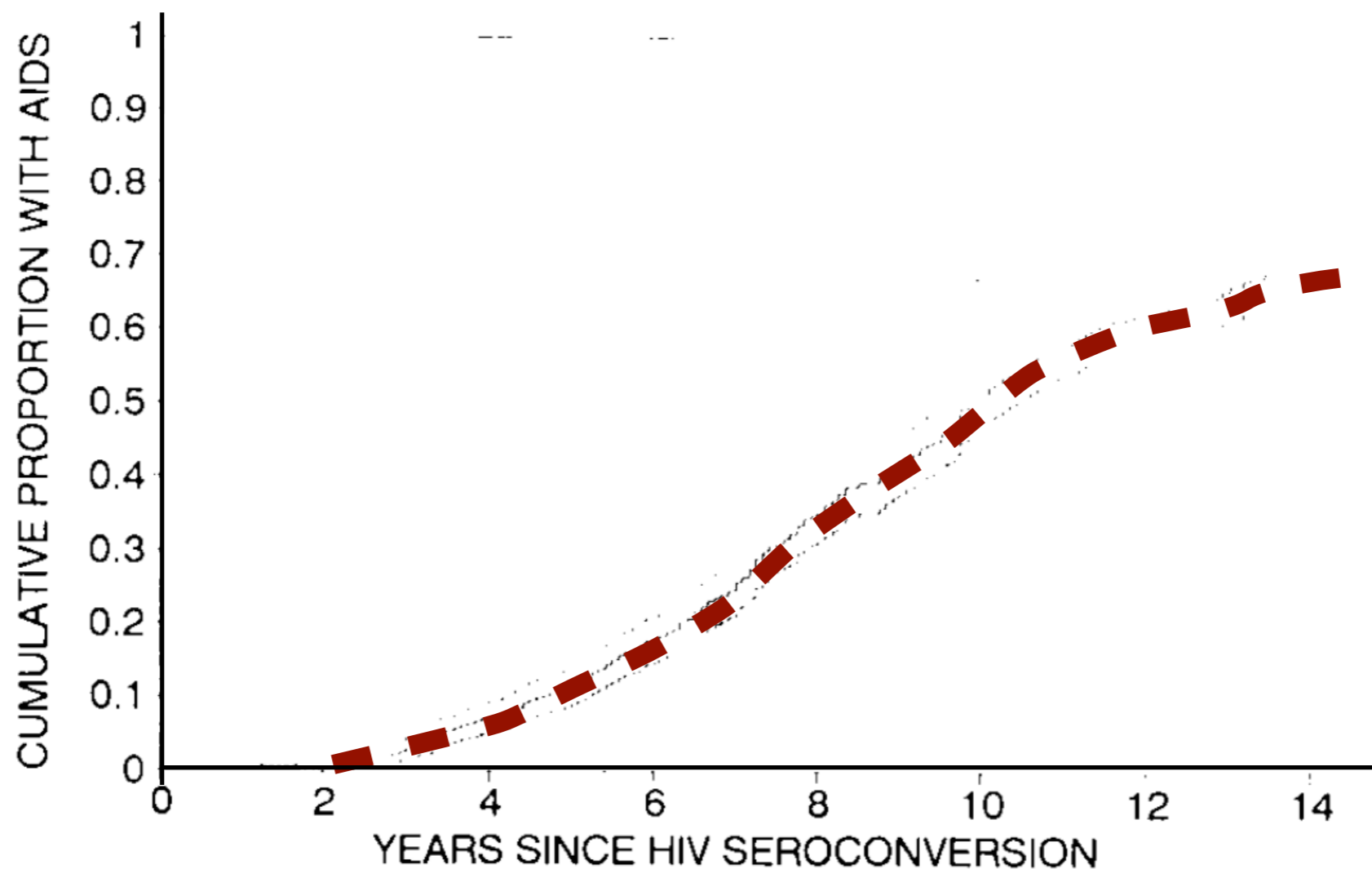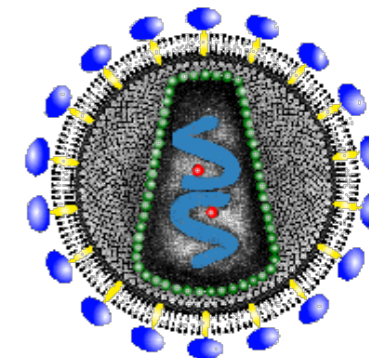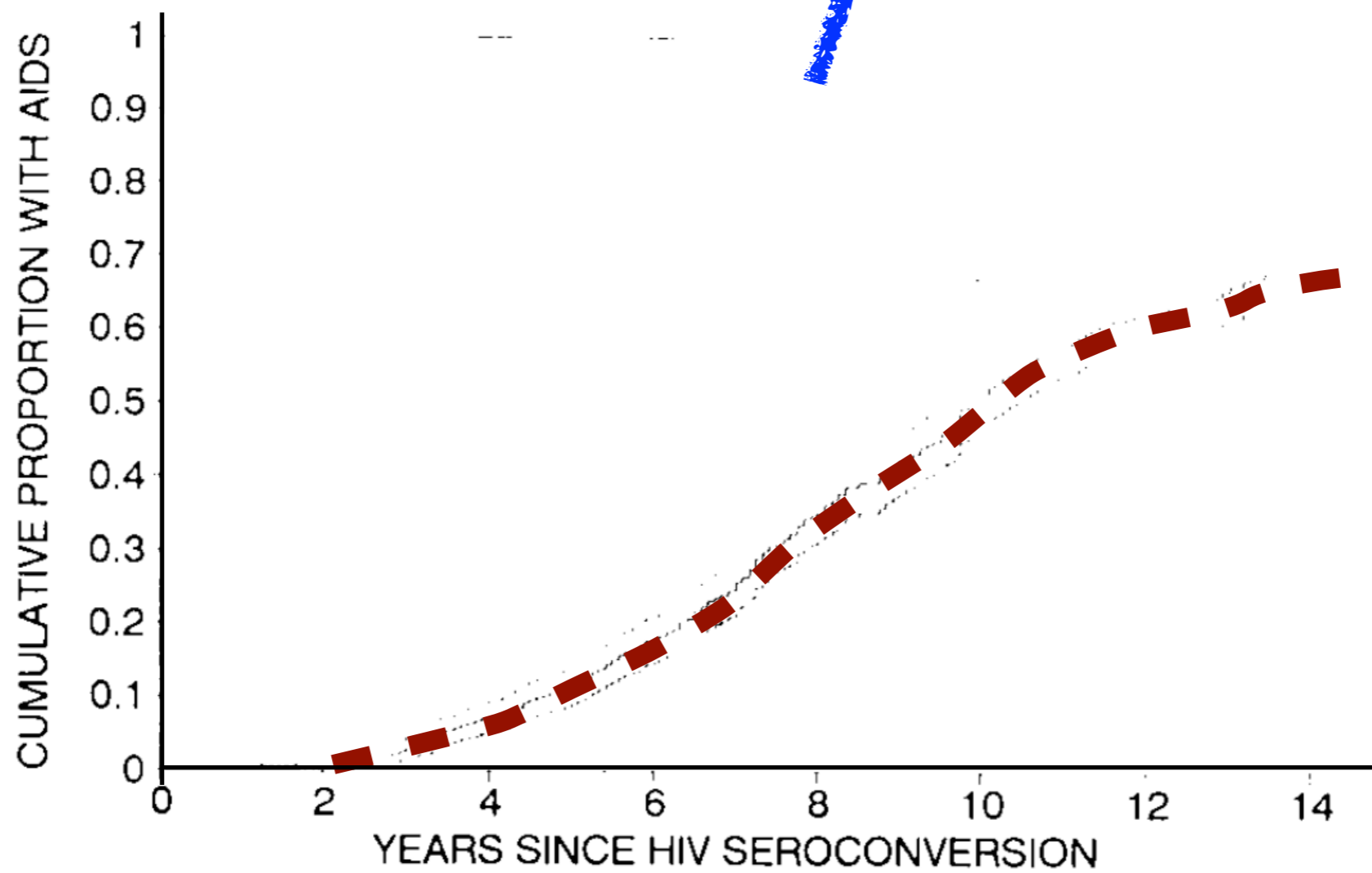# Why does HIV harm its host?
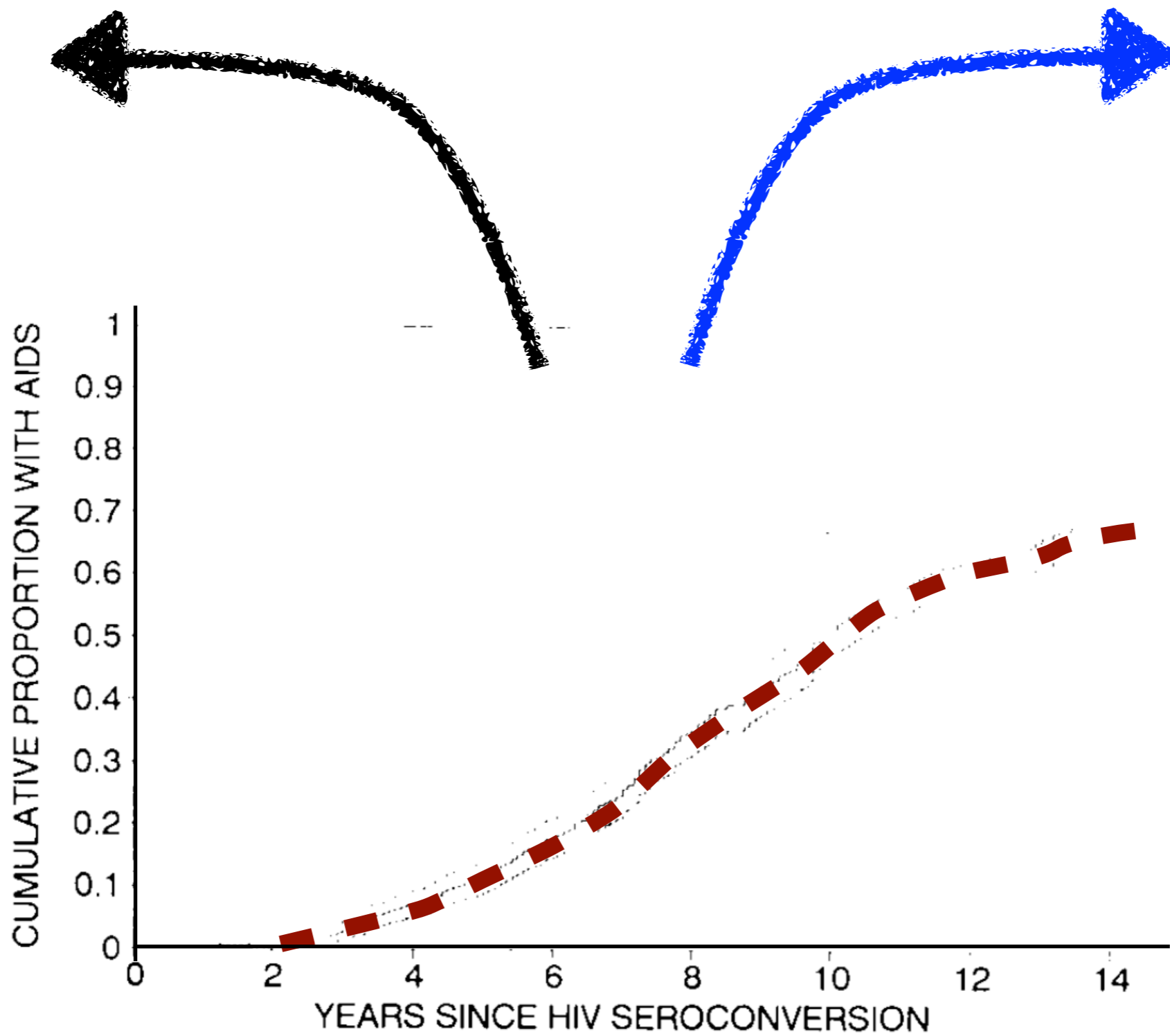
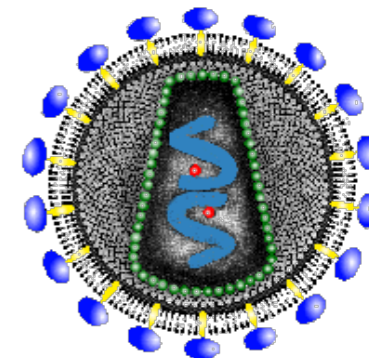Samuel Alizon
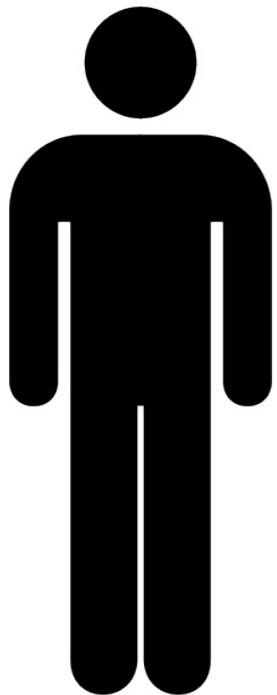
Ecological and Molecular mOdelling of infecTIONS

Lyon, 10-11 Dec 2014

MIVEGEC — UMR CNRS 5290 — Montpellier, France
samuel.alizon@cnrs.fr

Buchbinder *et al.* (1994, *AIDS*)

CUMULATIVE PROPORTION WITH AIDS

YEARS SINCE HIV SEROCONVERSION

Buchbinder *et al.* (1994, *AIDS*)

CUMULATIVE PROPORTION WITH AIDS

YEARS SINCE HIV SEROCONVERSION

Buchbinder *et al.* (1994, *AIDS*)
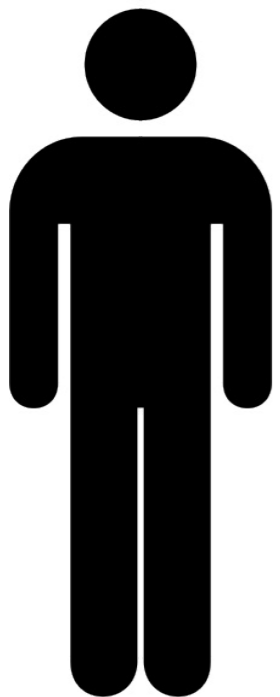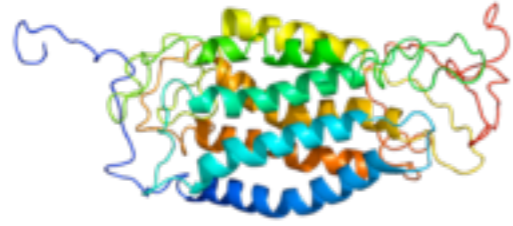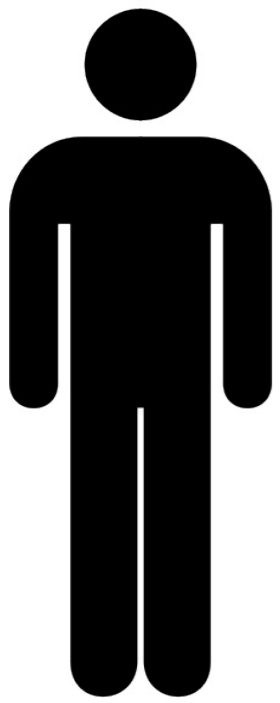
CCR5Δ32 deletion confers resistance to HIV infection

Liu *et al.* (1996, *Cell*)

CCR5Δ32 deletion confers resistance to HIV infection

Liu *et al.* (1996, *Cell*)

A handful of SNPs explains ~22% of the variance in spVL
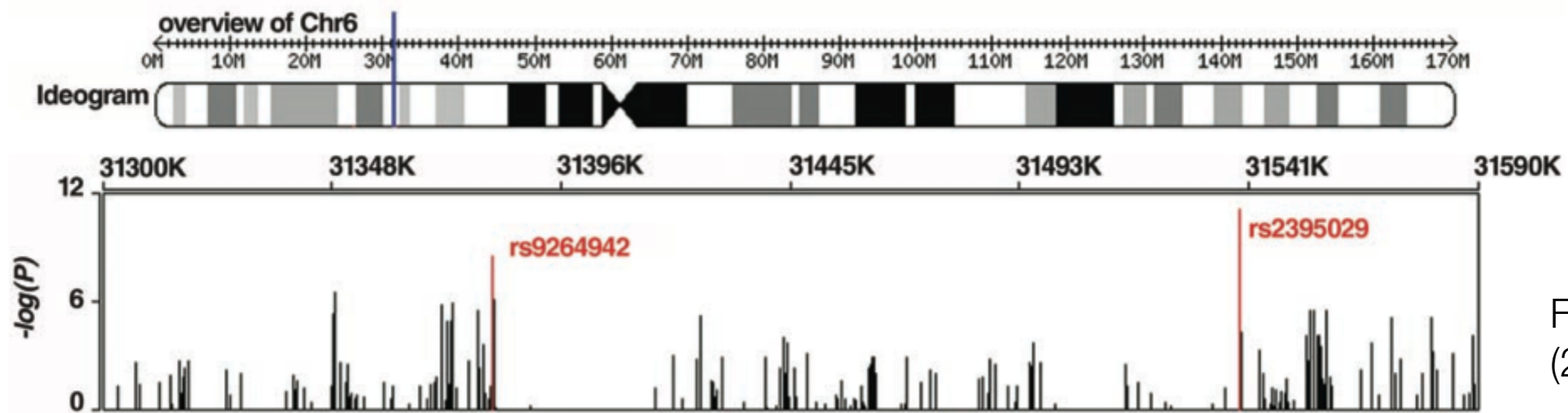
Fellay *et al.* (2007, *Science*)
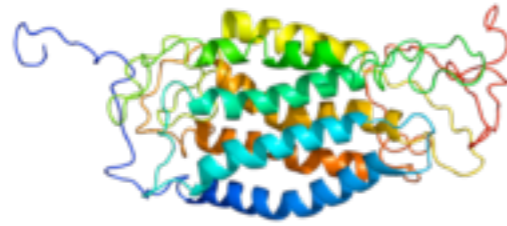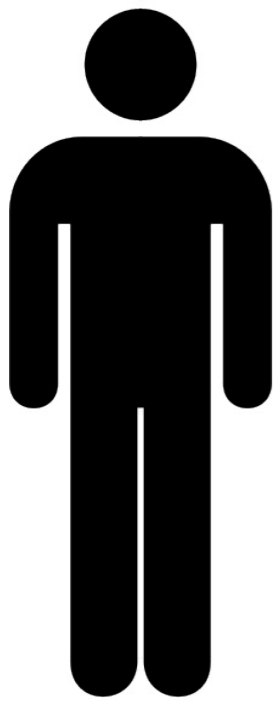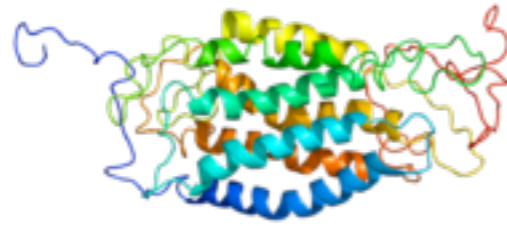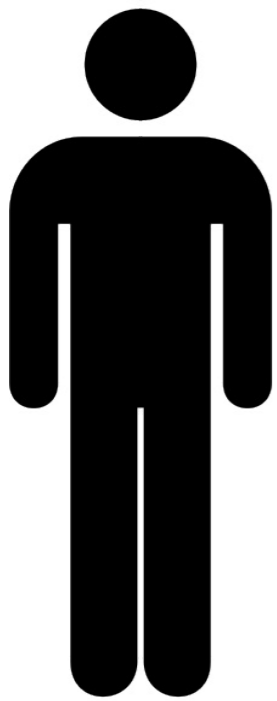
CCR5Δ32 deletion confers resistance to HIV infection

Liu *et al.* (1996, *Cell*)
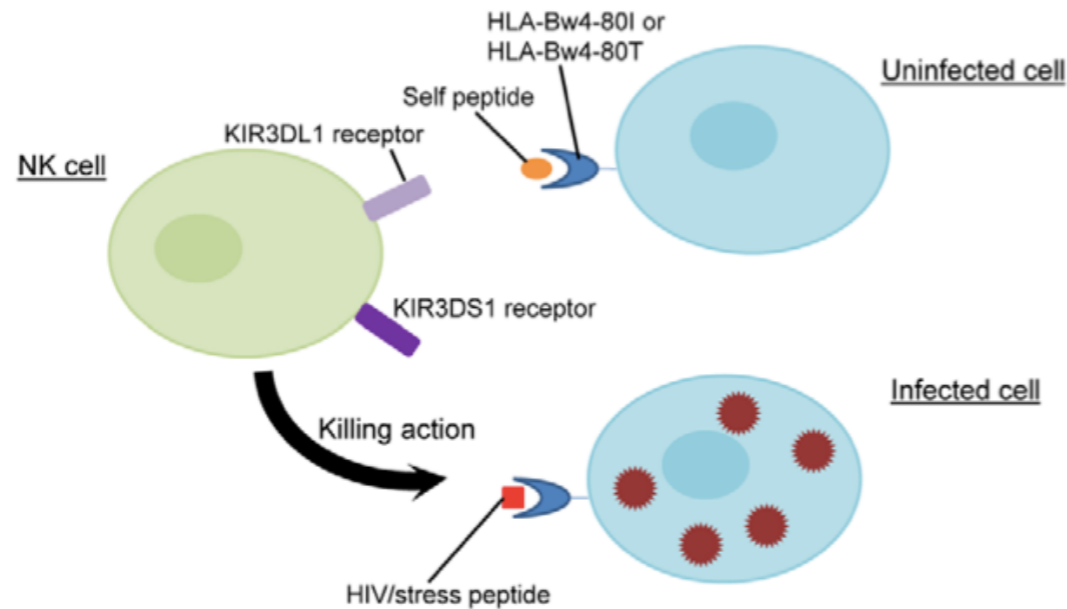
KIR copy numbers inversely correlate with viral load

Pelak *et al.* (2011, *PLoS Biol*)

A handful of SNPs explains ~22% of the variance in spVL
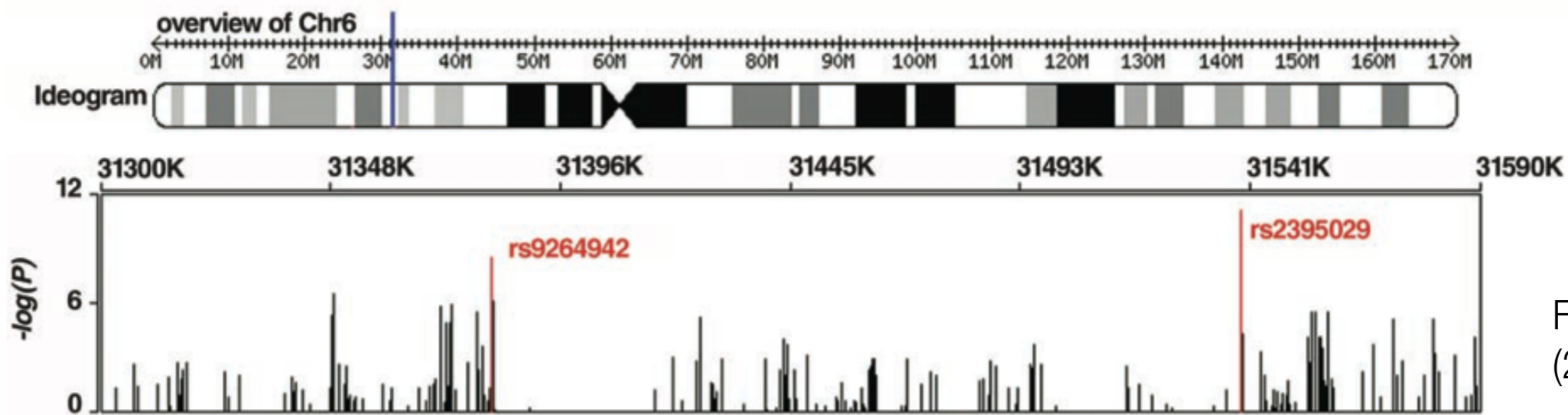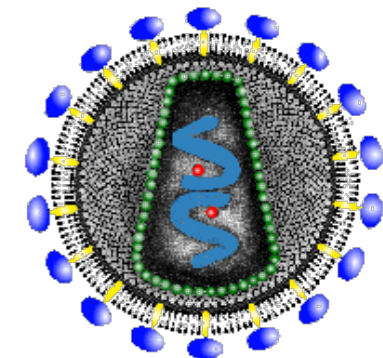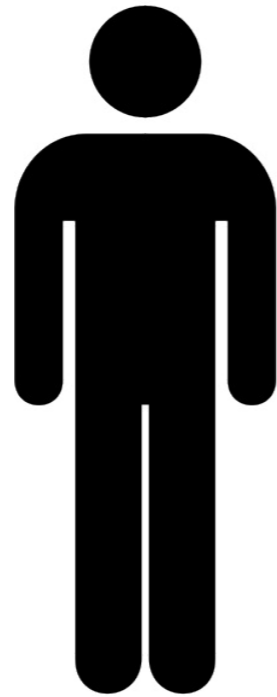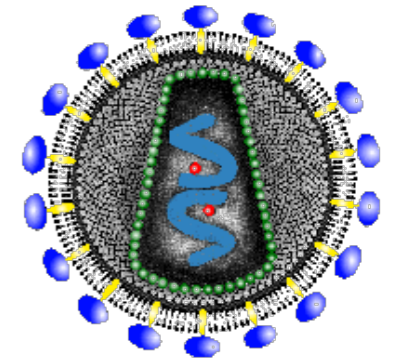
Fellay *et al.* (2007, *Science*)

✝ within 3 years
following infection
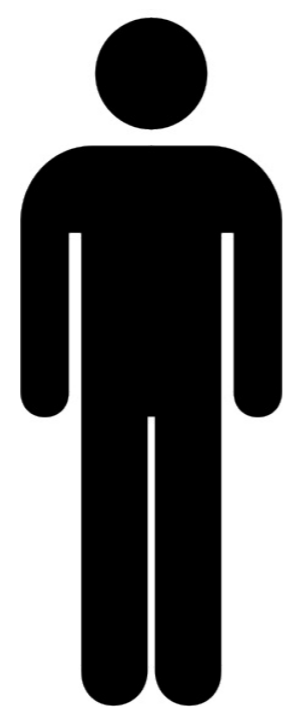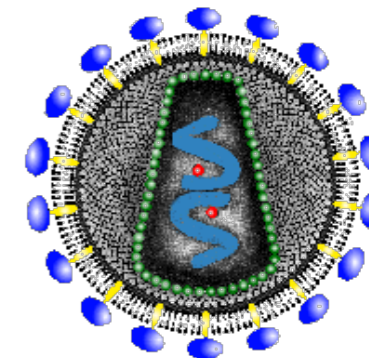
infection

✝ within 3 years
following infection
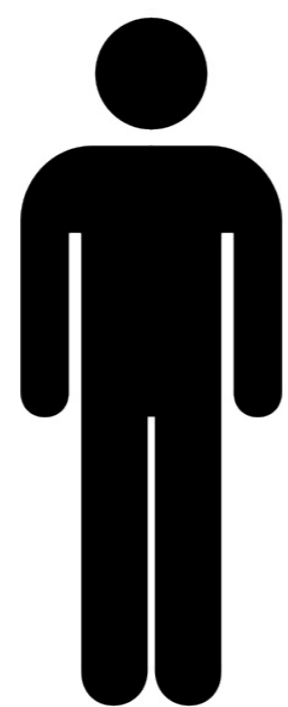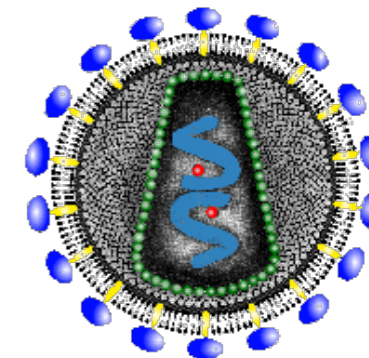
infection

† within 3 years
following infection

???

HIV infection course
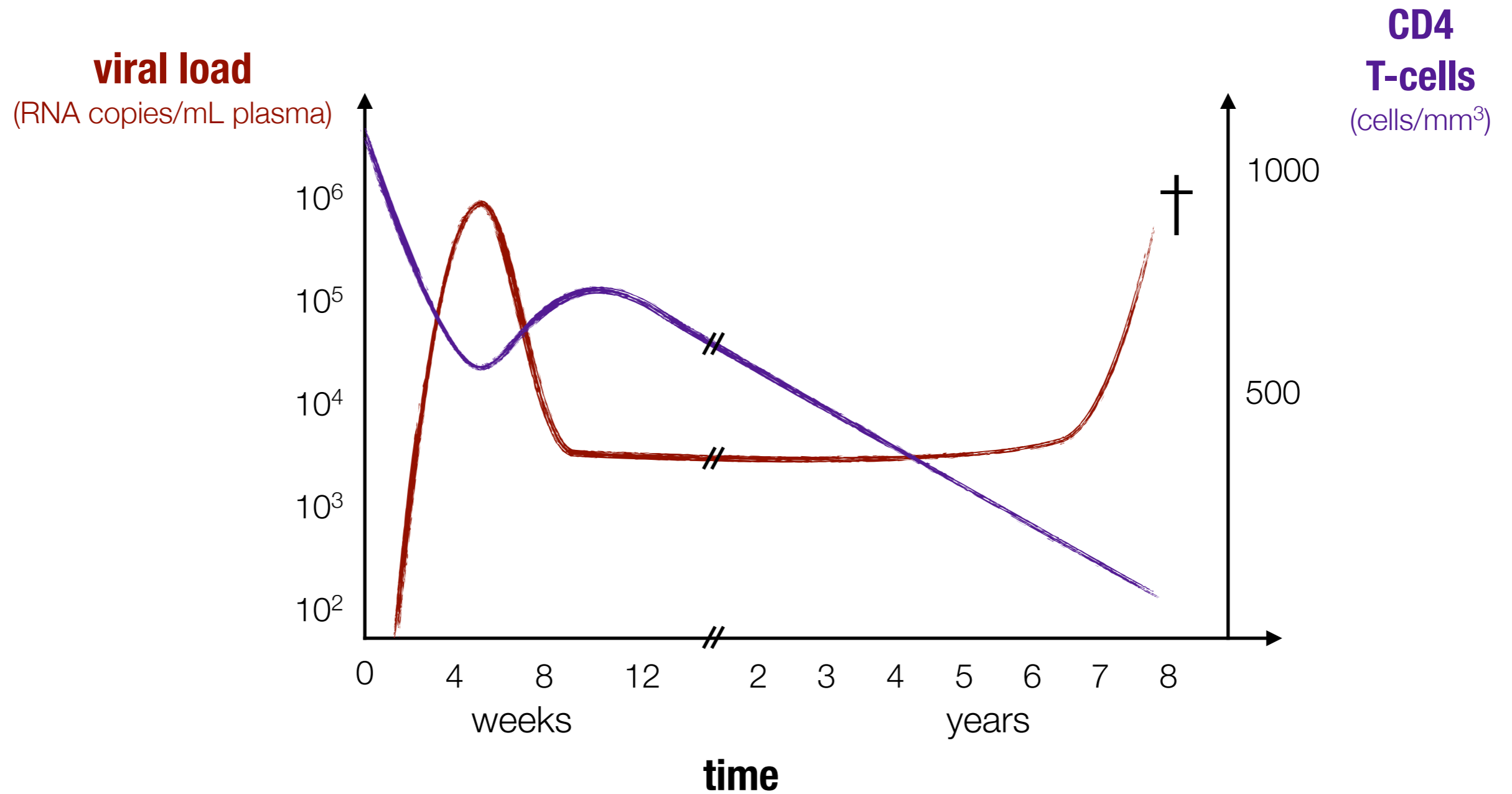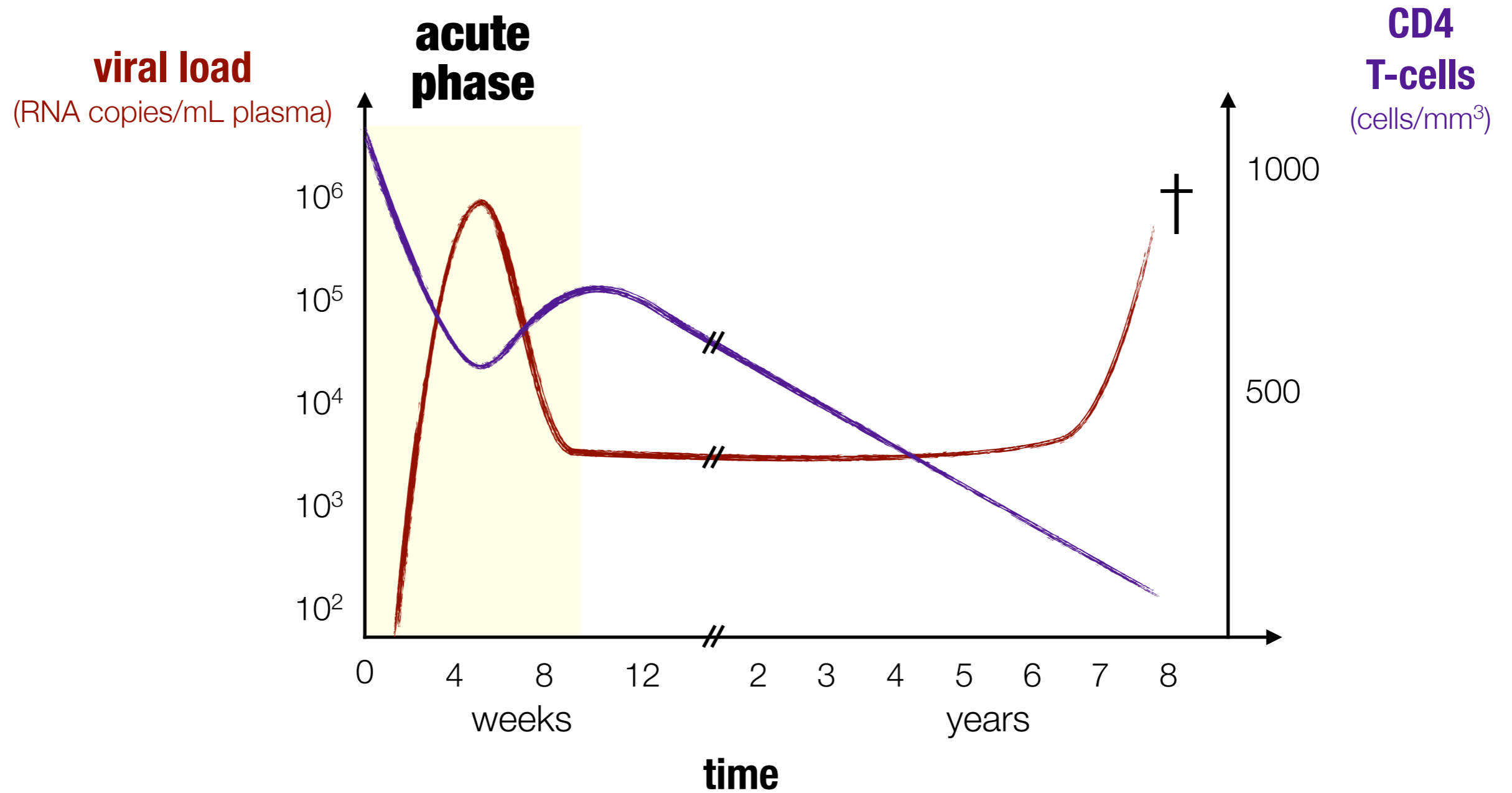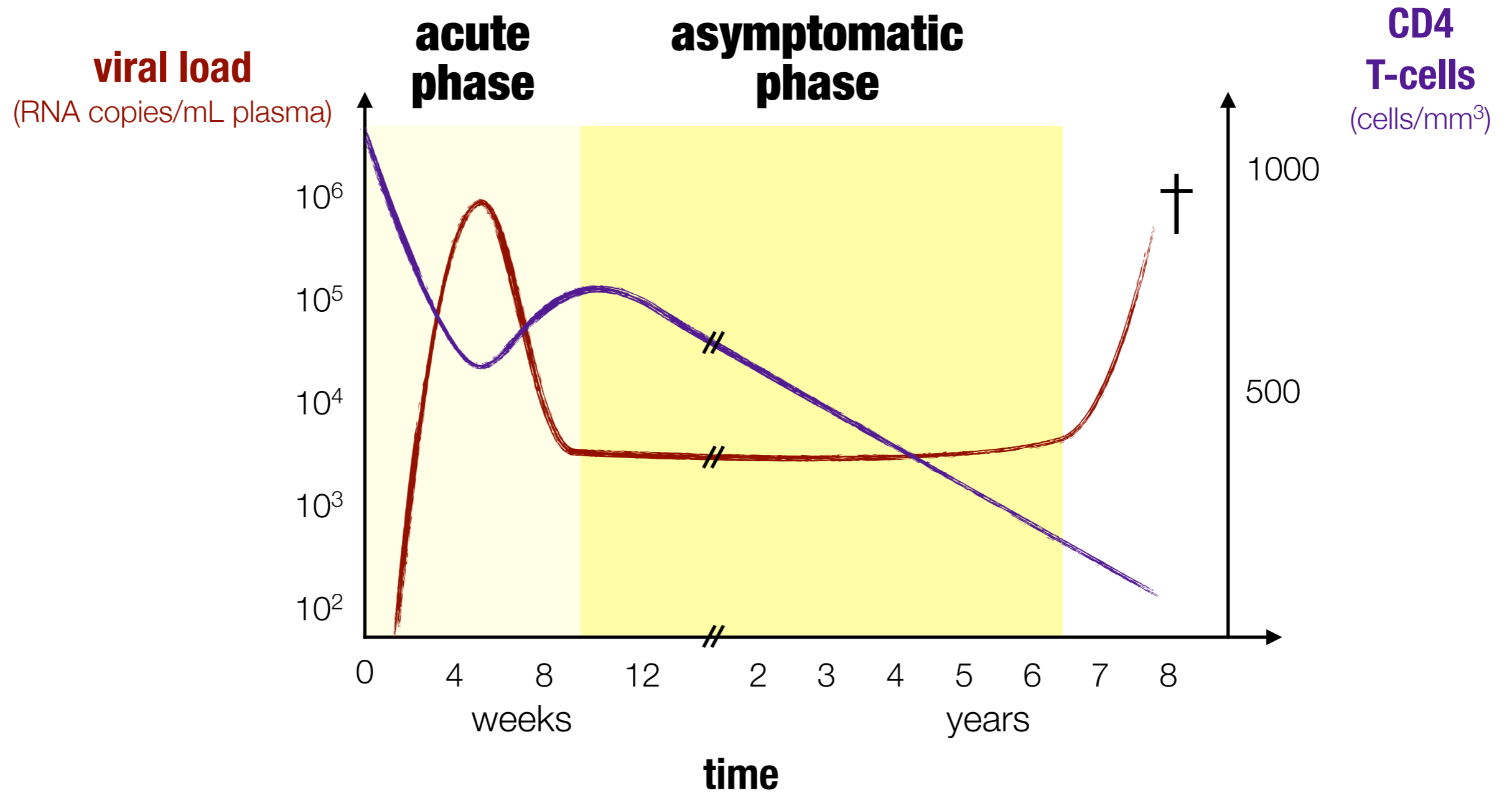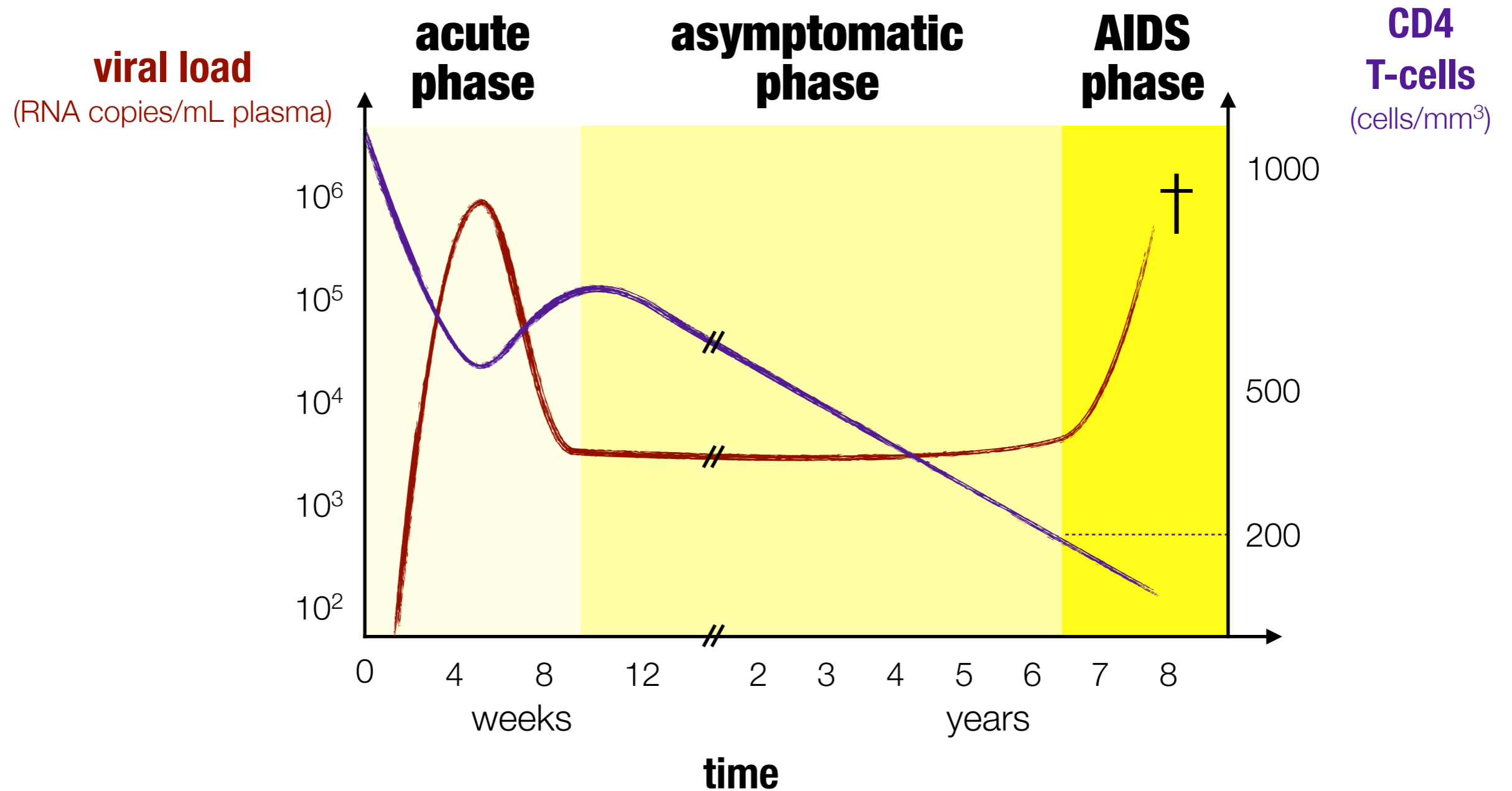
HIV infection course

HIV infection course

HIV infection course

HIV infection course

HIV infection course

HIV infection course

HIV infection course

HIV infection course

viral load
(RNA copies/mL plasma)

CD4
T-cells
(cells/mm^3)

predictors of virulence

CD4 decline slope

set-point viral load

10^6

10^5

10^4

10^3

10^2

1000

500

200

0   4   8   12   2   3   4   5   6   7   8

weeks   years

time

Mellors *et al.* (1996, *Science*)
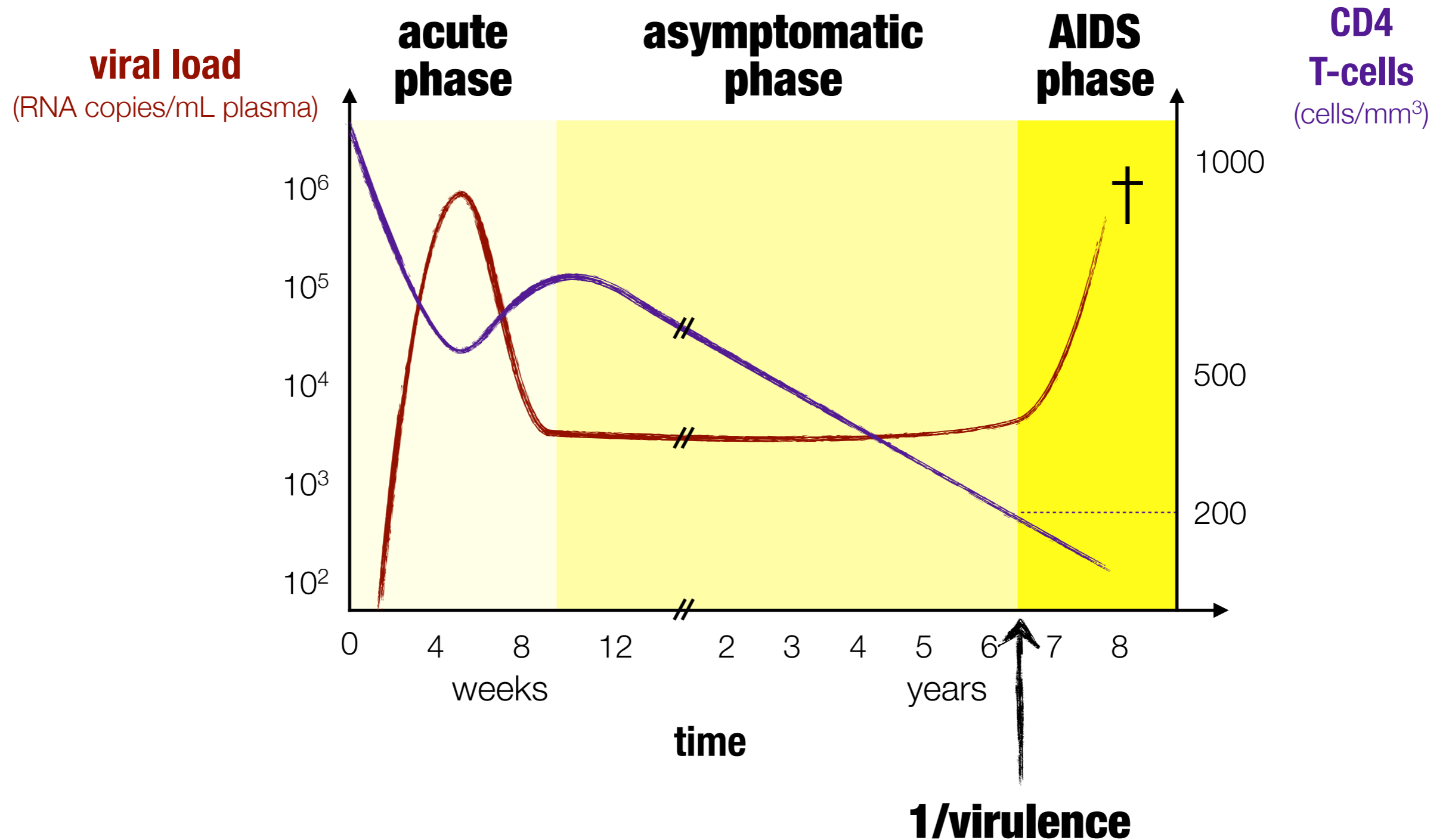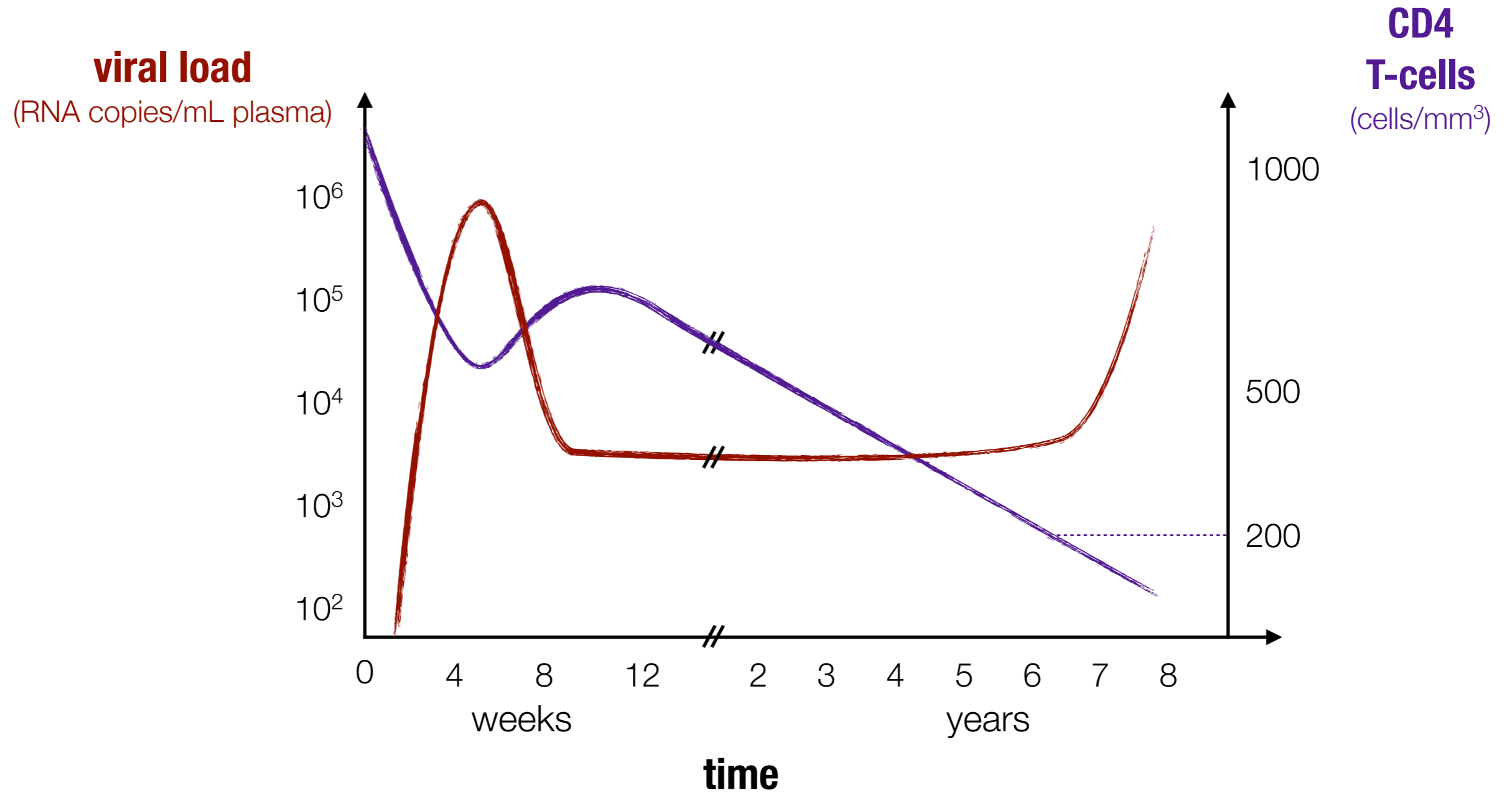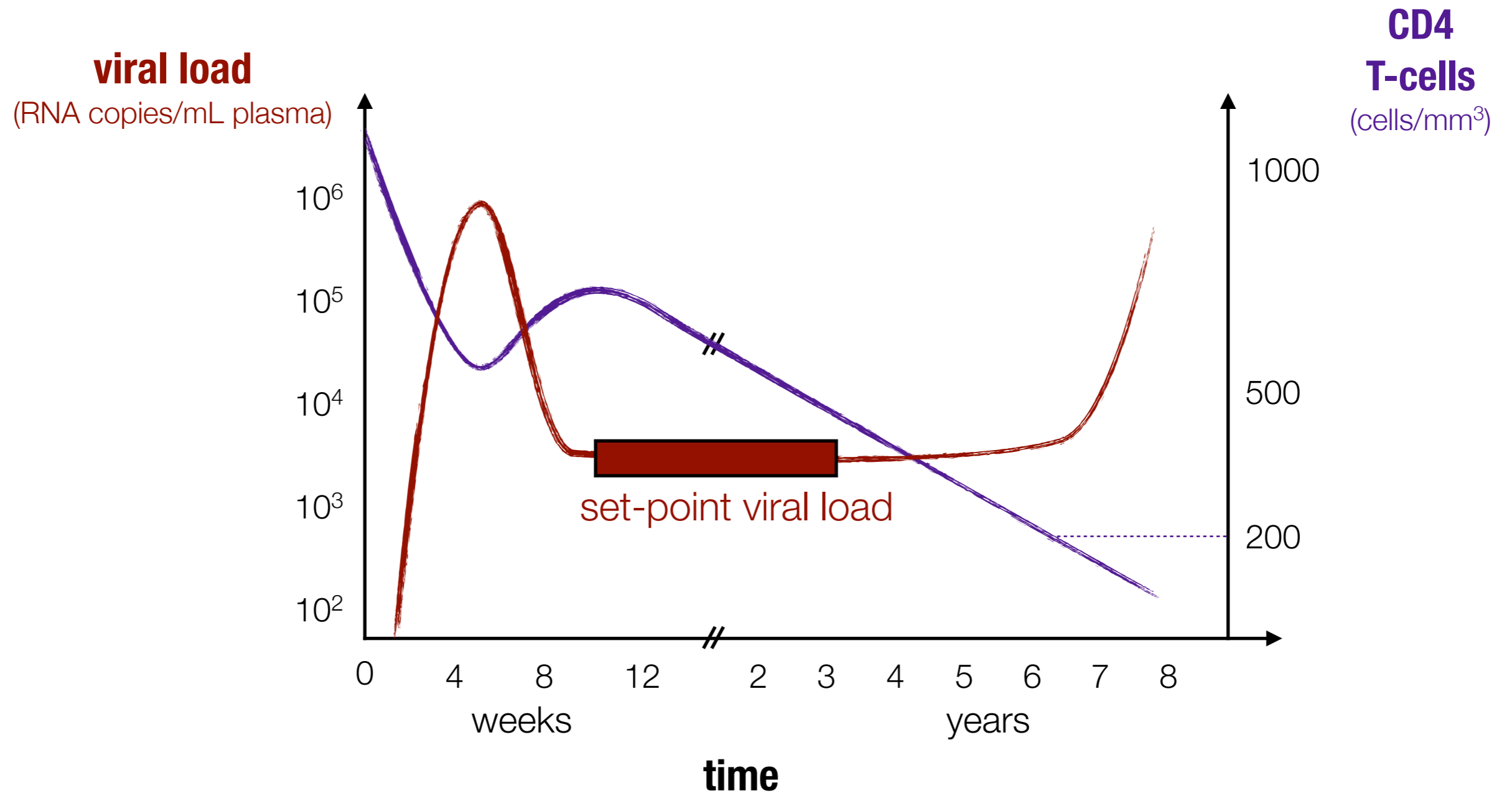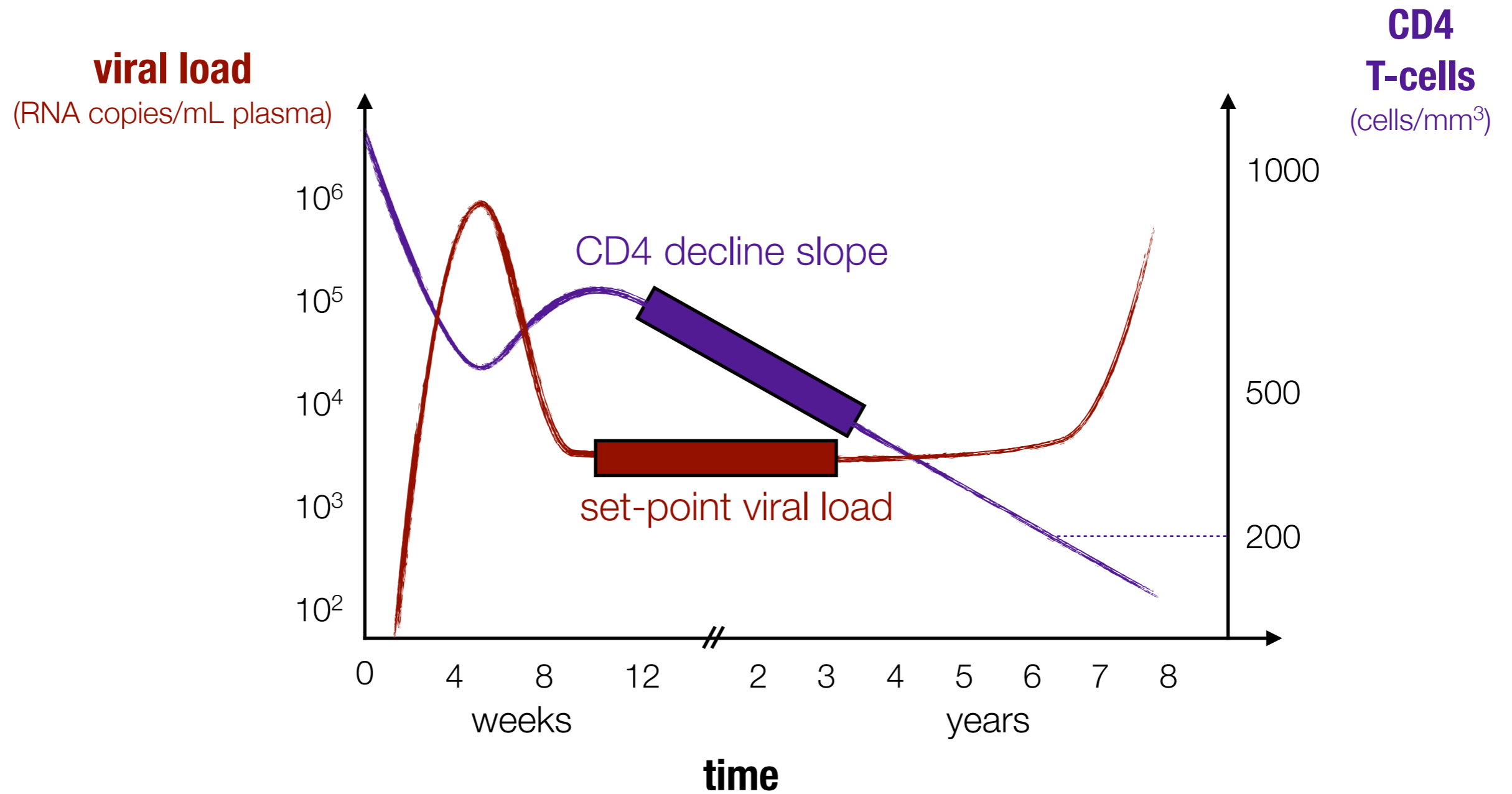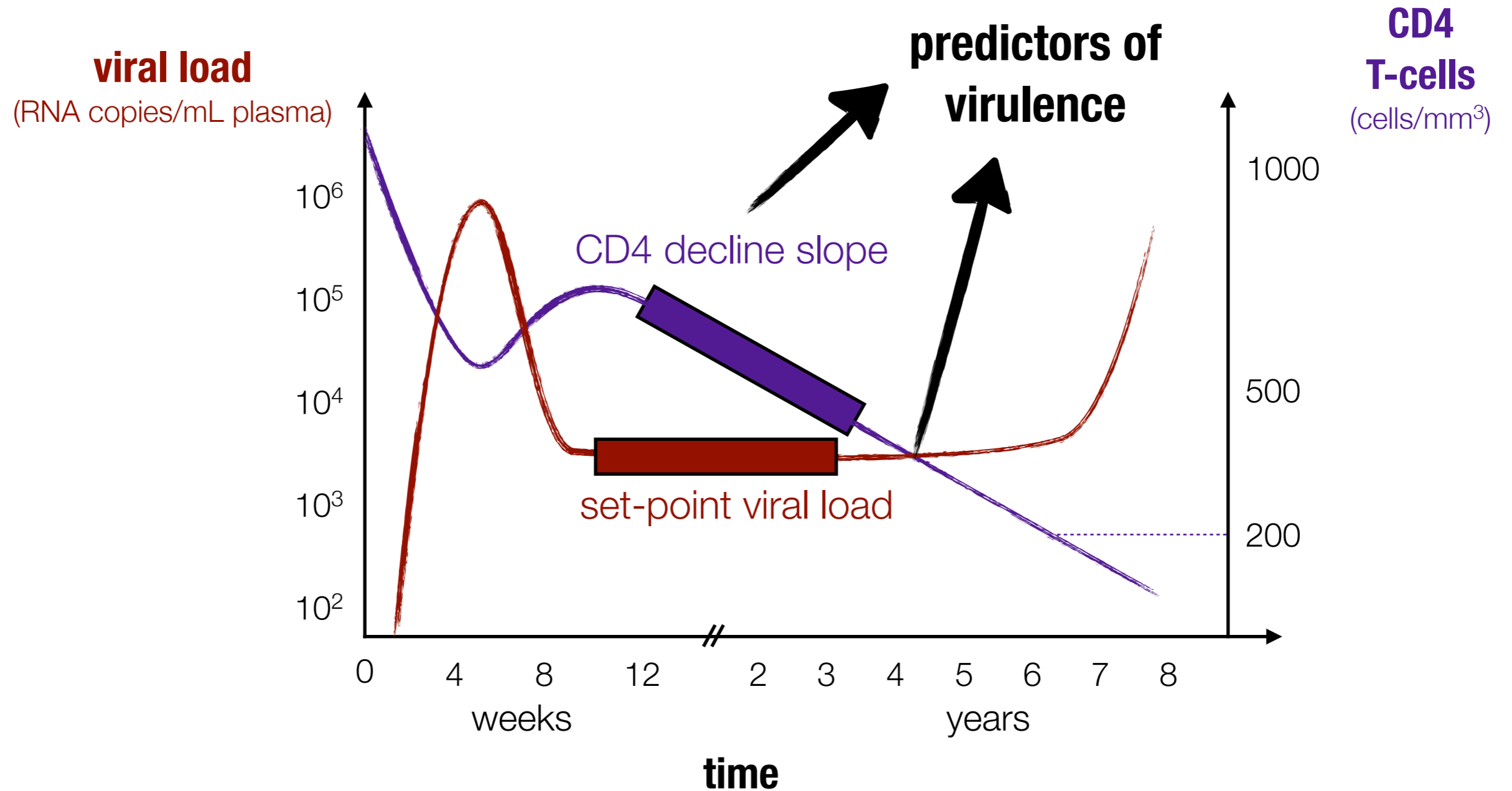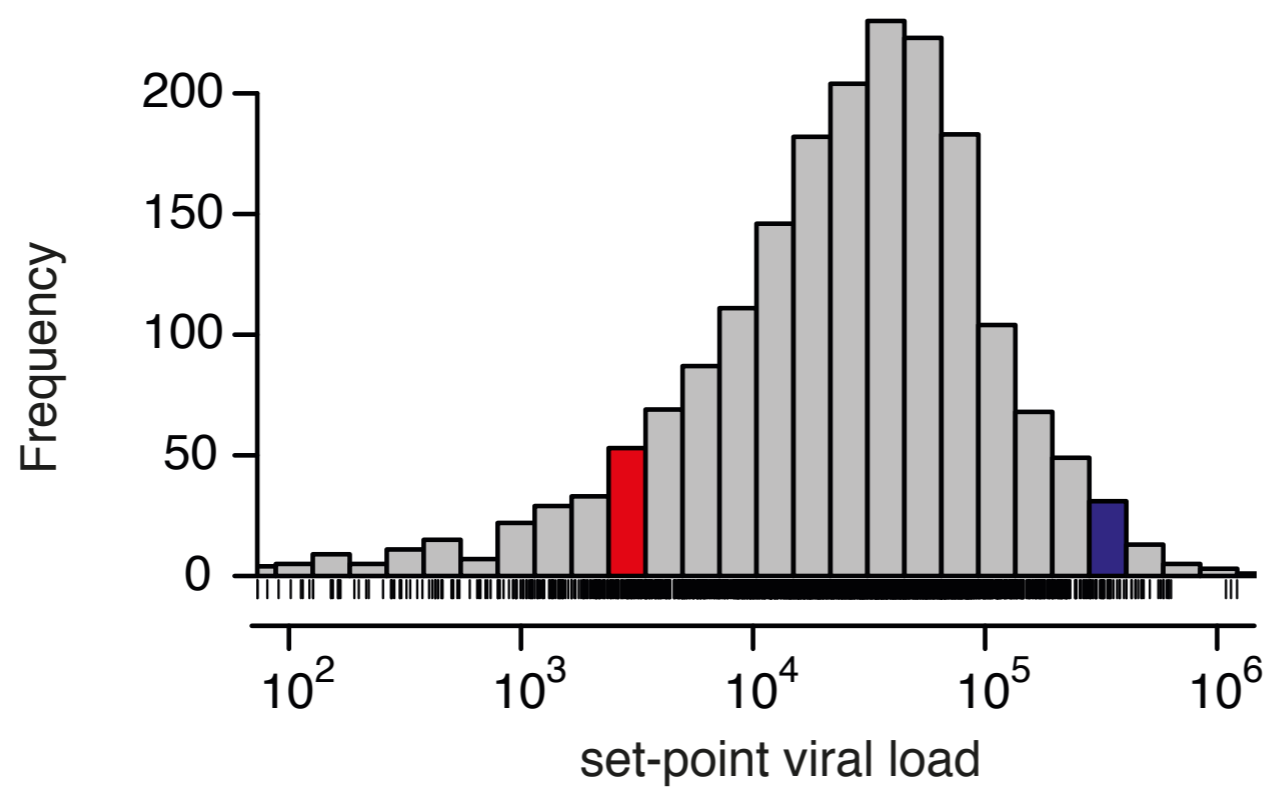
Virus load variations

Fraser *et al.*
(2014, *Science*)

Virus load variations

Frequency

200
150
100
50
0

set-point viral load

10^2 10^3 10^4 10^5 10^6

in 4µl of blood

Fraser *et al.*
(2014, *Science*)

Adjusted (for age and sex) $r = 0.206$, $P = 0.030$.

○ MTF (n = 72)
● FTM (n = 43)

$r = 0.196$, $P = 0.036$

y-axis: $\log_{10}$ HIV-1 RNA in recipients

x-axis: $\log_{10}$ HIV-1 RNA in donors

Tang *et al.* (2004, *AIDS Res. Hum. Retroviruses*)

Adjusted (for age and sex) $r = 0.206$, $P = 0.030$.

○ MTF (n = 72)
● FTM (n = 43)

$r = 0.196$, $P = 0.036$

log$_{10}$ HIV-1 RNA in recipients

log$_{10}$ HIV-1 RNA in donors

Tang *et al.* (2004, *AIDS Res. Hum. Retroviruses*)

Transcription couple data problem

- Very little data on transmission chains!

- Infections outside the couple

- Cannot control for within-host evolution

Hollingsworth *et al.* (2010, *PLoS Path*)

Transmission chain & phylogeny

patient 0 (p_0) — sample A

p_4 — sample B
p_6 — unsampled

p_3 — sample C
p_5 — unsampled

p_1 — sample D
p_2 — unsampled

2012   2013   2014

Emma
Saulnier's
PhD

Transcription chain & phylogeny

patient 0 (p₀) — sample A

p₄   sample B / unsampled

p₆

p₃   sample C / unsampled

p₅

p₁   sample D / unsampled

p₂

2012 2013 2014
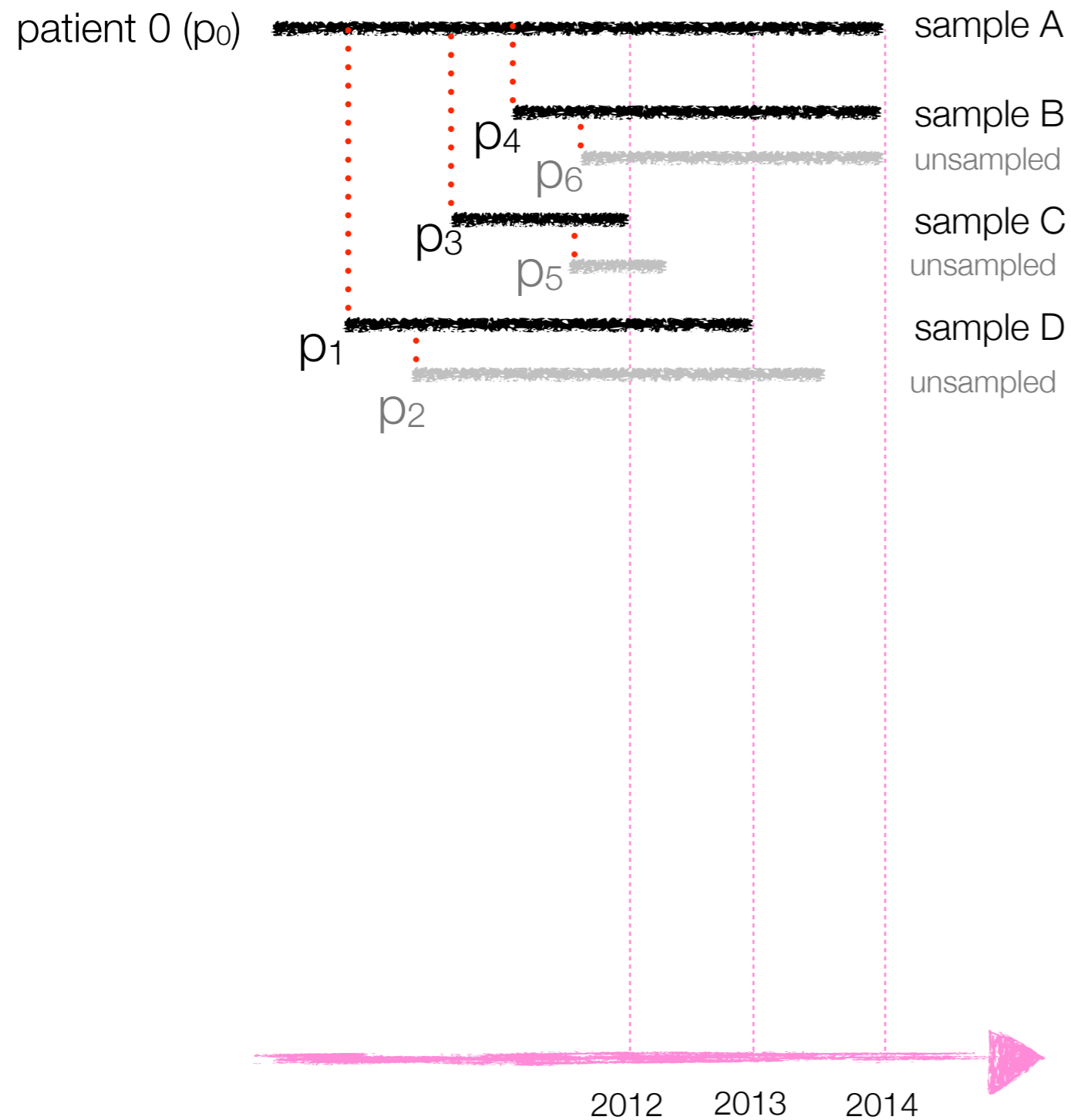
Emma
Saulnier's
PhD

Transmission chain & phylogeny

patient 0 (p_0) — sample A

p_4

p_6 — sample B, unsampled

p_3 — sample C

p_5 — unsampled

p_1 — sample D

p_2 — unsampled

pathogen A (p_0 or p_1 or p_3 or p_4?)

pathogen B (p_0 or p_1 or p_3 or p_4?)

pathogen C (p_0 or p_1 or p_3 or p_4?)

pathogen D (p_0 or p_1?)

2012   2013   2014
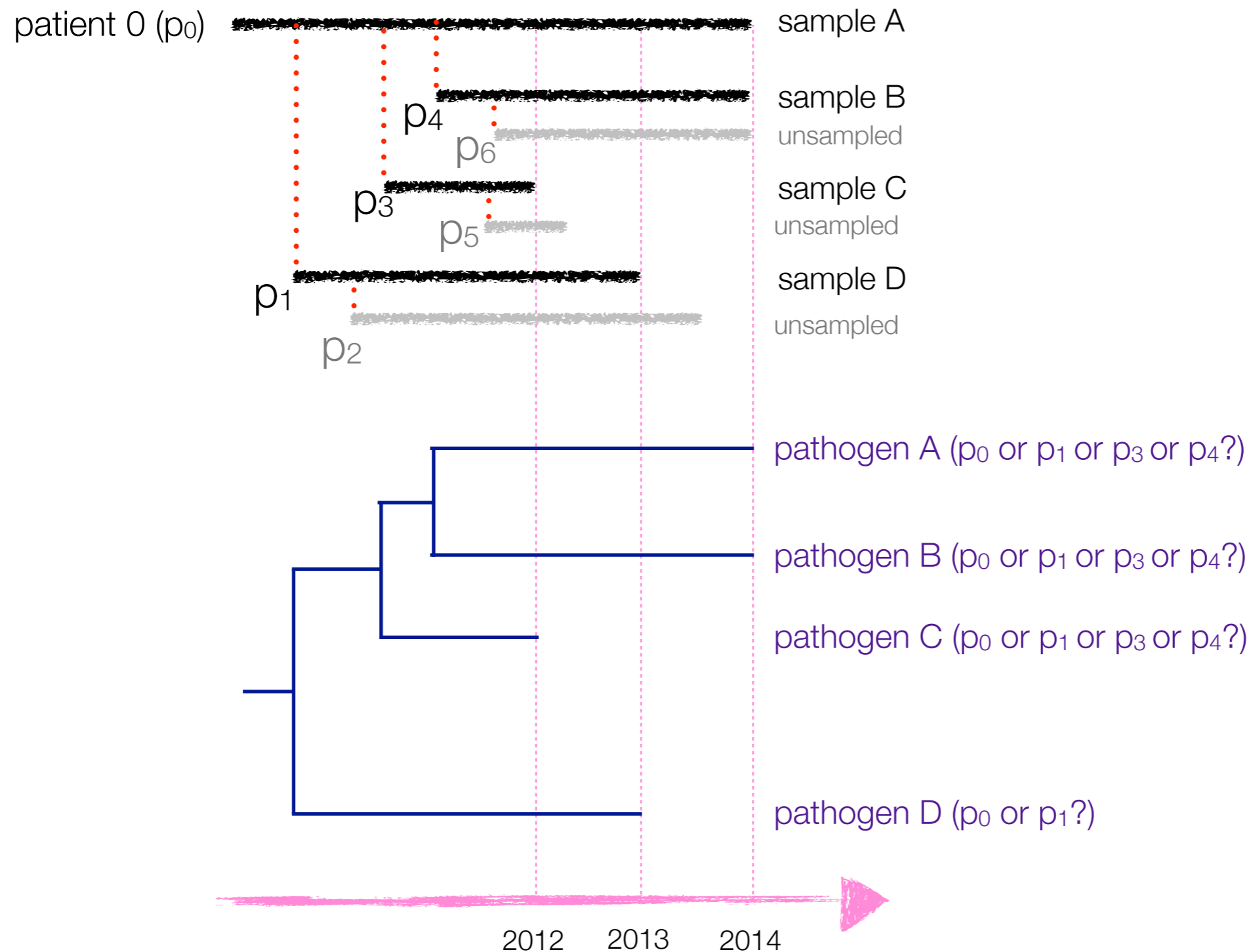
Emma Saulnier's PhD

Phylogeny of infections

Alizon *et al.* (2010, *PLoS Path*)
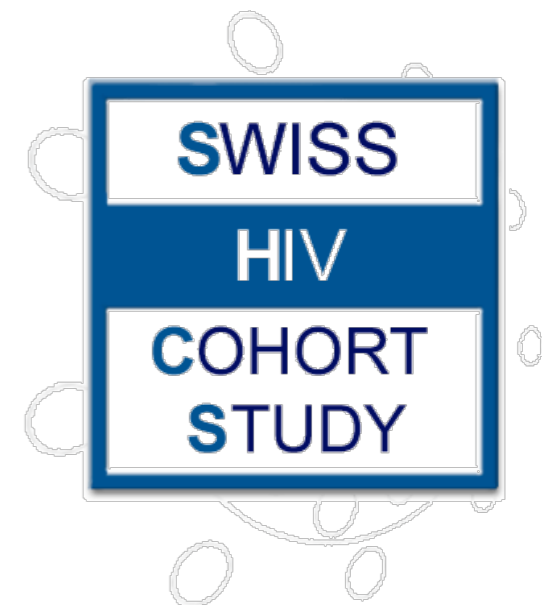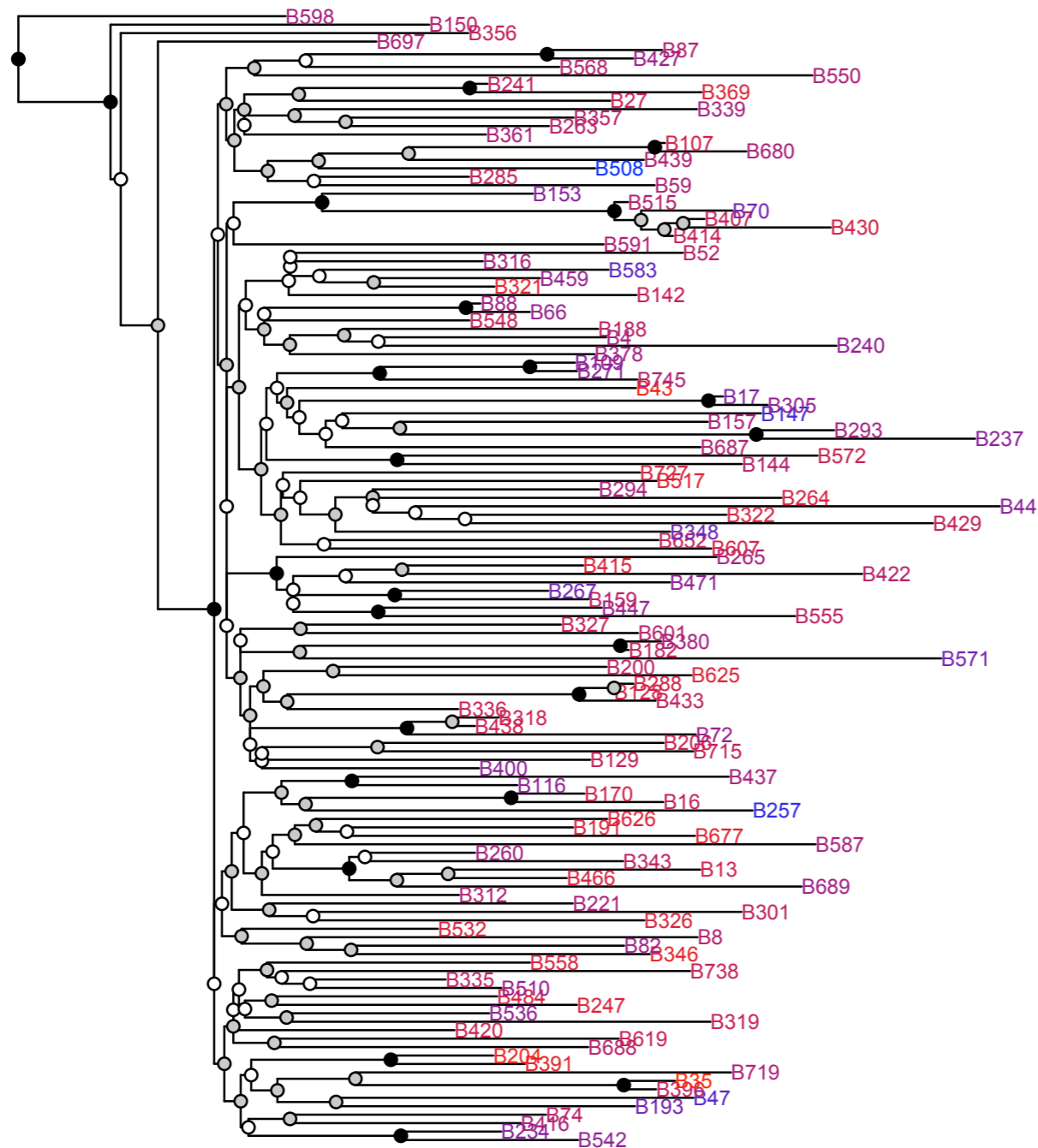
Phylogeny of infections
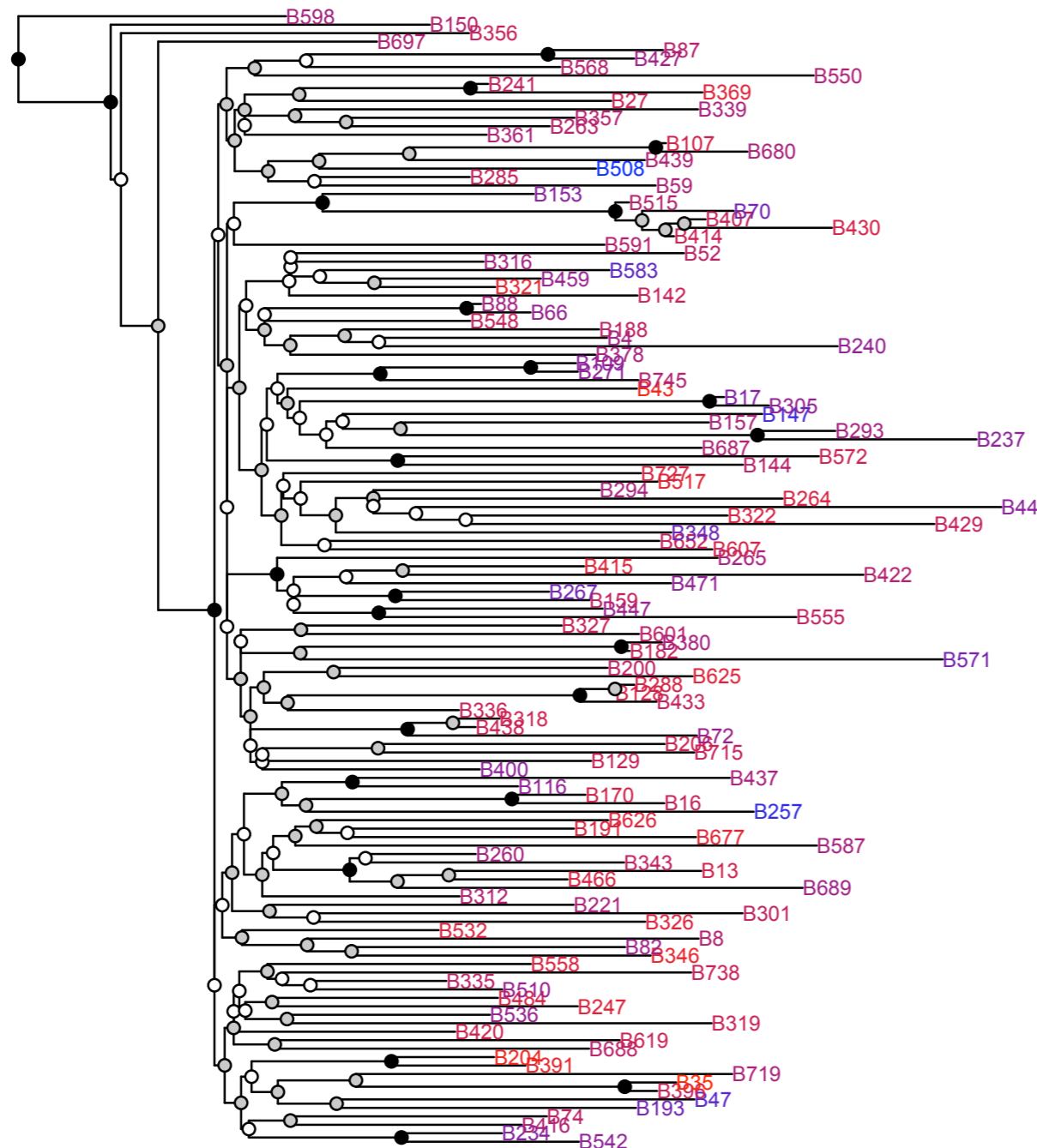
Proximity in the phylogeny reflects proximity in the transmission chain...

Alizon *et al.* (2010, *PLoS Path*)

Phylogeny of infections

Alizon *et al.* (2010, *PLoS Path*)

ln(SPVL)
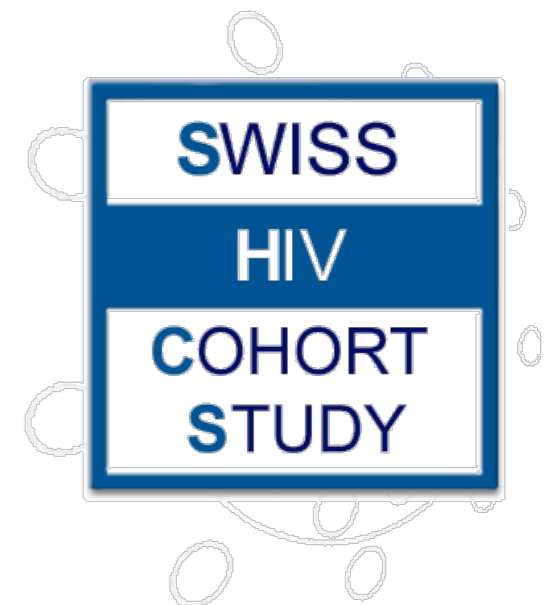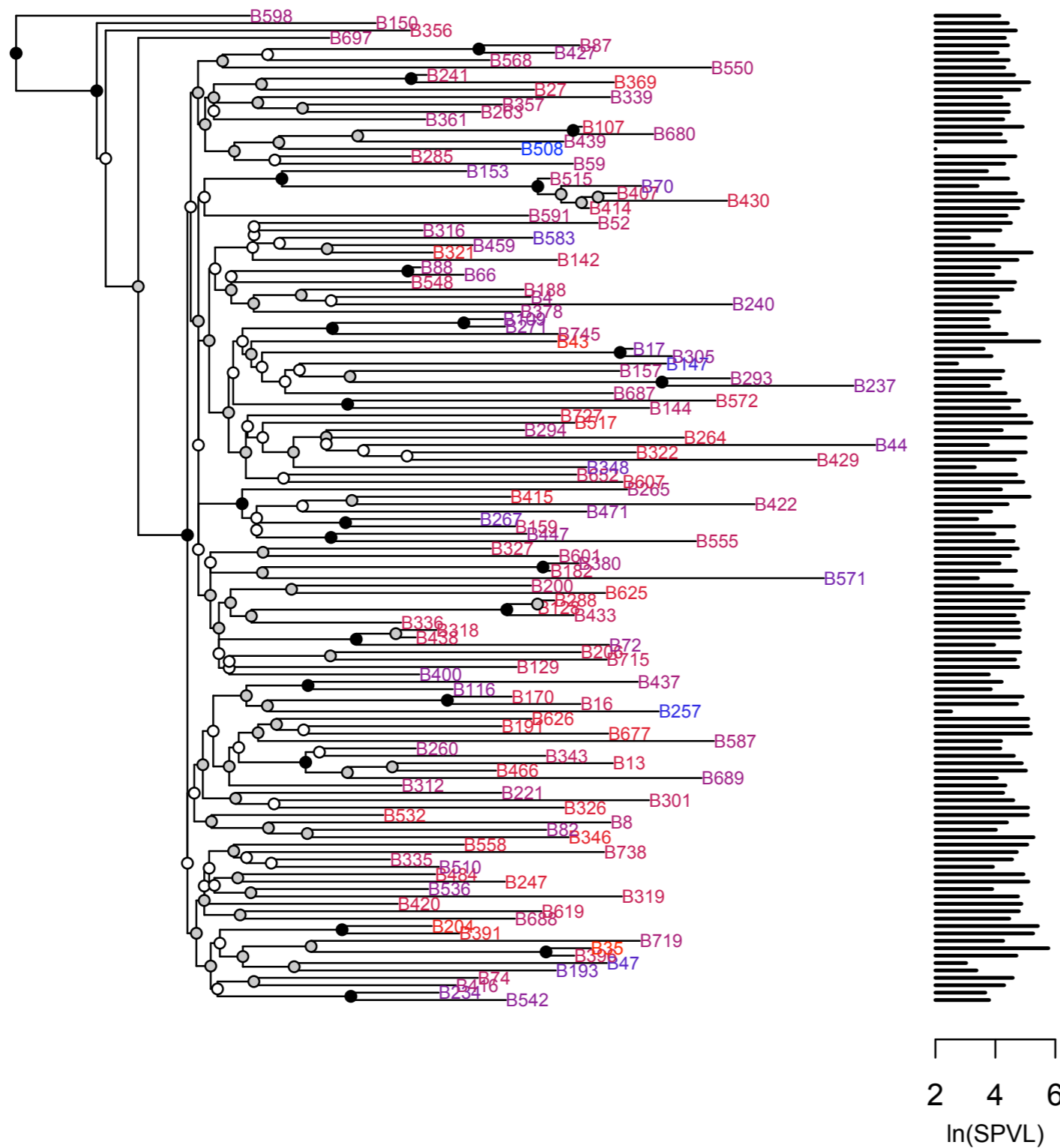
Phylogeny of infections
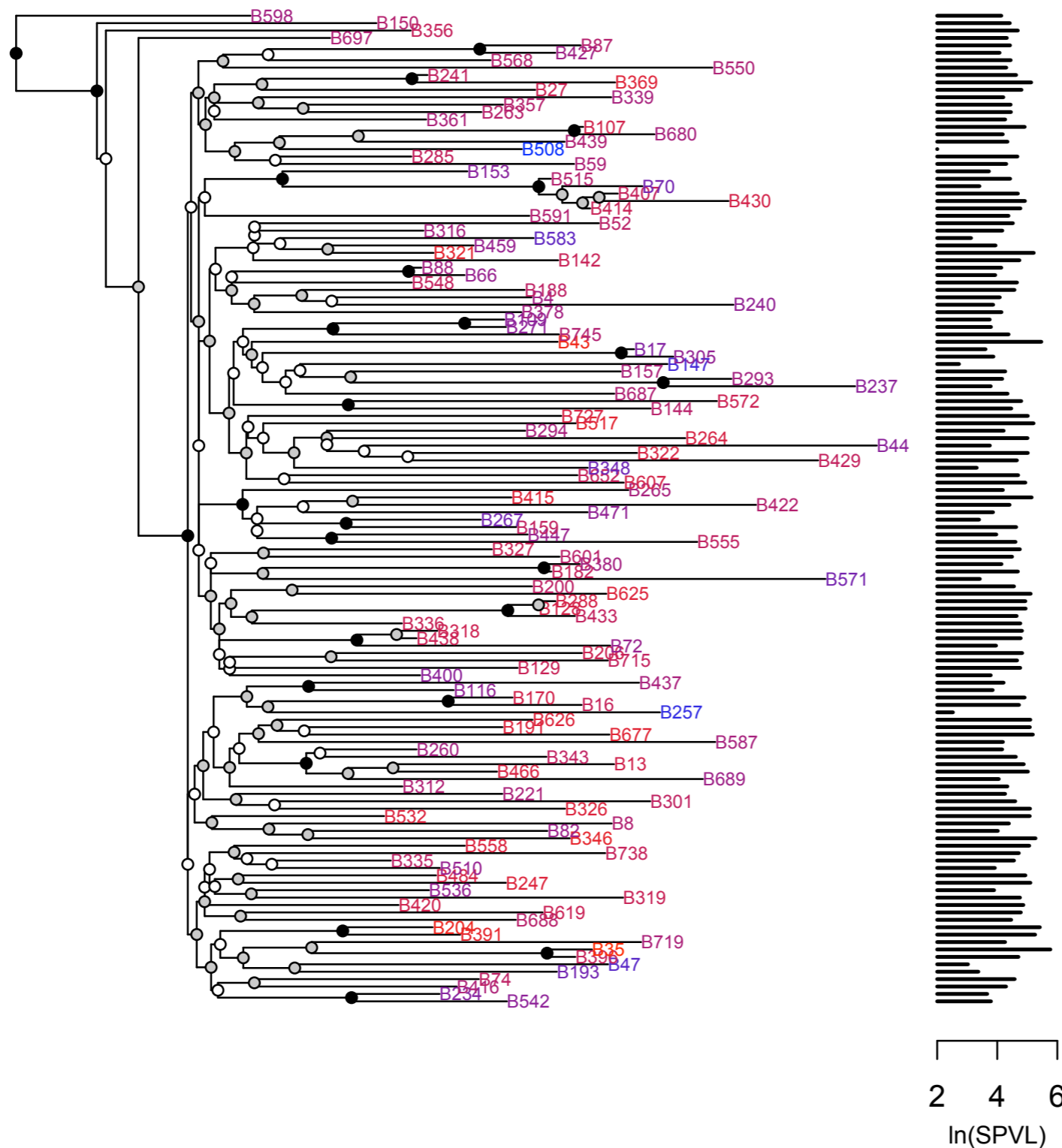
Do patients close in the phylogeny have similar trait values?

Alizon *et al.* (2010, *PLoS Path*)
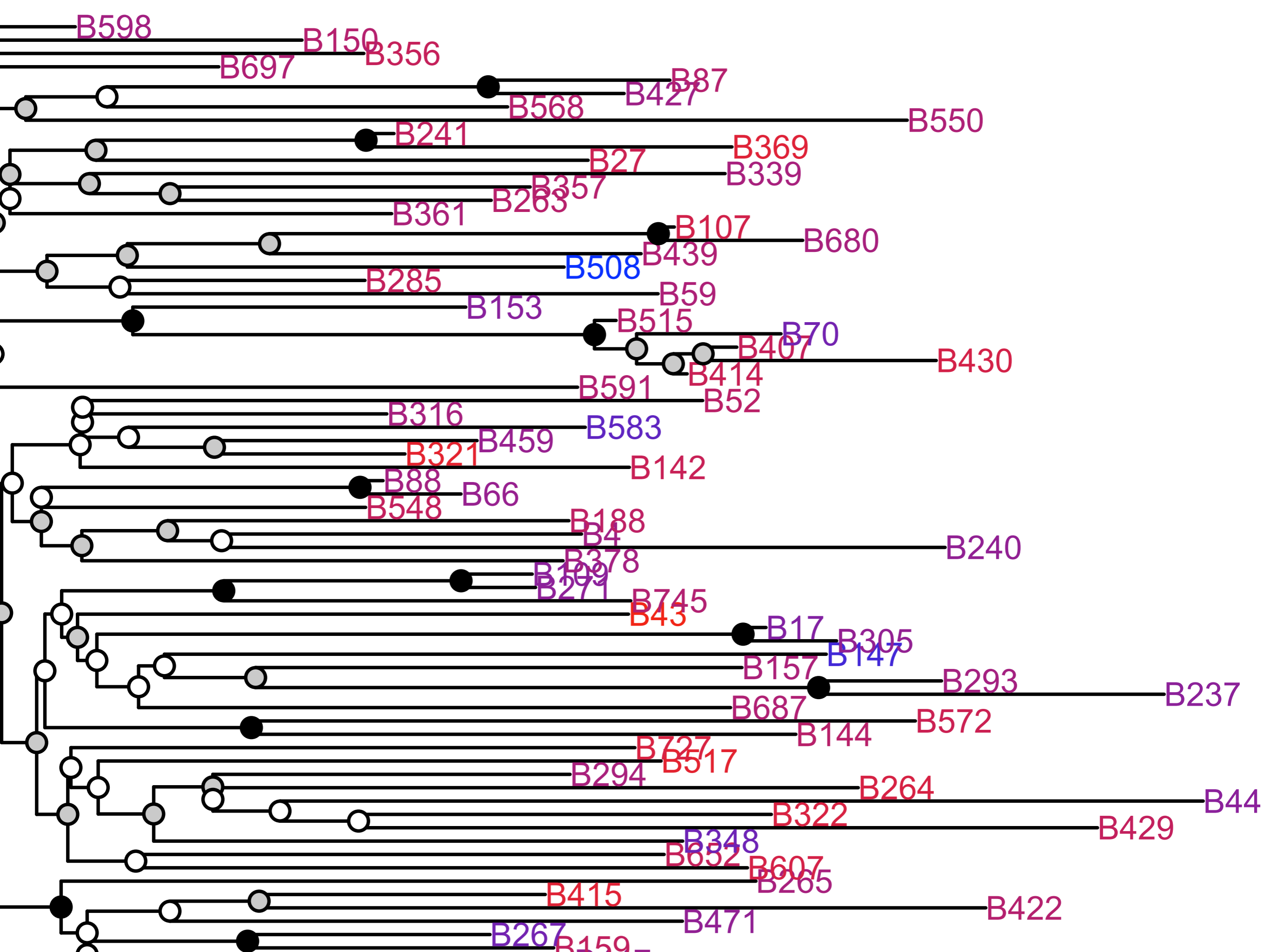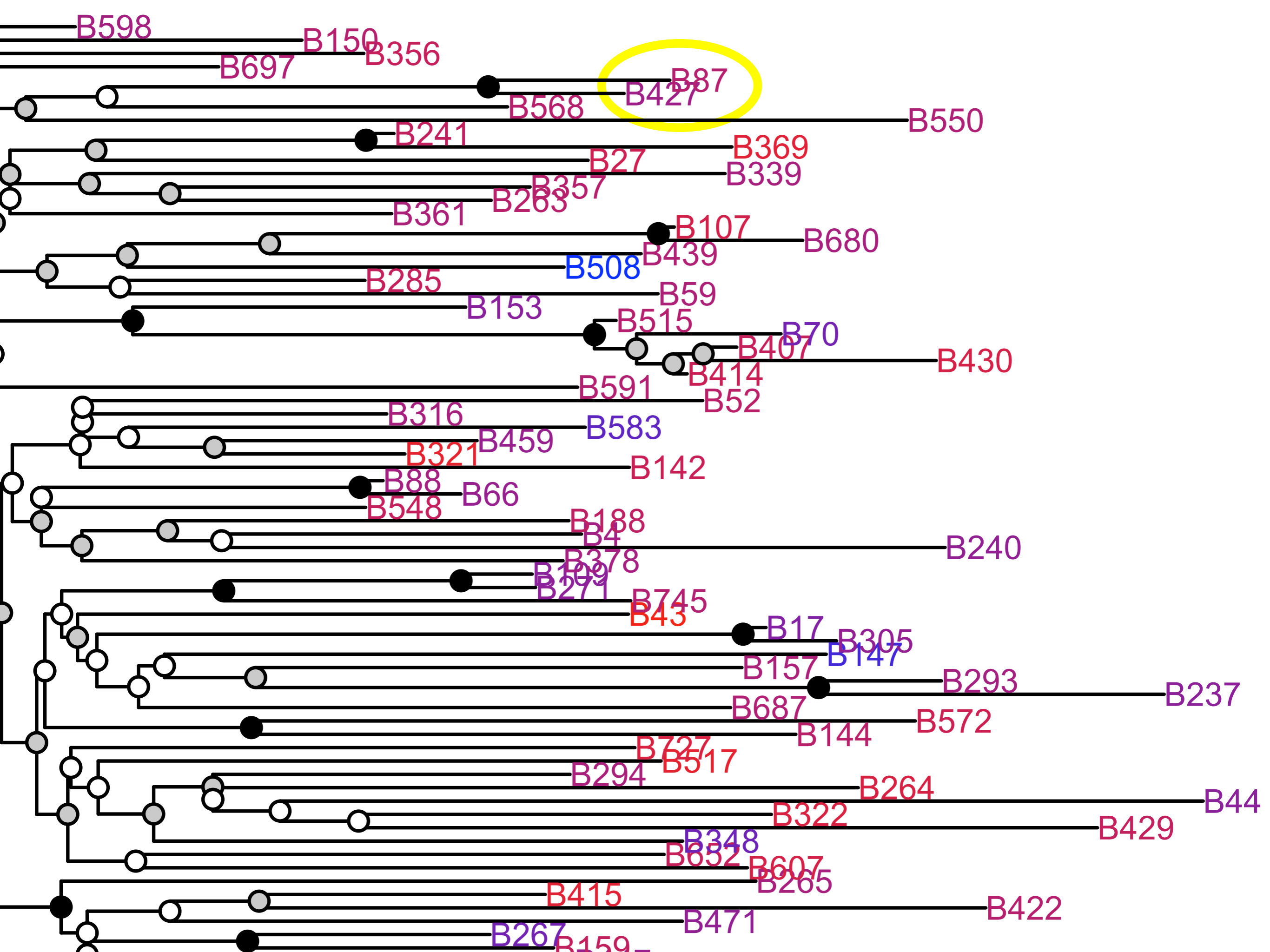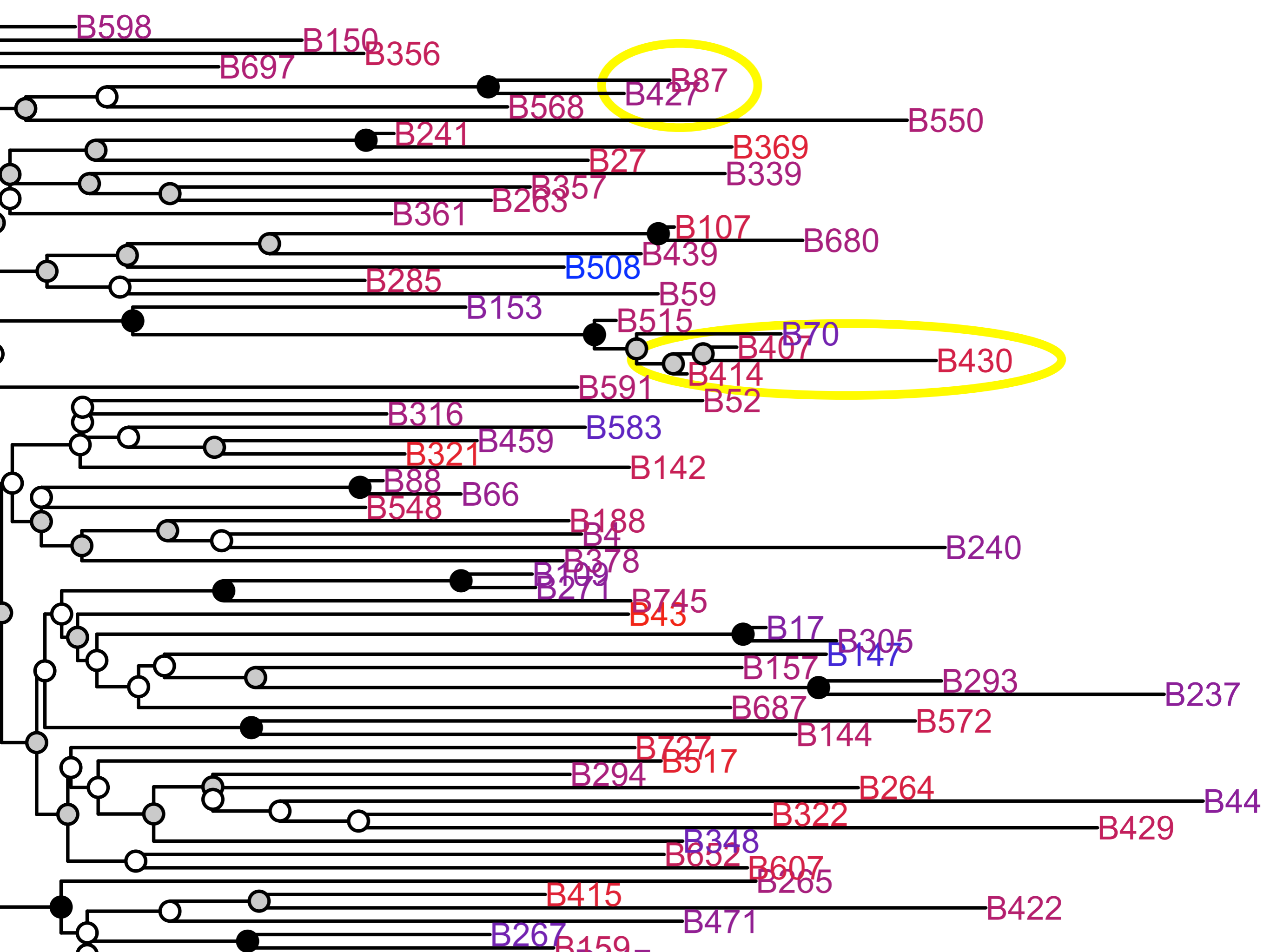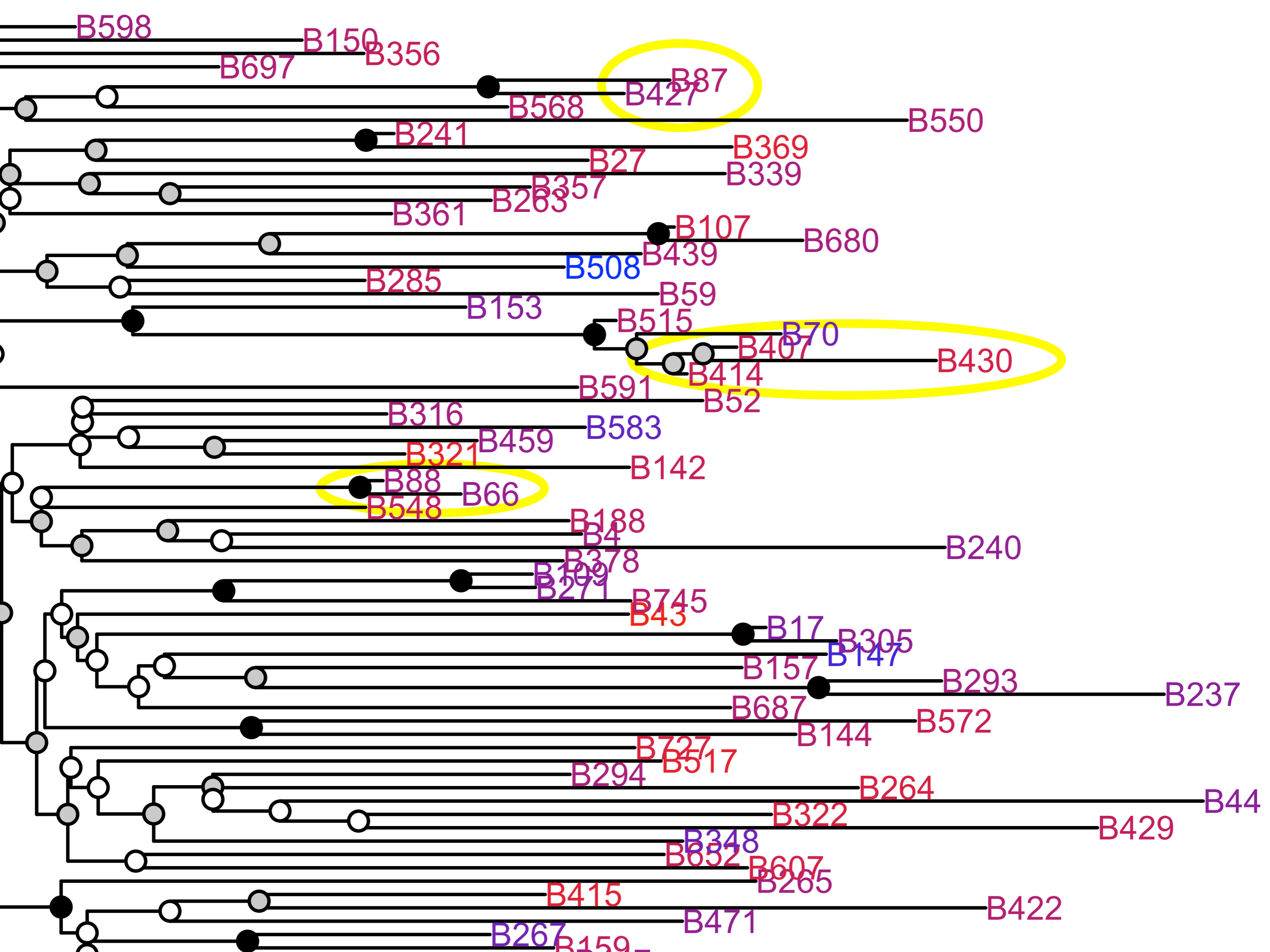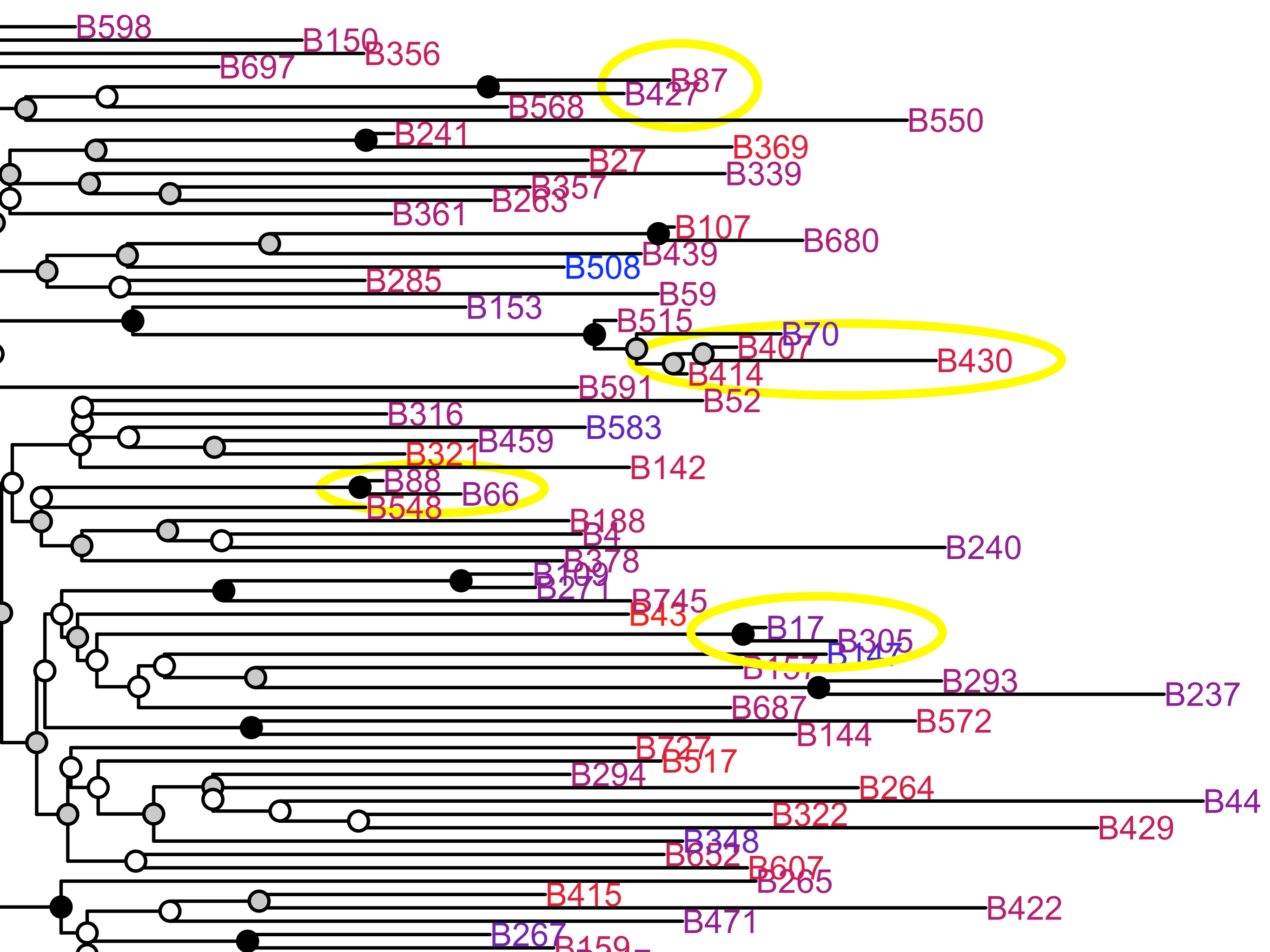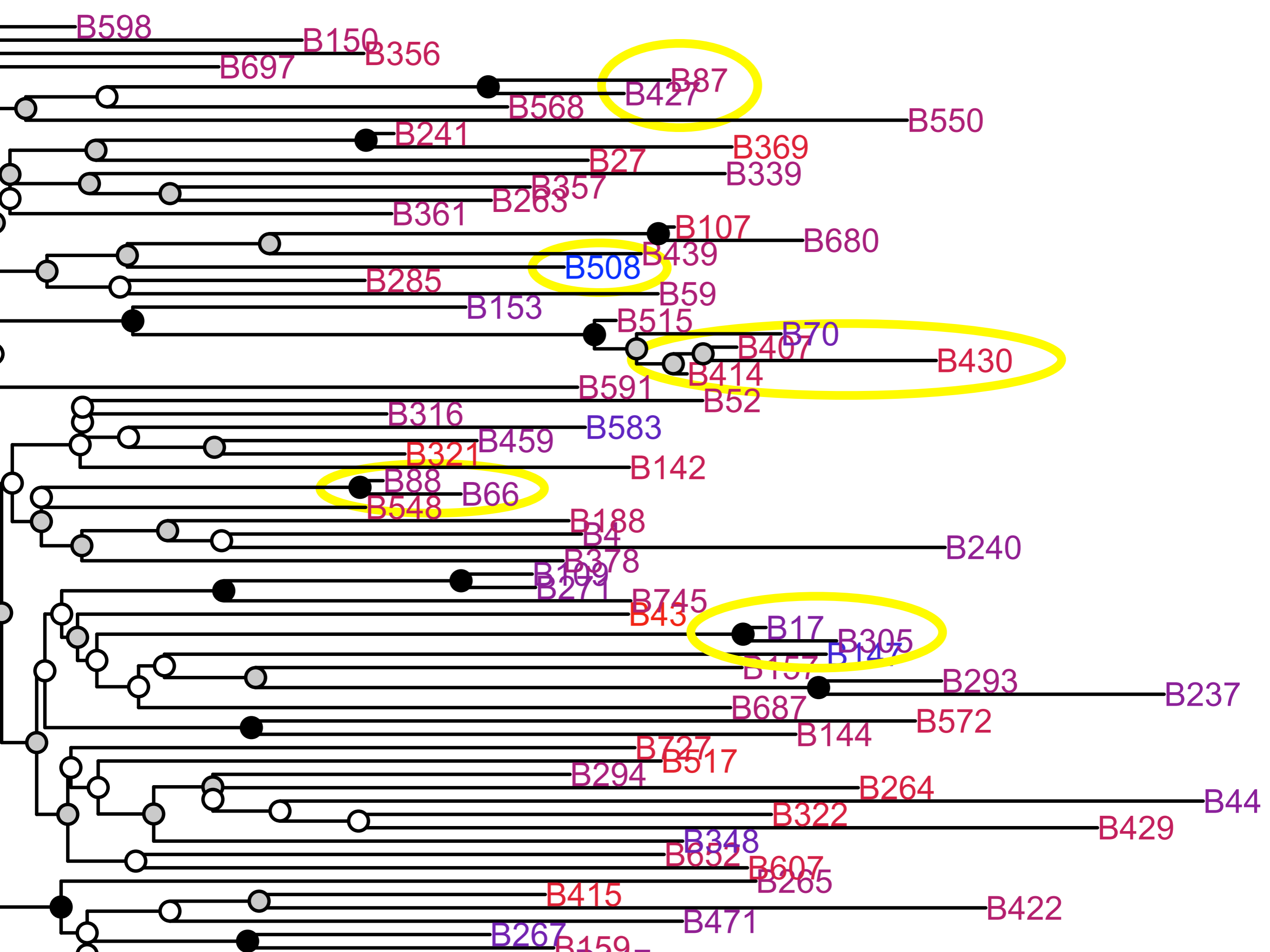
Phylogenetic comparative method
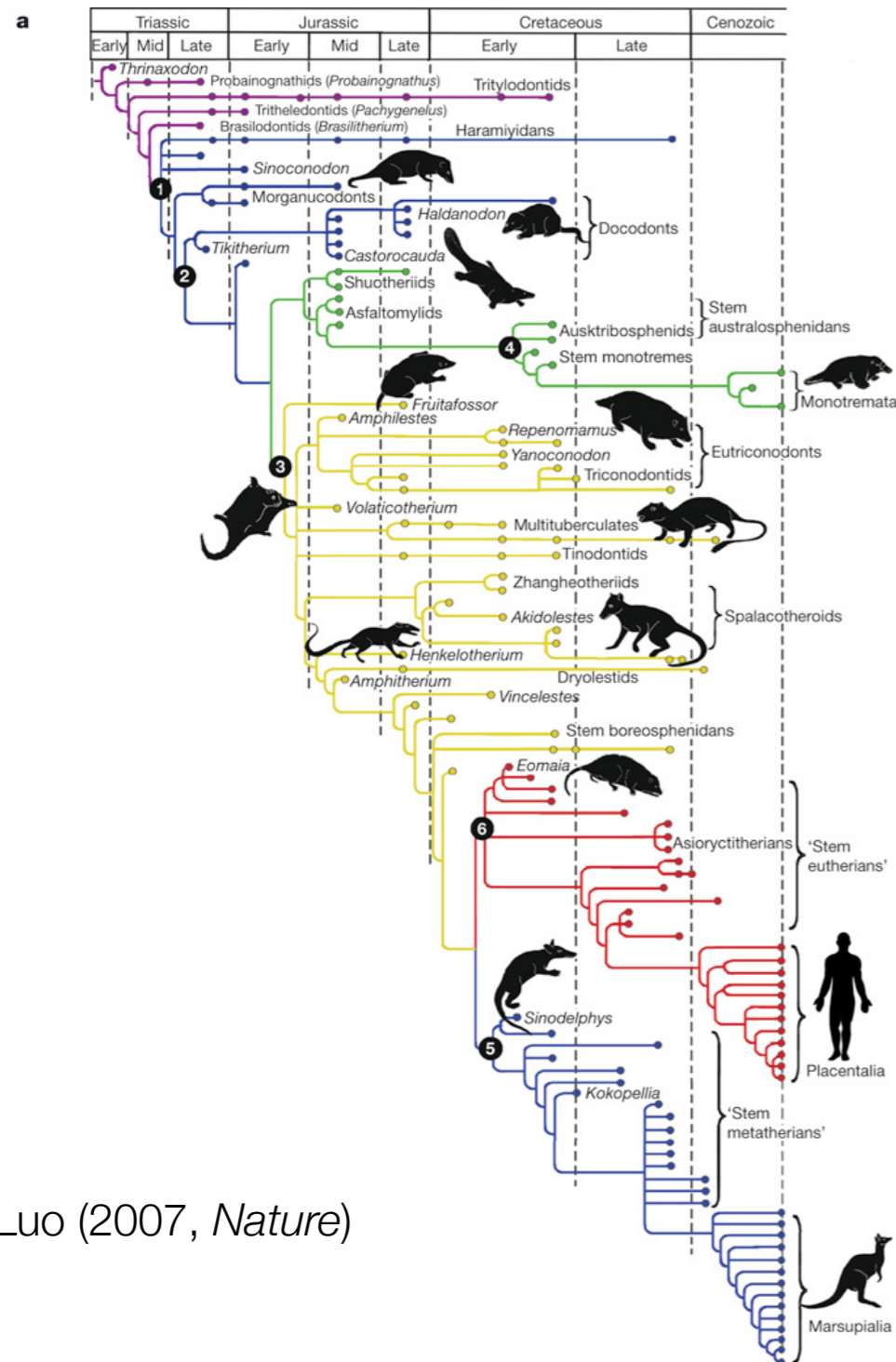
Luo (2007, *Nature*)

Felsenstein (1985, *Am. Nat.*)
Felsenstein (2002)

Phylogenetic comparative method
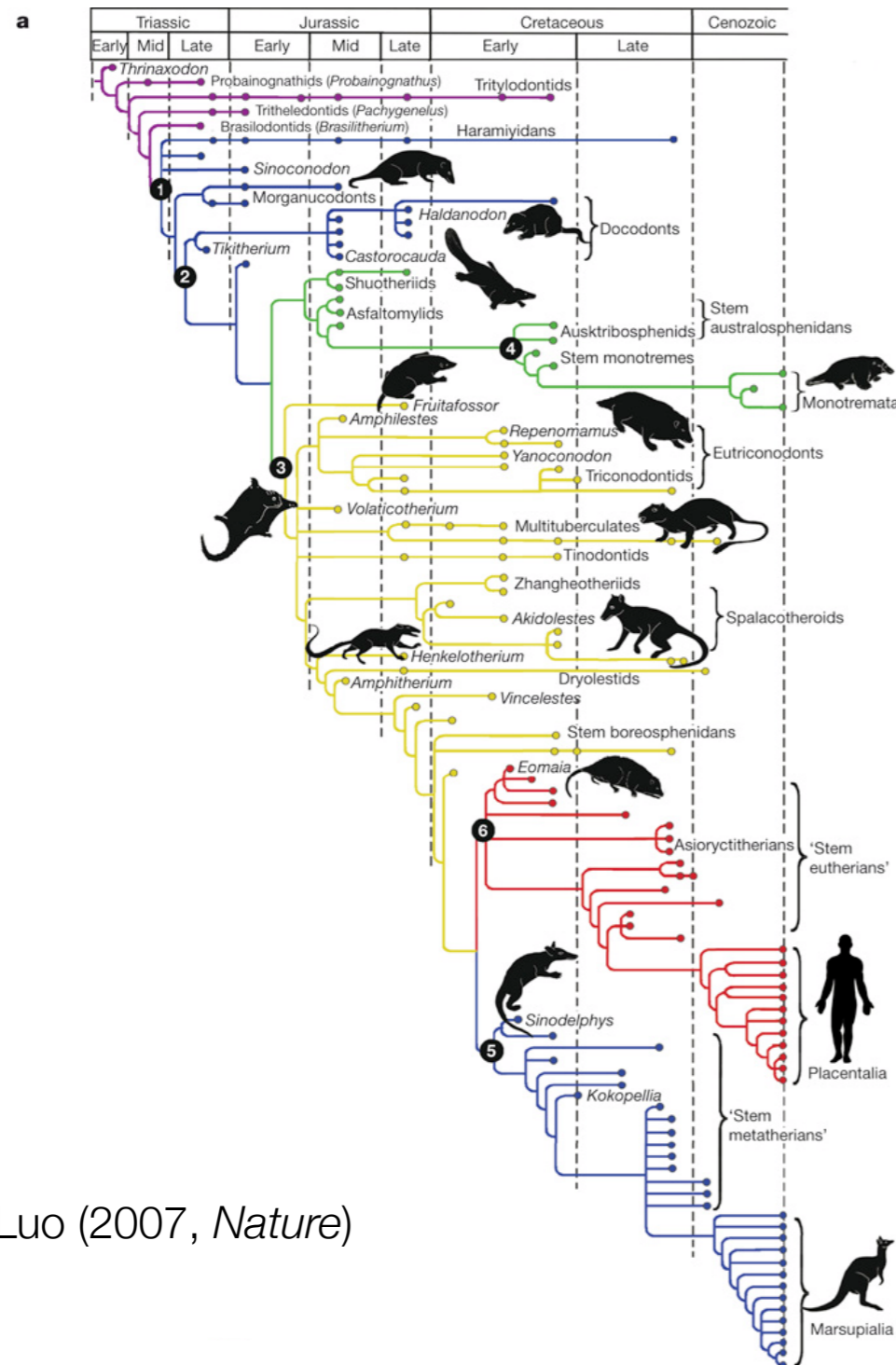
Luo (2007, *Nature*)

Phylogenetic signal estimates how the phylogeny explains variations in a quantitative trait among species

Felsenstein (1985, *Am. Nat.*)
Felsenstein (2002)

Phylogenetic comparative method

d_2

$t2$

d_1

$t1$

d_3

$t3$

d_4

$t4$

Felsenstein (1985, *Am. Nat.*)

Phylogenetic comparative method

Felsenstein (1985, *Am. Nat.*)
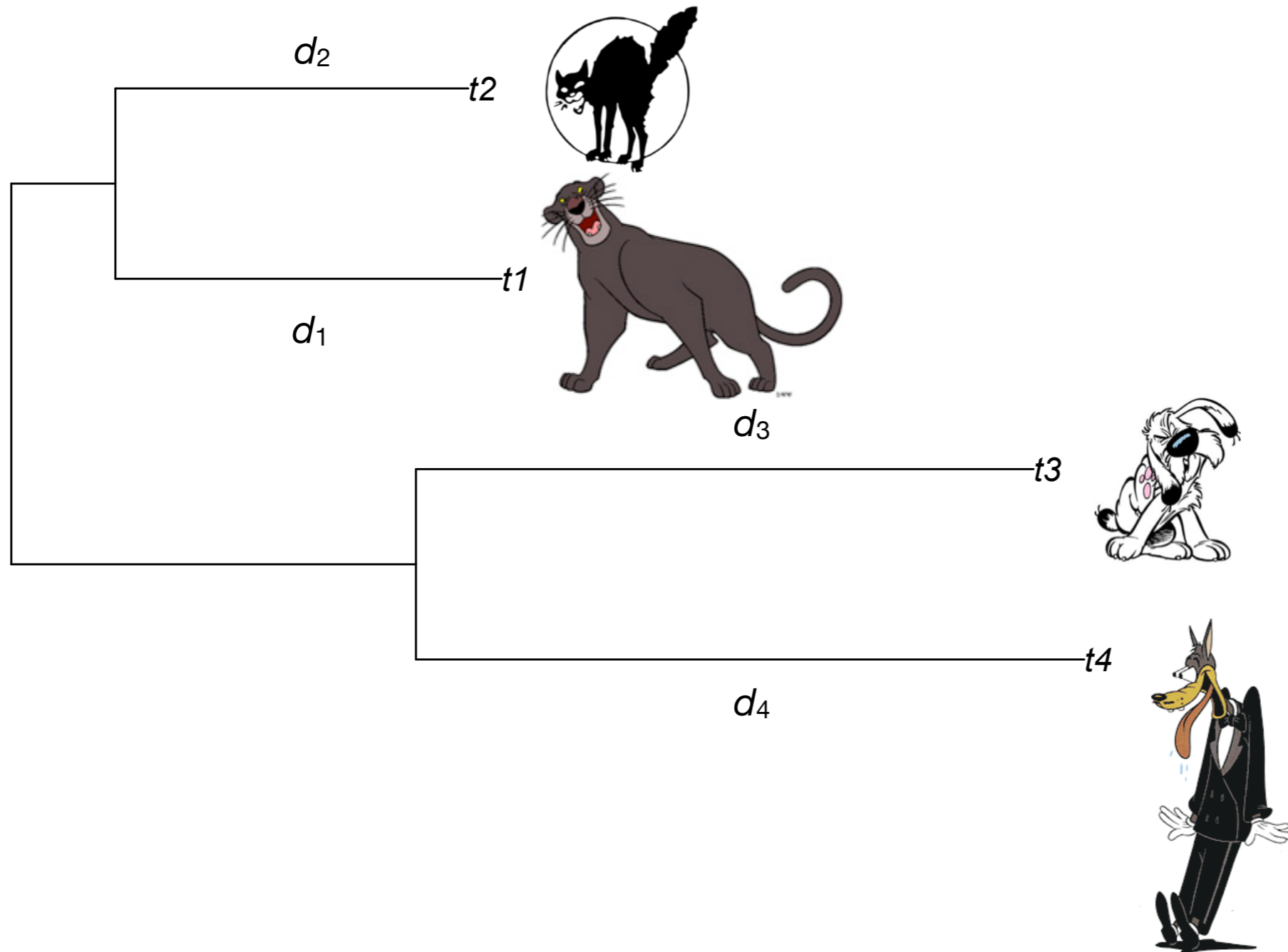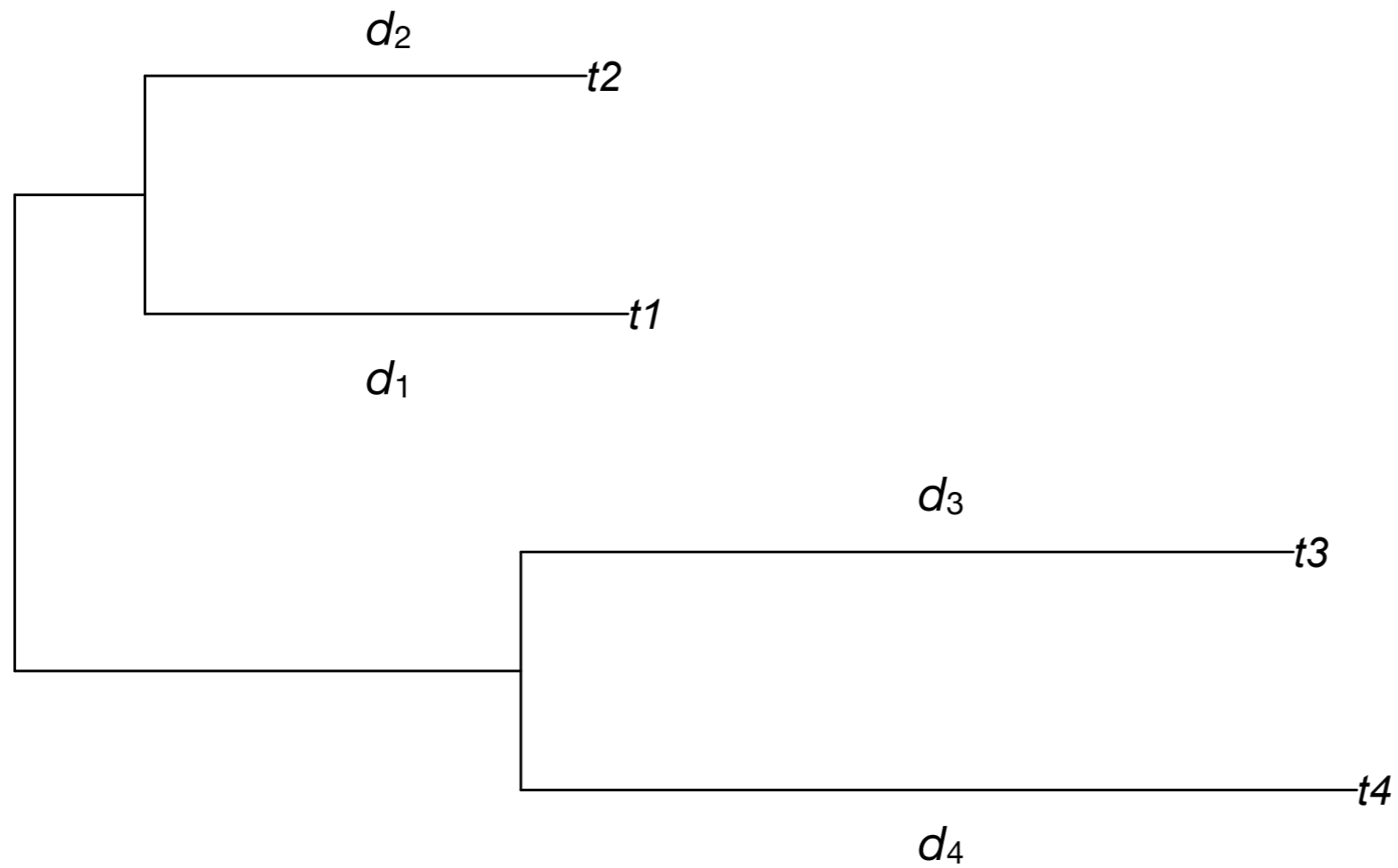
Phylogenetic comparative method
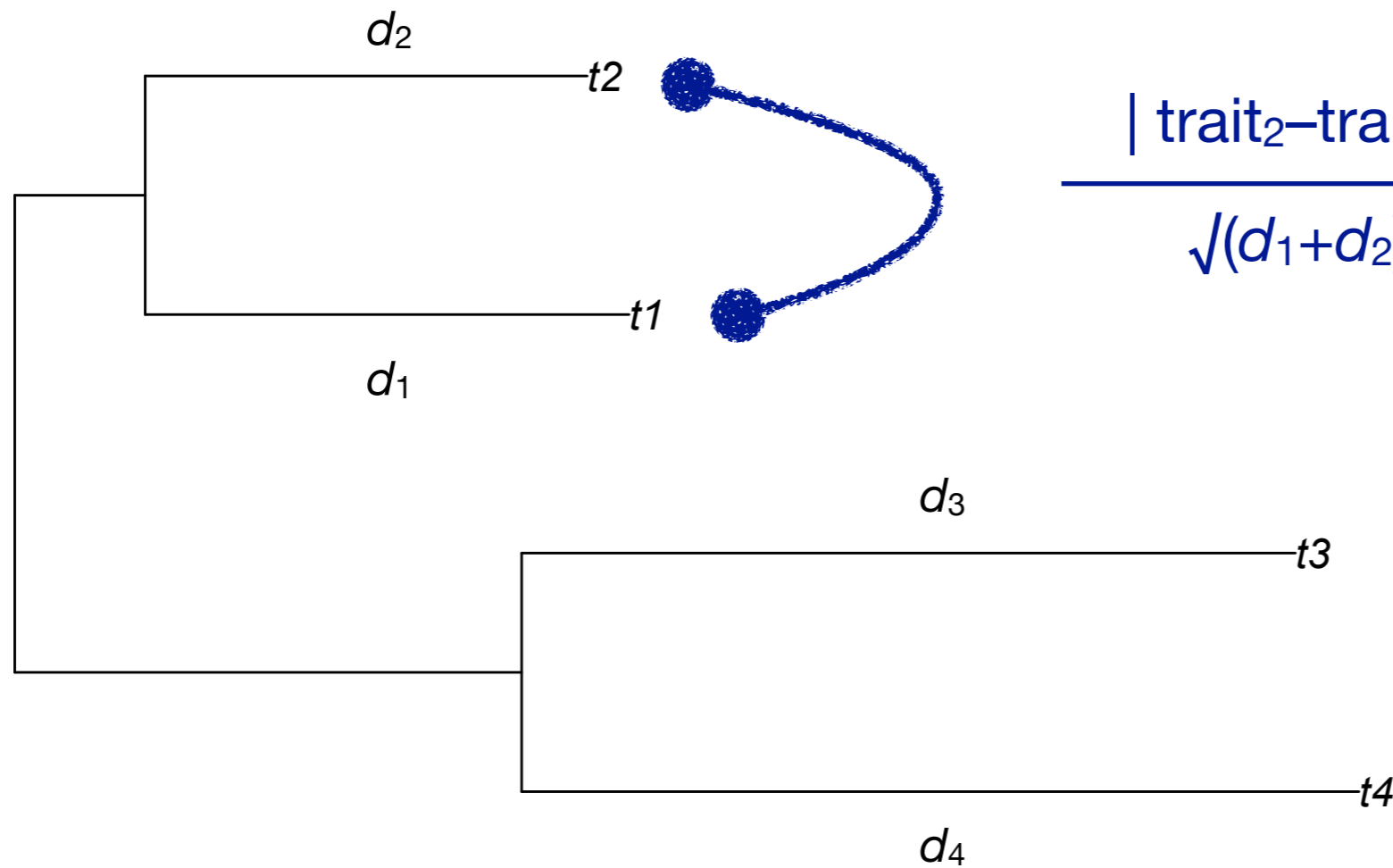
$$\frac{\mid \text{trait}_2 - \text{trait}_1 \mid}{\sqrt{(d_1 + d_2)}} = \text{contrast A}$$

Felsenstein (1985, *Am. Nat.*)

Phylogenetic comparative method

Felsenstein (1985, *Am. Nat.*)

Phylogenetic comparative method

contraste A

contraste C

contraste B

t2

t1

t3

t4

Felsenstein (1985, *Am. Nat.*)

Phylogenetic comparative method

d = 1
high signal

d = 0.7

low signal d = 0.2

Freckleton *et al.* (2002, *Am Nat*)
Blomberg *et al.* (2003, *Evolution*)
Shirreff *et al.* (2013, *EMPH*)

Phylogenetic comparative method

- Low variance in contrasts indicates that infections close in the phylogeny have similar traits

d = 1
high signal

d = 0.7

low signal d = 0.2

Freckleton *et al.* (2002, *Am Nat*)
Blomberg *et al.* (2003, *Evolution*)
Shirreff *et al.* (2013, *EMPH*)

Phylogenetic comparative method

- Low variance in contrasts indicates that infections close in the phylogeny have similar traits

- This measure is known as **phylogenetic signal**

d = 1
high signal

d = 0.7

low signal d = 0.2

Freckleton *et al.* (2002, *Am Nat*)
Blomberg *et al.* (2003, *Evolution*)
Shirreff *et al.* (2013, *EMPH*)

Phylogenetic comparative method

- Low variance in contrasts indicates that infections close in the phylogeny have similar traits

- This measure is known as **phylogenetic signal**

- Phylogeny explains trait distribution **assuming brownian motion** evolution

d = 1
high signal

d = 0.7

low signal d = 0.2

Freckleton *et al.* (2002, *Am Nat*)
Blomberg *et al.* (2003, *Evolution*)
Shirreff *et al.* (2013, *EMPH*)

Phylogenetic comparative method

- Low variance in contrasts indicates that infections close in the phylogeny have similar traits

- This measure is known as **phylogenetic signal**

- Phylogeny explains trait distribution **assuming brownian motion** evolution

- **Genetic heritability** can be inferred from phylogenetic signal using ABC methods

d = 1
high signal

d = 0.7

low signal d = 0.2

Freckleton *et al.* (2002, *Am Nat*)
Blomberg *et al.* (2003, *Evolution*)
Shirreff *et al.* (2013, *EMPH*)

Phylogenetic comparative method

- Low variance in contrasts indicates that infections close in the phylogeny have similar traits

- This measure is known as **phylogenetic signal**

- Phylogeny explains trait distribution **assuming brownian motion** evolution

- **Genetic heritability** can be inferred from phylogenetic signal using ABC methods

high signal
d = 1

d = 0.7

low signal   d = 0.2

Freckleton *et al.* (2002, *Am Nat*)
Blomberg *et al.* (2003, *Evolution*)
Shirreff *et al.* (2013, *EMPH*)

Set-point virus load is partly 'heritable'

Alizon *et al.* (2010, *PLoS Path*)
Shirreff *et al.* (2013, *EMPH*)

Set-point virus load is partly 'heritable'

Up to 56% of the variance in set-point virus load can be explained by virus factors

Alizon *et al.* (2010, *PLoS Path*)
Shirreff *et al.* (2013, *EMPH*)

Set-point virus load is partly 'heritable'

N couples	Heritability	Study (reference)	Country, subtype(s)	Adjustments
97	36% (6 to 66%)	Hollingsworth et al. (17)	Uganda, mostly A, D, and recombinants	Age, sex, subtype, symptomatic genital ulcer disease (GUD)
141	44% (19 to 69%)	Lingappa et al. (18)	Partners in Prevention (14 sites in East and Southern Africa), diverse subtypes	Age, sex, subtype, sexually transmitted infection, GUD, circumcision, hormonal contraceptive use, source partner characteristics
195	26% (8 to 44%)	Yue et al. (19)	Zambia, mostly C	Age, sex, HLA, HLA sharing between partners
433	33% (20 to 46%)	Overall summary estimate (weighted by standard error)		

Fraser et al. (2014, *Science*)

Set-point virus load is partly 'heritable'

N couples	Heritability	Study (reference)	Country, subtype(s)	Adjustments
97	36% (6 to 66%)	Hollingsworth et al. (17)	Uganda, mostly A, D, and recombinants	Age, sex, subtype, symptomatic genital ulcer disease (GUD)
141	44% (19 to 69%)	Lingappa et al. (18)	Partners in Prevention (14 sites in East and Southern Africa), diverse subtypes	Age, sex, subtype, sexually transmitted infection, GUD, circumcision, hormonal contraceptive use, source partner characteristics
195	26% (8 to 44%)	Yue et al. (19)	Zambia, mostly C	Age, sex, HLA, HLA sharing between partners
433	33% (20 to 46%)	Overall summary estimate (weighted by standard error)		

Fraser et al. (2014, *Science*)

Why shouldn't HIV virulence be 'heritable'?

Why shouldn't HIV virulence be 'heritable'?

patient 1
patient 2
patient 3
patient 4
patient 5
patient 6
patient 7
patient 8
patient 9

Rambaut et al. (2004) **Nature Reviews | Genetics**

Why shouldn't HIV virulence be 'heritable'?

Salmonella typhimurium (Bacteria)

Ebert (1998, *Science*)
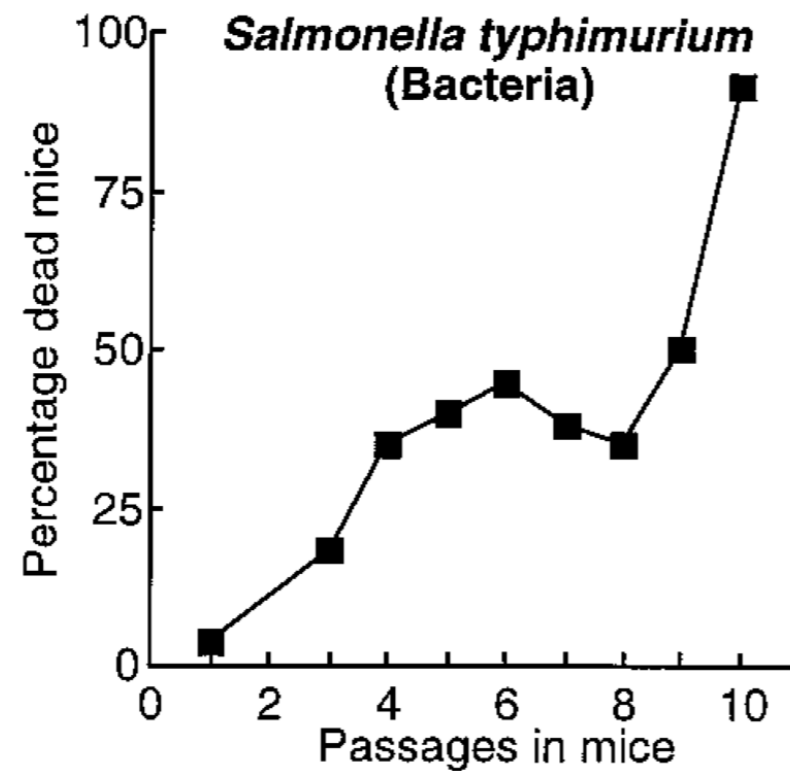
Why shouldn't HIV virulence be 'heritable'?
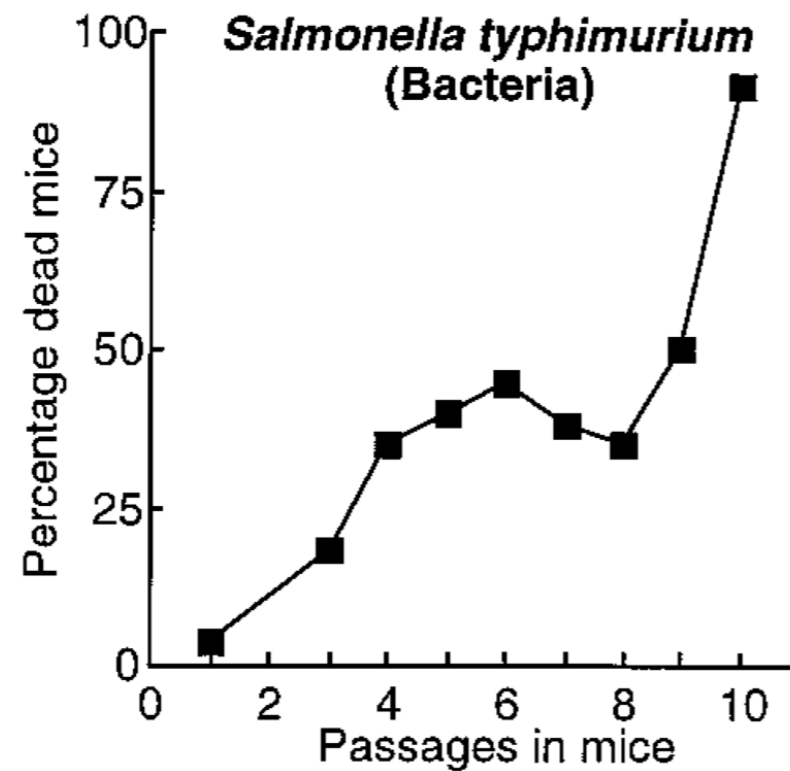
**Within-host competition
often favors
more virulent strains**

Ebert (1998, *Science*)

1) Rugged fitness landscape?

Hinkley et al. (2011, *Nat Genet*)
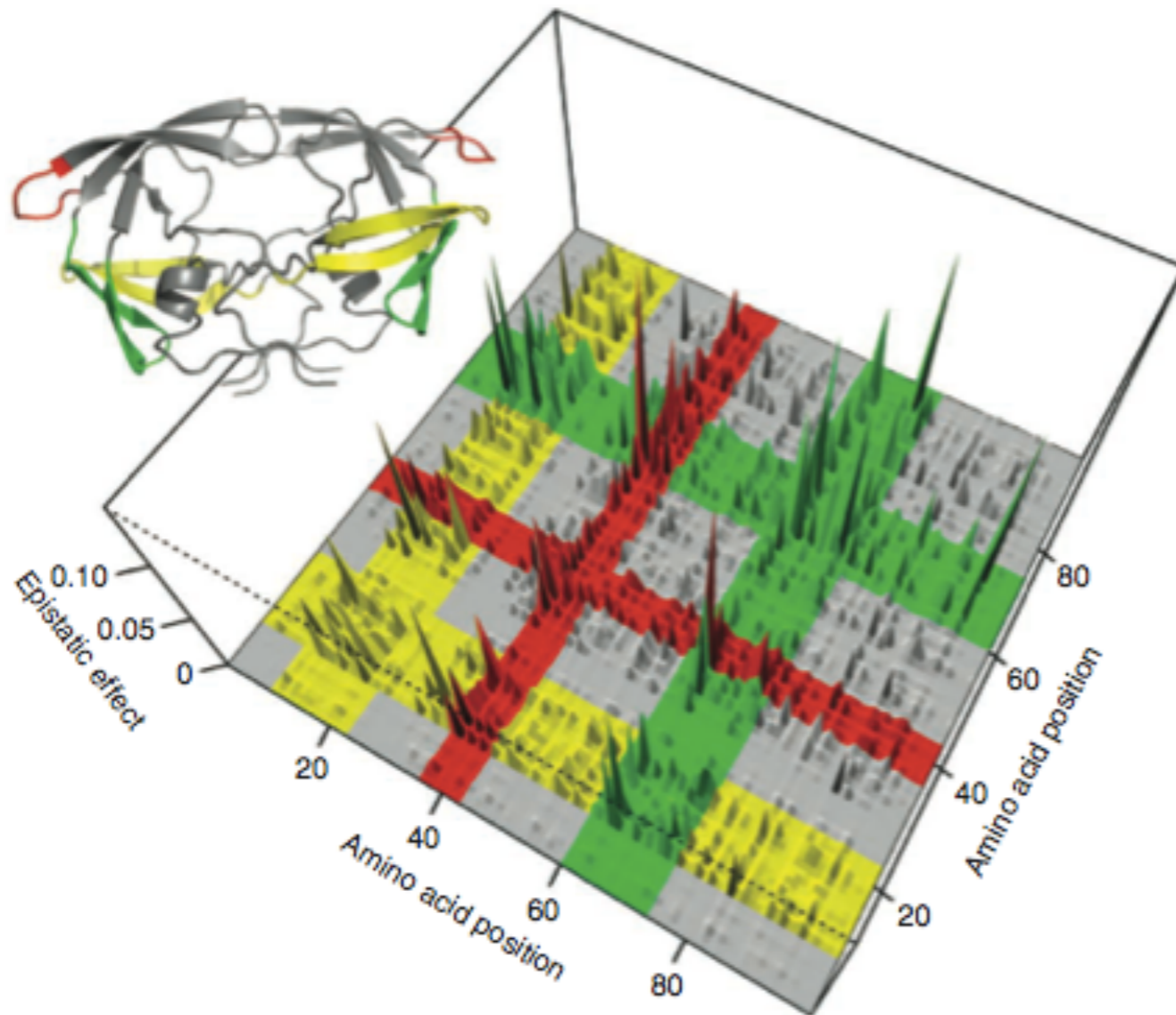
1) Rugged fitness landscape?
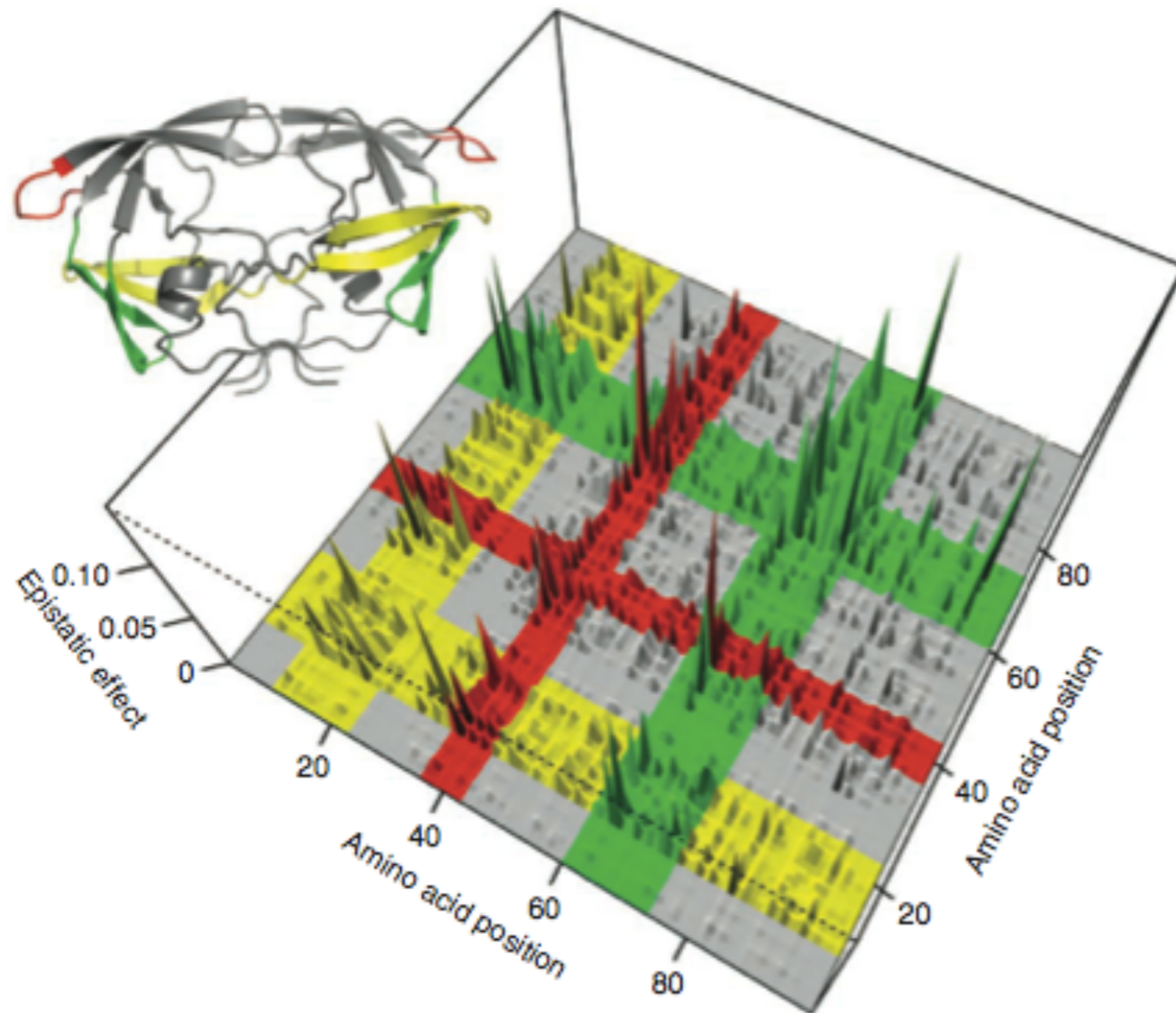
Virulence evolution is slow due to strong (epistatic) constraints
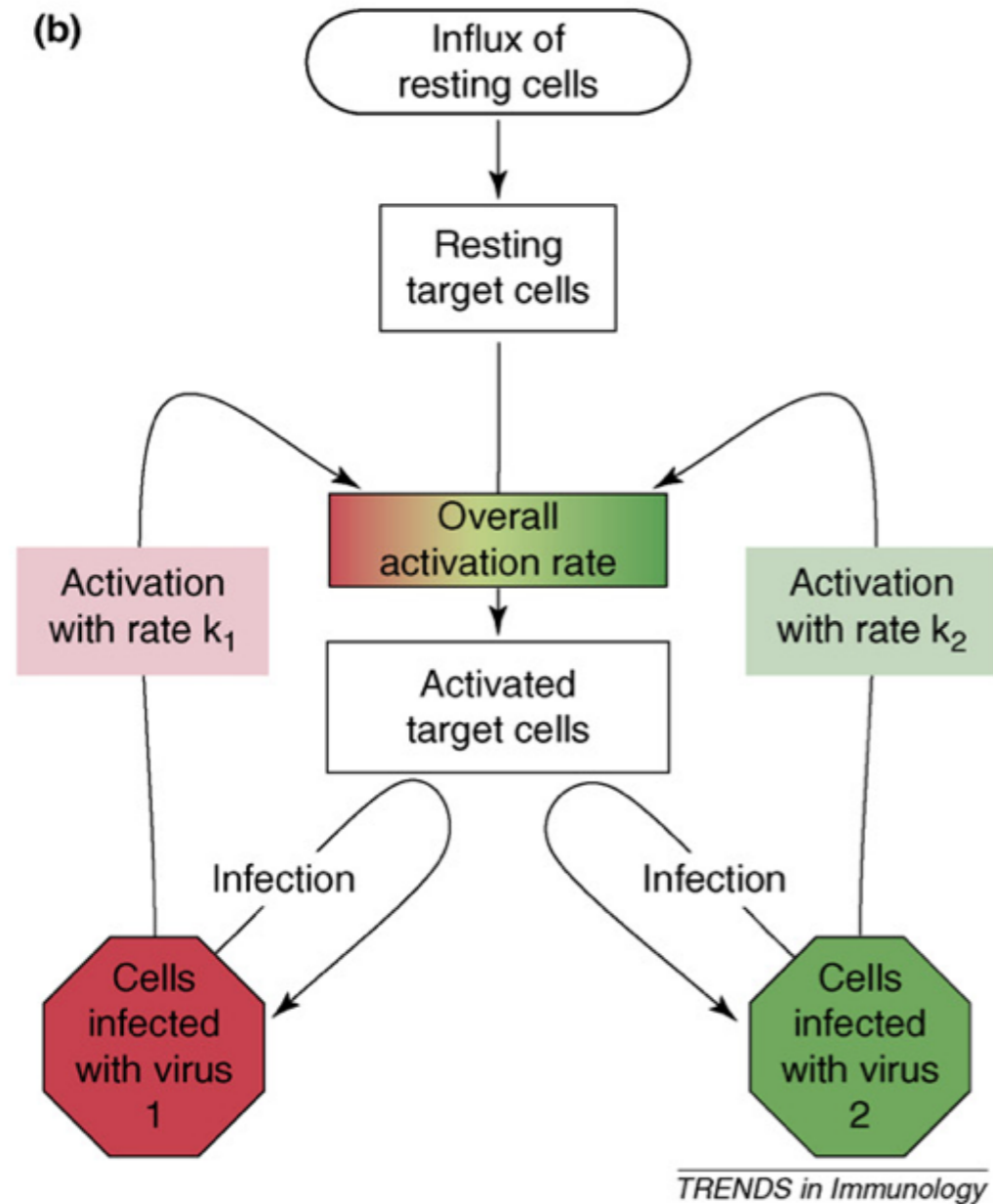
Hinkley et al. (2011, *Nat Genet*)

2) Virulence as a 'public goods'

(b)

Influx of resting cells

Resting target cells

Overall activation rate

Activation with rate k_1

Activation with rate k_2

Activated target cells

Infection

Infection

Cells infected with virus 1

Cells infected with virus 2

TRENDS in Immunology

Bonhoeffer et al. (2003, *Trends Microbiol*)
Bartha et al. (2008, *Trends Immunol*)
Hool et al. (2013, *Epidemics*)

2) Virulence as a 'public goods'

(b)

Influx of resting cells

Resting target cells

Activation with rate k_1

Overall activation rate

Activation with rate k_2

Activated target cells

Infection

Infection

Cells infected with virus 1

Cells infected with virus 2

TRENDS in Immunology
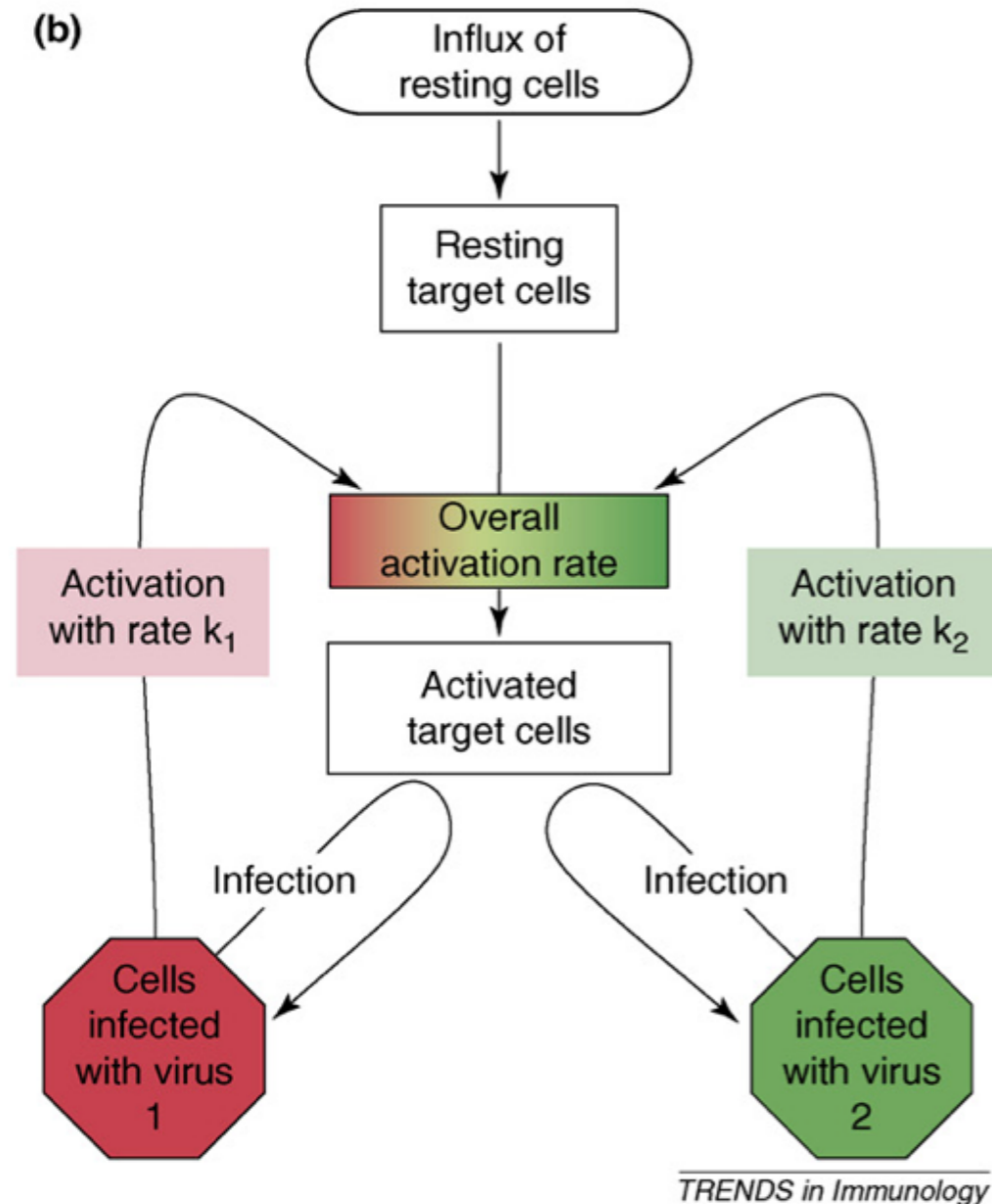
More virulent strains are less competitive

(they pay a cost to activate target cells that benefit to all strains)

Bonhoeffer et al. (2003, *Trends Microbiol*)
Bartha et al. (2008, *Trends Immunol*)
Hool et al. (2013, *Epidemics*)

3) Within-host evolution is a 'dead-end'
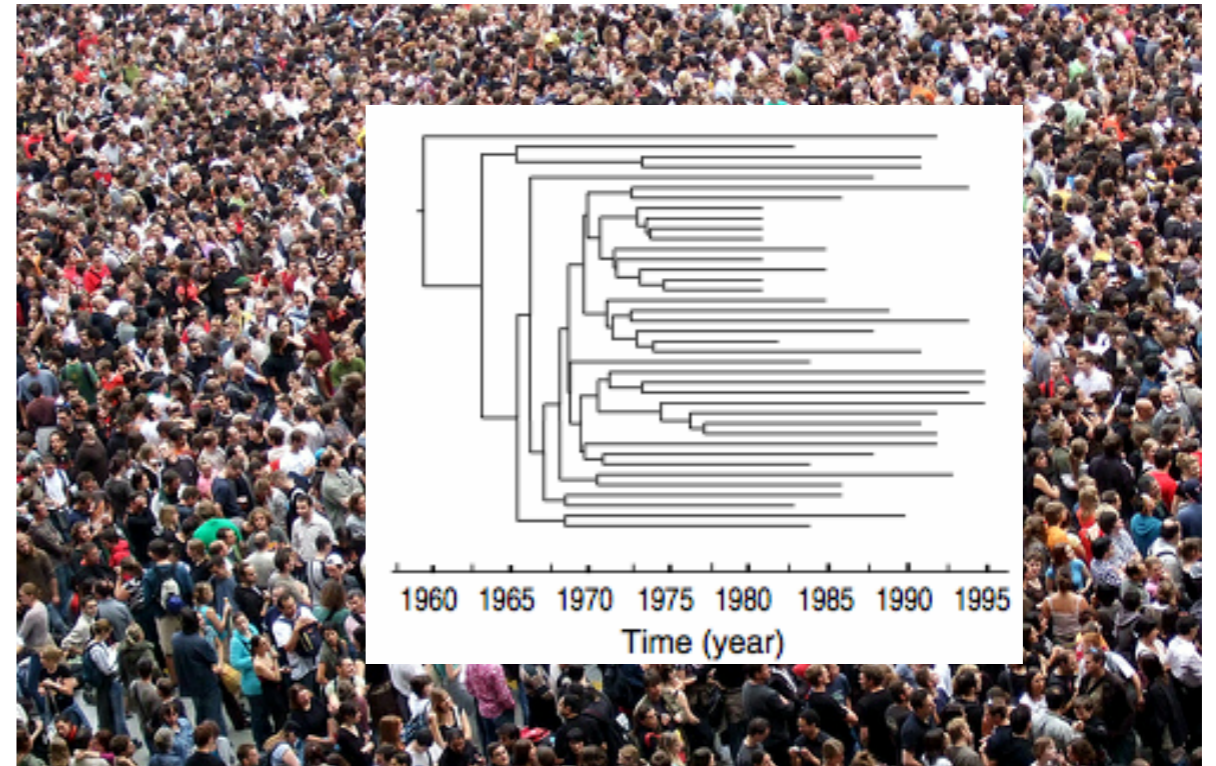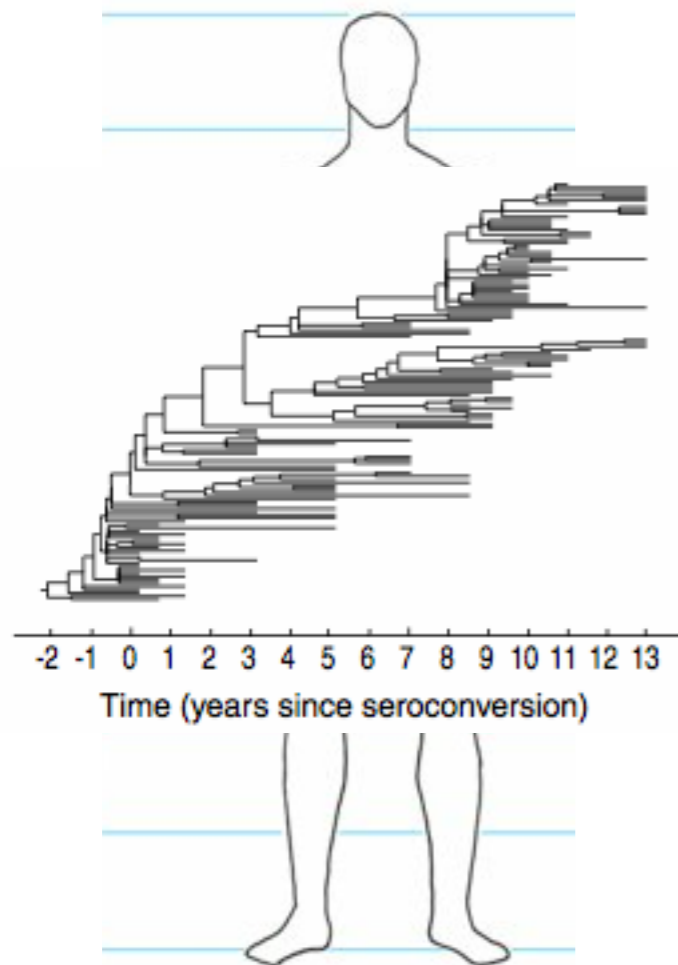
© James Cridland

3) Within-host evolution is a 'dead-end'

Time (years since seroconversion)

© James Cridland

Time (year)
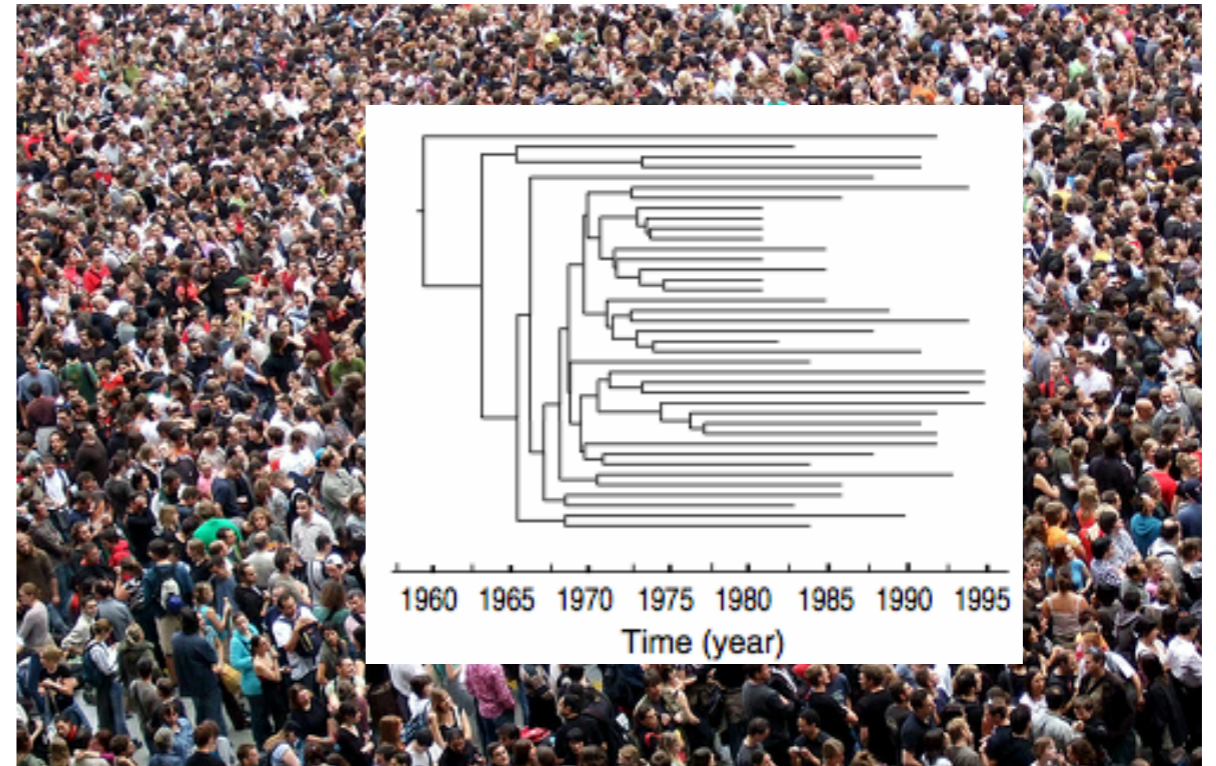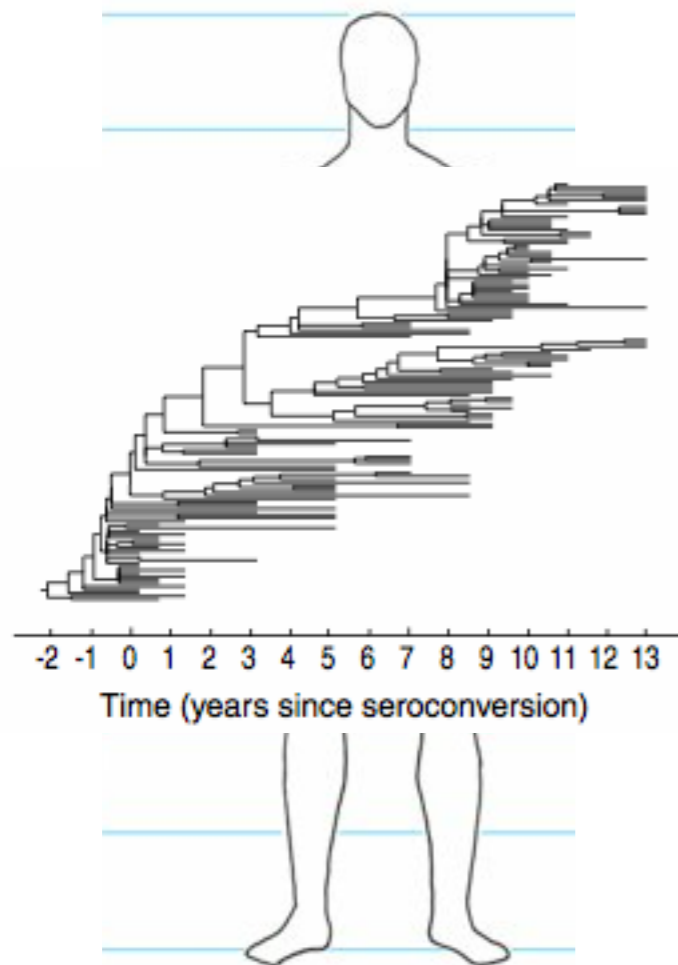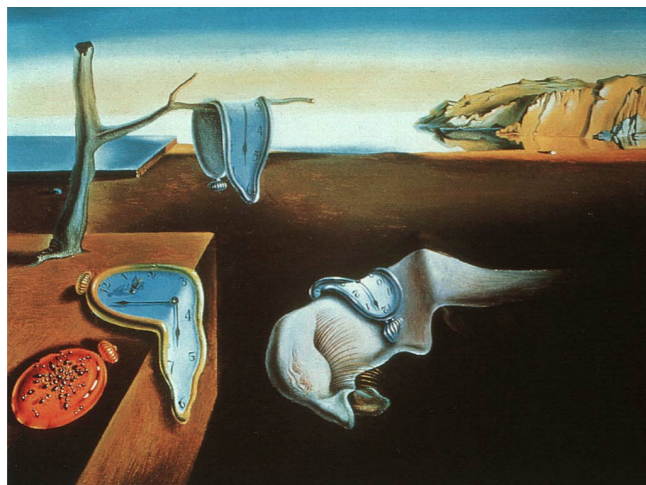
Measure HIV substitution rates at different levels

3) Within-host evolution is a 'dead-end'
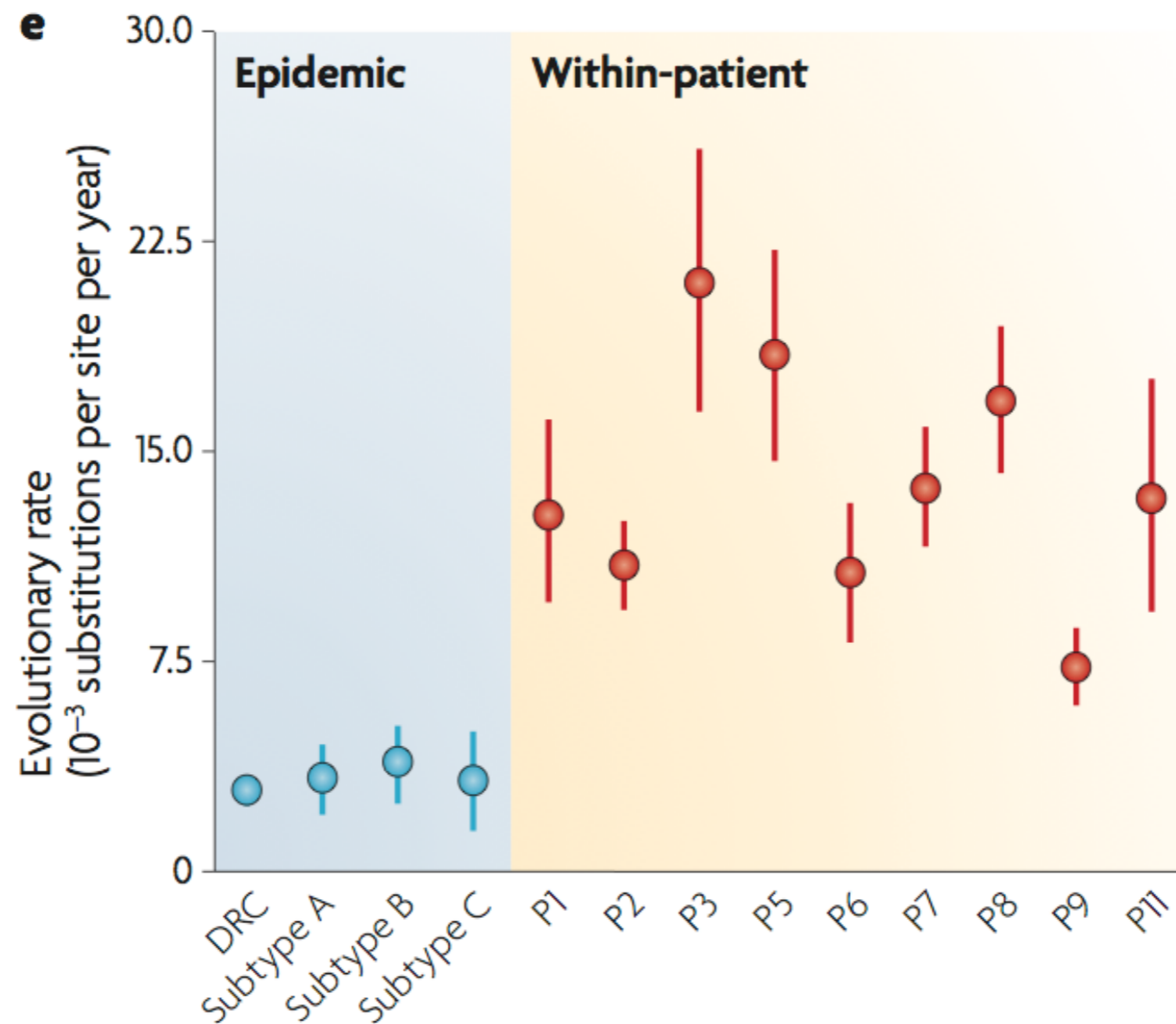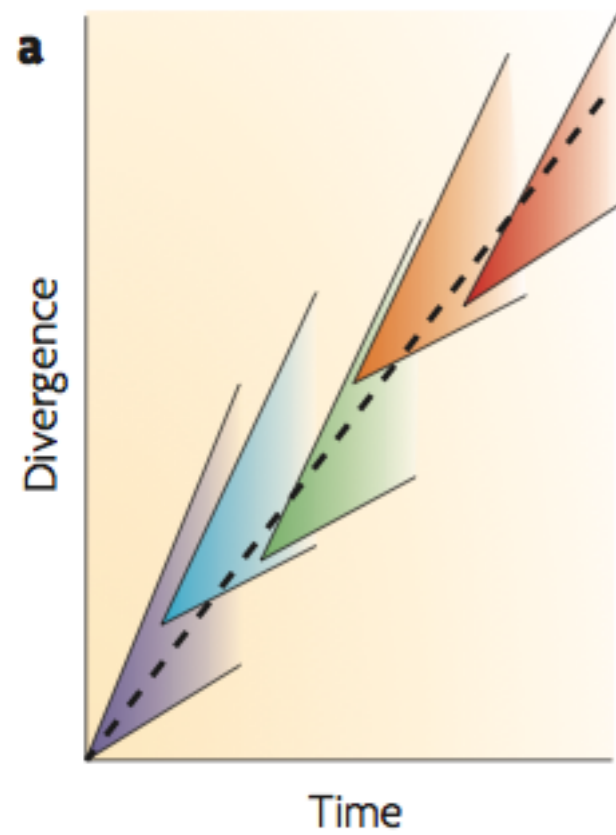
Pybus & Rambaut (2009, *Nat. Rev. Genet.*)

Why would evolutionary rates differ?

Pybus & Rambaut (2009, *Nat. Rev. Genet.*)

Why would evolutionary rates differ?

Store & retrieve

'less evolved' virions
are more transmissible

Stage specific

heterogeneity in
transmission timing
+ variable ER

Pybus & Rambaut (2009, *Nat. Rev. Genet.*)
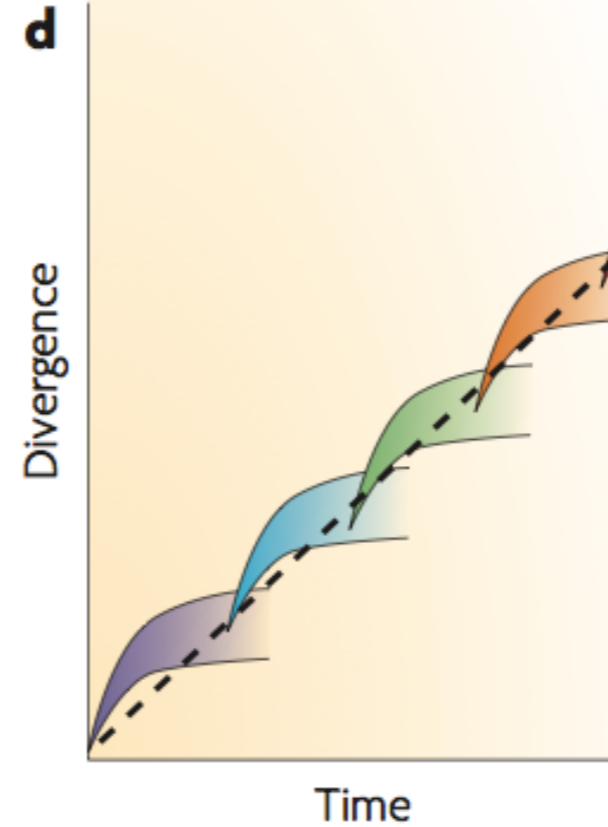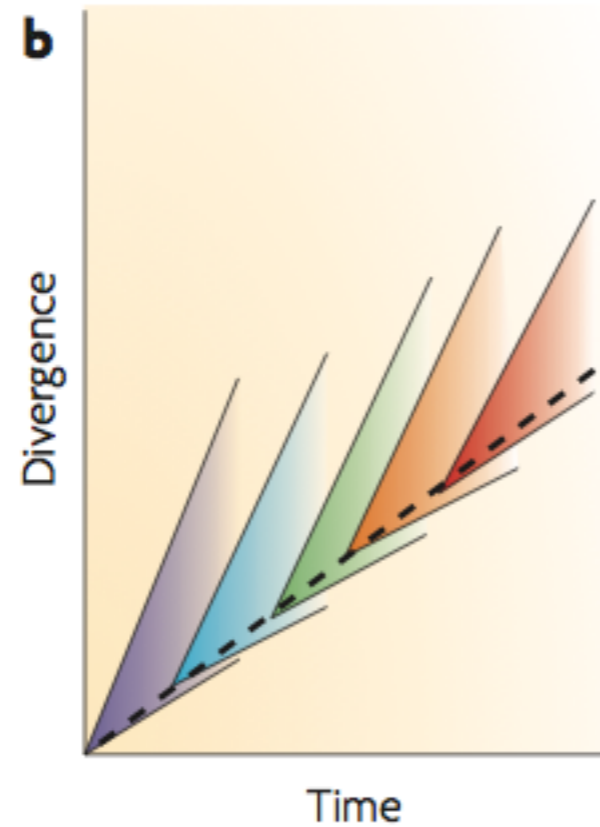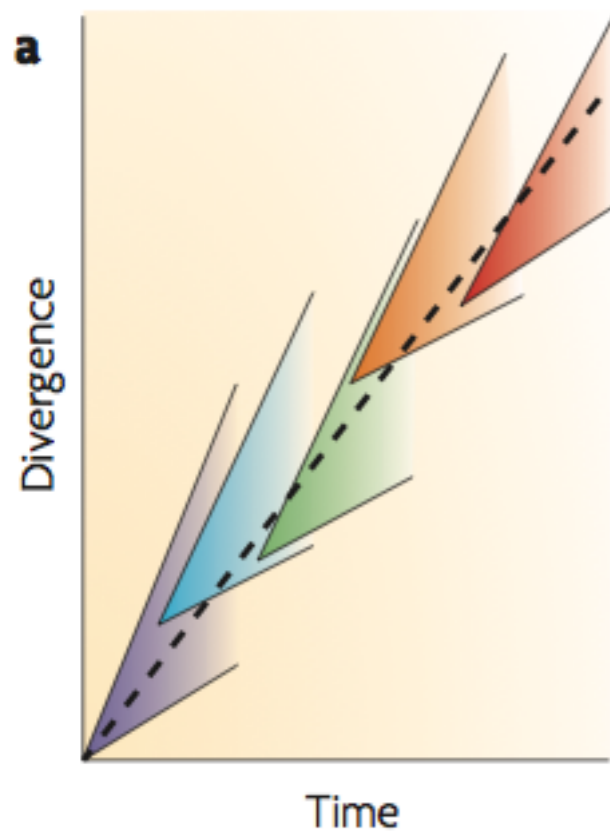
Why would evolutionary rates differ?

Store & retrieve
'less evolved' virions
are more transmissible
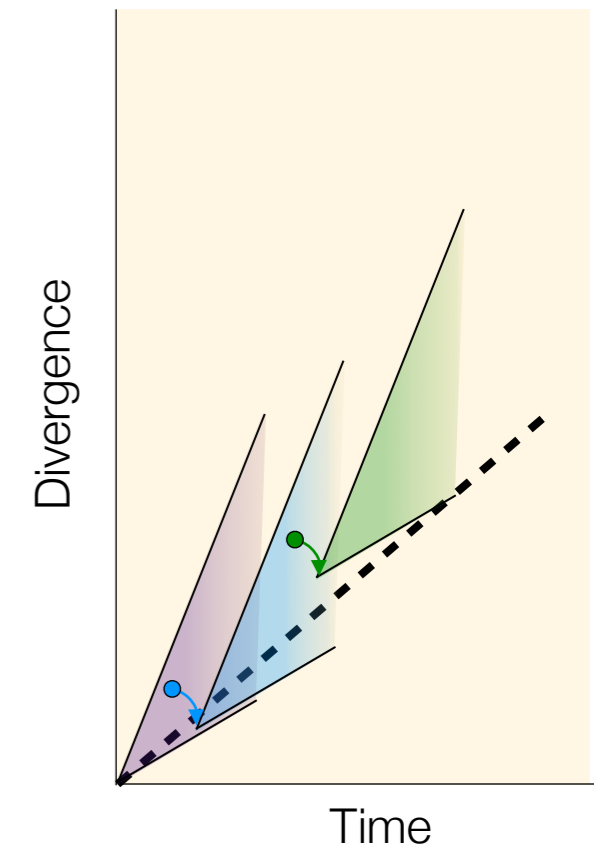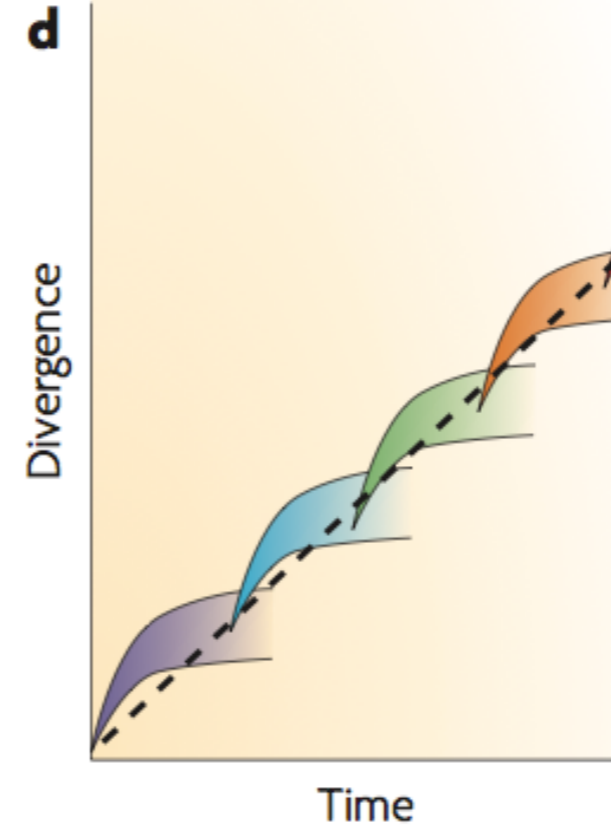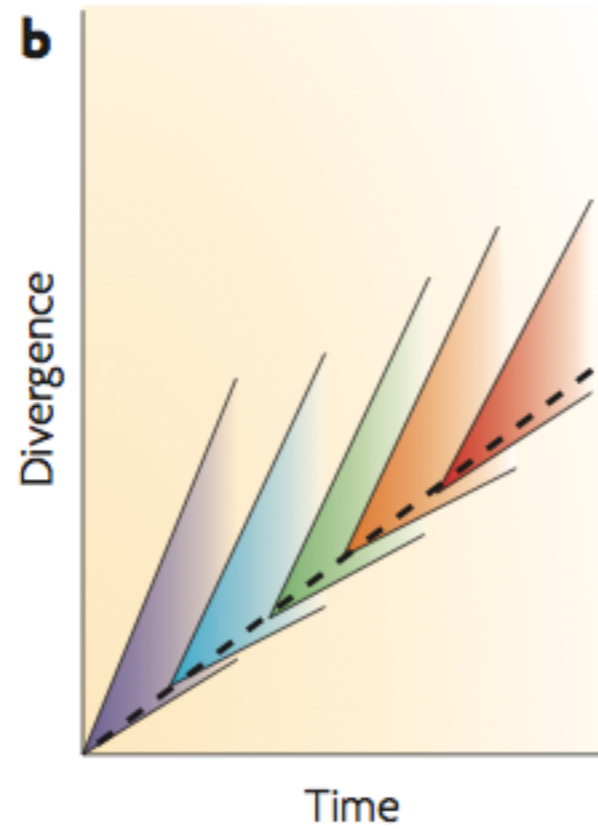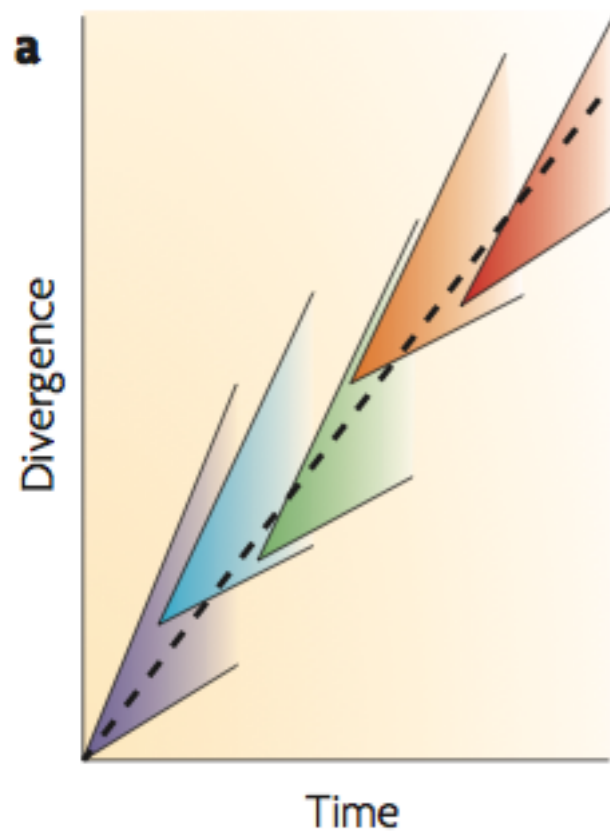
Stage specific
heterogeneity in
transmission timing
+ variable ER

Adapt & revert

Pybus & Rambaut (2009, *Nat. Rev. Genet.*)

Variation among regions?

Variation among regions?

Variation among regions?

Variation among regions?

Alizon & Fraser (2013, *Retrovirology*)

A

GAG | VIF | TAT | NEF | VPR | VPU | REV | LTR | POL | ENV1a 1b 1c 1d | ENV2

B

Log₁₀(ER)

-1
-2
-3

C

Log₁₀(ER)

-1
-2
-3

2000   4000   6000   8000   10000

Genomic position

BH evolutionary rates are lower than
WH rates throughout the HIV genome

Alizon & Fraser (2013, *Retrovirology*)

A

GAG | VIF | TAT | NEF
VPR | VPU | REV | LTR
POL | ENV1a | 1b | 1c | 1d | ENV2

B

Log$_{10}$(ER)

C

Log$_{10}$(ER)

Genomic position

BH evolutionary rates are lower than WH rates throughout the HIV genome

Supports the store and retrieve hypothesis

Alizon & Fraser (2013, *Retrovirology*)

Virulence is (partly) heritable, it affects infection fitness…
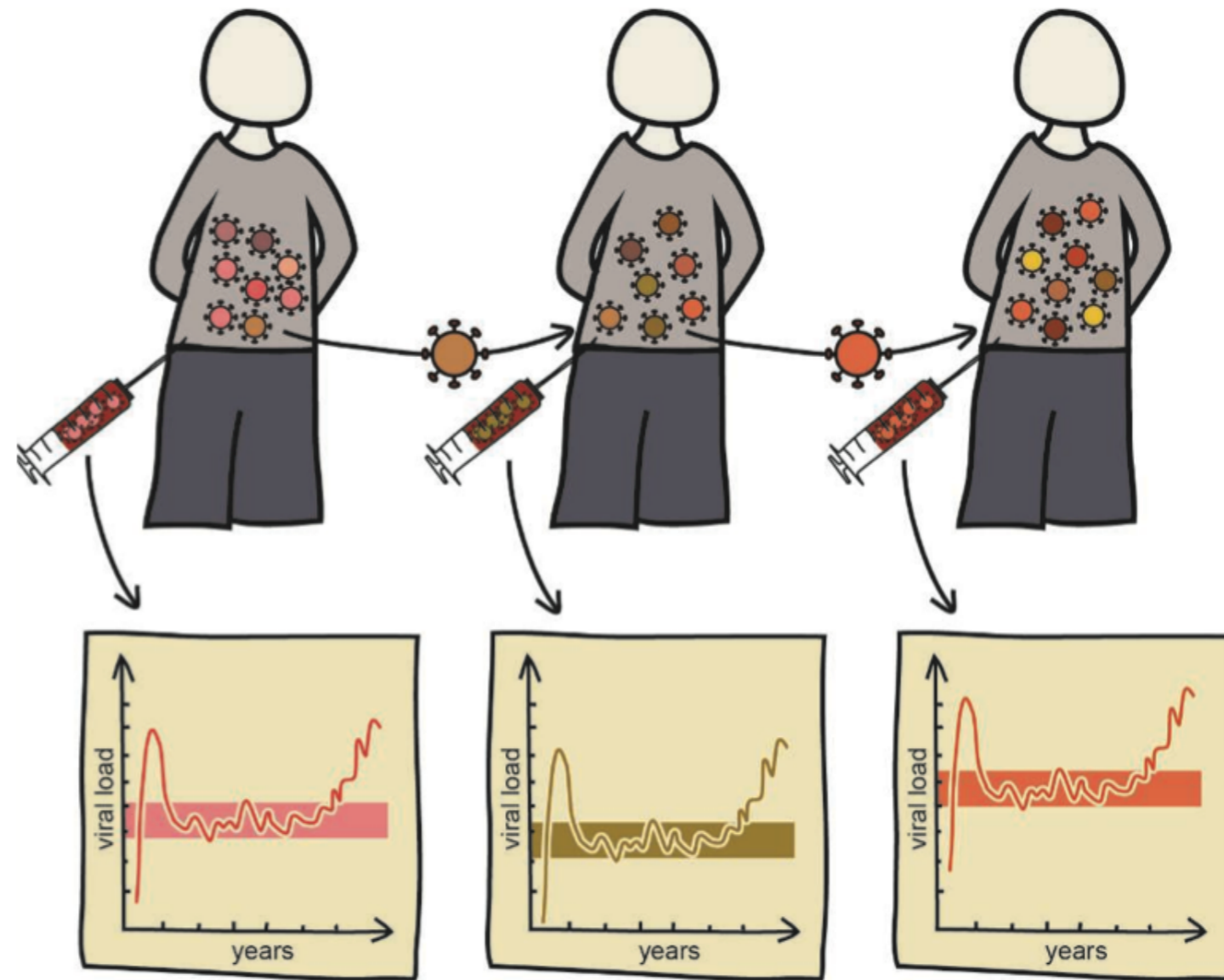
How does it evolve?

1) More virulent strains are more competitive WH

2) More virulent strains are more transmitted

1) More virulent strains are more competitive WH

2) More virulent strains are more transmitted
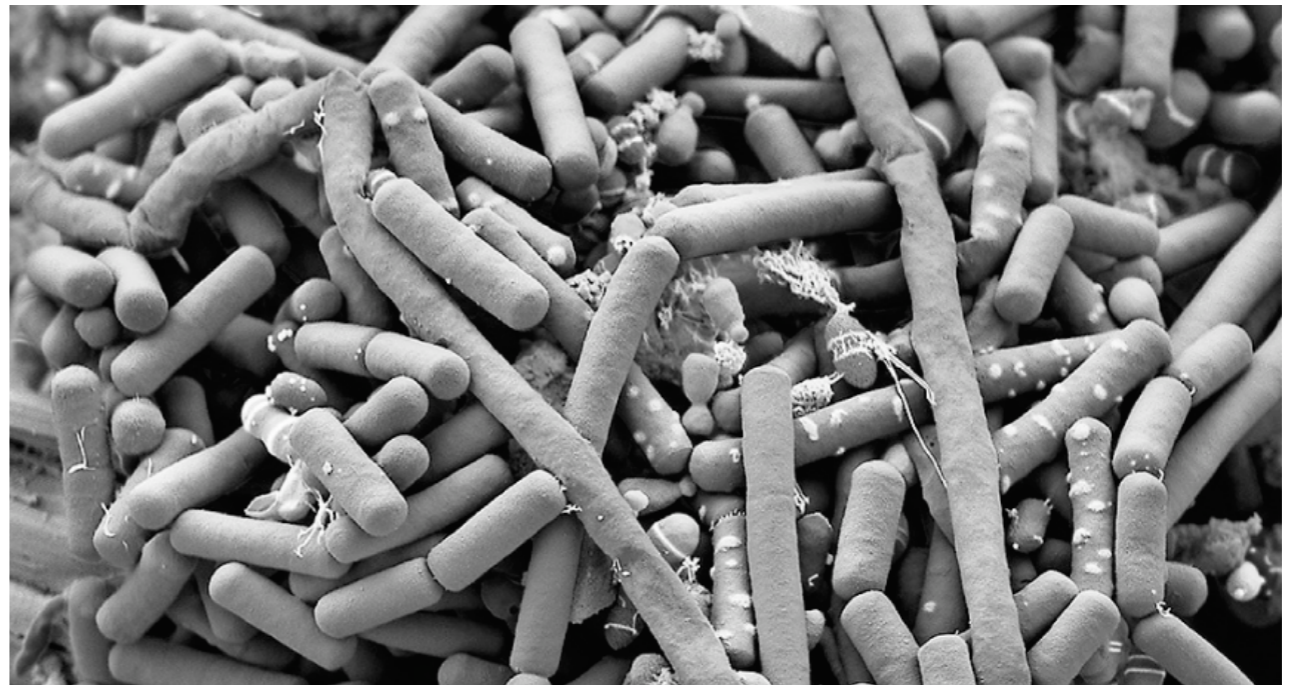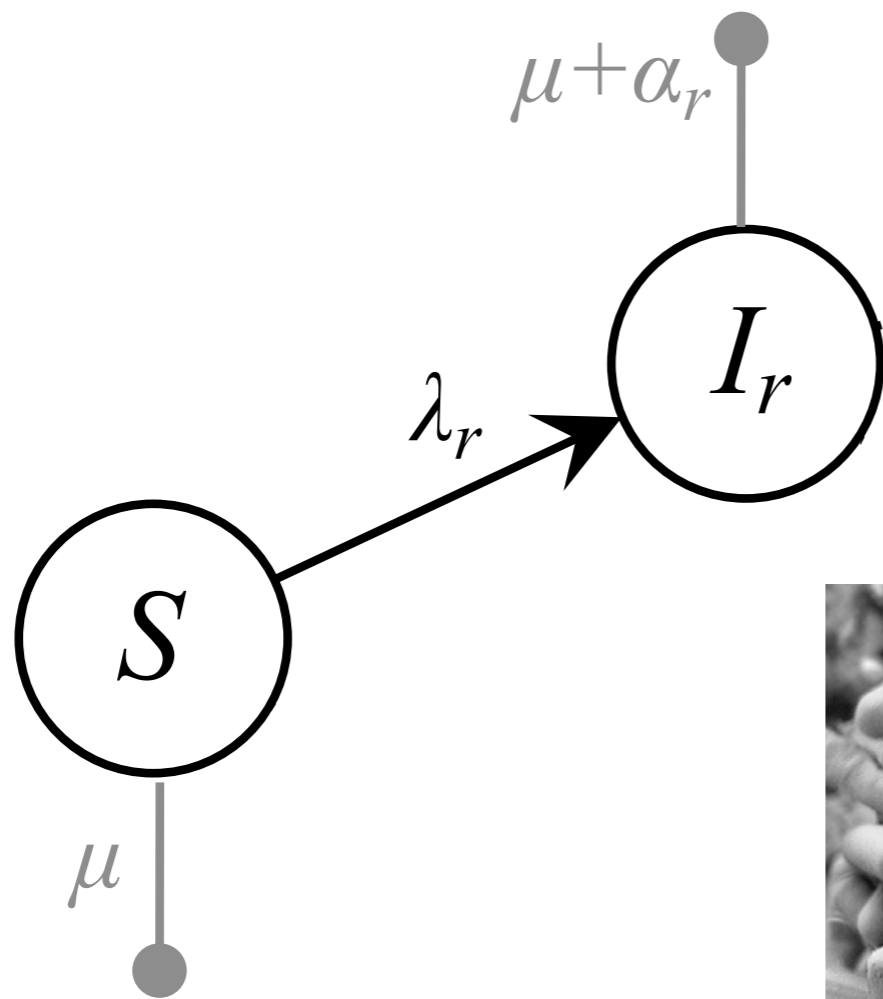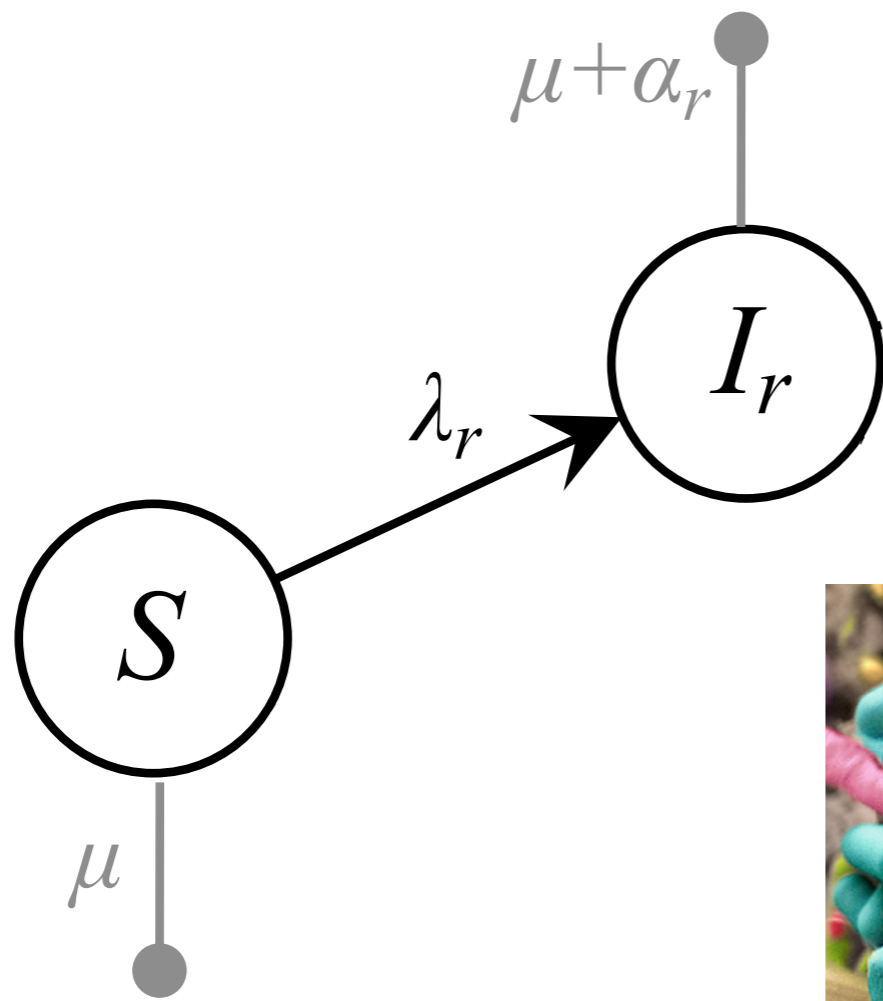
Coinfections are the rule rather than the exception
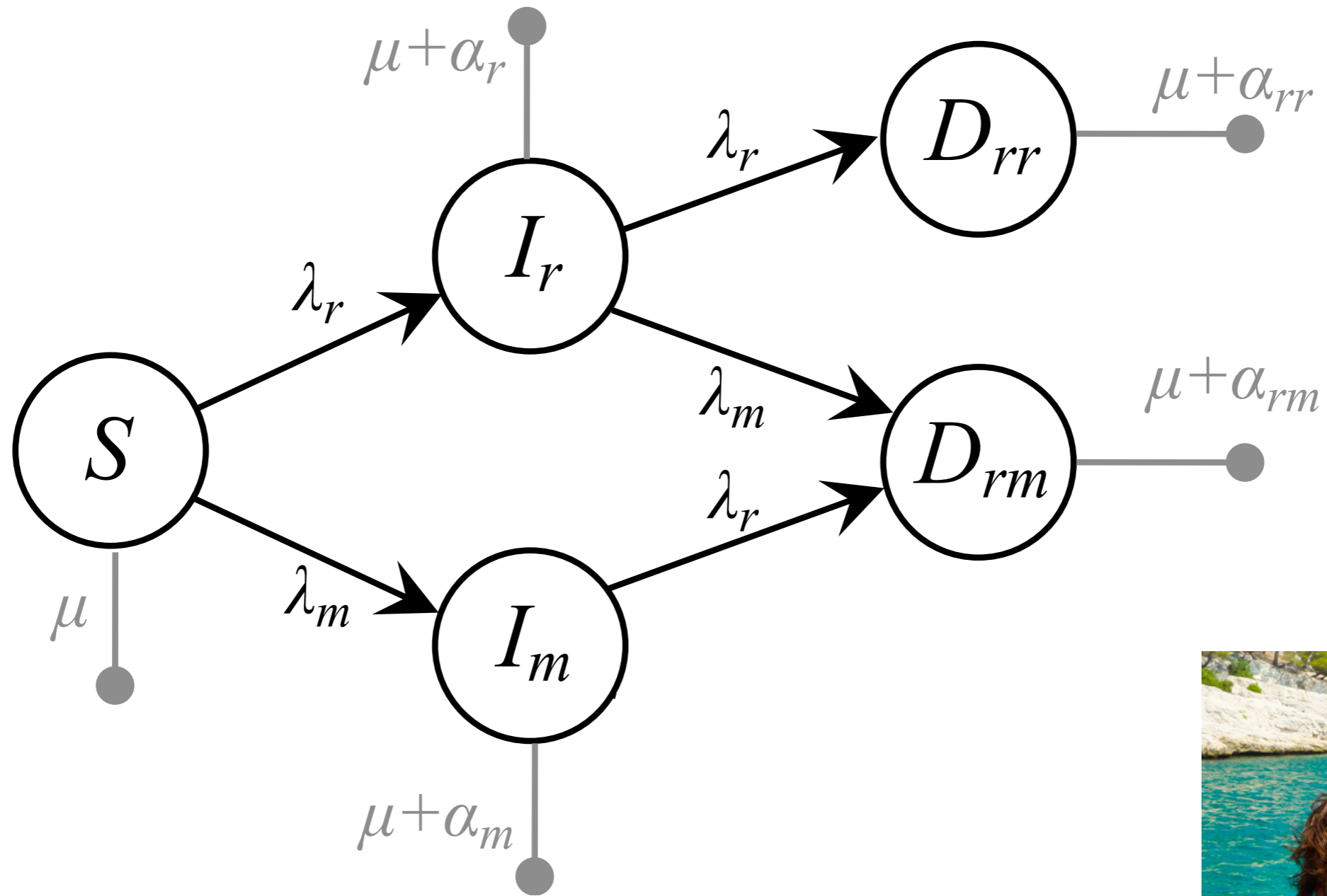
Mircea Sofonea's PhD

Coinfections are the rule rather than the exception
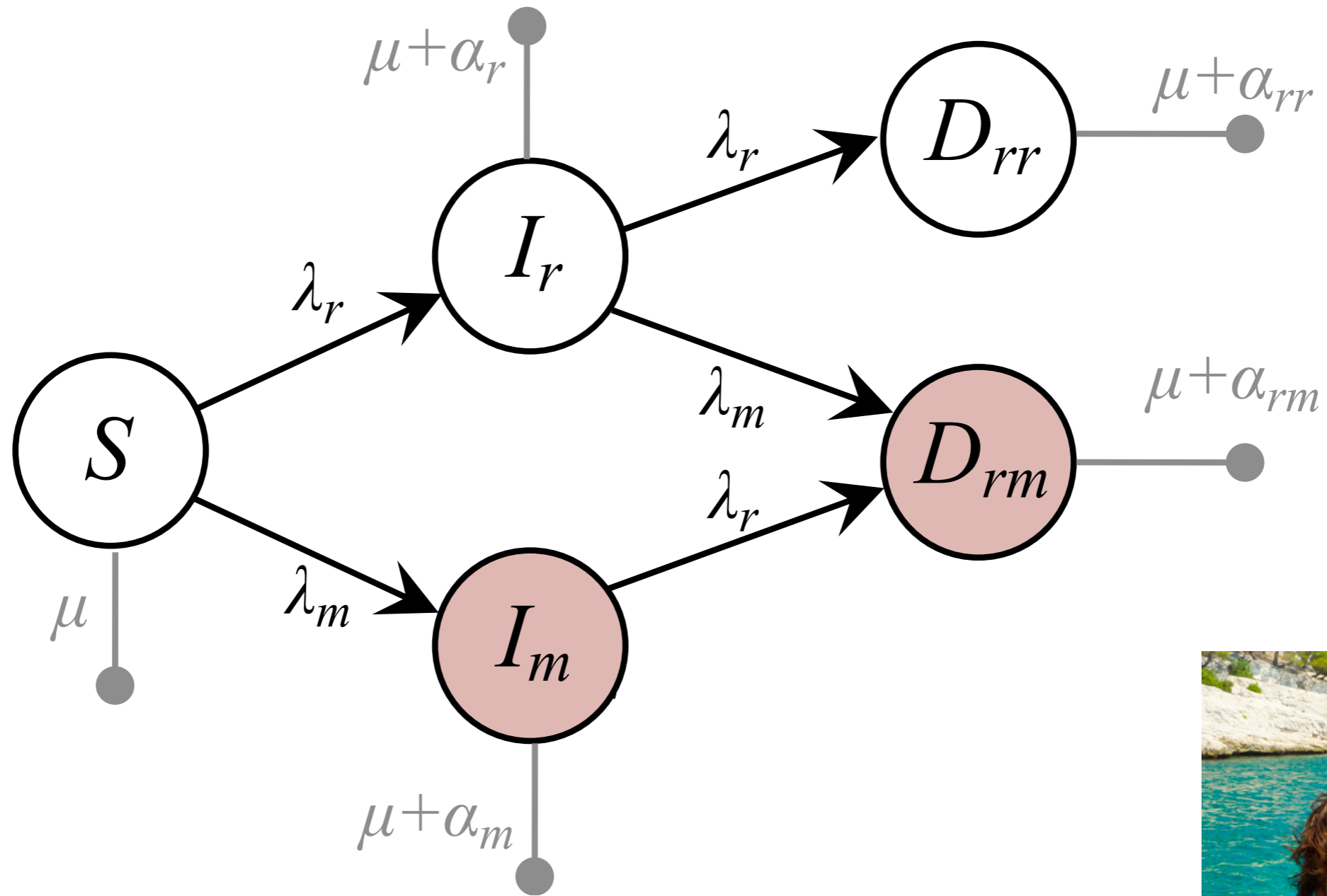
Mircea Sofonea's PhD

1) More virulent strains are more competitive WH

2) More virulent strains are more transmitted

Variability of set-point virus load
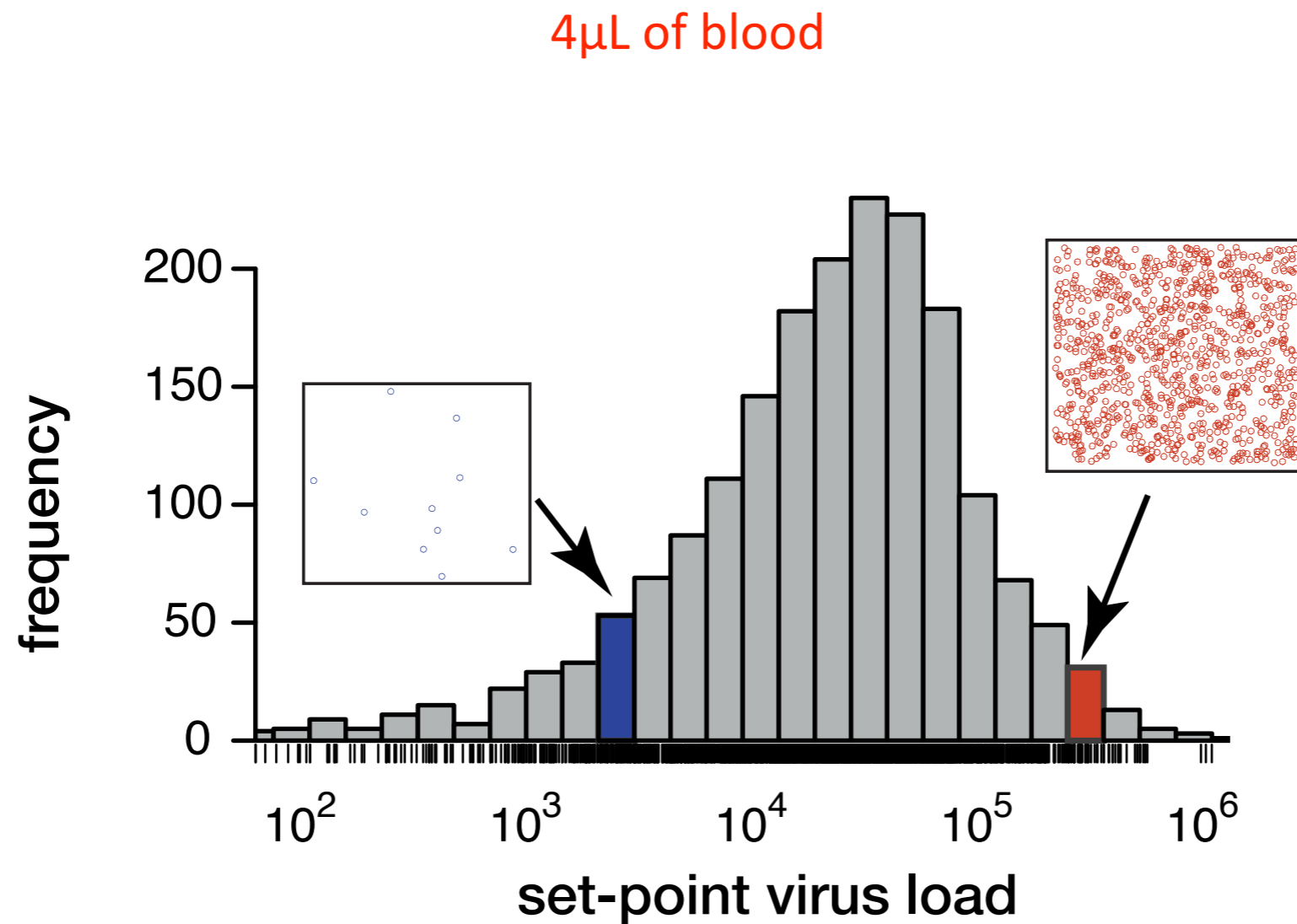
4µL of blood

Fraser *et al.* (2007, *PNAS*)
Fraser *et al.* (2014, *Science*)

Effect on infection fitness

Fraser *et al.* (2007, *PNAS*)
Fraser *et al.* (2014, *Science*)

Effect on infection fitness

cost: virulence decreases infection duration

Fraser *et al.* (2007, *PNAS*)
Fraser *et al.* (2014, *Science*)

Effect on infection fitness

Fraser *et al.* (2007, *PNAS*)
Fraser *et al.* (2014, *Science*)
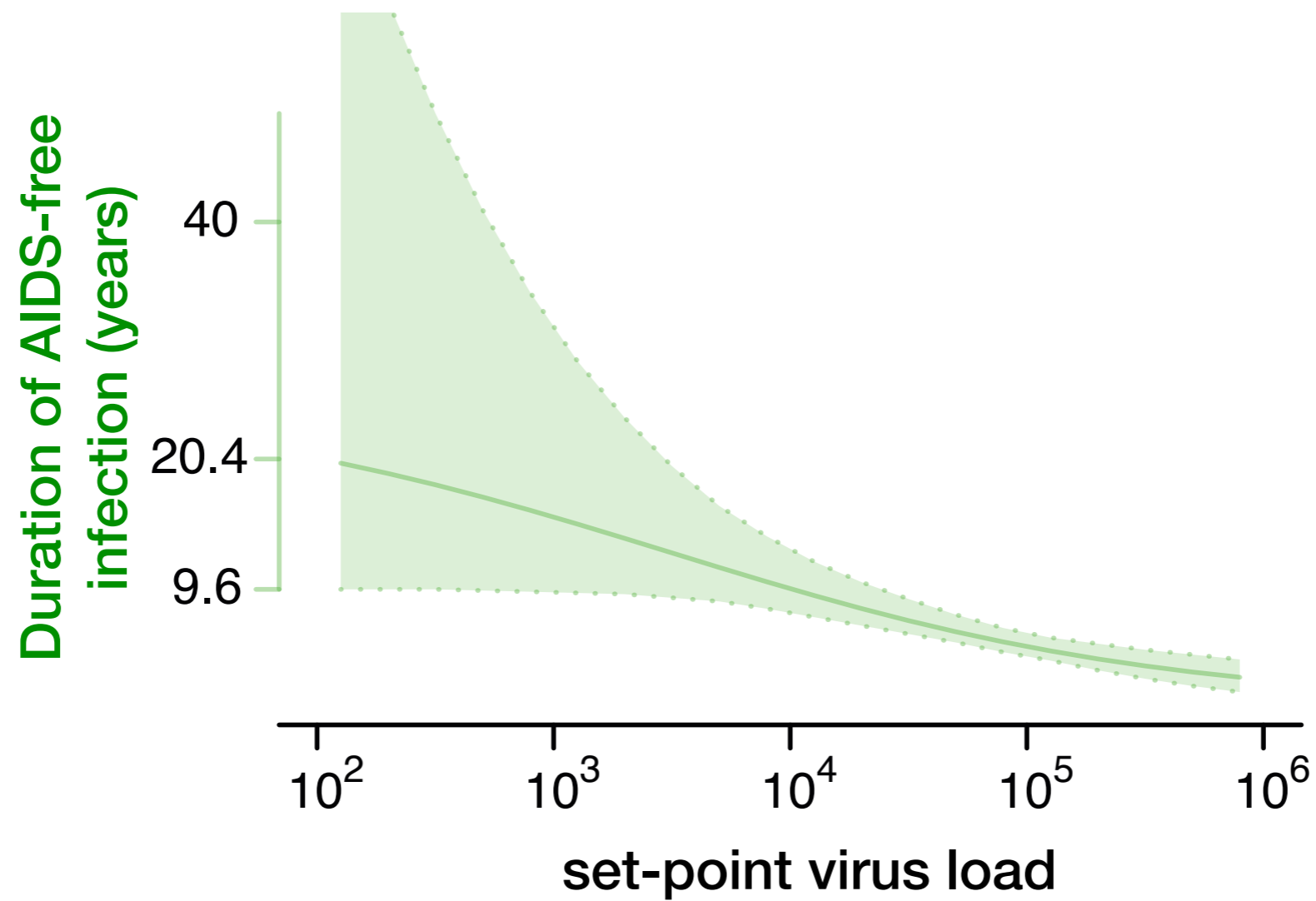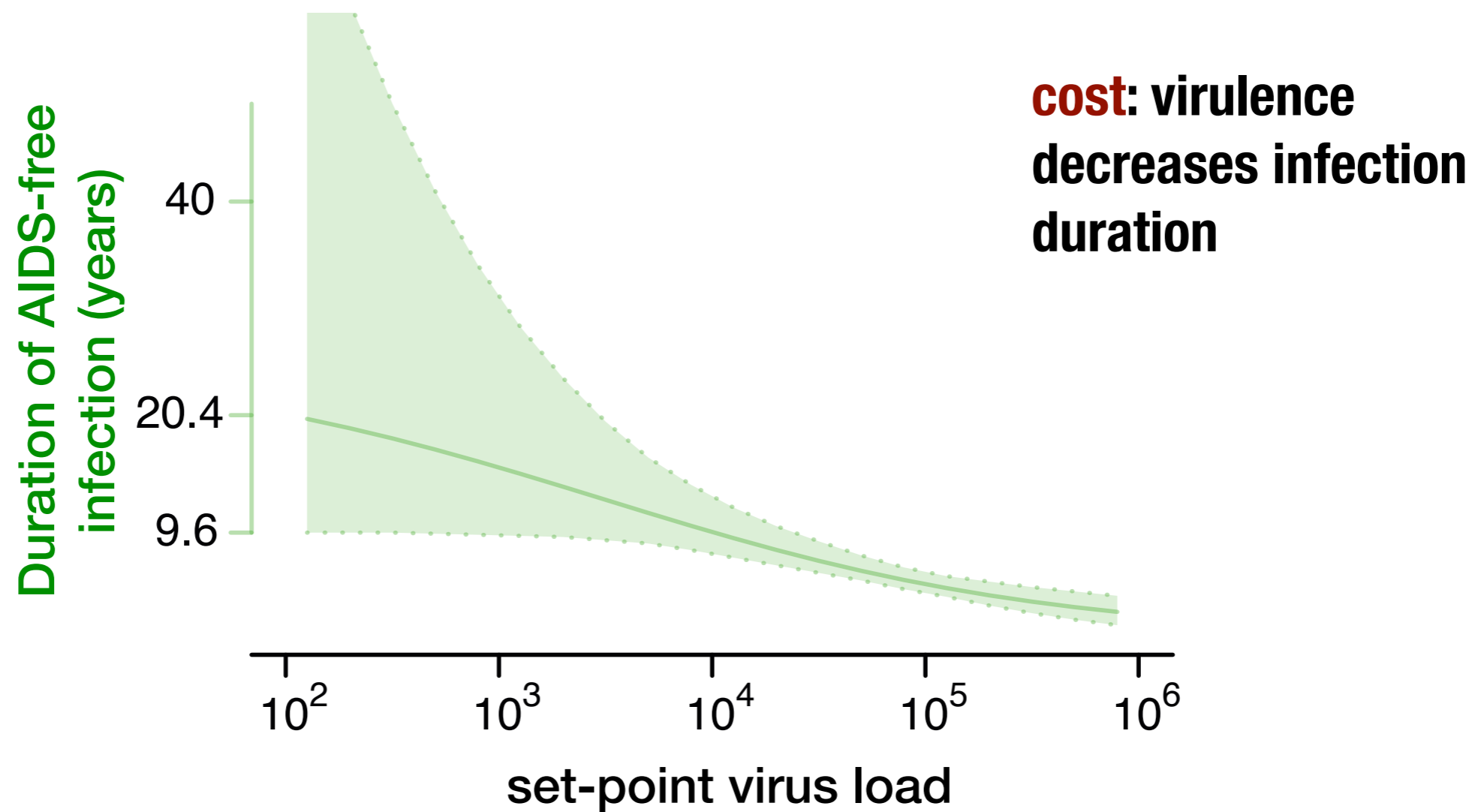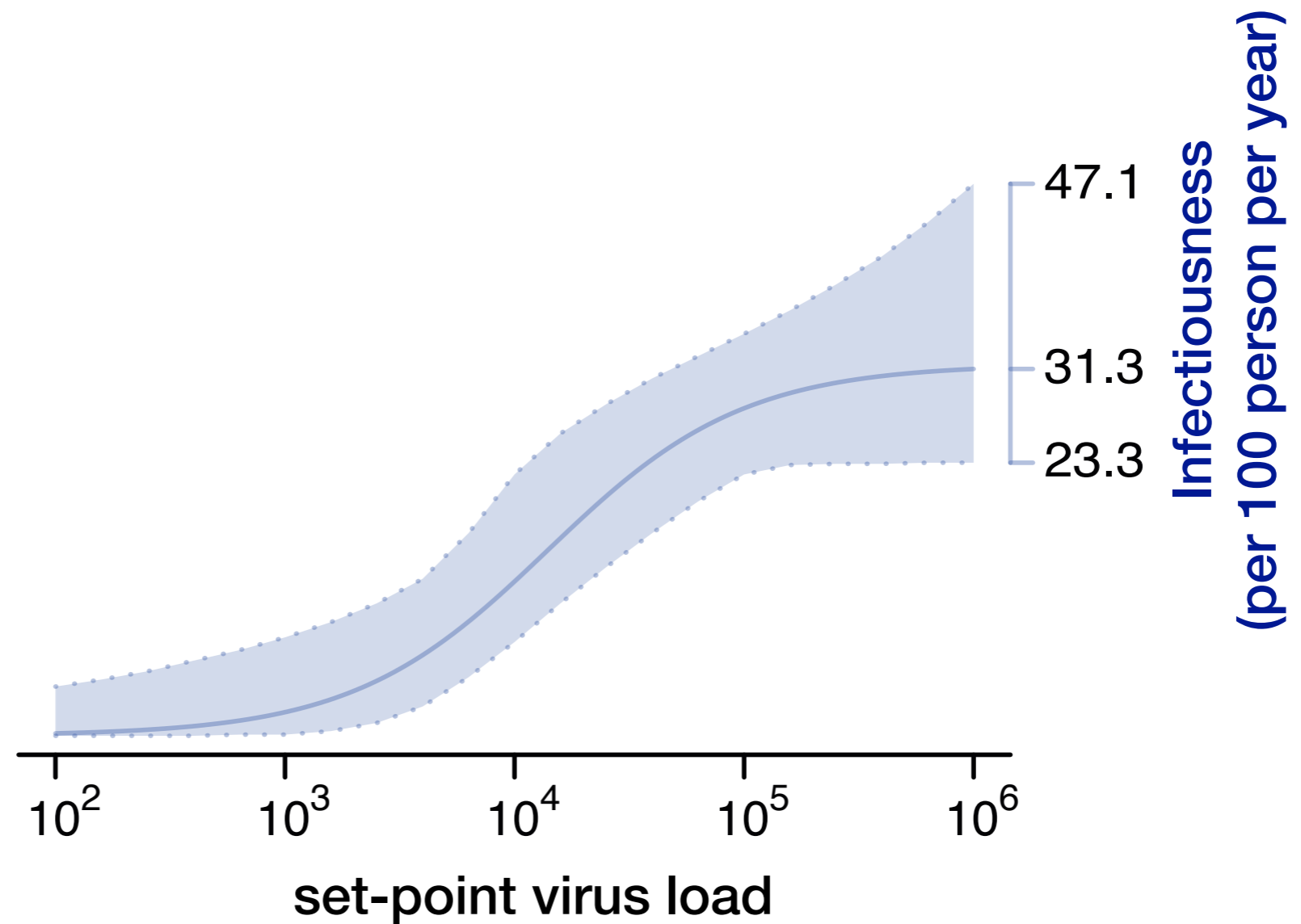
Effect on infection fitness

Benefit: virulence
increases
transmission rate
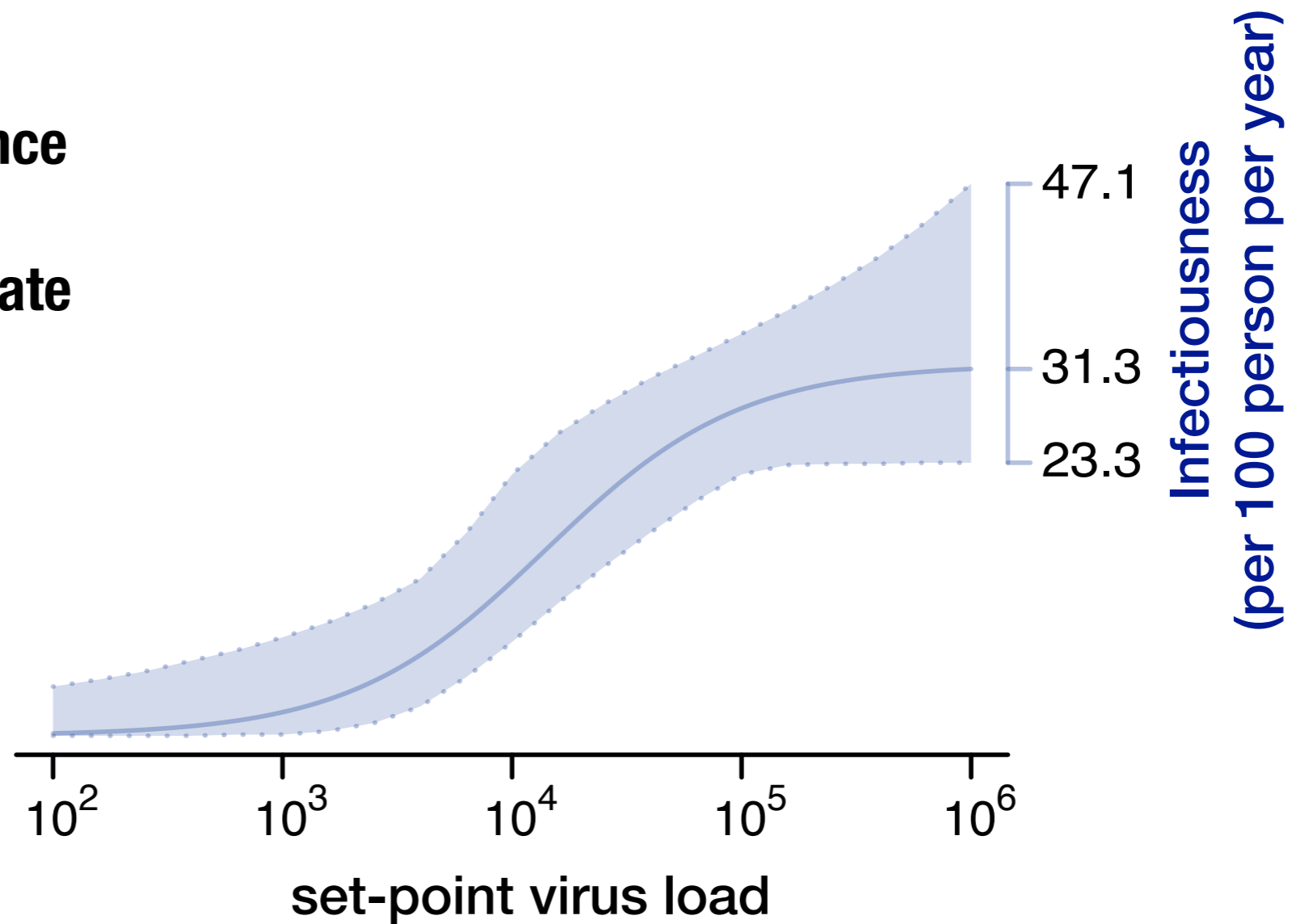
Fraser *et al.* (2007, *PNAS*)
Fraser *et al.* (2014, *Science*)

Effect on infection fitness

Fraser *et al.* (2007, *PNAS*)
Fraser *et al.* (2014, *Science*)

Effect on infection fitness

trade-off!

Duration of AIDS-free infection (years)

Infectiousness (per 100 person per year)

set-point virus load

Fraser *et al.* (2007, *PNAS*)
Fraser *et al.* (2014, *Science*)

Effect on infection fitness

Fraser *et al.* (2007, *PNAS*)
Fraser *et al.* (2014, *Science*)
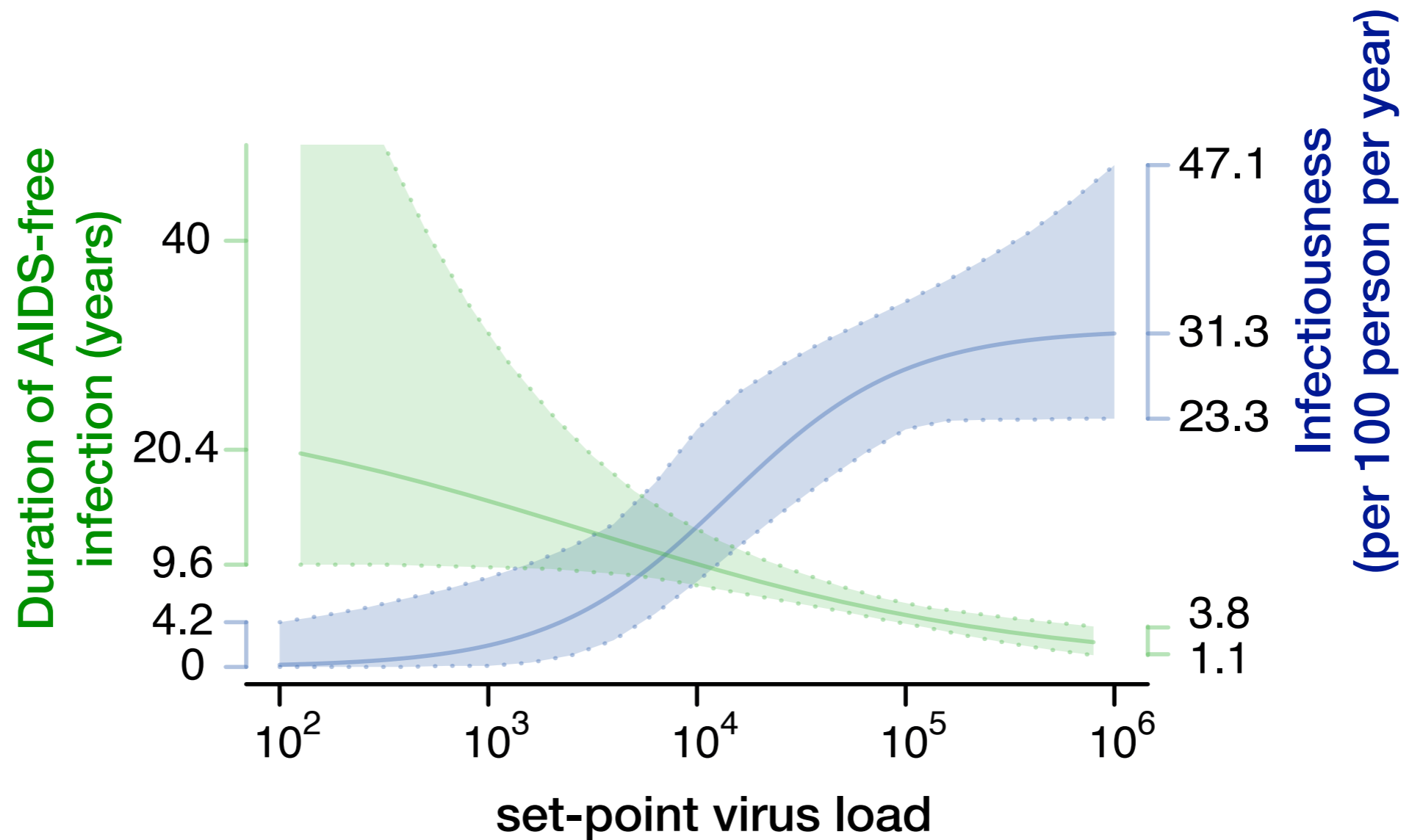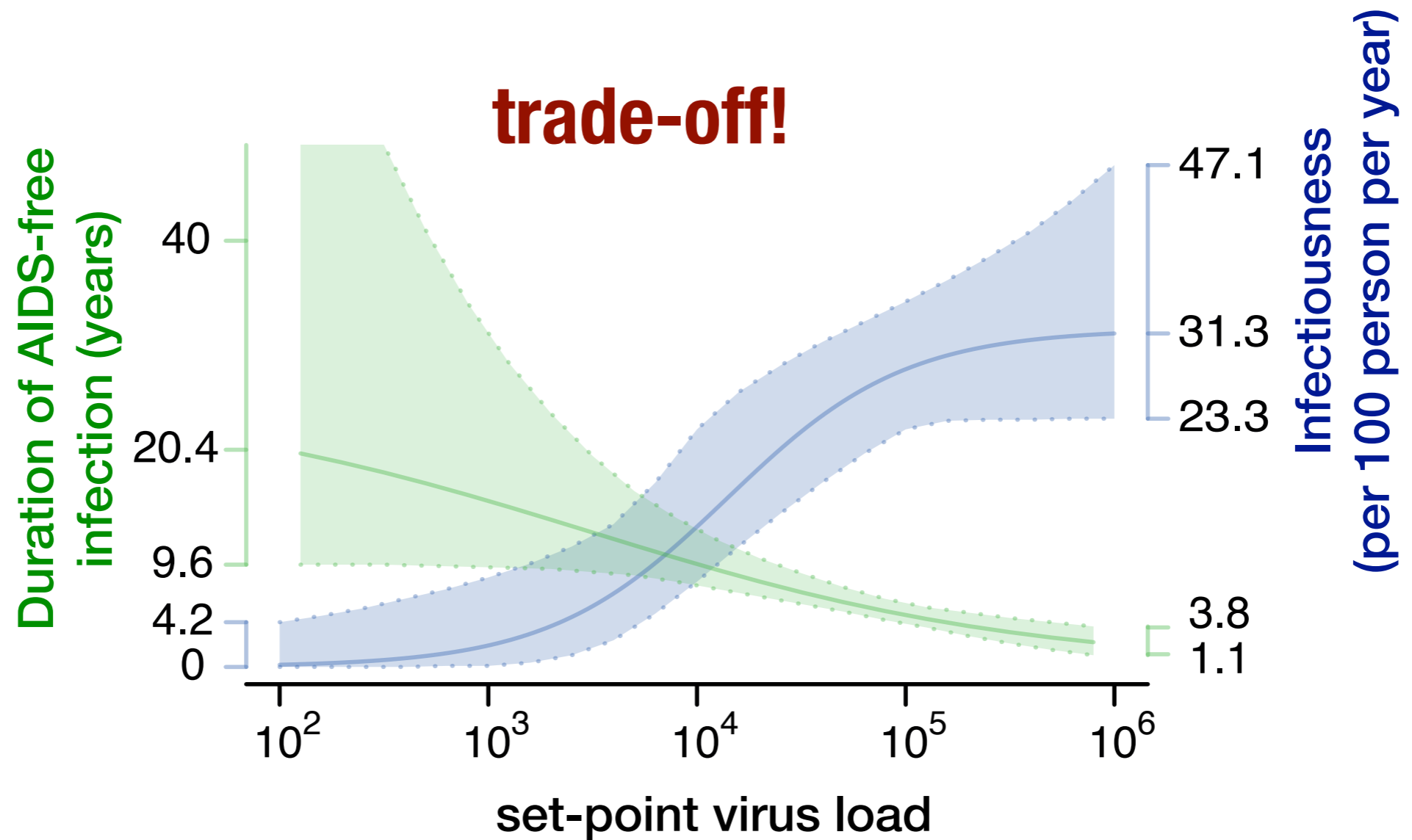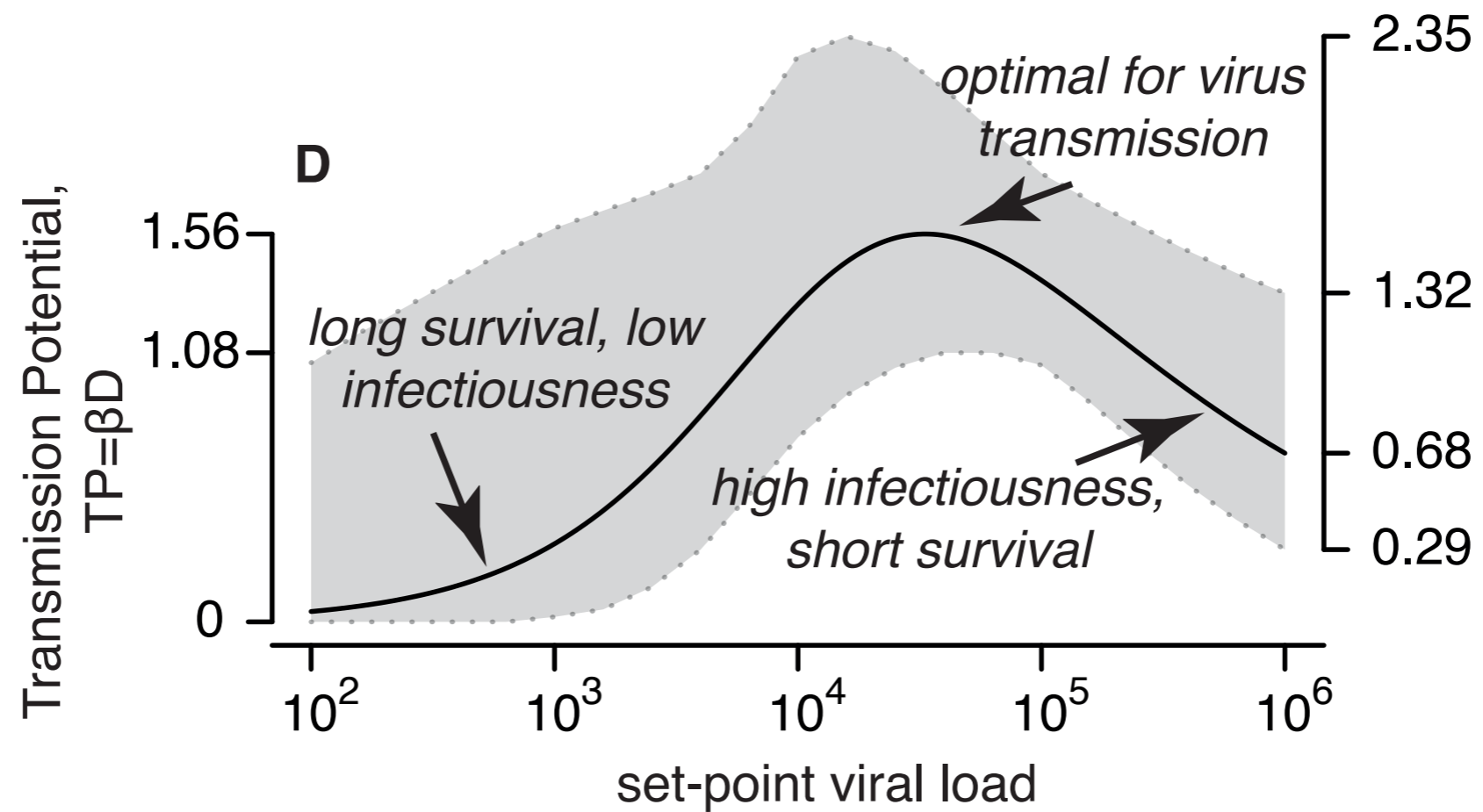
Why do parasites harm their host?
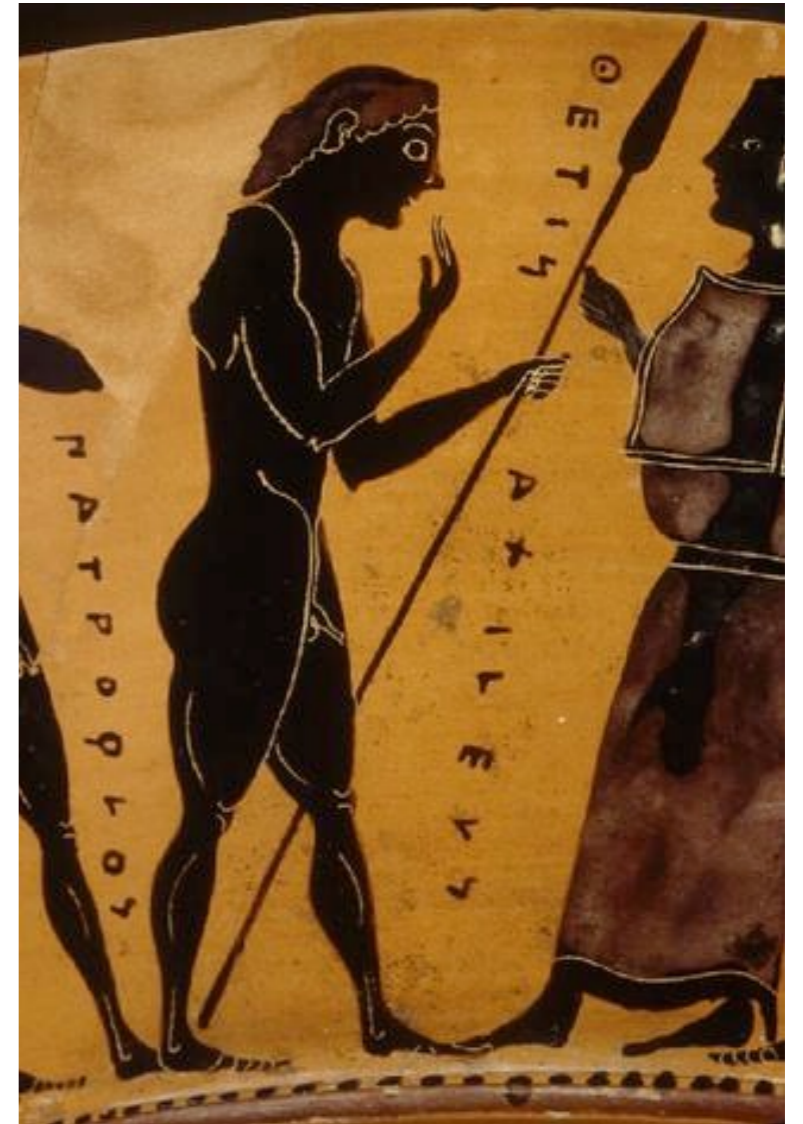
Anderson & May (1982, *Parasitology*)
Ewald (1994)
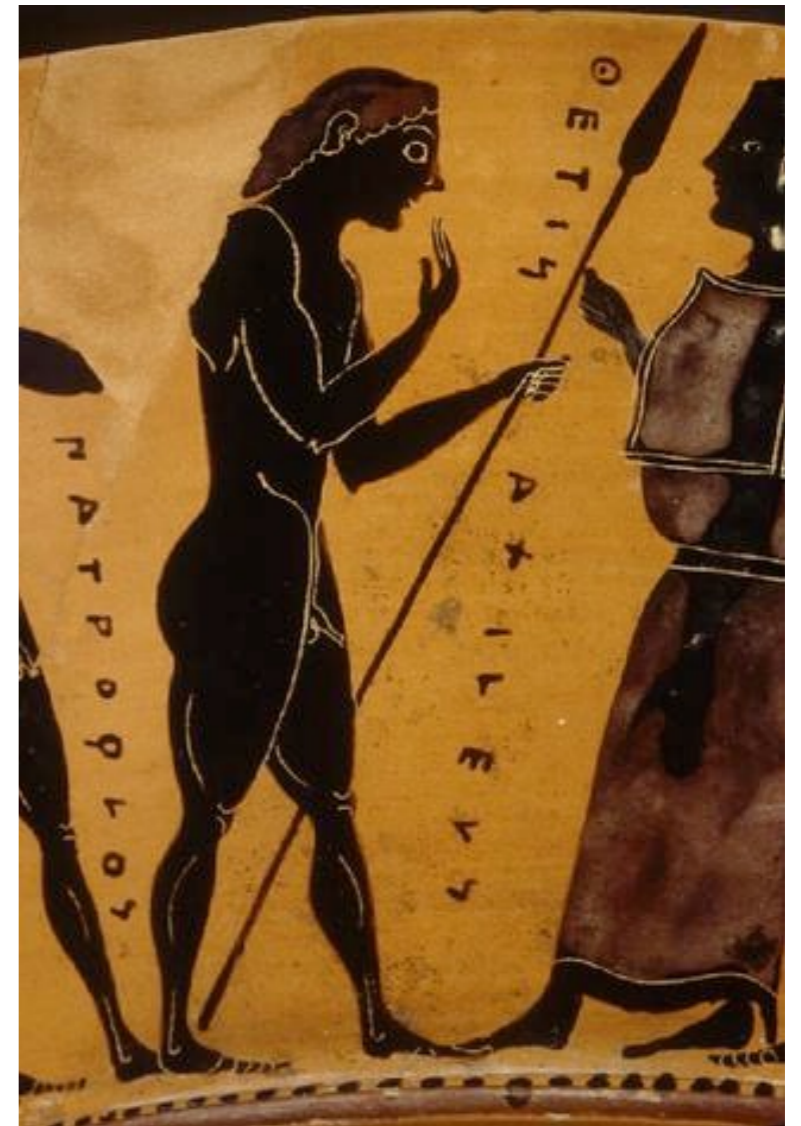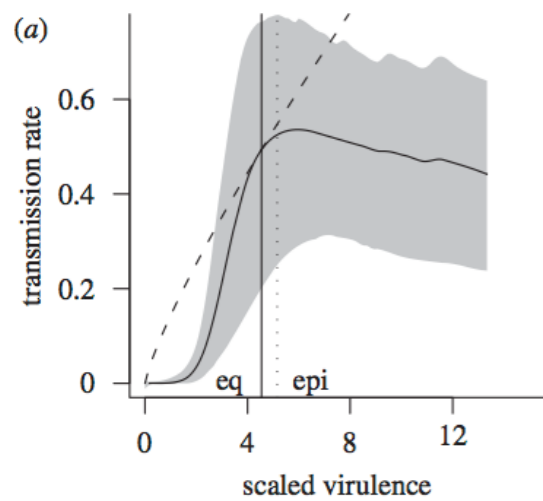
Why do parasites harm their host?

Bolker *et al.*
(2010, *J R Soc Interface*)

Anderson & May (1982, *Parasitology*)
Ewald (1994)

Why do parasites harm their host?

Bolker *et al.*
(2010, *J R Soc Interface*)

Doumayrou et al. (2013, *Evolution*)
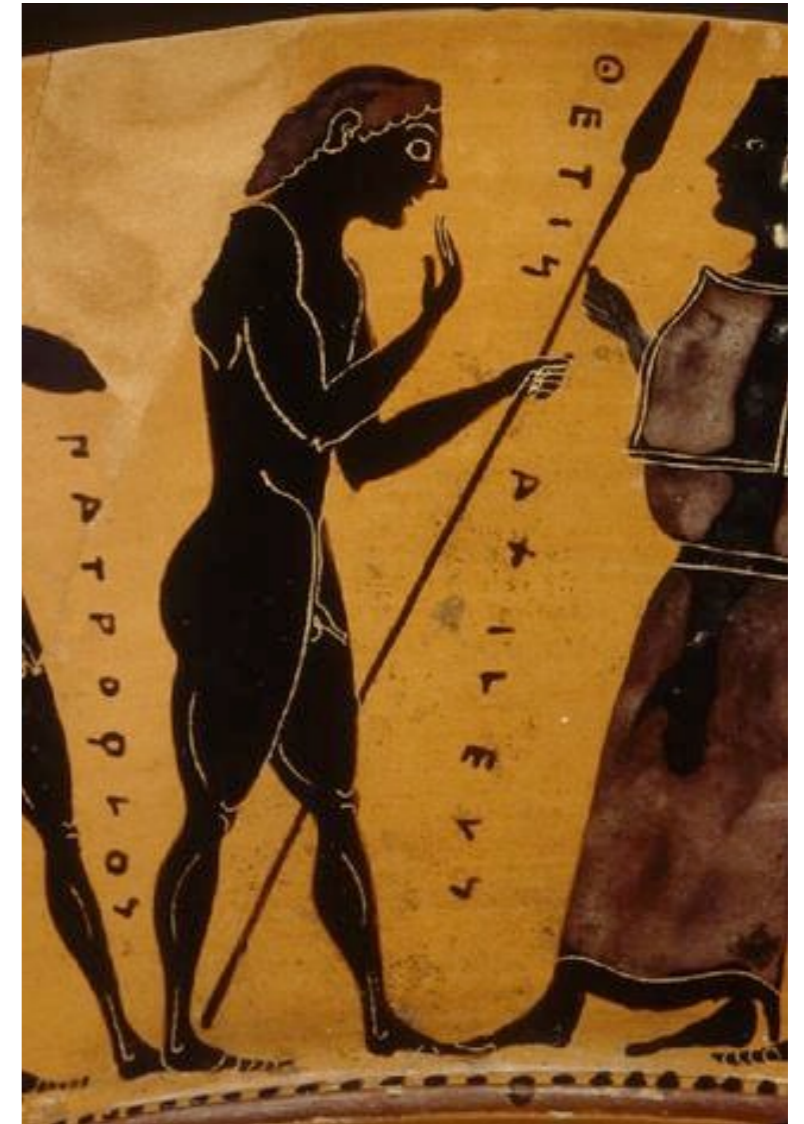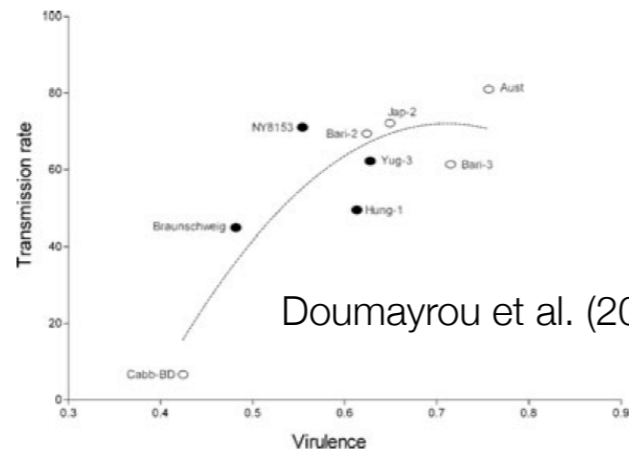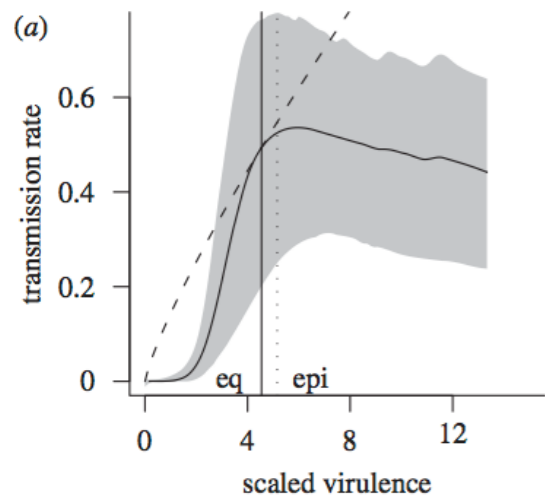
Anderson & May (1982, *Parasitology*)
Ewald (1994)

Why do parasites harm their host?

Bolker *et al.*
(2010, *J R Soc Interface*)
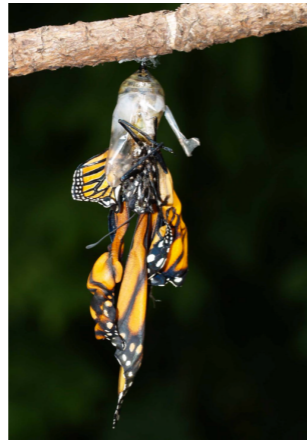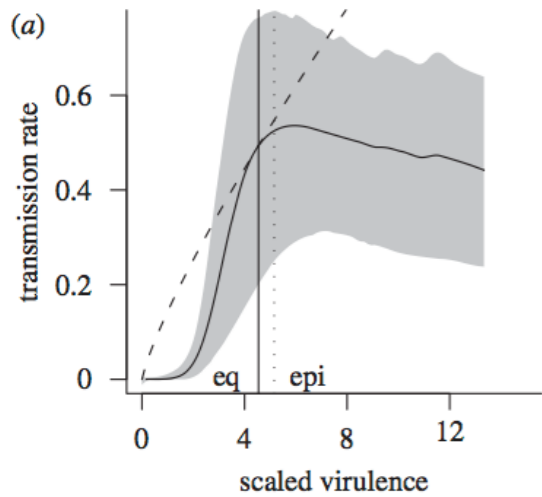
De Roode *et al.* (2008, *PNAS*)
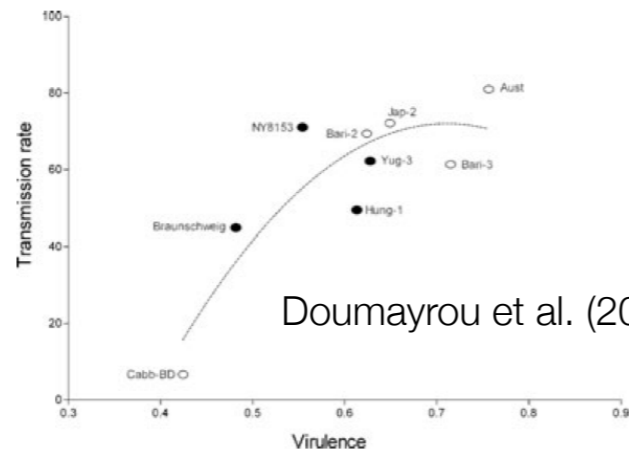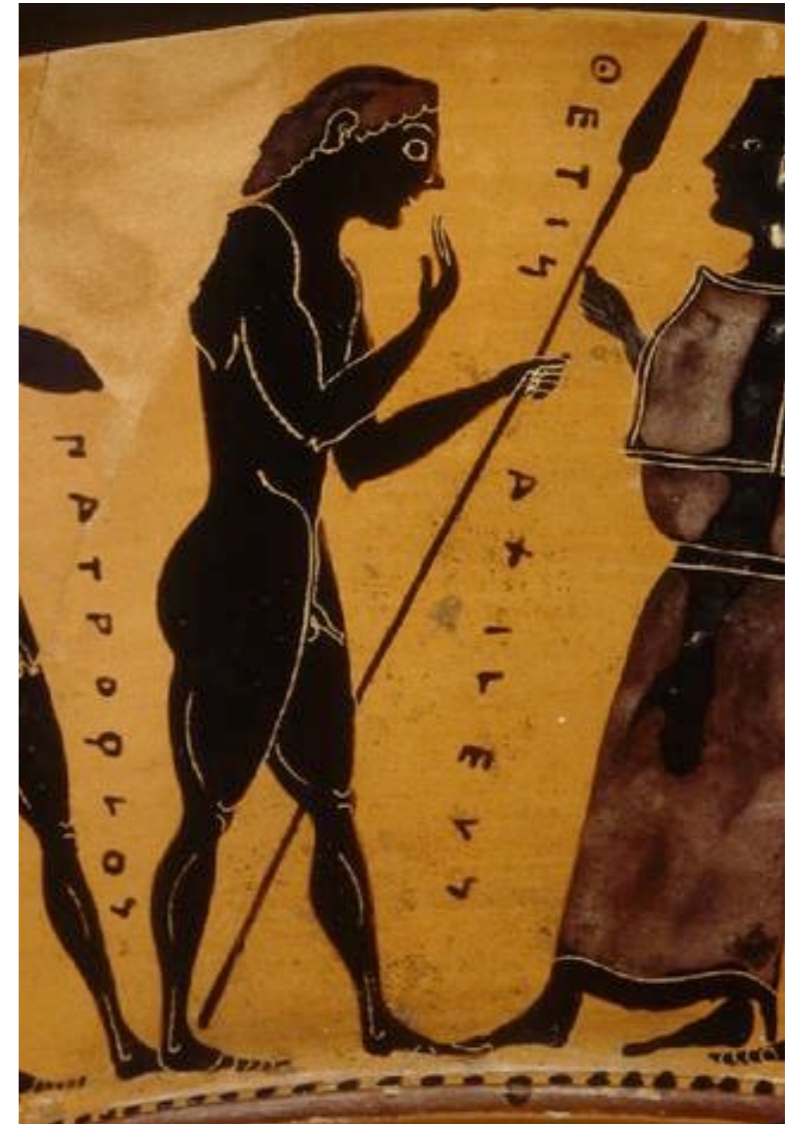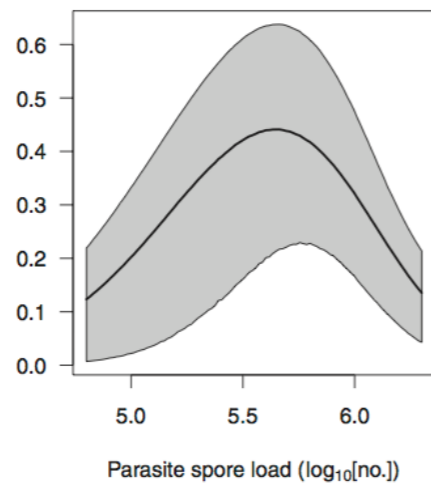
Doumayrou et al. (2013, *Evolution*)

Anderson & May (1982, *Parasitology*)
Ewald (1994)

Why do parasites harm their host?

Bolker *et al.*
(2010, *J R Soc Interface*)
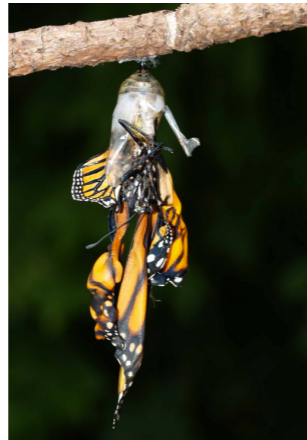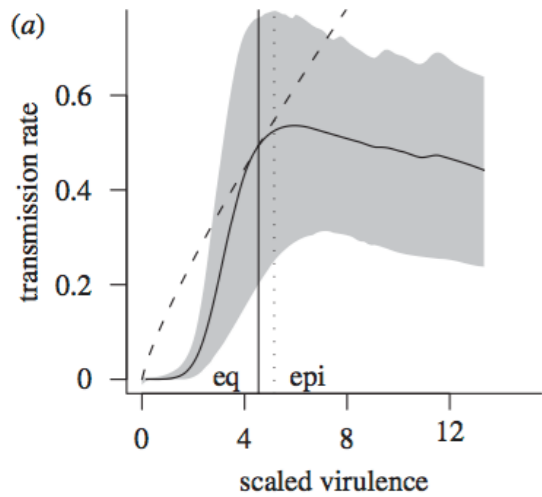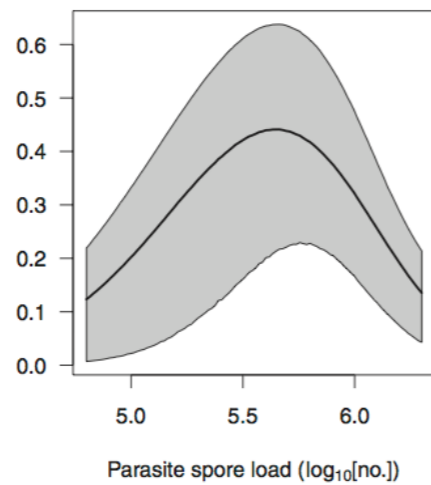
De Roode *et al.* (2008, *PNAS*)   Jensen *et al.* (2006, *PLoS Biol*)
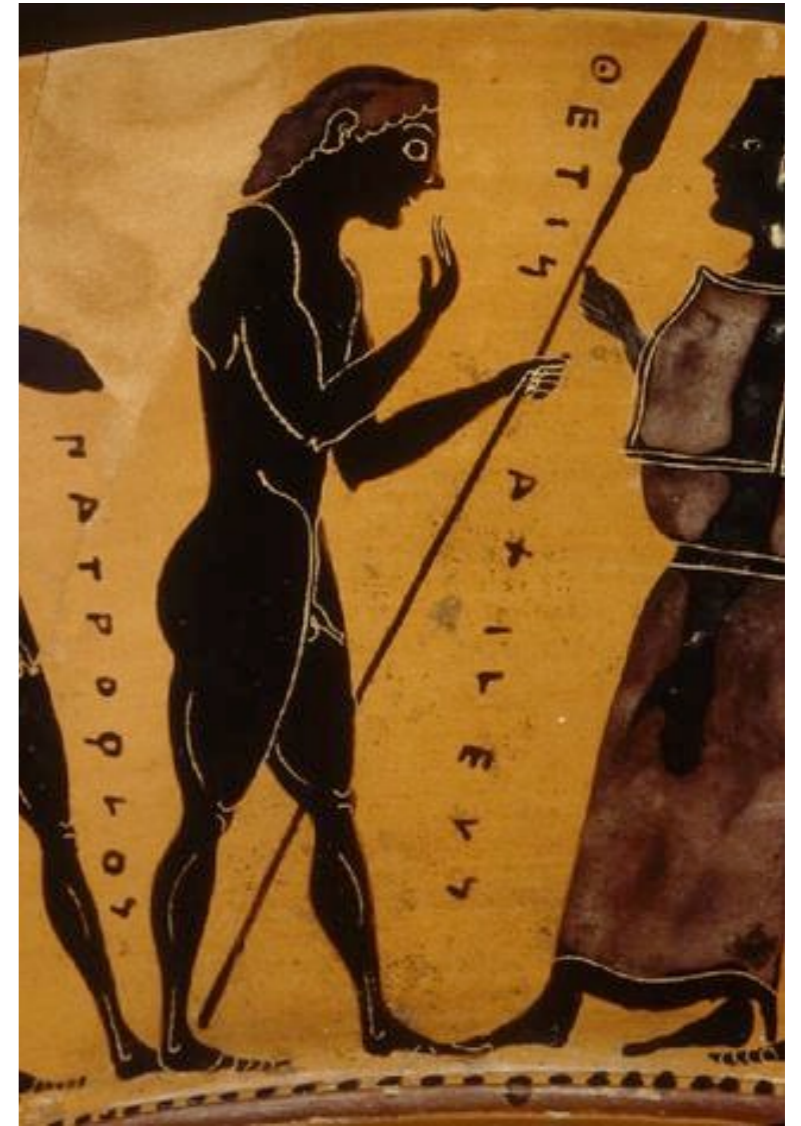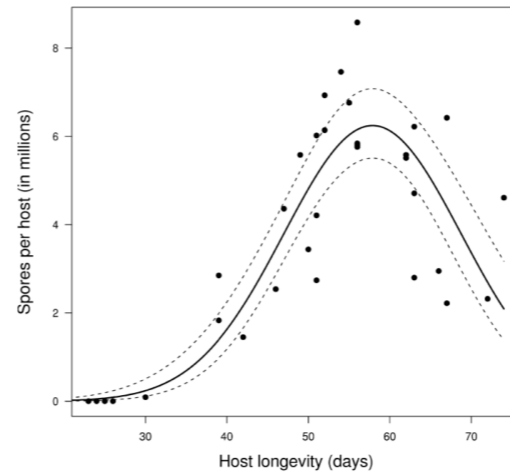
Doumayrou et al. (2013, *Evolution*)

Anderson & May (1982, *Parasitology*)
Ewald (1994)
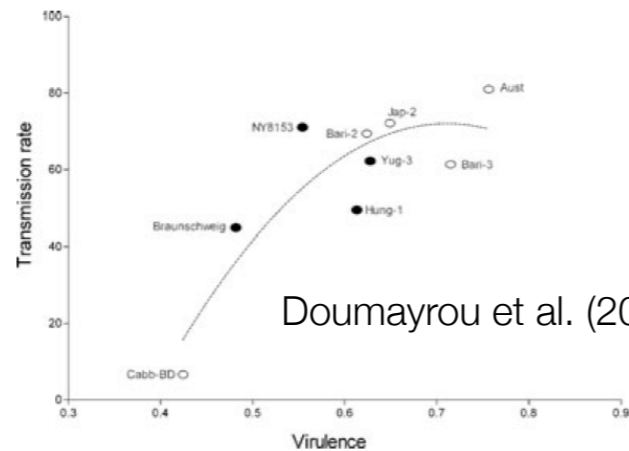
HIV virulence evolution (Switzerland)
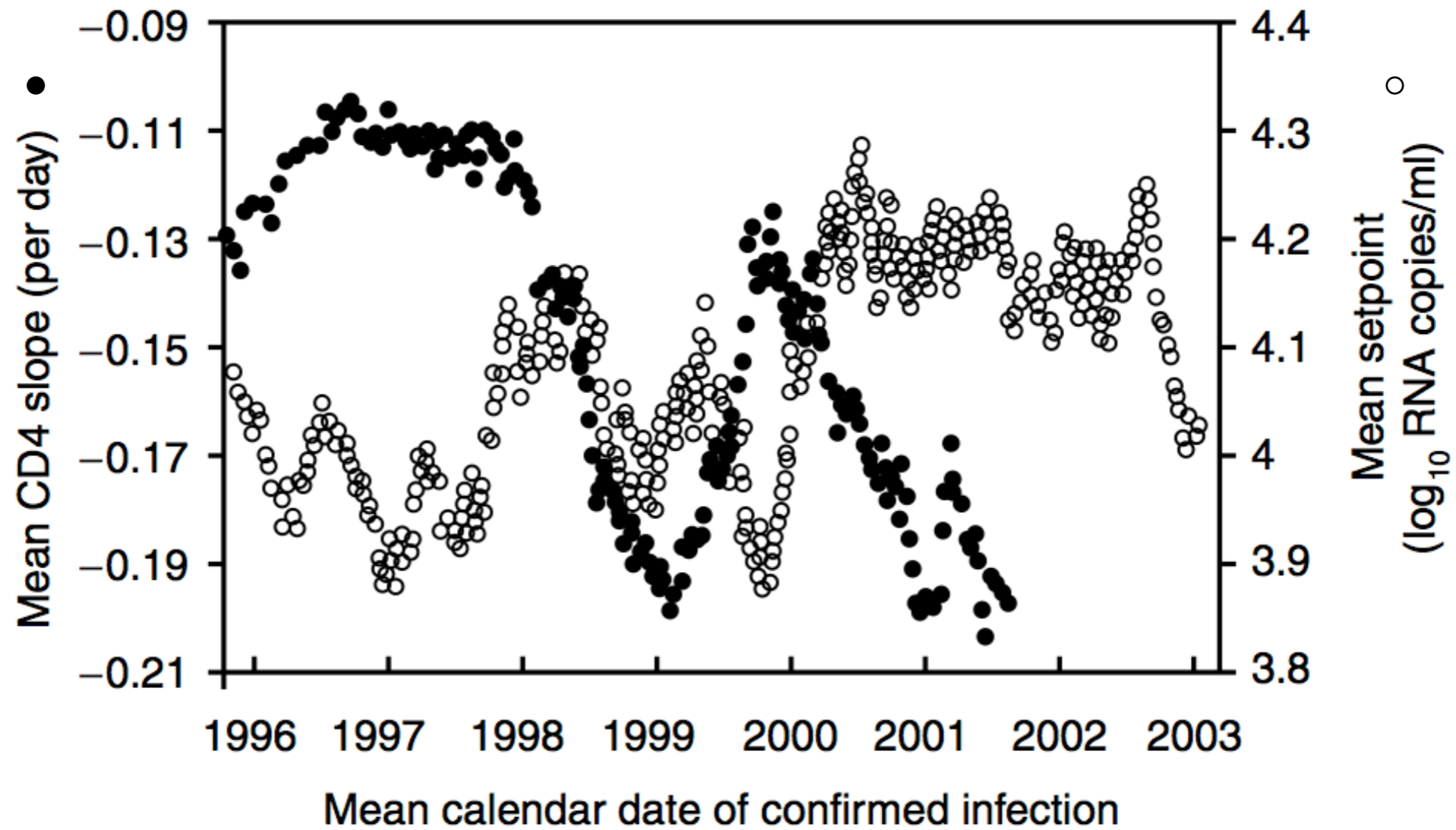
Müller *et al.* (2006, *AIDS*)

HIV virulence evolution (USA)
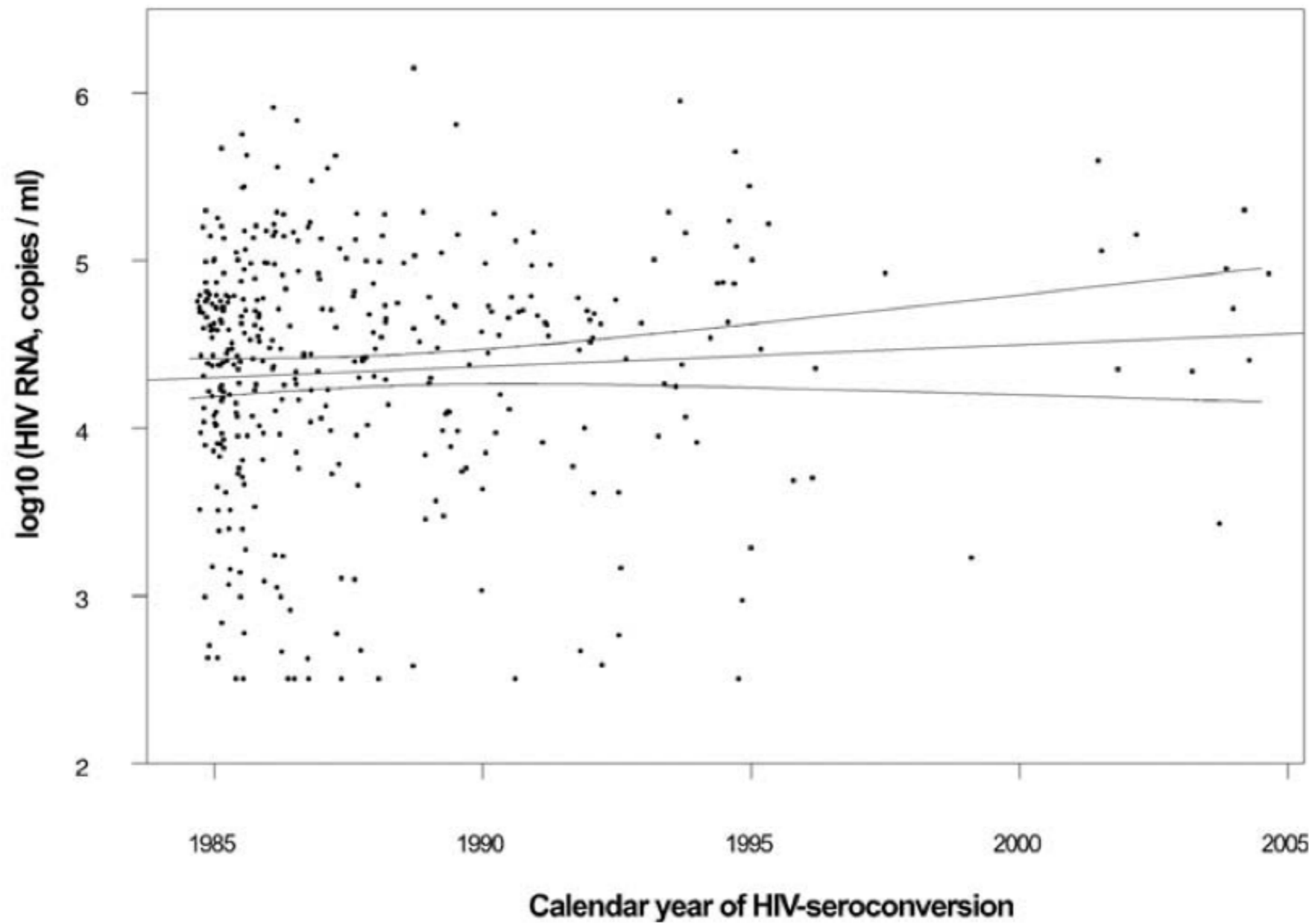
Herbeck *et al.* (2008, *PLoS ONE*)
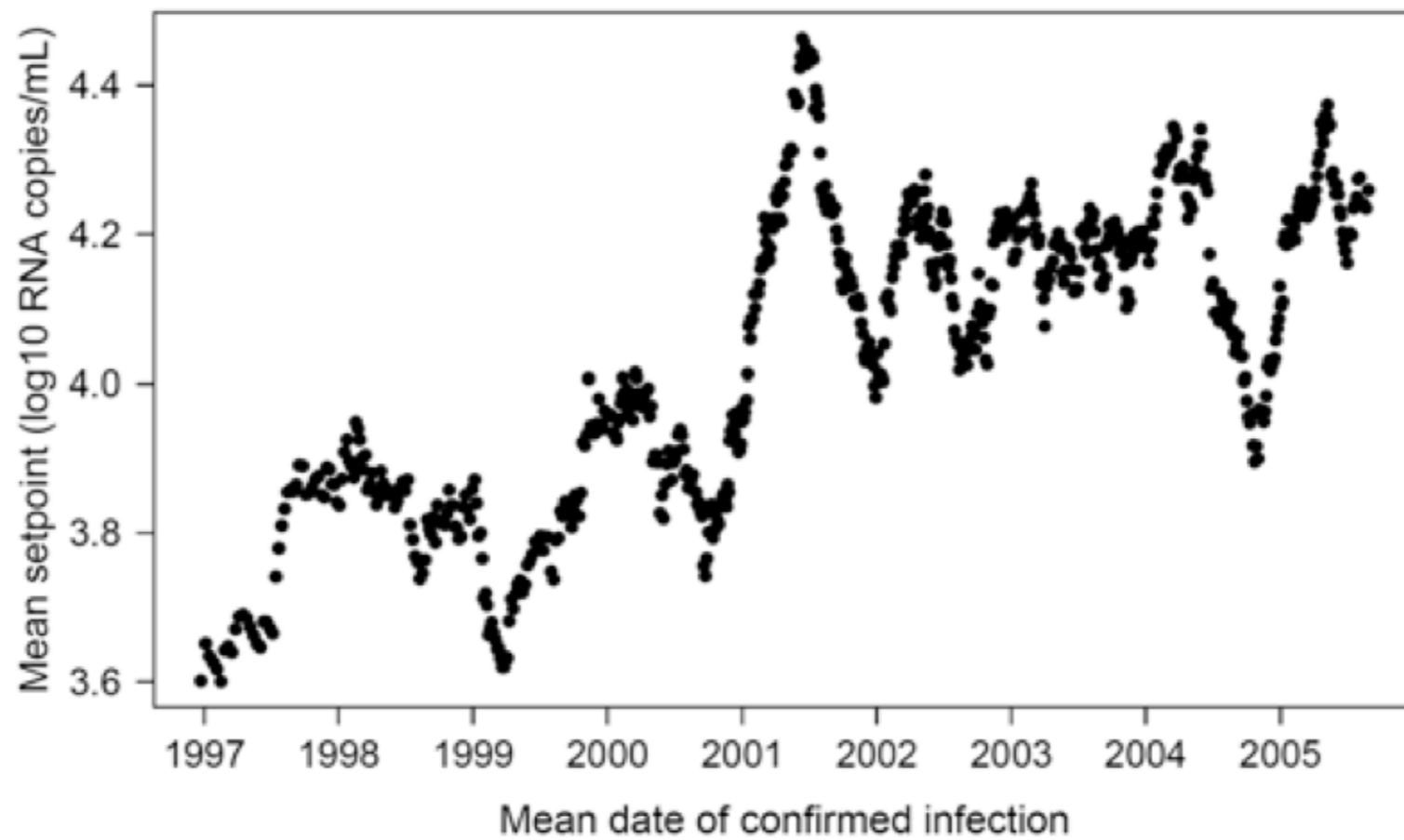
HIV virulence evolution (Italy)

Müller *et al.* (2009, *PLoS Path*)

HIV virulence evolution (meta-analysis)

Herbeck et al. (2012, *AIDS*)

HIV virulence evolution (meta-analysis)

Summary CD4⁺ T-cell count trends by year

Slope = 0.044
P = 1.5e–04

Summary plasma viral load trends by year

Slope = −0.00037
P = 2.4e–04

spVL has increased but is now stabilising
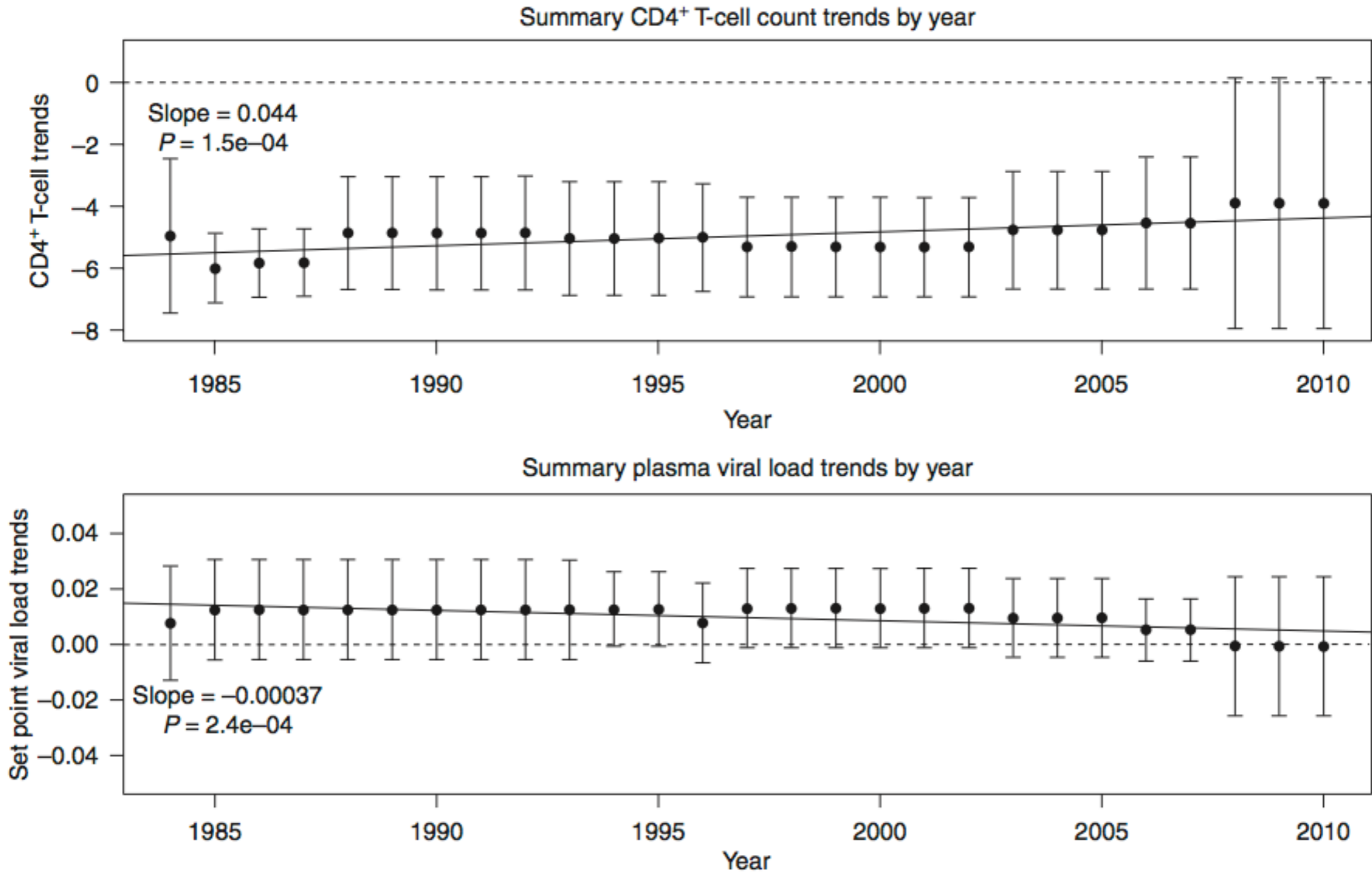
Herbeck et al. (2012, *AIDS*)
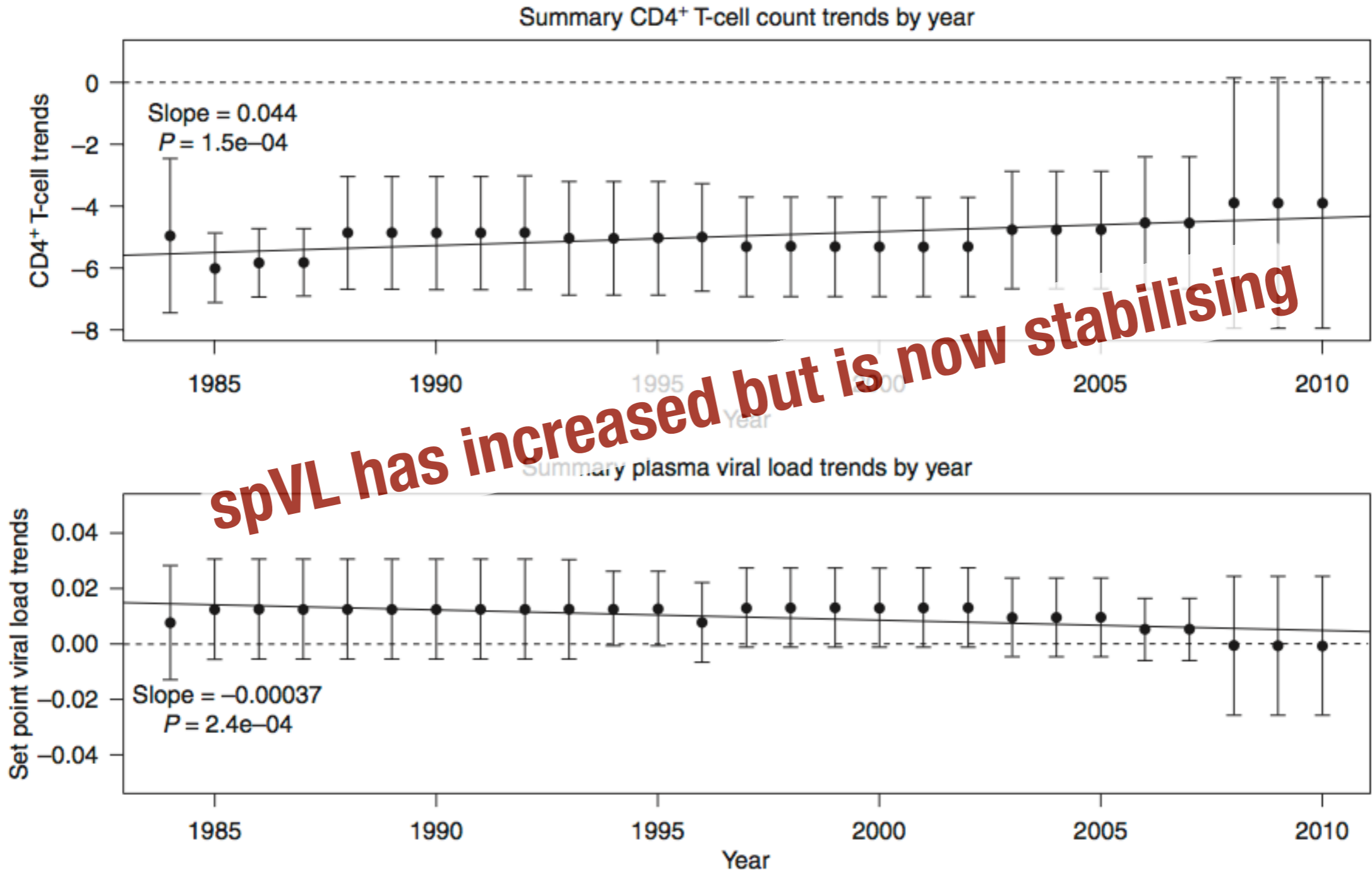
Evolution towards the ESS?
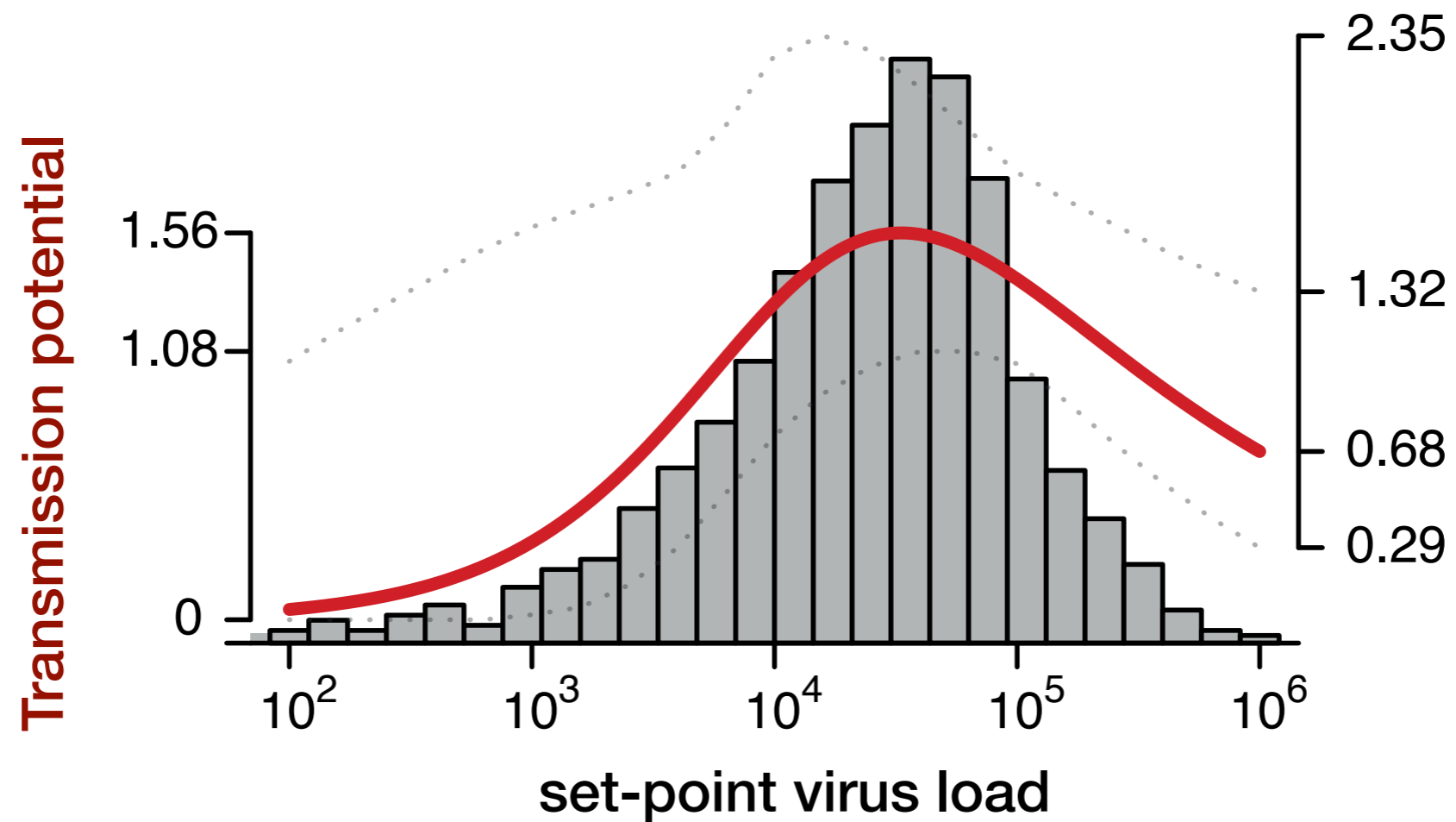
HIV virulence evolution (The Netherlands)

Gras *et al.* (2009, *PLoS ONE*)
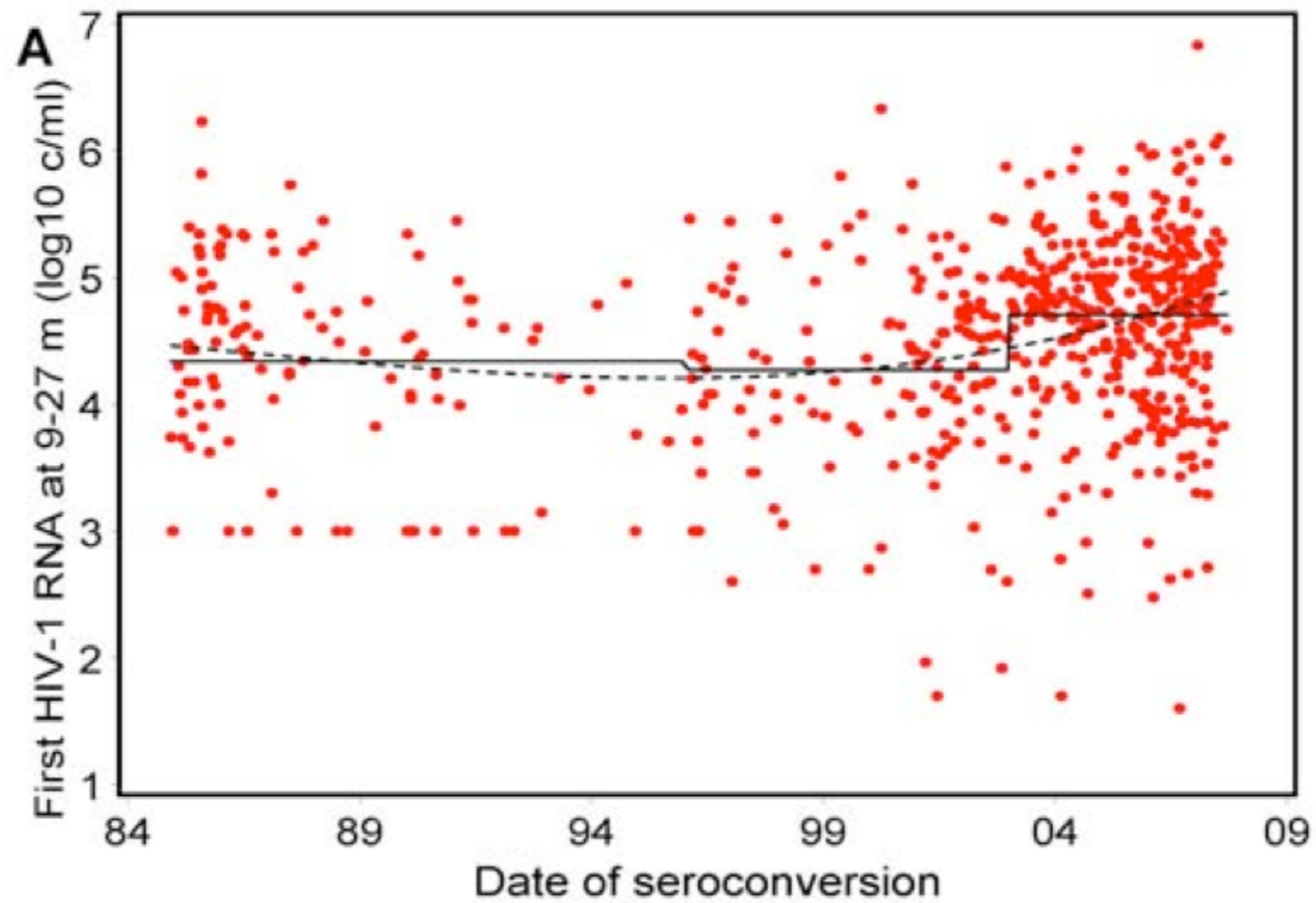
HIV virulence evolution and HAART
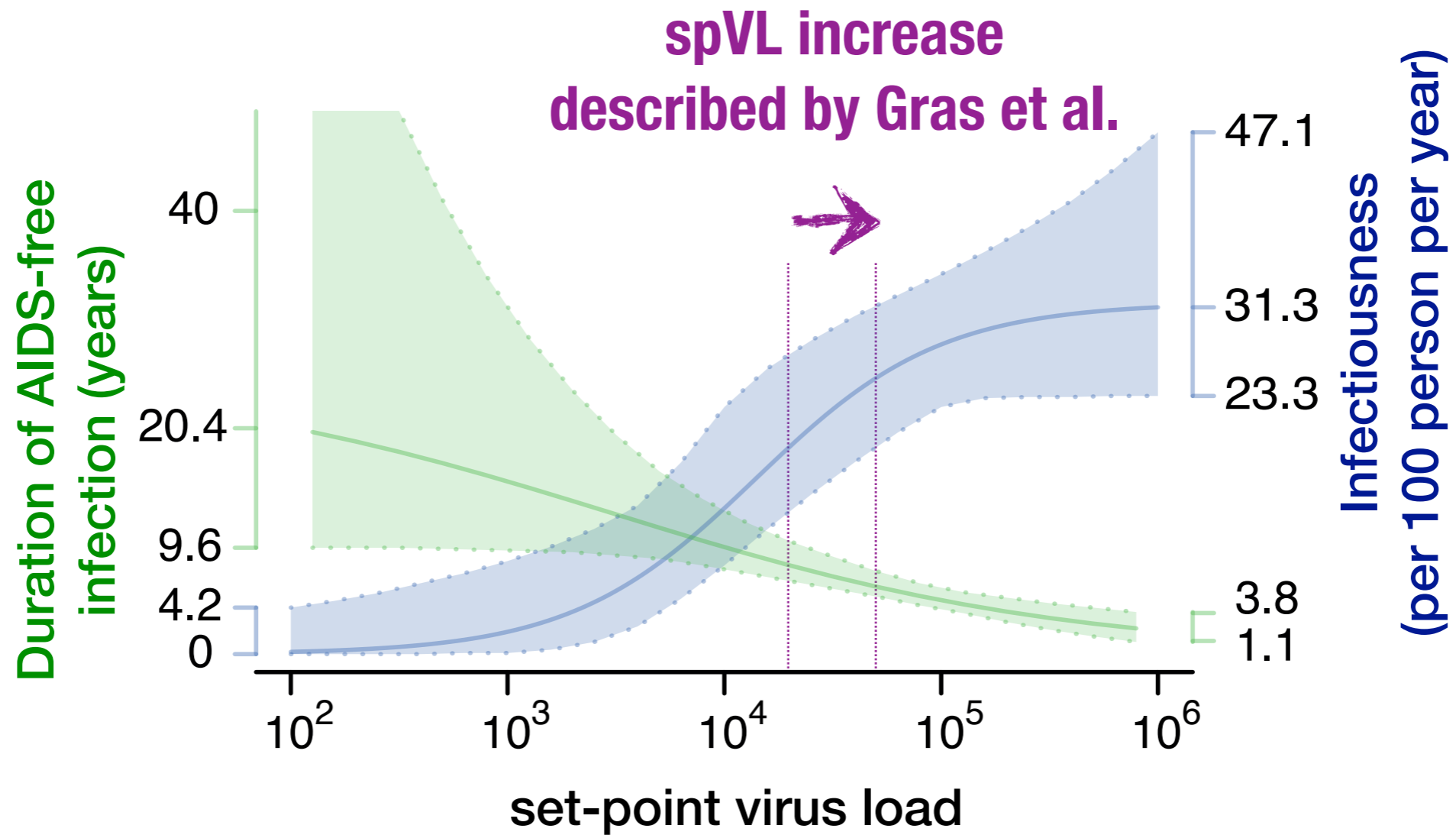
Évolution et traitements
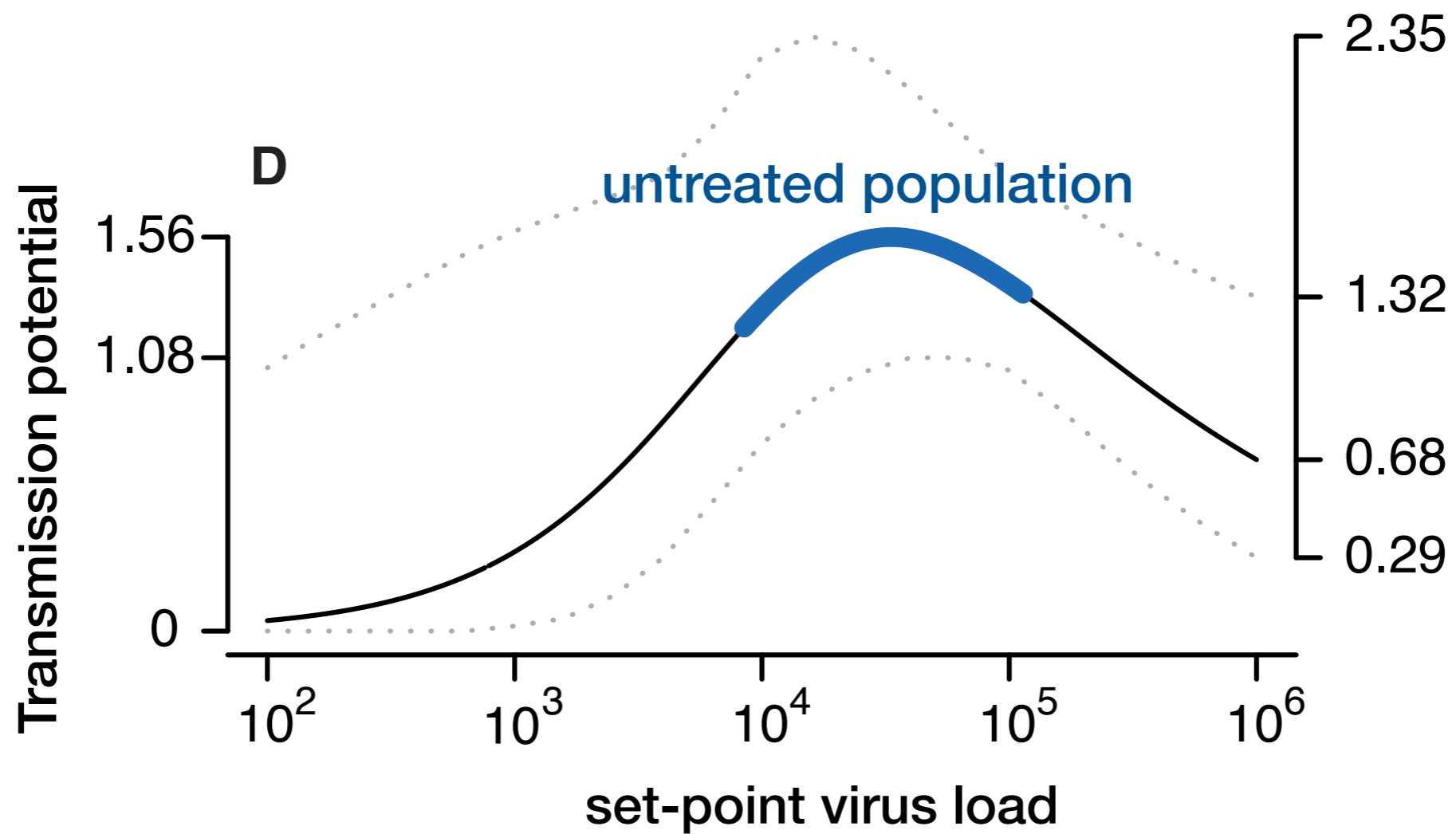
Évolution et traitements
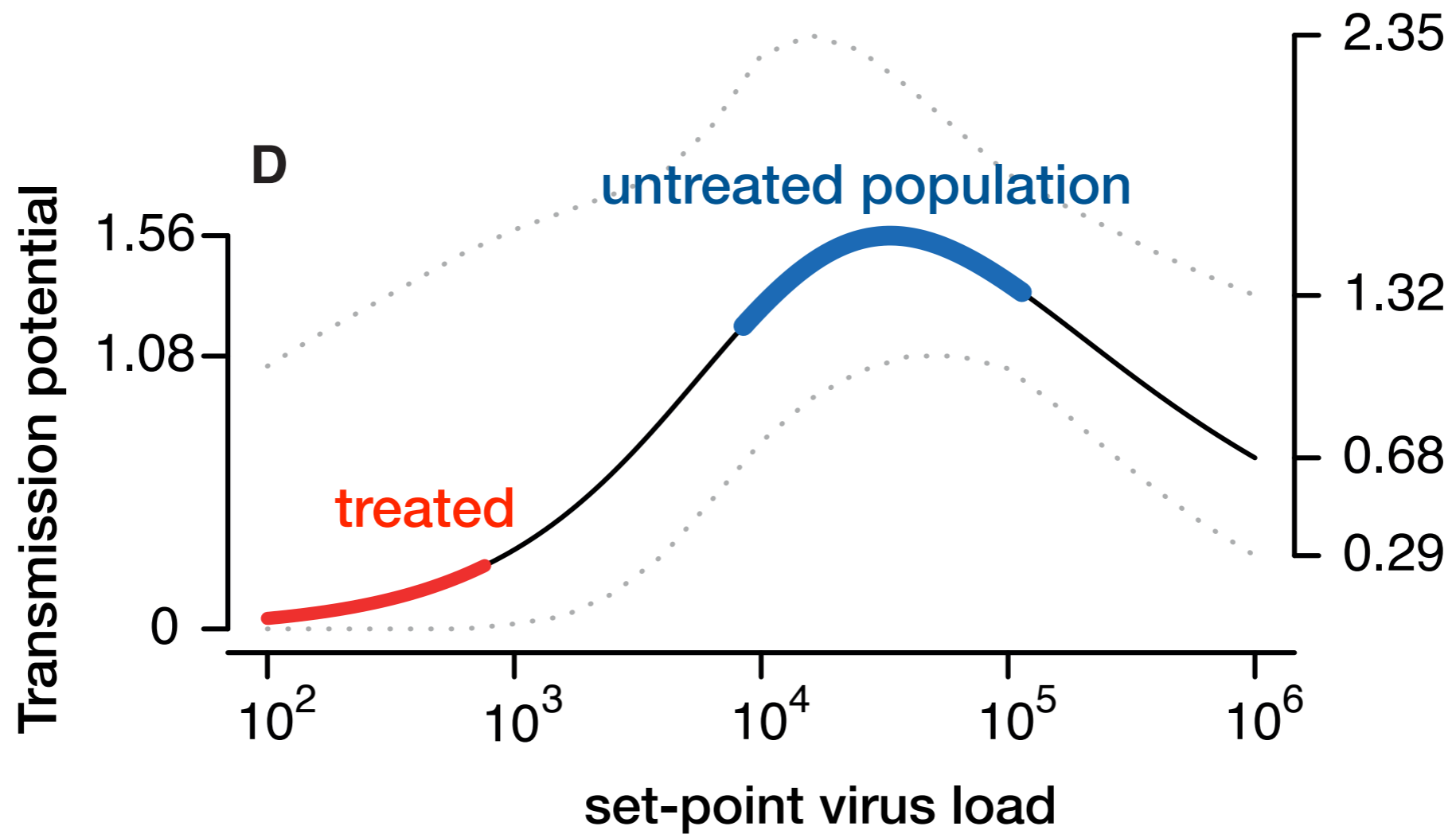
Évolution et traitements
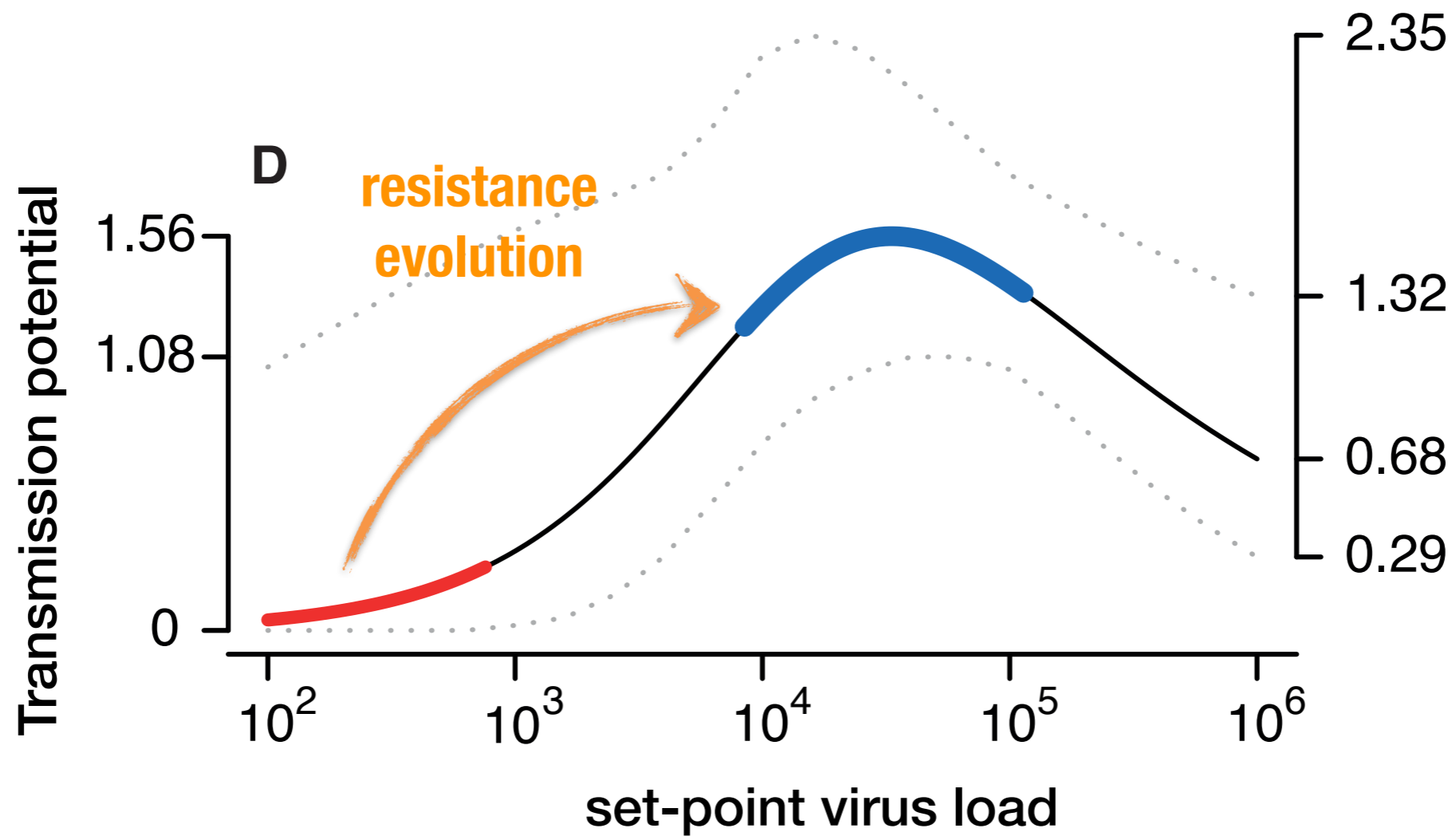
Eugene Geidelberg's MSc

Évolution et traitements

D

virus load
evolution

Transmission potential

1.56

1.08

0

10^2 10^3 10^4 10^5 10^6

set-point virus load

2.35

1.32

0.68

0.29

see also van Baalen (1998, *Proc B*)
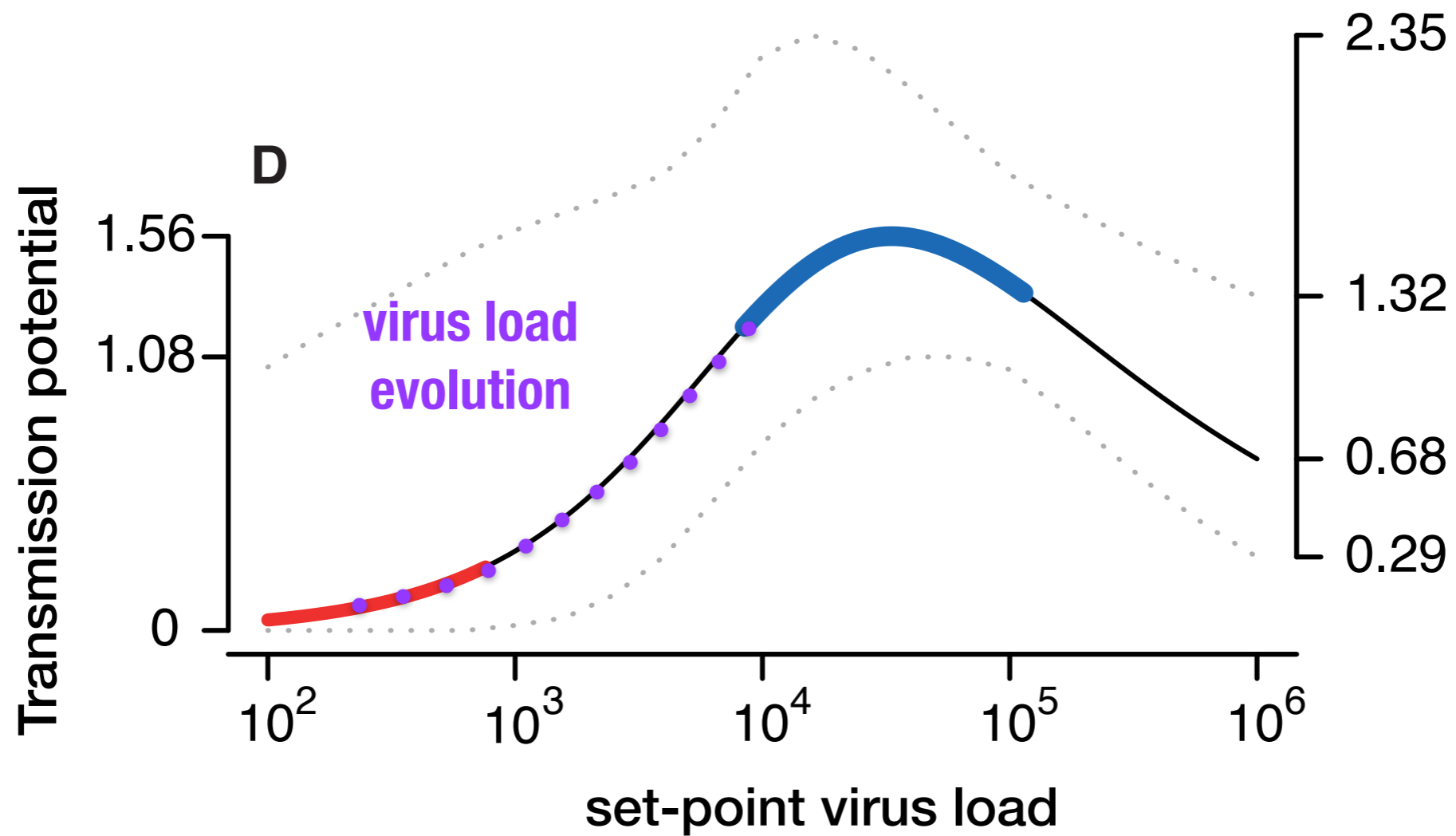
Thanks!

ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Imperial College
London

MEME

SWISS
HIV
COHORT
STUDY

Christophe Fraser
(& George Shirreff)

Tsukushi Kamiya
(& Nicole Mideo)

Sebastian Bonhoeffer
(& Tanja Stadler,
Roger Kouyos, …)

cnrs
INSB
institut des sciences biologiques

Inserm

ATIP avenir
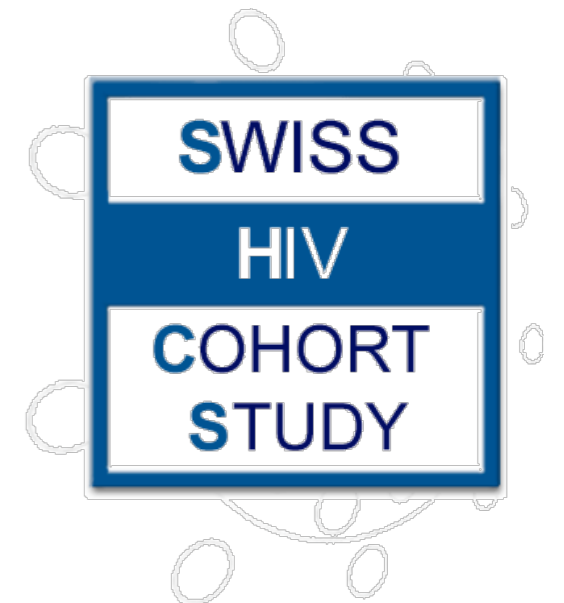
In summary…

- HIV virulence is partly '**heritable**' from one infection to the next

- 3 explanatory hypotheses: the WH fitness landscape is rugged, virulent strains are less competitive, within-host evolution is a 'dead-end'

- **Immunosuppression** can favor more virulent strains

- HIV-1 virulence seems **adaptive** for the virus (increased transmission)

- Is the evolution towards the ESS affected by **public health policies**?

MIVEGEC — UMR CNRS 5290 — Montpellier, France
samuel.alizon@cnrs.fr

Annexes

Coinfection, immunosuppression and virulence

ESS immunosuppression

ESS virulence

Kamiya, Mideo, Alizon (*in prep*)

All three together...

λ (Pagel 1994 *Proc. B*, Freckleton et al. 2002 *Am. Nat.*)

λ (Pagel 1994 *Proc. B*, Freckleton et al. 2002 *Am. Nat.*)

	A	B
species 1	2.51	0.38
species 2	2.80	0.89
species 3	2.23	0.88
species 4	2.32	1.60
species 5	1.64	0.33
species 6	1.18	1.42
species 7	0.58	0.31

λ (Pagel 1994 *Proc. B*, Freckleton et al. 2002 *Am. Nat.*)

V

	species 1	species 2	species 3	species 4	species 5	species 6	species 7
species 1	1.44	1.26	0.20	0.20	0	0	0
species 2	1.26	1.44	0.20	0.20	0	0	0
species 3	0.20	0.20	1.44	1.42	0	0	0
species 4	0.20	0.20	1.44	1.44	0	0	0
species 5	0	0	0	0	1.44	0.07	0.07
species 6	0	0	0	0	0.07	1.44	1.21
species 7	0	0	0	0	0.07	1.21	1.44

	A	B
species 1	2.51	0.38
species 2	2.80	0.89
species 3	2.23	0.88
species 4	2.32	1.60
species 5	1.64	0.33
species 6	1.18	1.42
species 7	0.58	0.31

Tree branch lengths: 0.18, 1.06, 0.18, 0.20, 0.02, 1.22, 0.02, 1.37, 0.07, 0.23, 1.14, 0.23

λ (Pagel 1994 *Proc. B*, Freckleton et al. 2002 *Am. Nat.*)

$\times \lambda$

V

	species 1	species 2	species 3	species 4	species 5	species 6	species 7
species 1	1.44	1.26	0.20	0.20	0	0	0
species 2	1.26	1.44	0.20	0.20	0	0	0
species 3	0.20	0.20	1.44	1.42	0	0	0
species 4	0.20	0.20	1.44	1.44	0	0	0
species 5	0	0	0	0	1.44	0.07	0.07
species 6	0	0	0	0	0.07	1.44	1.21
species 7	0	0	0	0	0.07	1.21	1.44

		A	**B**
0.18	species 1	2.51	0.38
0.18	species 2	2.80	0.89
0.02	species 3	2.23	0.88
0.02	species 4	2.32	1.60
1.37	species 5	1.64	0.33
0.23	species 6	1.18	1.42
0.23	species 7	0.58	0.31

1.06
0.20
1.22
0.07
1.14

λ (Pagel 1994 *Proc. B*, Freckleton et al. 2002 *Am. Nat.*)

- λ is a weight on the off-diagonal terms of $\mathbf{V}$ (distances between taxa)

- $\lambda=0$: evolution independent of the phylogeny

- $\lambda=1$: Brownian motion prediction on the phylogeny

$\times\lambda$

$\mathbf{V}$

	species 1	species 2	species 3	species 4	species 5	species 6	species 7
species 1	1.44	1.26	0.20	0.20	0	0	0
species 2	1.26	1.44	0.20	0.20	0	0	0
species 3	0.20	0.20	1.44	1.42	0	0	0
species 4	0.20	0.20	1.44	1.44	0	0	0
species 5	0	0	0	0	1.44	0.07	0.07
species 6	0	0	0	0	0.07	1.44	1.21
species 7	0	0	0	0	0.07	1.21	1.44

	A	B
species 1	2.51	0.38
species 2	2.80	0.89
species 3	2.23	0.88
species 4	2.32	1.60
species 5	1.64	0.33
species 6	1.18	1.42
species 7	0.58	0.31

Tree branch lengths: 0.18, 1.06, 0.18, 0.20, 0.02, 1.22, 0.02, 1.37, 0.07, 0.23, 1.14, 0.23

λ (Pagel 1994 *Proc. B*, Freckleton et al. 2002 *Am. Nat.*)

- λ is a weight on the off-diagonal terms of **V** (distances between taxa)

- $\lambda=0$: evolution independent of the phylogeny

- $\lambda=1$: Brownian motion prediction on the phylogeny

$\times\lambda$

V

	species 1	species 2	species 3	species 4	species 5	species 6	species 7
species 1	1.44	1.26	0.20	0.20	0	0	0
species 2	1.26	1.44	0.20	0.20	0	0	0
species 3	0.20	0.20	1.44	1.42	0	0	0
species 4	0.20	0.20	1.44	1.44	0	0	0
species 5	0	0	0	0	1.44	0.07	0.07
species 6	0	0	0	0	0.07	1.44	1.21
species 7	0	0	0	0	0.07	1.21	1.44

		A	B
0.18	species 1	2.51	0.38
1.06			
0.18	species 2	2.80	0.89
0.20			
0.02	species 3	2.23	0.88
1.22			
0.02	species 4	2.32	1.60
1.37	species 5	1.64	0.33
0.07			
0.23	species 6	1.18	1.42
1.14			
0.23	species 7	0.58	0.31

(c)

Trait B

Trait A

Log-likelihood difference from ML

λ (Freckleton et al. 2002, Am. Nat.)

- Trait evolution from an ancestral state (α) under Brownian model:

$$y_i = \alpha + \sum_{j=1}^{T_i} \epsilon_{i,j} t_{i,j}$$

- For n species undergoing independent Brownian motion, $\mathbf{y}$ has a multinomial proba density:

$$p(\mathbf{y}) = \frac{1}{(2\pi\sigma^2 t)^{n/2}} \exp\left[-\frac{(\mathbf{y} - \alpha\mathbf{X})^T (\mathbf{y} - \alpha\mathbf{X})}{2\sigma^2 t}\right]$$

- However, species evolution is not independent. If species share a common ancestor at time t_a:

$$\mathrm{cov}(y_i, y_j) = \sigma^2 t_a$$

λ

- Trait evolution from an ancestral state (α) under Brownian model:

$$y_i = \alpha + \sum_{j=1}^{T_i} \epsilon_{i,j} t_{i,j}$$

- For n species undergoing independent Brownian motion, $\mathbf{y}$ has a multinomial proba density:

$$p(\mathbf{y}) = \frac{1}{(2\pi\sigma^2 t)^{n/2}} \exp\left[-\frac{(\mathbf{y} - \alpha\mathbf{X})^T (\mathbf{y} - \alpha\mathbf{X})}{2\sigma^2 t}\right]$$

- However, species evolution is not independent. If species share a common ancestor at time t_a:

$$\mathrm{cov}(y_i, y_j) = \sigma^2 t_a$$

- If V is the $n \times n$ variance-covariance matrix that describes the phylogeny

$$p(\mathbf{y}) = \frac{1}{(2\pi\sigma^2 t)^{n/2}} |\mathbf{V}|^{-0.5} \exp\left[-\frac{1}{2\sigma^2} (\mathbf{y} - \mathbf{X}\alpha)^T \mathbf{V}^{-1} (\mathbf{y} - \mathbf{X}\alpha)\right]$$

λ

- Trait evolution from an ancestral state (α) under Brownian model:

$$y_i = \alpha + \sum_{j=1}^{T_i} \epsilon_{i,j} t_{i,j}$$

- For n species undergoing independent Brownian motion, $\mathbf{y}$ has a multinomial proba density:

$$p(\mathbf{y}) = \frac{1}{(2\pi\sigma^2 t)^{n/2}} \exp\left[-\frac{(\mathbf{y} - \alpha\mathbf{X})^T (\mathbf{y} - \alpha\mathbf{X})}{2\sigma^2 t}\right]$$

- However, species evolution is not independent. If species share a common ancestor at time t_a:

$$\mathrm{cov}(y_i, y_j) = \sigma^2 t_a$$

- If V is the $n \times n$ variance-covariance matrix that describes the phylogeny

$$p(\mathbf{y}) = \frac{1}{(2\pi\sigma^2 t)^{n/2}} |\mathbf{V}|^{-0.5} \exp\left[-\frac{1}{2\sigma^2} (\mathbf{y} - \mathbf{X}\alpha)^T \mathbf{V}^{-1} (\mathbf{y} - \mathbf{X}\alpha)\right]$$

K (Blomberg et al. 2003, Evolution)

- K is based on the MSE (or on the variance in IC)

- The amount of phylogenetic signal is given by the ratio between MSE from data points (MSE_0) and MSE derived from the phylogeny (MSE)

- The ratio is a function of $\mathbf{V}$, the variance-covariance matrix

$$\text{MSE}_0 = \frac{(\mathbf{X} - \hat{a})^T (\mathbf{X} - \hat{a})}{n - 1}$$

$$\mathbf{U} = \mathbf{D}\mathbf{X}$$

$$\text{MSE} = \frac{(\mathbf{U} - \hat{a})^T (\mathbf{U} - \hat{a})}{n - 1}$$

$$\mathbf{D}\,\mathbf{V}\,\mathbf{D}^T = \mathbf{I}$$

$$\frac{\text{MSE}_0}{\text{MSE}} = \frac{1}{n - 1}\left(\text{tr}(\mathbf{V}) - \frac{n}{\sum\sum \mathbf{V}^{-1}}\right)$$

K (Blomberg et al. 2003, Evolution)

- K is based on the MSE (or on the variance in IC)

$$\mathrm{MSE}_0 = \frac{(\mathbf{X} - \hat{a})^T (\mathbf{X} - \hat{a})}{n - 1}$$

- The amount of phylogenetic signal is given by the ratio between MSE from data points (MSE_0) and MSE derived from the phylogeny (MSE)

$$\mathrm{MSE} = \frac{(\mathbf{U} - \hat{a})^T (\mathbf{U} - \hat{a})}{n - 1}$$

$$\mathbf{U} = \mathbf{D}\mathbf{X}$$

$$\mathbf{D}\,\mathbf{V}\,\mathbf{D}^T = \mathbf{I}$$

- The ratio is a function of $\mathbf{V}$, the variance-covariance matrix

$$\frac{\mathrm{MSE}_0}{\mathrm{MSE}} = \frac{1}{n-1} \left(\mathrm{tr}(\mathbf{V}) - \frac{n}{\sum \sum \mathbf{V}^{-1}} \right)$$

- To correct for tree structure and tree size, the ratio is corrected by the expected value of K:

$$K = \mathrm{observed}\frac{\mathrm{MSE}_0}{\mathrm{MSE}} \bigg/ \mathrm{expected}\frac{\mathrm{MSE}_0}{\mathrm{MSE}}$$

Heritability and phylogenetic signal

$$h^2 = \frac{\text{VAR(G)}}{\text{VAR(P)}}$$
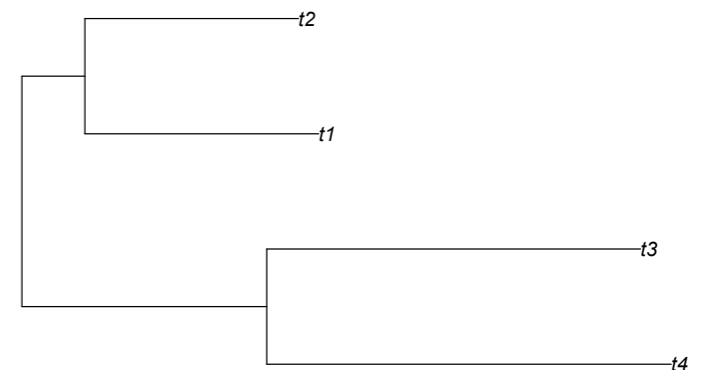
Heritability and phylogenetic signal

- Heritability is the proportion of the genetic variance in the phenotypic variance in a **population**

$$h^2 = \frac{\mathrm{VAR(G)}}{\mathrm{VAR(P)}}$$

Heritability and phylogenetic signal

- Heritability is the proportion of the genetic variance in the phenotypic variance in a **population**
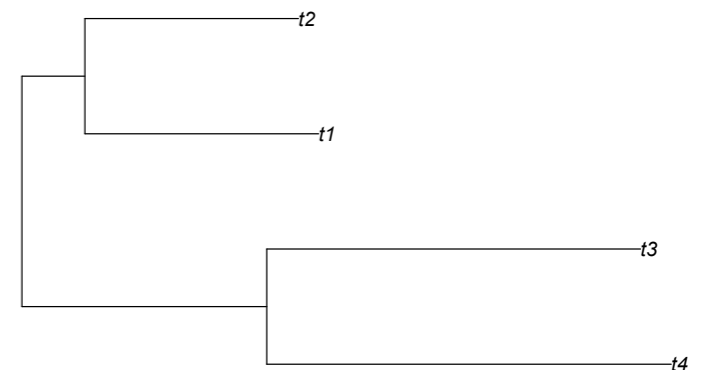
$$h^2 = \frac{\text{VAR(G)}}{\text{VAR(P)}}$$

- Phylogenetic signal is usually defined for **species** (not populations)...

Heritability and phylogenetic signal

- Heritability is the proportion of the genetic variance in the phenotypic variance in a **population**

$$h^2 = \frac{\text{VAR}(\text{G})}{\text{VAR}(\text{P})}$$

- Phylogenetic signal is usually defined for **species** (not populations)...

- ... but phylogenies of infections are built over a population of infected patients

Heritability and phylogenetic signal

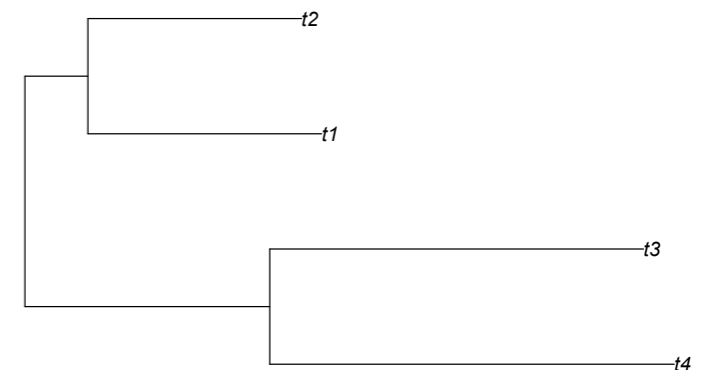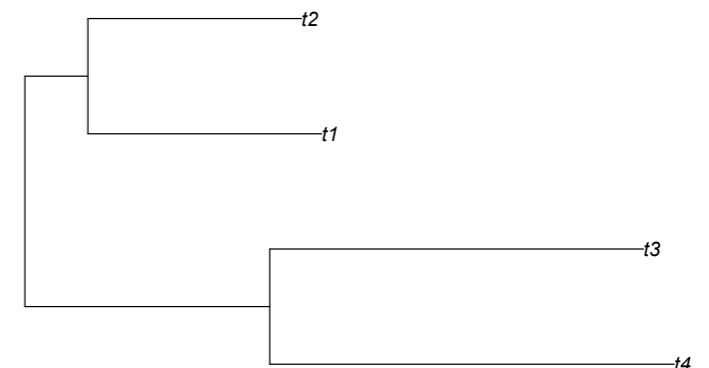- Heritability is the proportion of the genetic variance in the phenotypic variance in a **population**

$$h^2 = \frac{\text{VAR(G)}}{\text{VAR(P)}}$$

- Phylogenetic signal is usually defined for **species** (not populations)...

- ... but phylogenies of infections are built over a population of infected patients
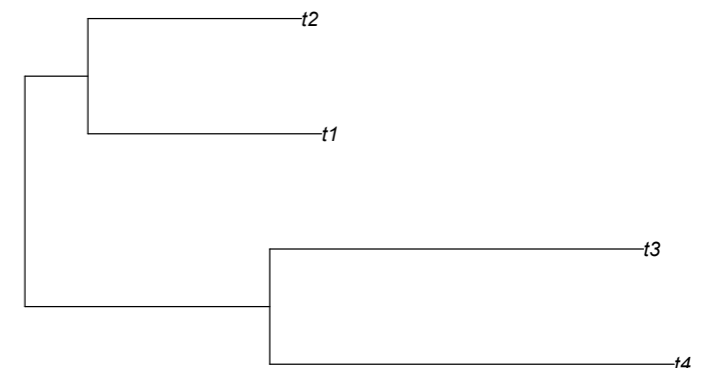
- One can also show the equivalence between the two using Lynch's phylogenetic mixed model (Housworth et al. 2004, Am. Nat.)

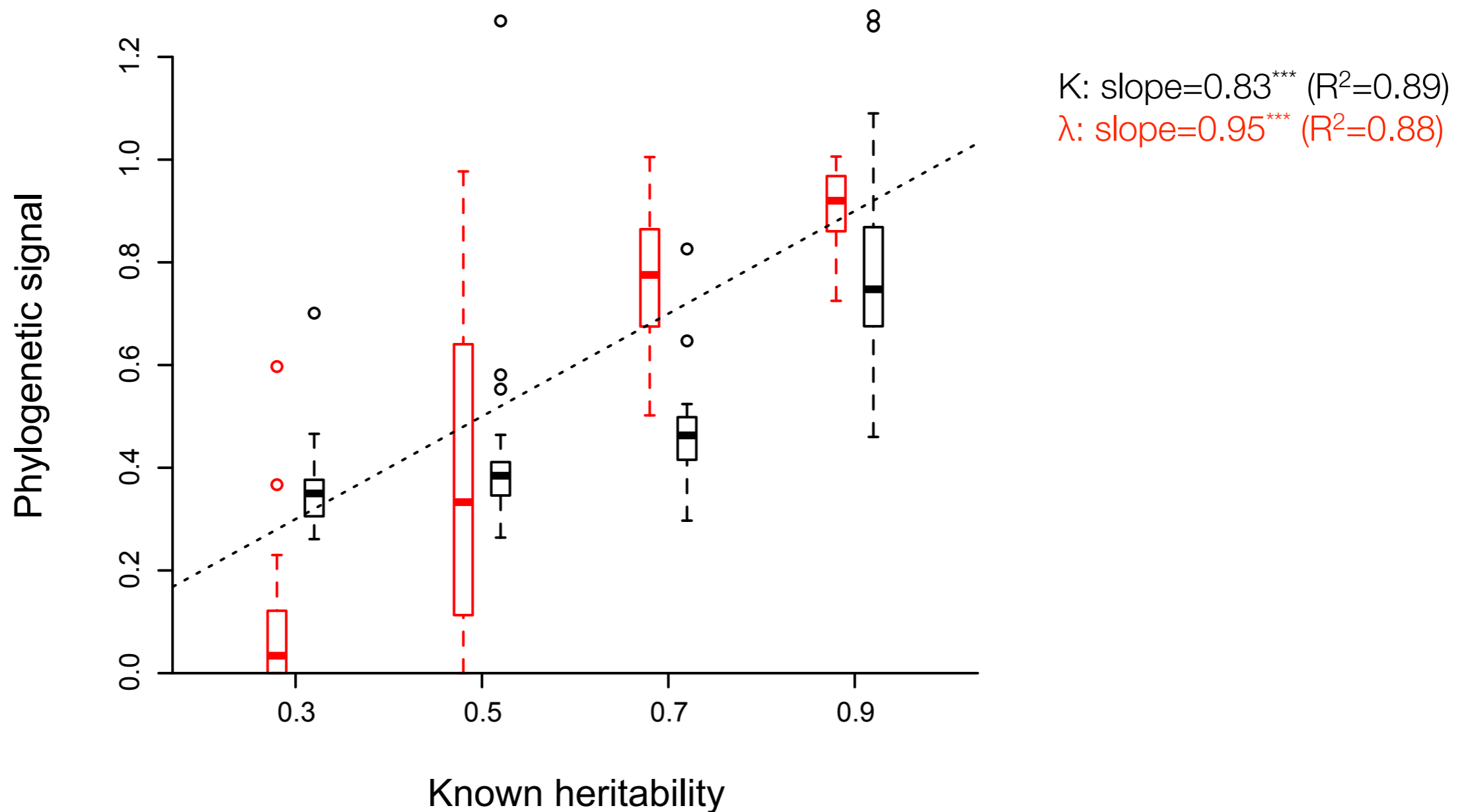# Heritability and phylogenetic signal

- Create phylogenies on which traits evolve with a heritability h^2

$$x_n = h^2\, x_{n-1} + (1 - h^2)\, y$$

Heritability and phylogenetic signal

- Create phylogenies on which traits evolve with a heritability h^2

$$x_n = h^2 \, x_{n-1} + (1 - h^2) \, y$$

K: slope=0.83*** (R^2=0.89)
λ: slope=0.95*** (R^2=0.88)

Phylogenetic signal estimators

- K (Blomberg et al. 2003, Evolution)

  - combines independent contrasts and a randomisation test (p-value)

  - robust

  - limited for large trees

  - estimated on a ML tree

- Pagel's λ (Freckleton et al. 2002, Am. Nat)

  - based on a maximum likelihood approach

  - sensitive to small variations in the phylogeny

  - estimated on a set of trees obtained from a Bayesian inference

Other estimators…
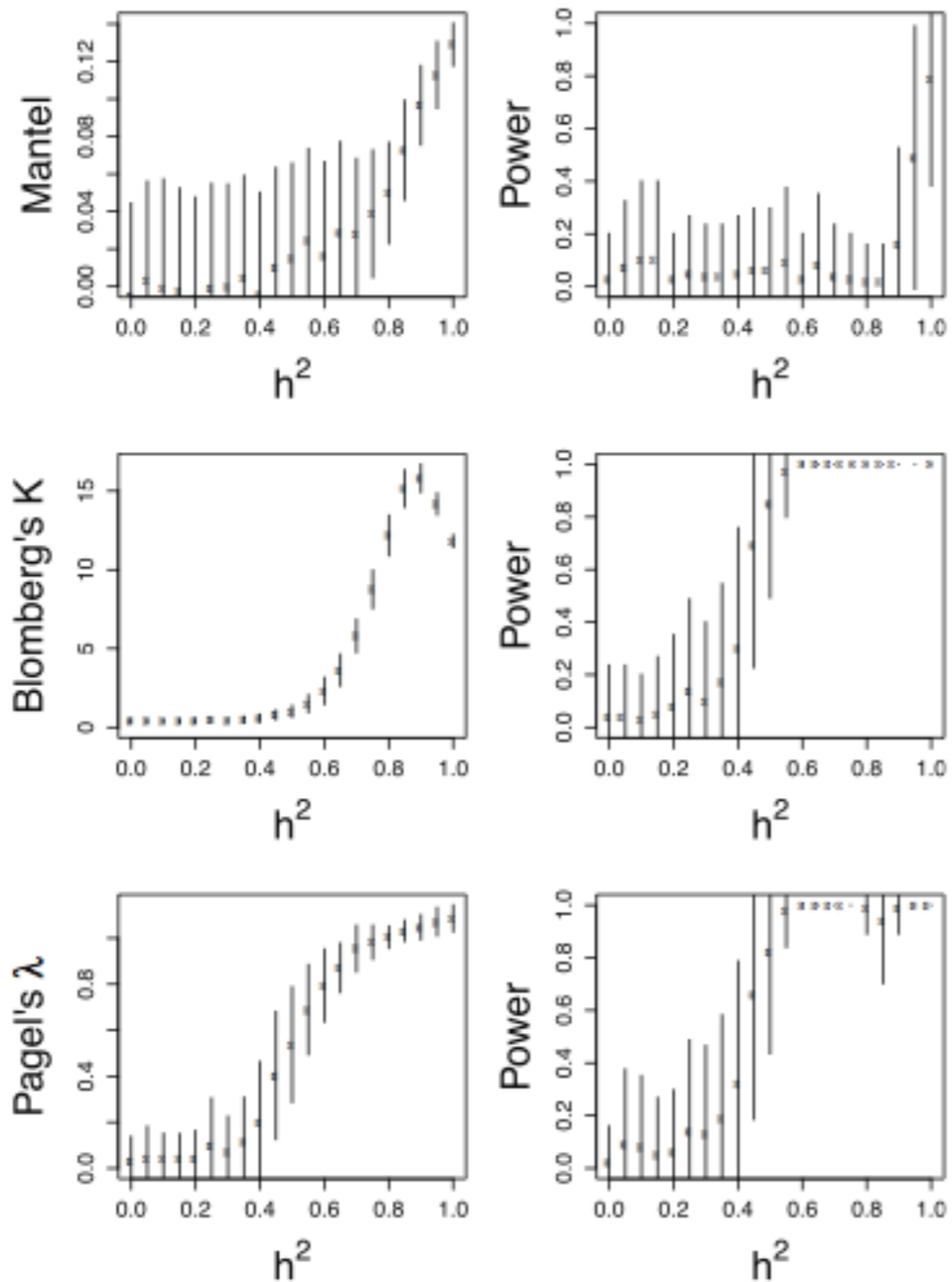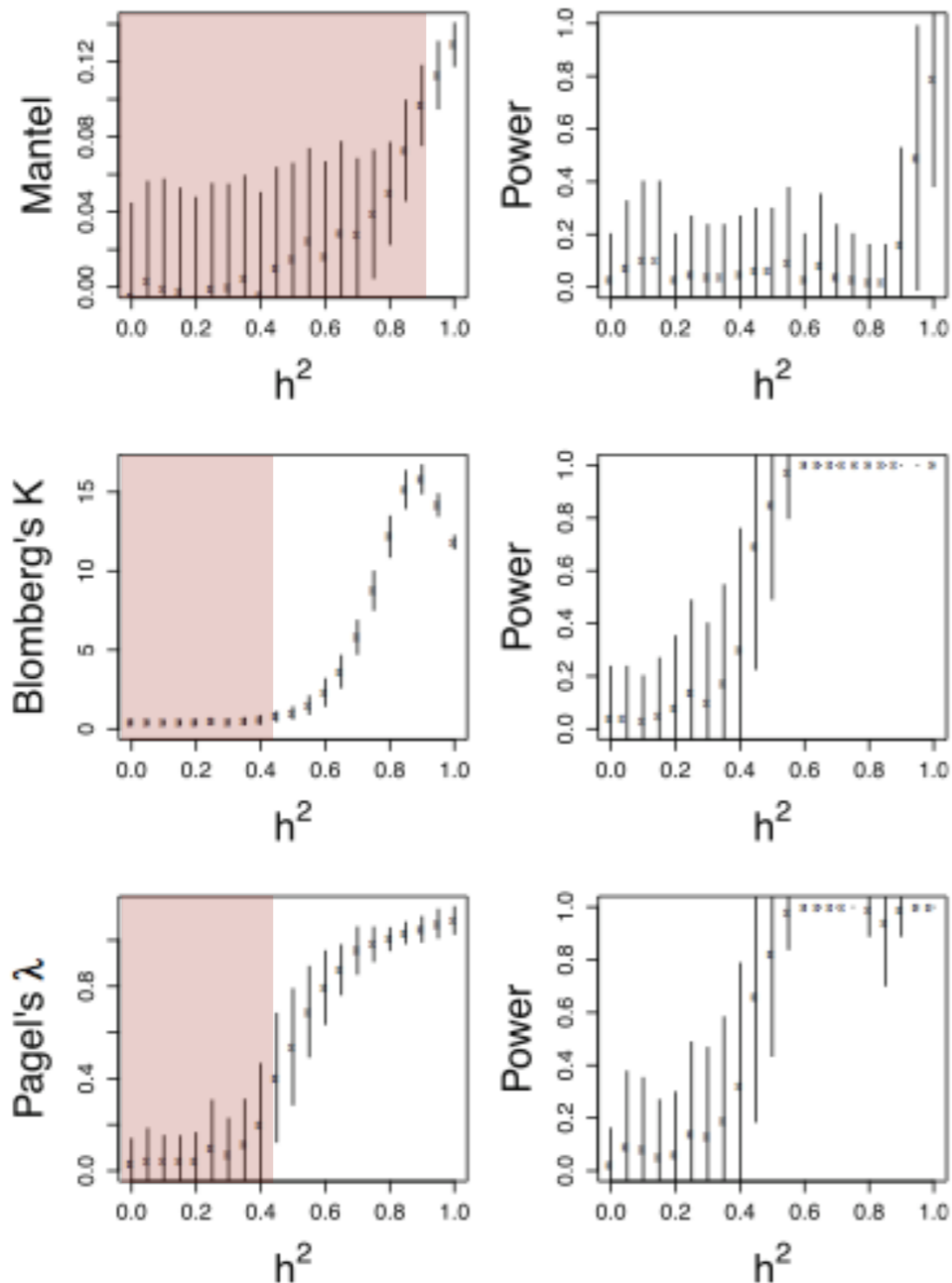

The method can only detect high levels of heritability (at least ~40%)

Shirreff et al. (2013, *EMPH*)

Other estimators…

The method can only detect high levels of heritability (at least ~40%)

Shirreff et al. (2013, *EMPH*)
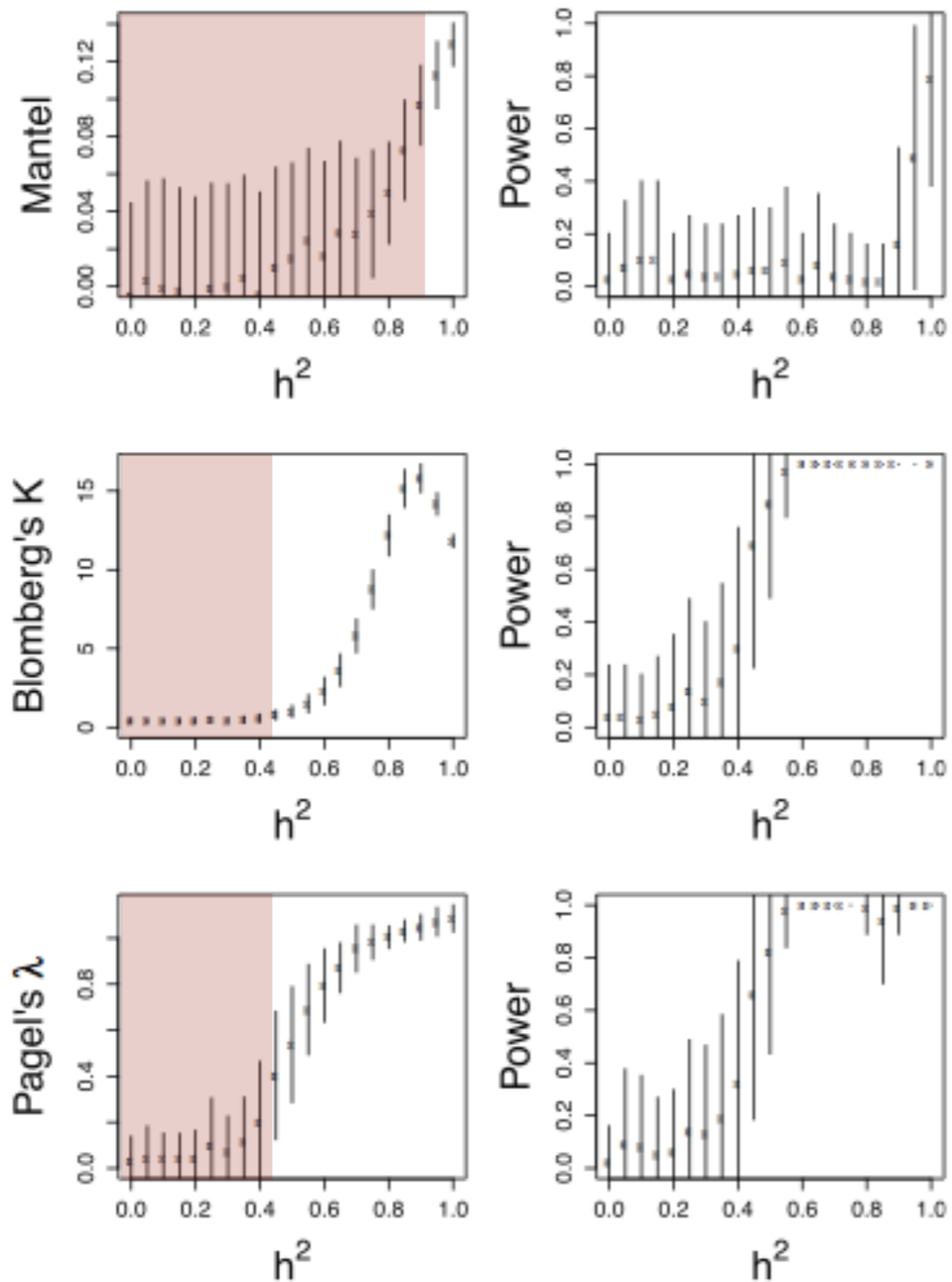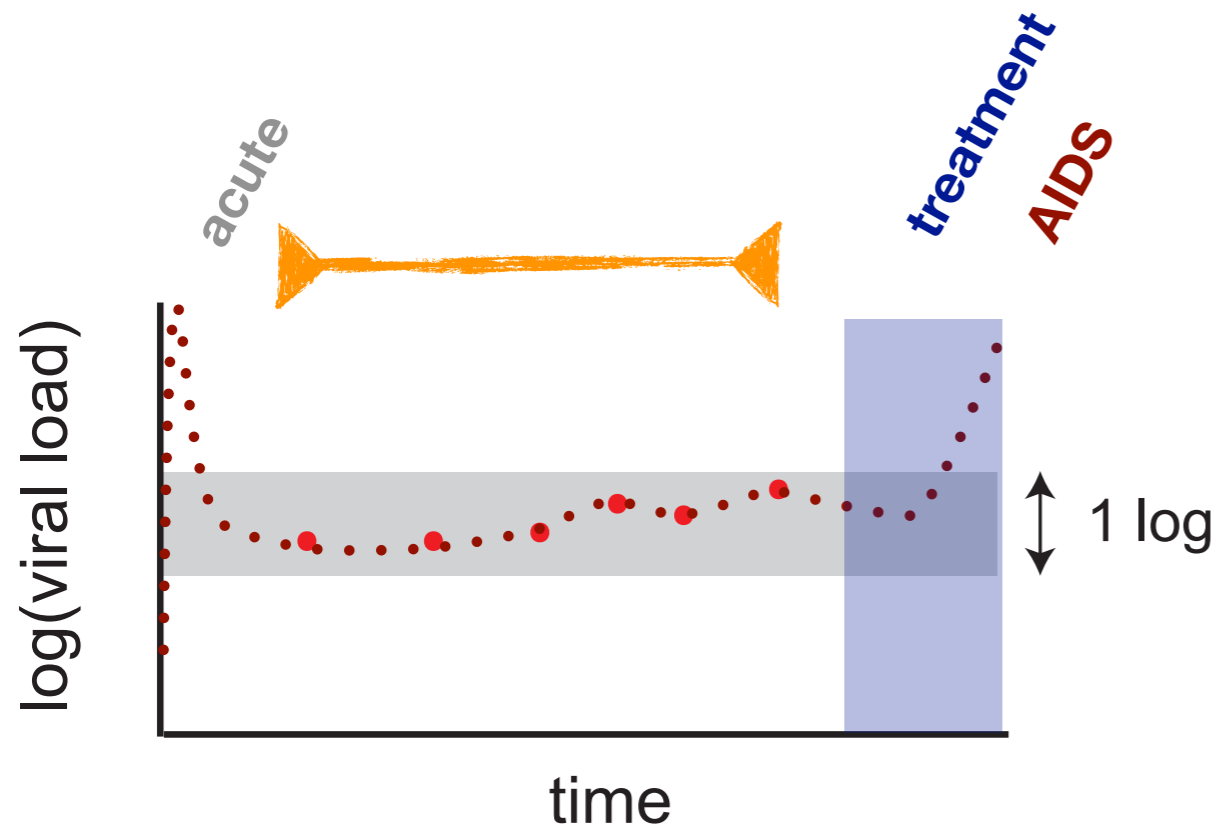
Other estimators…

The method can only detect high levels of heritability (at least ~40%)

Not done for PMM (Housworth et al. 2004) and for d (Blomberg et al. 2003).

Shirreff et al. (2013, *EMPH*)

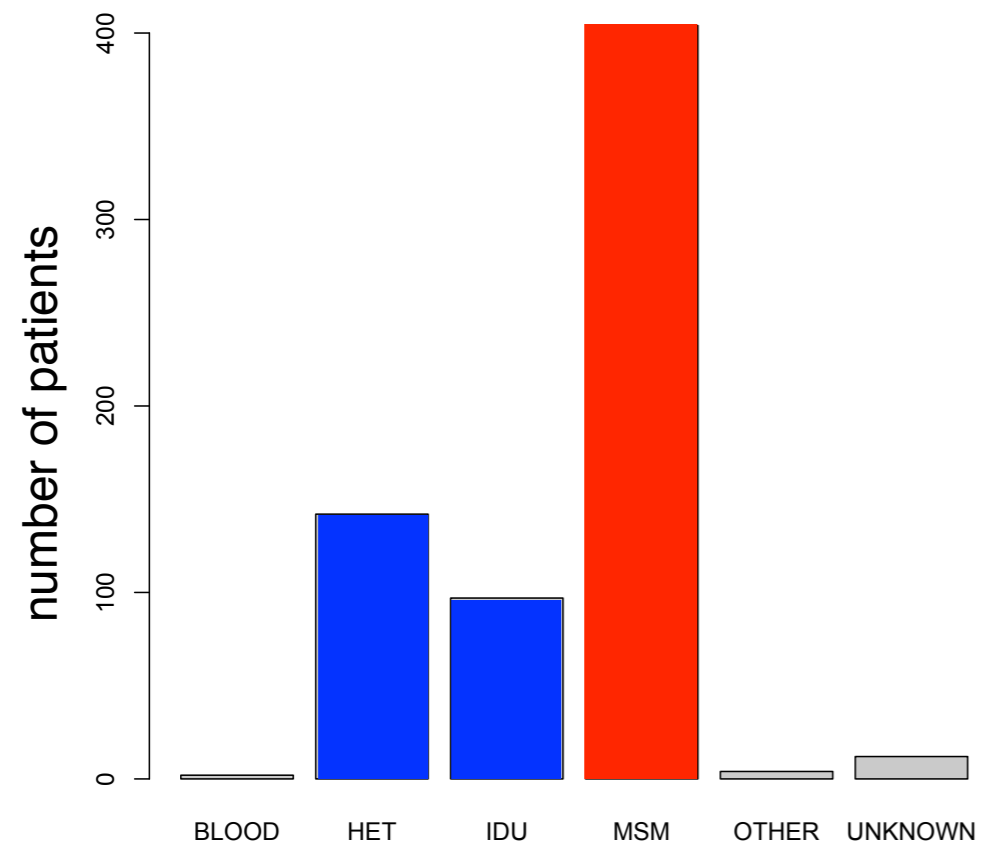# Set-point virus load

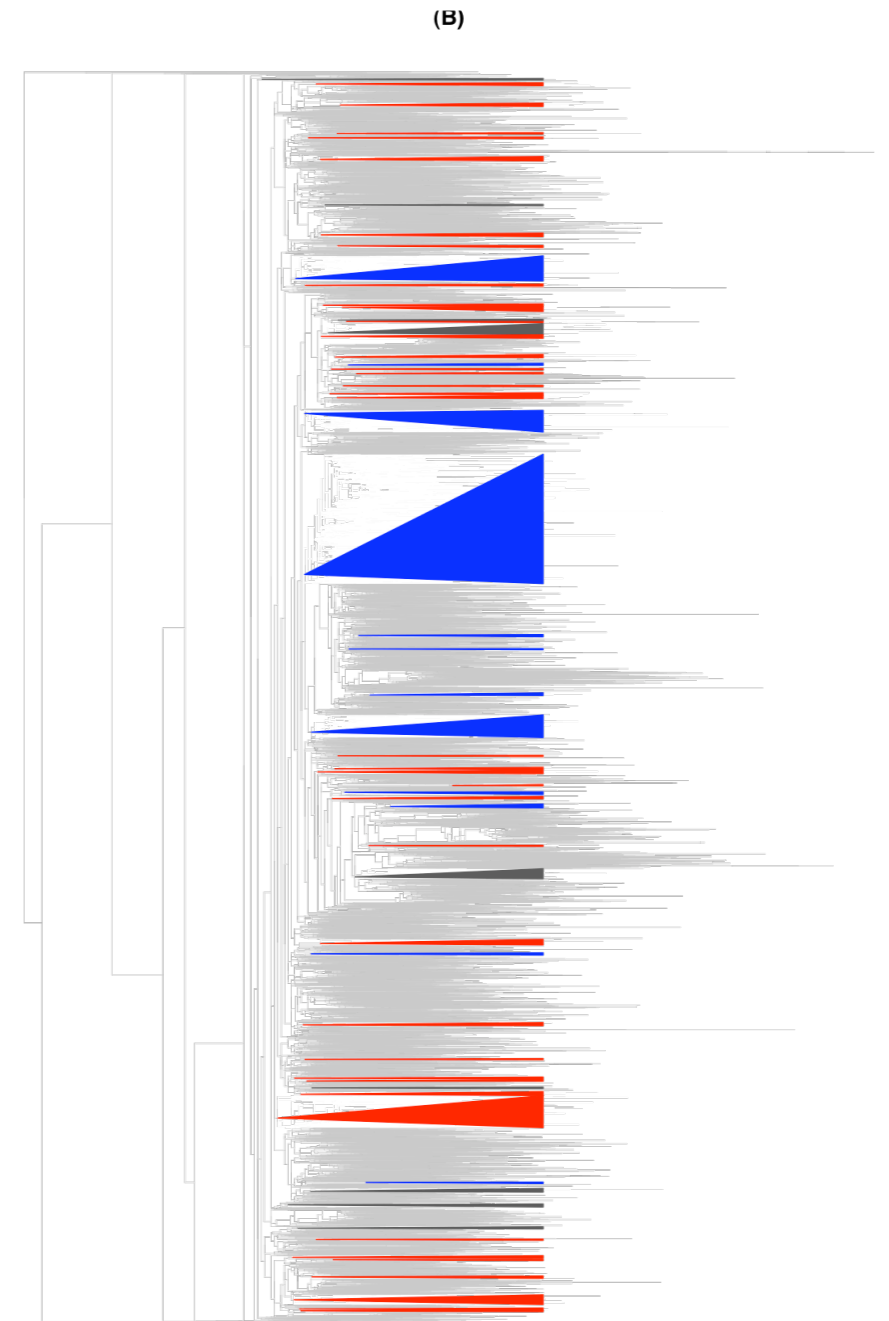**strict spVL criterion**

liberal spVL criterion
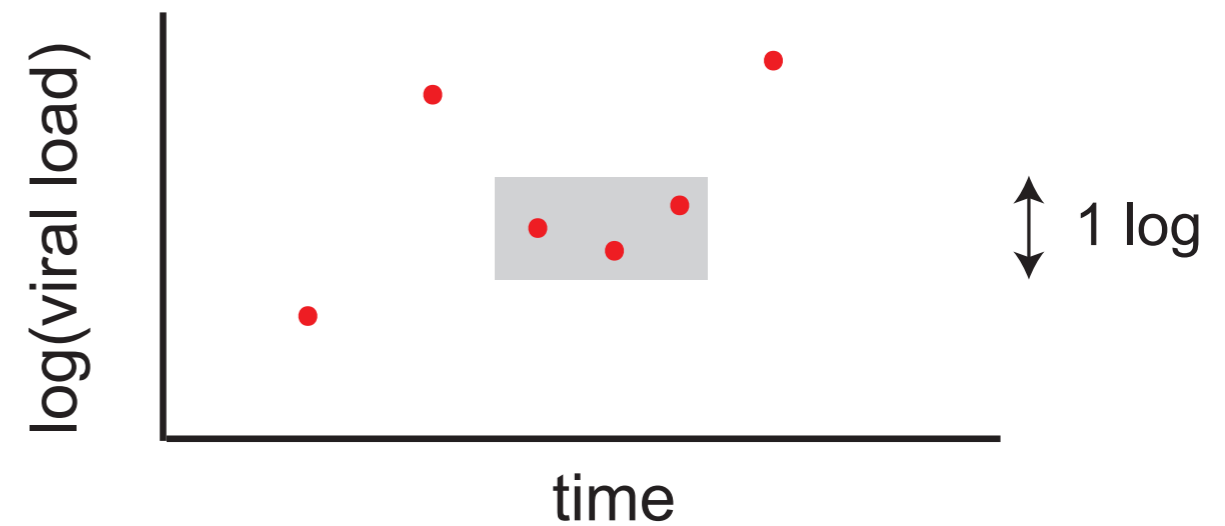
Fellay et al. (2007, Science)
Fellay et al. (2009, PLoS Genet.)

Transcription groups

Transmission groups

Kouyos et al. (2010, JID)

Transcription groups

number of patients

400
300
200
100
0

BLOOD HET IDU MSM OTHER UNKNOWN
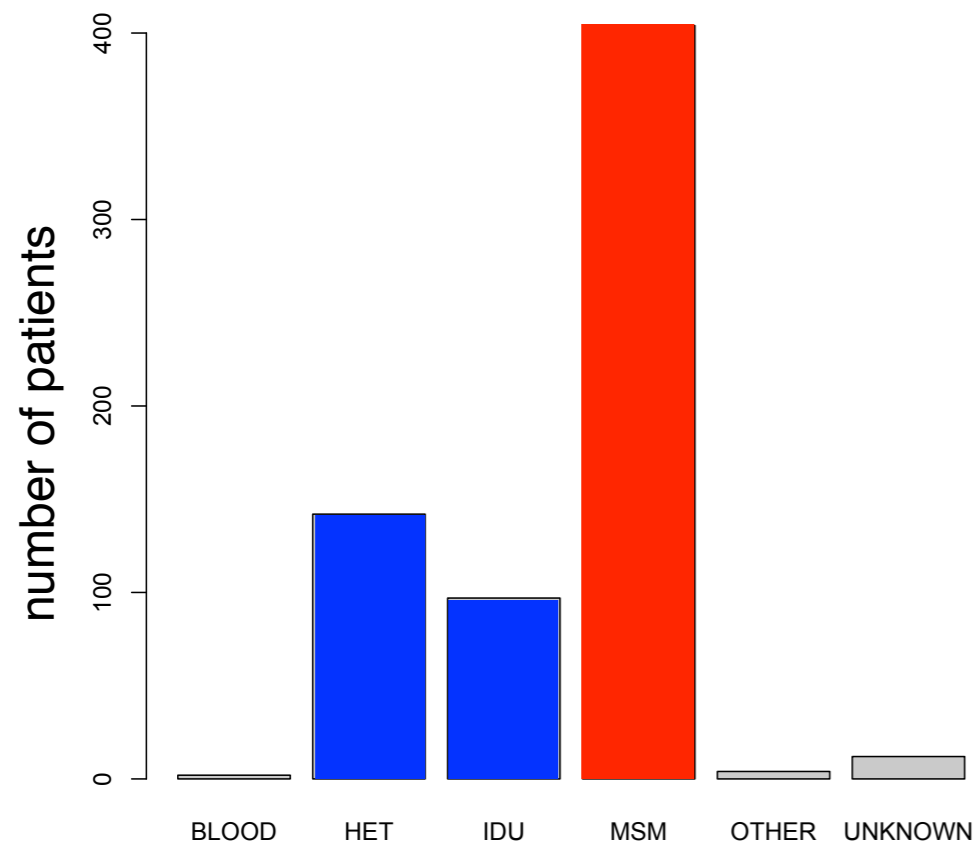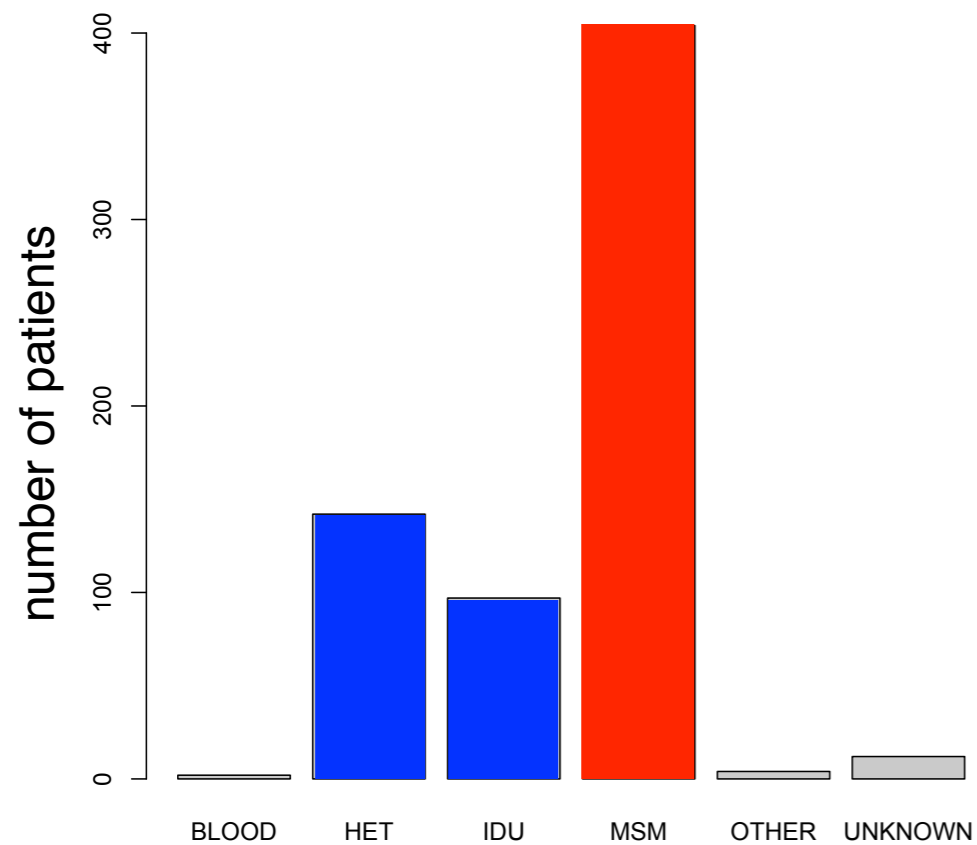
The phylogeny should be closer to the transmission chain for MSMs

(B)

Kouyos et al. (2010, JID)

Virus control over spVL

mean=0.41   median=0.51

number of phylogenies

phylogenetic signal (λ)

SWISS HIV COHORT STUDY

Alizon et al. (2010, *PLoS Path*)
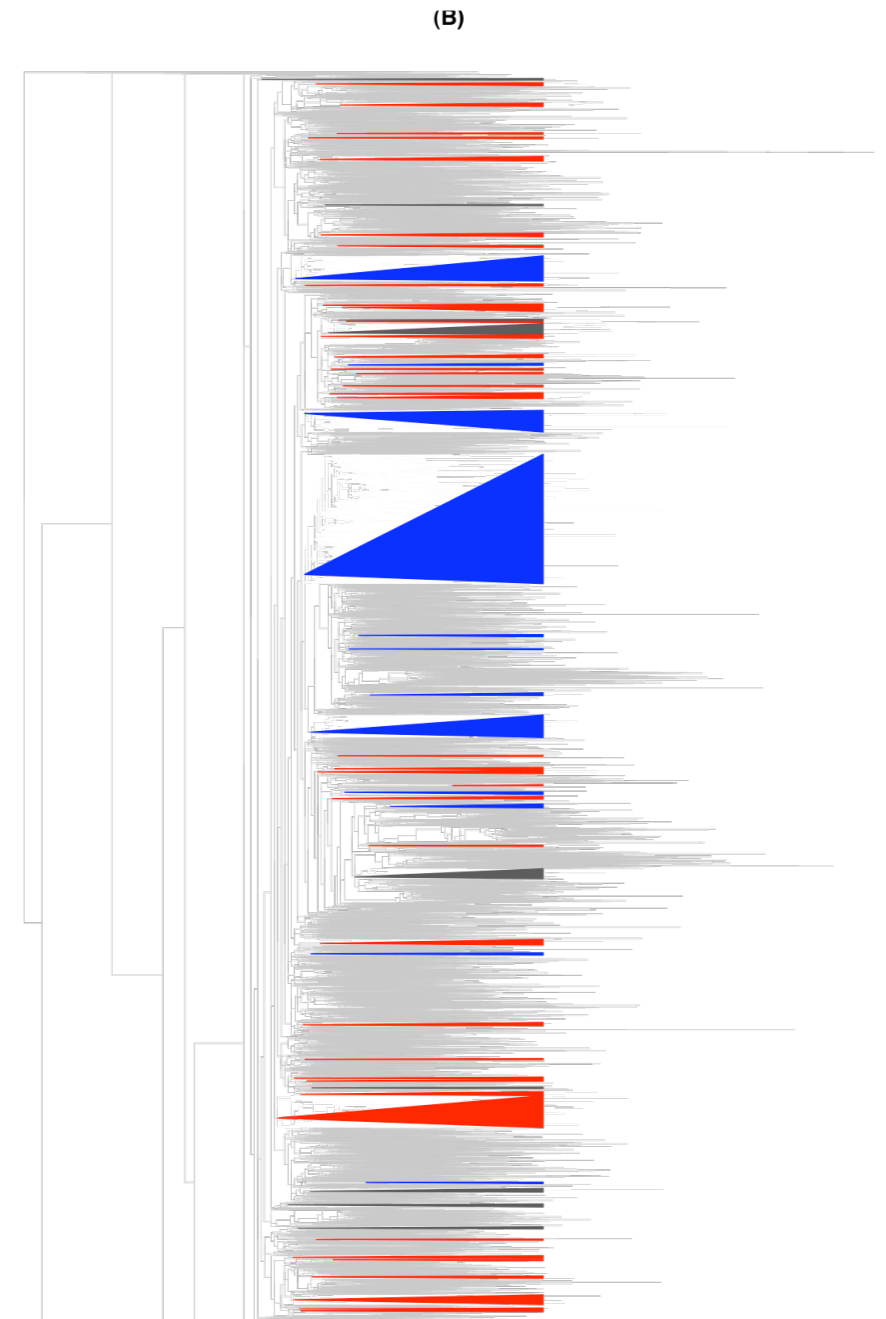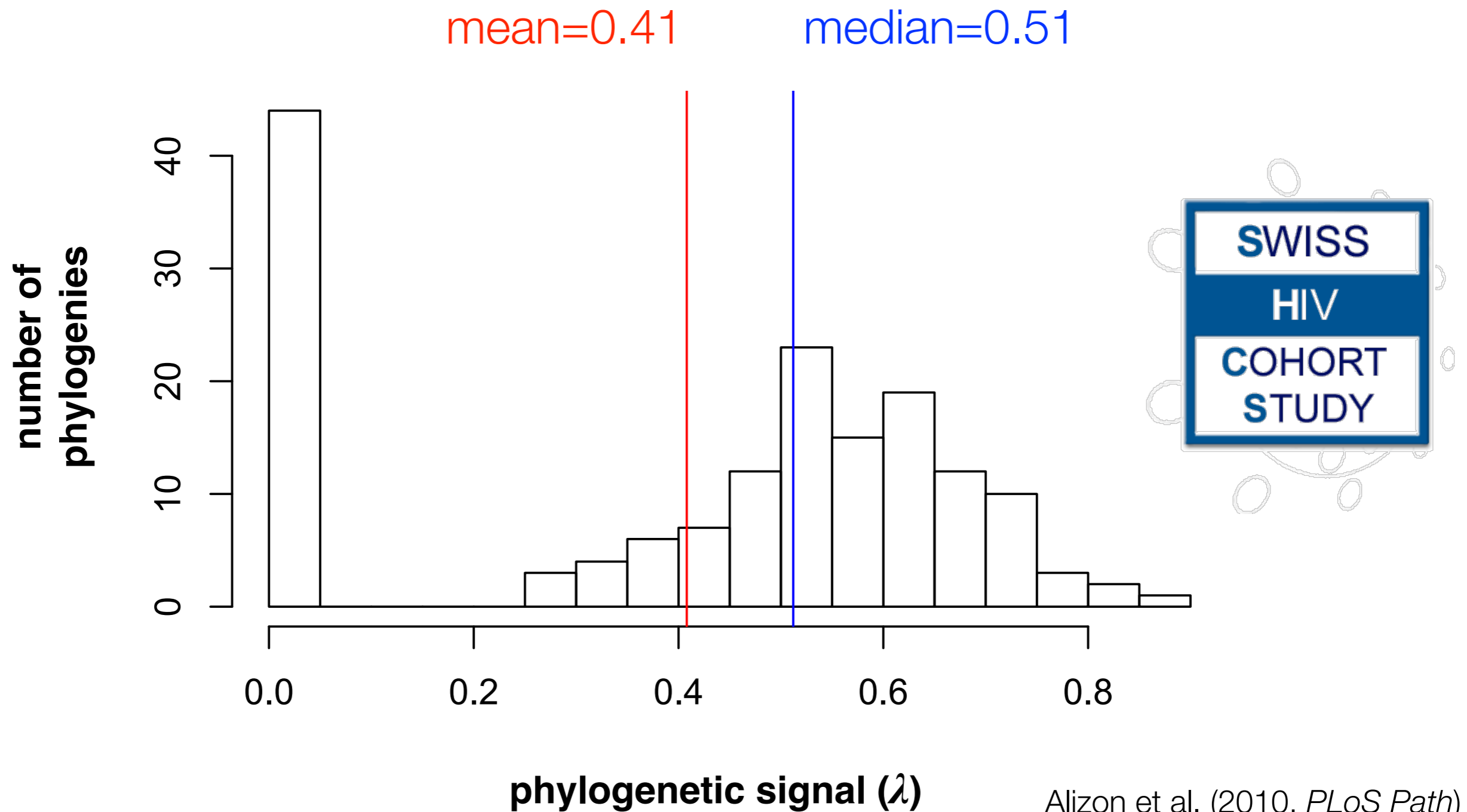
Virus control over spVL

	Pagel's λ median 160 trees no branch length
MSM strict (*n*=134)	0.51
strict (*n*=230)	0.17
MSM all (*n*=404)	0.13
all (*n*=661)	—
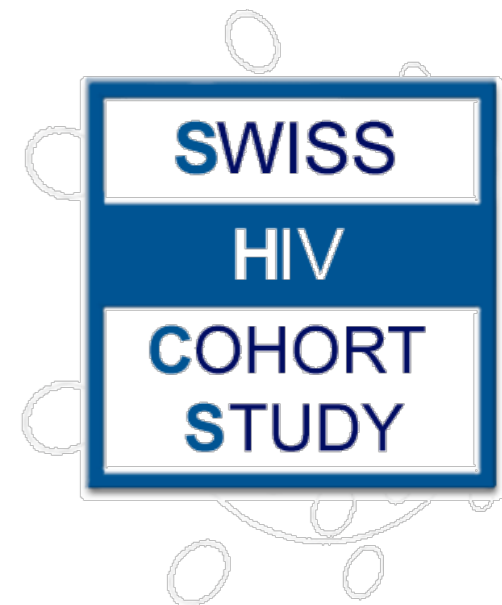
Pagel (1994, *Proc B*)
Blomberg et al. (2003, *Evolution*)
Alizon et al. (2010, *PLoS Path*)

Virus control over spVL

	Pagel's λ median 160 trees no branch length	Blomberg's K p-val by randomisation branch length brownian motion
MSM strict (*n*=134)	**0.51**	**0.59*****
strict (*n*=230)	**0.17**	0.03
MSM all (*n*=404)	**0.13**	**0.09***
all (*n*=661)	—	0.002
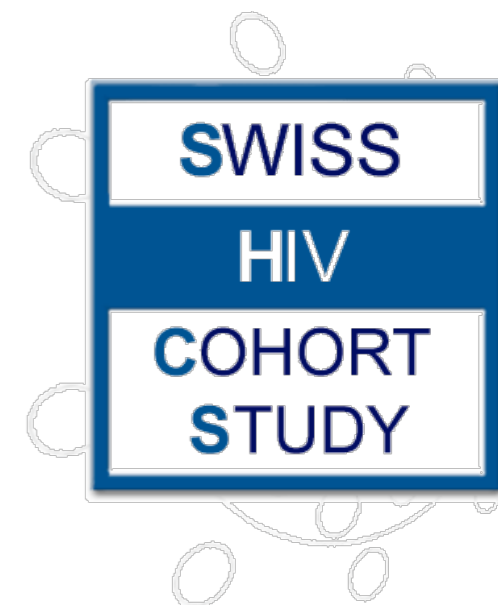
Pagel (1994, *Proc B*)
Blomberg et al. (2003, *Evolution*)
Alizon et al. (2010, *PLoS Path*)

SWISS
HIV
COHORT
STUDY

Virus control over spVL

	Pagel's λ median 160 trees no branch length	**Blomberg's K** p-val by randomisation branch length brownian motion	**Blomberg's d** p-val by randomisation branch length Ornstein-Uhlenbeck
MSM strict (*n*=134)	**0.51**	**0.59*****	**0.72*****
strict (*n*=230)	**0.17**	0.03	**0.25***
MSM all (*n*=404)	**0.13**	**0.09***	0.0001
all (*n*=661)	—	0.002	

SWISS HIV COHORT STUDY

Pagel (1994, *Proc B*)
Blomberg et al. (2003, *Evolution*)
Alizon et al. (2010, *PLoS Path*)