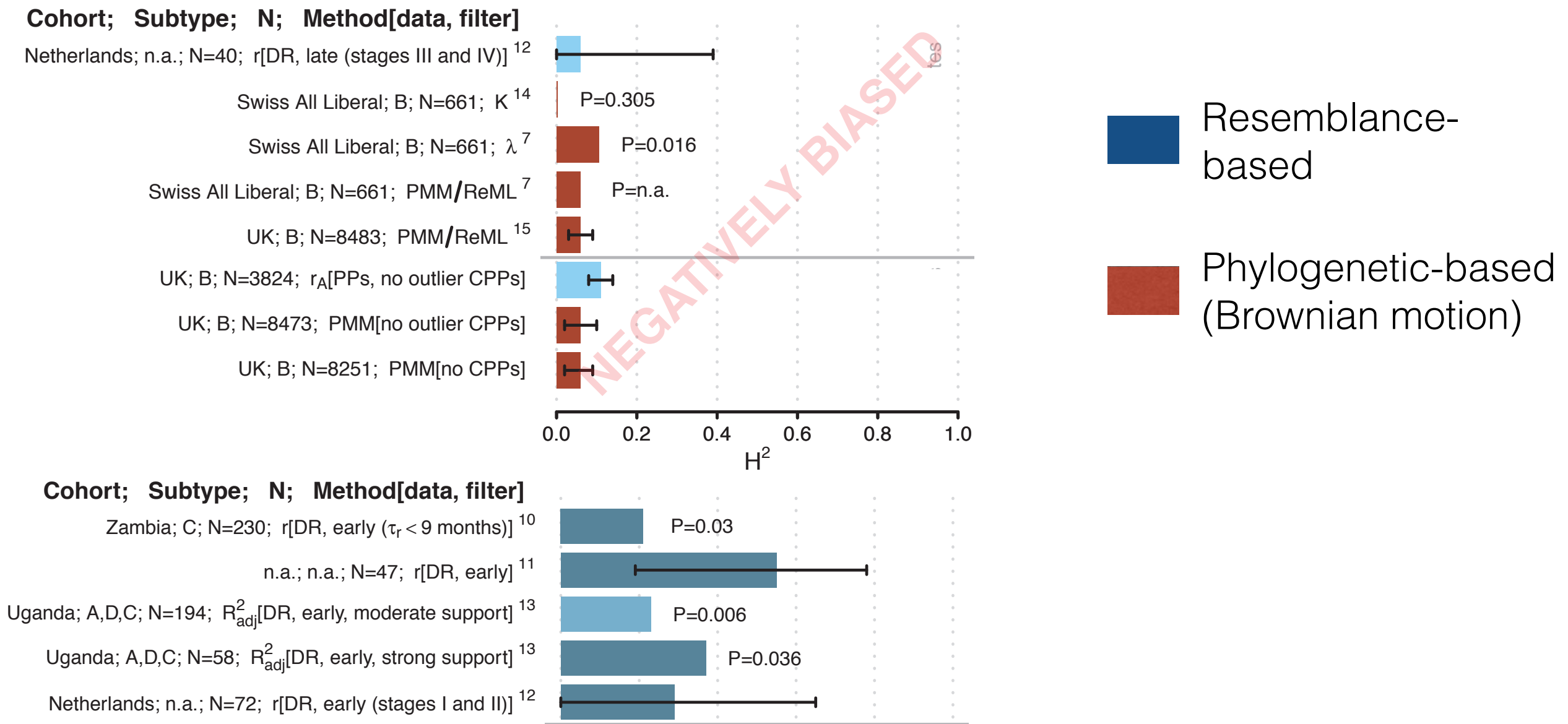Within-host evolution

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)

PMM

Selection on the trait is ignored due to assuming Brownian motion:

‣ **negative bias** (as selection gives rise to less genotypic variation than expected by Brownian motion)

Previous results

H^2

Cohort; Subtype; N; Method[data, filter]

Netherlands; n.a.; N=40; r[DR, late (stages III and IV)] [12]

Swiss All Liberal; B; N=661; K [14]   P=0.305

Swiss All Liberal; B; N=661; λ [7]   P=0.016

Swiss All Liberal; B; N=661; PMM/ReML [7]   P=n.a.

UK; B; N=8483; PMM/ReML [15]

UK; B; N=3824; r_A[PPs, no outlier CPPs]

UK; B; N=8473; PMM[no outlier CPPs]

UK; B; N=8251; PMM[no CPPs]

0.0   0.2   0.4   0.6   0.8   1.0
H^2

Resemblance-based

Phylogenetic-based (Brownian motion)

Cohort; Subtype; N; Method[data, filter]

Zambia; C; N=230; r[DR, early ($\tau_r < 9$ months)] [10]   P=0.03

n.a.; n.a.; N=47; r[DR, early] [11]

Uganda; A,D,C; N=194; R^2_{adj}[DR, early, moderate support] [13]   P=0.006

Uganda; A,D,C; N=58; R^2_{adj}[DR, early, strong support] [13]   P=0.036

Netherlands; n.a.; N=72; r[DR, early (stages I and II)] [12]

What is the true value of H^2?

7

H² estimators overcoming the biases

Resem-blance based

Anova-CPP *(extending PP by Shirreff et al., 2012)*

▸ Input is a phylogenetic tree; closest phylogenetic pairs (CPP) are determined

▸ Anova on CPPs to determine how much more similar they are to each other than across pairs

Phylo-genetic based

POUMM *(generalizing PMM to selection)*

▸ Input is a phylogenetic tree

▸ Assumptions:

- Ornstein-Uhlenbeck process for genotypic trait evolution

- Contribution from host is drawn from a normal distribution

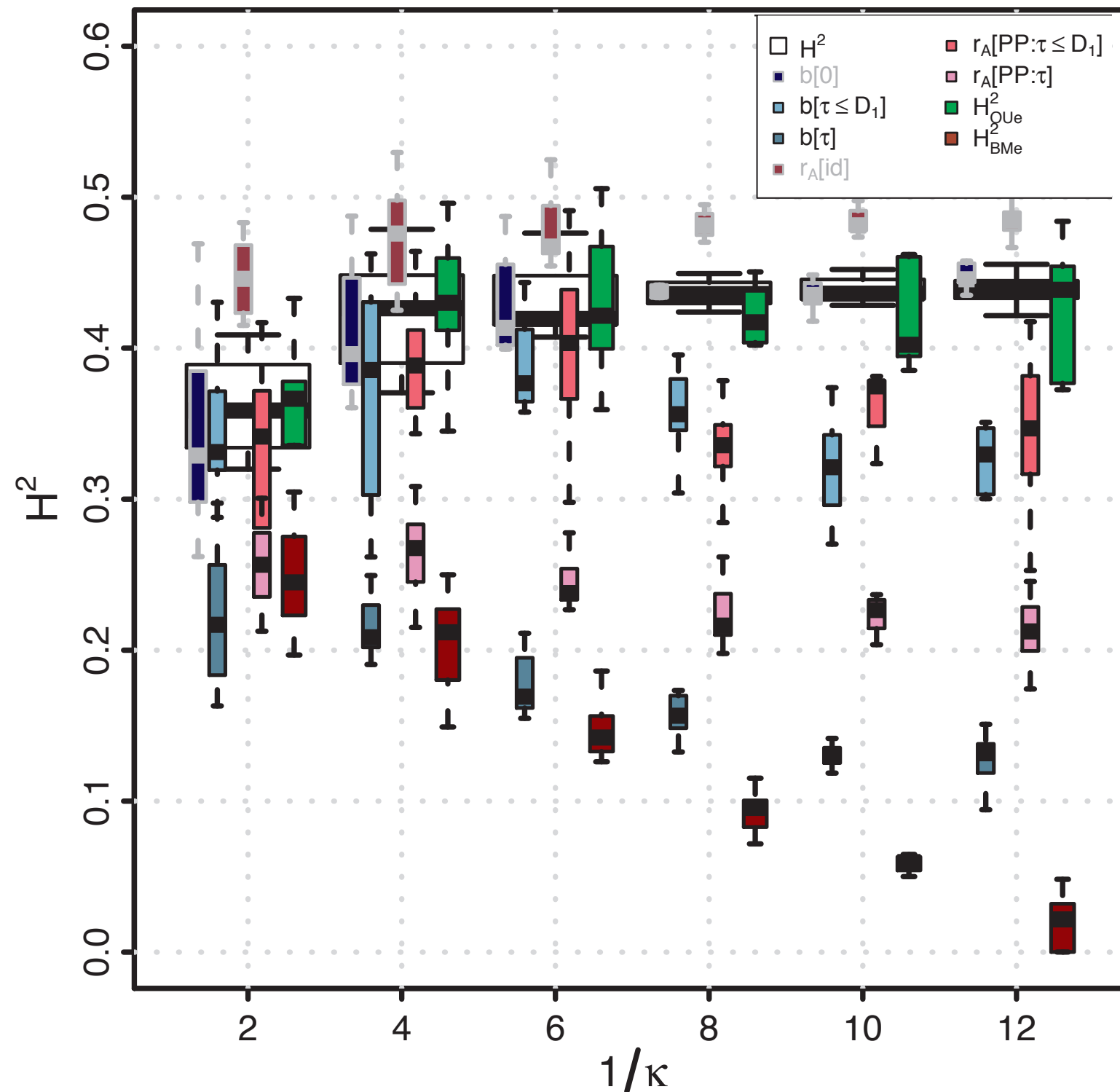▸ Maximum-likelihood estimation of relative contribution of genotype

POUMM
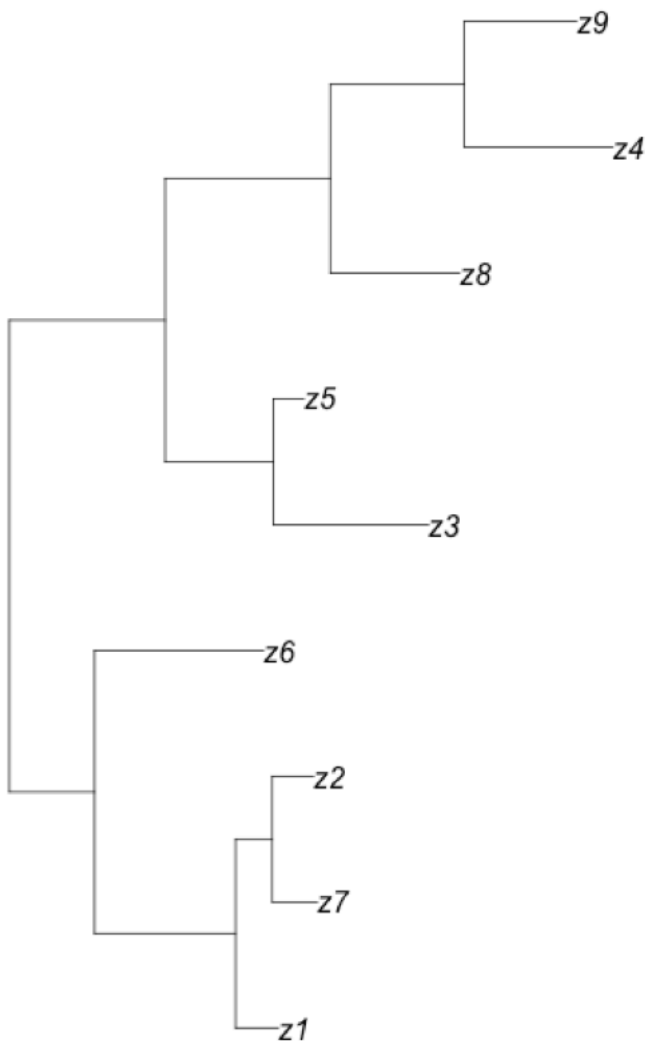
Tip trait z = G + e

‣ G: genotypic contribution, evolves according to an Ornstein-Uhlenbeck process with parameters $(\Theta, \alpha, \sigma^2)$

‣ e: host contribution, assumed to be drawn from a normal distribution $N(0, \sigma_e^2)$

We estimated $(\Theta, \alpha, \sigma^2, \sigma_e^2)$ and thus

$H^2 = 1 - \sigma_e^2/\sigma^2(z)$
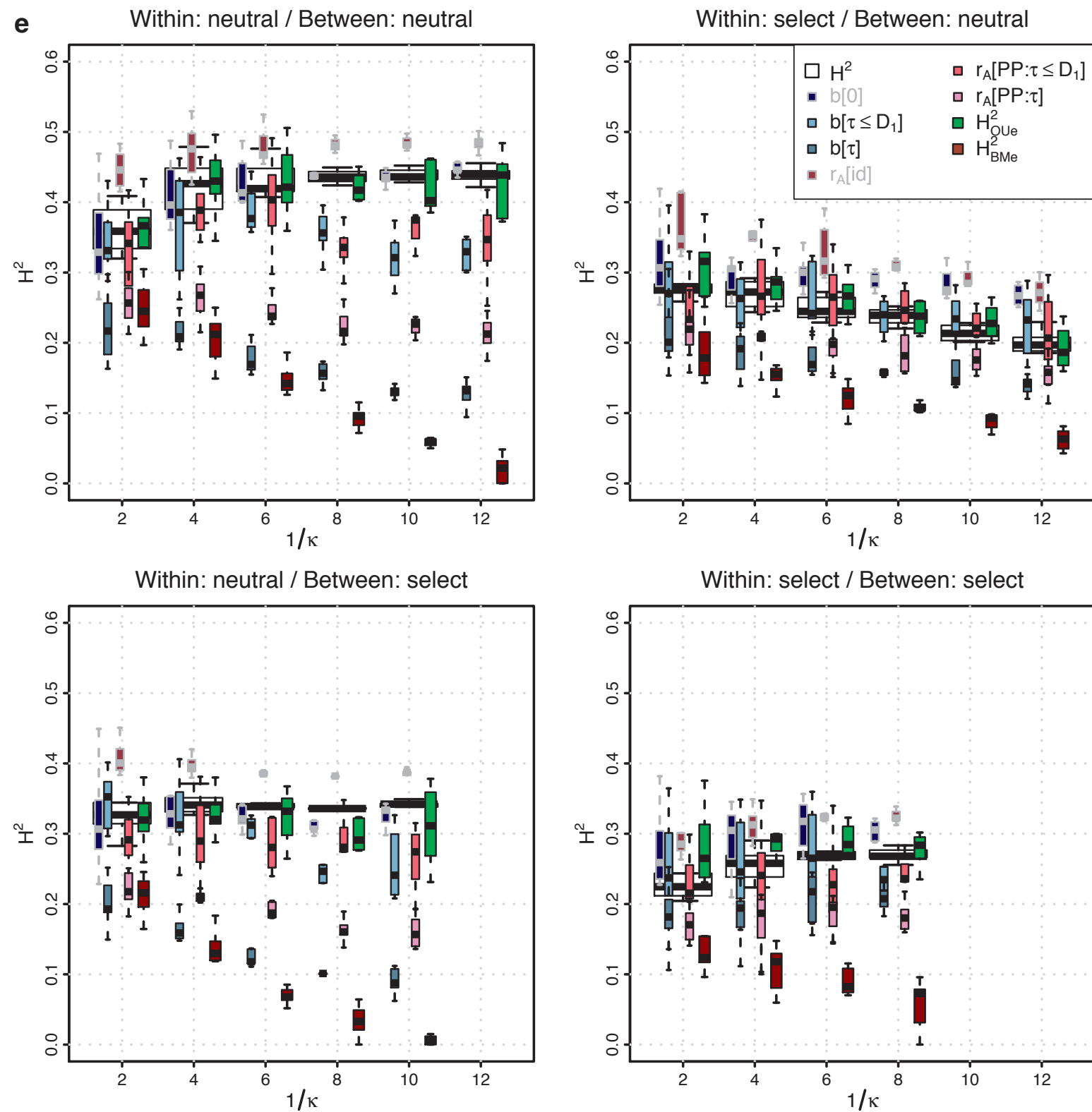
Simulation results

Resemblance-based:

b: DR

rA: Anova-CPP

Phylogenetic-based:

H^2_{OUe}: POUMM

H^2_{BMe}: PMM

Within: neutral / Between: neutral

H^2

$1/\kappa$

Within: neutral / Between: select

Within: select / Between: select

H^2

□ H^2		■ $r_A[PP:\tau \le D_1]$
■ $b[0]$		■ $r_A[PP:\tau]$
■ $b[\tau \le D_1]$		■ H^2_{OUe}
■ $b[\tau]$		■ H^2_{BMe}
■ $r_A[id]$		

□ H^2	
■ $b[0]$	
■ $b[\tau \le D_1]$	
■ $b[\tau]$	
■ $r_A[id]$	

$1/\kappa$ 10

Simulation results

Empirical results

H^2

Cohort; Subtype; N; Method[data, filter]

Netherlands; n.a.; N=40; r[DR, late (stages III and IV)] [12]

Swiss All Liberal; B; N=661; K [14]   P=0.305

Swiss All Liberal; B; N=661; λ [7]   P=0.016

Swiss All Liberal; B; N=661; PMM/ReML [7]   P=n.a.

UK; B; N=8483; PMM/ReML [15]

UK; B; N=3824; r_A[PPs, no outlier CPPs]

UK; B; N=8473; PMM[no outlier CPPs]

UK; B; N=8251; PMM[no CPPs]

NEGATIVELY BIASED

0.0   0.2   0.4   0.6   0.8   1.0

H^2

Resemblance-based

Phylogenetic-based (Brownian motion)

Phylogenetic-based (Ornstein-Uhlenbeck)

a

Cohort; Subtype; N; Method[data, filter]

Zambia; C; N=230; r[DR, early ($\tau_r < 9$ months)] [10]   P=0.03

n.a.; n.a.; N=47; r[DR, early] [11]

Uganda; A,D,C; N=194; R^2_{adj}[DR, early, moderate support] [13]   P=0.006

Uganda; A,D,C; N=58; R^2_{adj}[DR, early, strong support] [13]   P=0.036

Netherlands; n.a.; N=72; r[DR, early (stages I and II)] [12]

UK; B; N=222; r_A[CPPs, no outlier CPPs]

Virus determines at least 20-30% of lg(spVL) variation!

0.0   0.2   0.4   0.6   0.8   1.0

H^2

12

Estimating the host contribution for each patient

Tip trait z = G + e

▸ G: genotypic contribution, evolves according to an Ornstein-Uhlenbeck process with parameters $(\Theta, \alpha, \sigma^2)$

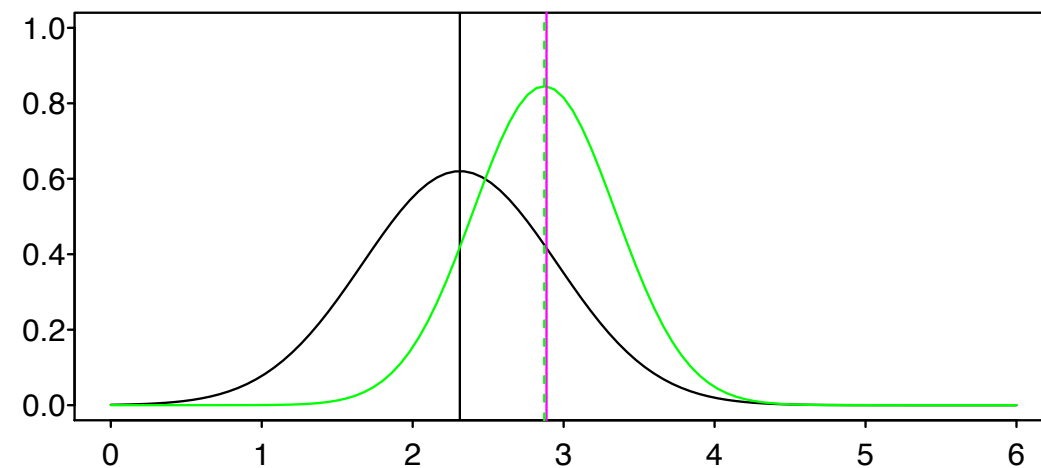▸ e: host contribution, assumed to be drawn from a normal distribution $N(0, \sigma_e^2)$

So far we estimated $(\Theta, \alpha, \sigma^2, \sigma_e^2)$ and thus

$H^2 = 1 - \sigma_e^2/\sigma^2(z)$

by implicitly integrating over all possible G and e (such that G + e = z) for each tip

Now we explicitly sample G and e in the estimation method!

Estimating the host contribution for each patient

——— z and z-N$(0,\sigma_e^{2)}$

——— true G = z-e (unknown)

We can separate z into virus contribution G and host contribution e -> **GWAS on G and e!**

Conclusions & Outlook

H^2 for pathogens

Resemblance-based: ANOVA-CPP

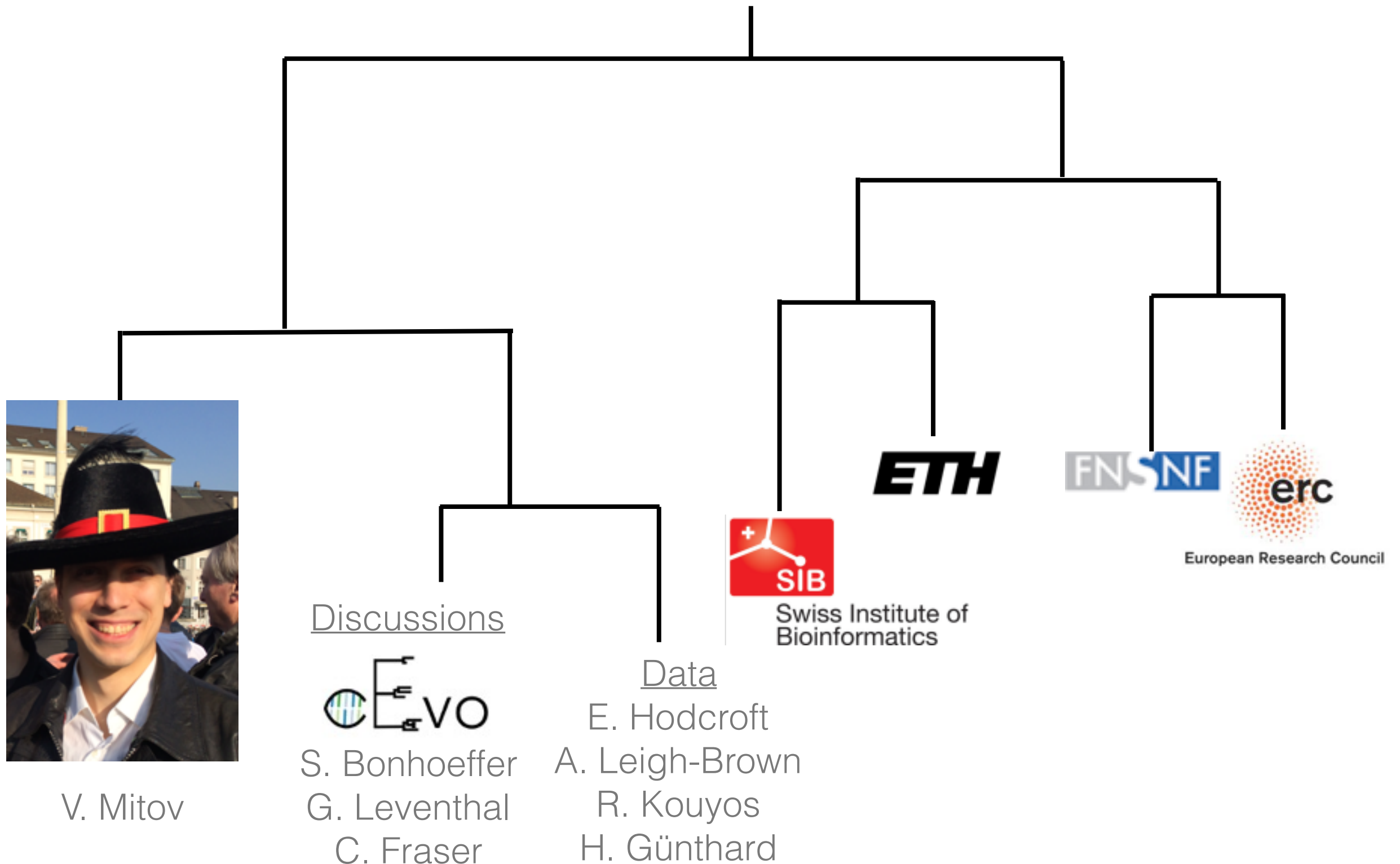Phylogenetic-based: POUMM

HIV virulence

Previous controversy due to methods

Now all data support >20% of lg(spVL) variation being explained by virus

Outlook

Disentangle host and virus contribution $z = G + e$ via POUMM. Then:
- GWAS to **identify mutations** on host / virus genotype using trait value G and e (rather than z)

Phylogeny of Acknowledgements

V. Mitov

Discussions
cEvo
S. Bonhoeffer
G. Leventhal
C. Fraser

Data
E. Hodcroft
A. Leigh-Brown
R. Kouyos
H. Günthard

16

References

7 Shirreff, G. *et al.* How effectively can HIV phylogenies be used to measure heritability? *Evolution, Medicine, and Public Health* **2013,** 209–224 (2013).

10 Tang, J. *et al.* HLA allele sharing and HIV type 1 viremia in seroconverting Zambians with known transmitting partners. *AIDS Res. Hum. Retroviruses* **20,** 19–25 (2004).

11 Hecht, F. M. *et al.* HIV RNA level in early infection is predicted by viral load in the transmission source. *AIDS* **24,** 941–945 (2010).

12 van der Kuyl, A. C., Jurriaans, S., Pollakis, G., Bakker, M. & Cornelissen, M. HIV RNA levels in transmission sources only weakly predict plasma viral load in recipients. *AIDS* **24,** 1607–1608 (2010).

13 Hollingsworth, T. D. *et al.* HIV-1 transmitting couples have similar viral load set-points in Rakai, Uganda. *PLoS Pathog.* **6,** e1000876 (2010).

14 Alizon, S. *et al.* Phylogenetic approach reveals that virus genotype largely determines HIV set-point viral load. *PLoS Pathog.* **6,** e1001123 (2010).

15 Hodcroft, E. *et al.* The Contribution of Viral Genotype to Plasma Viral Set-Point in HIV Infection. *PLoS Pathog.* **10,** e1004112 (2014).