Predictability of viral-host co-evolution

Aleksandra M Walczak

Laboratoire de Physique - ENS, CNRS

work with
Jacopo Marchi, ENS
Michael Lässig, Cologne
Andreas Mayer, ENS
Vijay Balasubramanian, UPenn
Thierry Mora, ENS

Prediction in living systems

Serena hitting a ball

SE Palmer et al

integration of
equations of motion
+
predictive information

$$I\left[\text{input}_{t-\tau}, \text{response}_t\right]$$
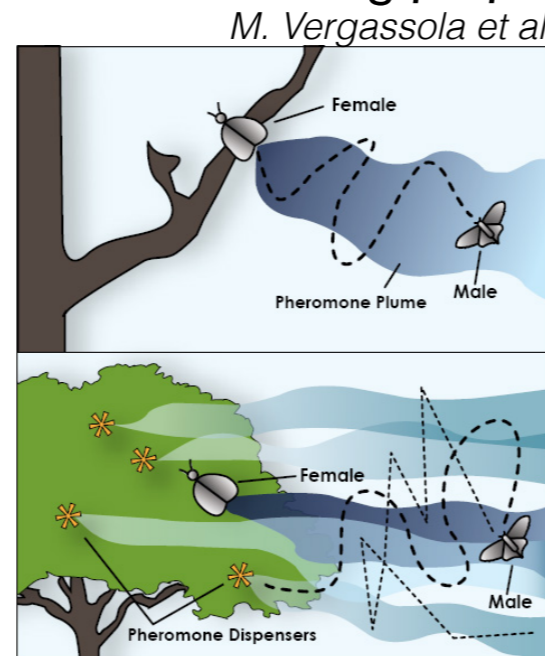
$$\partial_t^2 x = f(x, v) + \xi$$

everything predicts

Prediction in living systems

Serena hitting a ball   *male moths searching for females*

M. Vergassola et al

SE Palmer et al

observations ➡ probabilistic map of sources

→ maximize entropy reduction rate

$$\Delta S(\mathbf{r} \to \mathbf{r}_j) = P_t(\mathbf{r}_j)[-S]$$

exploitation - max likelihood search

motion:
reduce entropy

$$+ \sum_{k:\text{no de detections}} P(\text{k detections})\Delta S_{\text{k detections}}$$

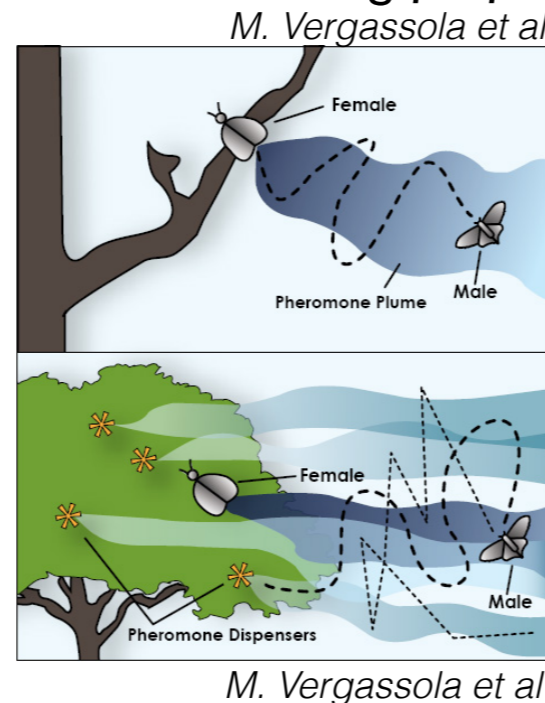exploration - gathering data

update

everything predicts
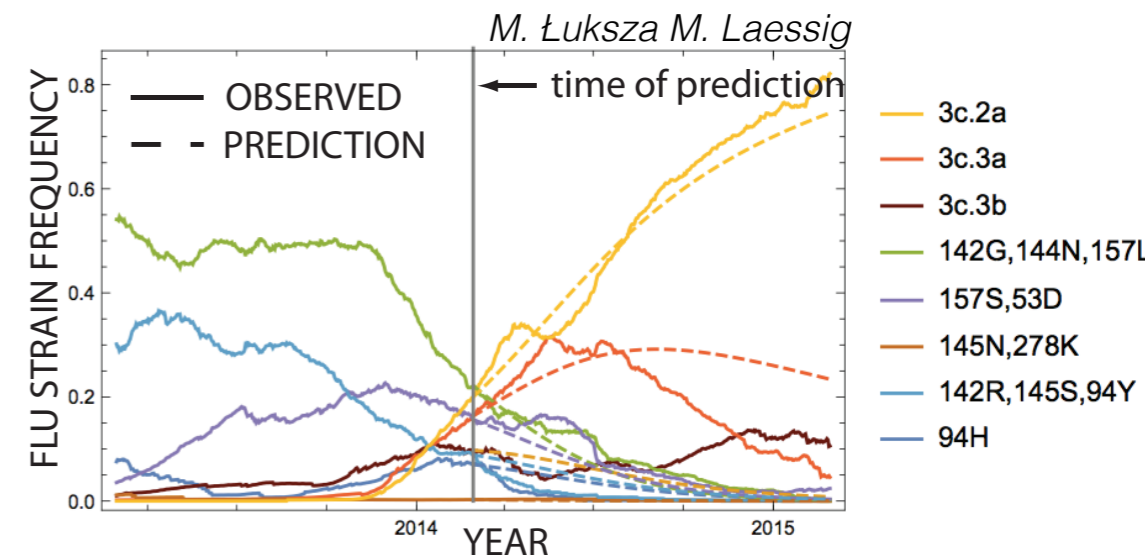
Prediction in living systems

Serena hitting a ball

male moths searching for females

flu strain evolution

SE Palmer et al

M. Vergassola et al

M. Łuksza M. Laessig

integration of many body
stochastic equations of motion
+
interaction with the environment

$$\hat{X}_\alpha(t+\Delta t) = \frac{1}{Z} \sum_{i \in \alpha} x_i(t) \exp\left[\int_t^{t+\Delta t} F_i(t') dt' \right]$$

initial frequency of mutant

growth rate

final frequency in lineage

sum over mutants in lineage

$$F_i(t) = f_{st}(\mathbf{a}_i) + \sum_\rho R_\rho(t) f_{ag}(H_{i\rho})$$

growth rate

estimated immune pressure

protein stability

everything predicts

Prediction in living systems

Serena hitting a ball

SE Palmer et al

male moths searching for females

M. Vergassola et al

Female
Pheromone Plume
Male

Female
Pheromone Dispensers
Male

flu strain evolution

M. Łuksza M. Laessig

OBSERVED — PREDICTION --- — time of prediction

FLU STRAIN FREQUENCY

2014   YEAR   2015

3c.2a
3c.3a
3c.3b
142G,144N,157L
157S,53D
145N,278K
142R,145S,94Y
94H
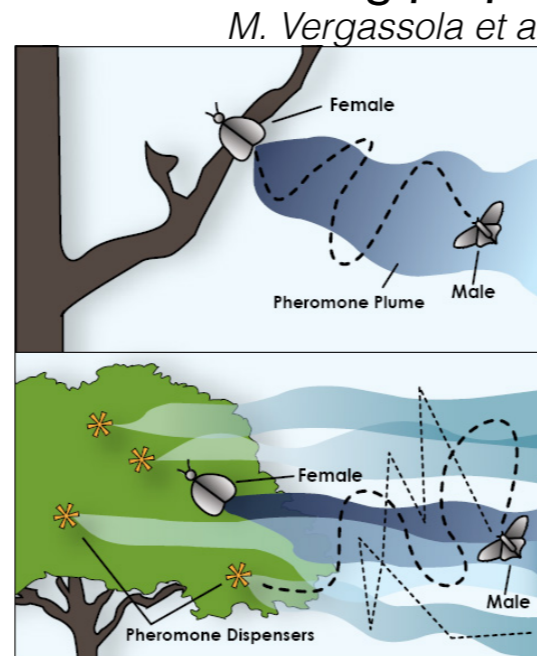
everything predicts

pathogens (viruses, bacteria)

FLU$_{EUROPE}^{2010}$ FLU$_{ASIA}^{2010}$

HSV^{2011}

FLU$_{EUROPE}^{2011}$ FLU$_{EUROPE}^{2012}$

B19^{2011}

FLU$_{ASIA}^{2012}$

B19^{2012}

COLD^{2011}

COLD^{2012}

time

receptor statistics

a b z x y c z y b a

a b z x a a a

immunity as prediction?
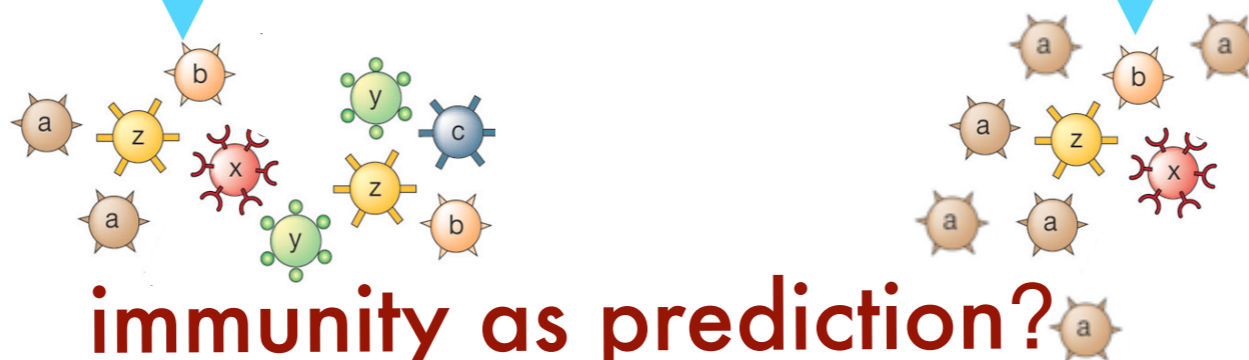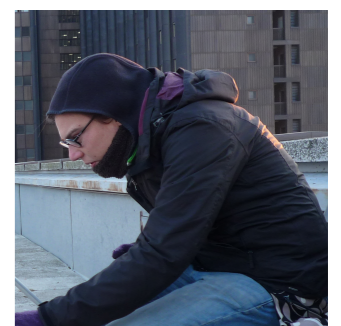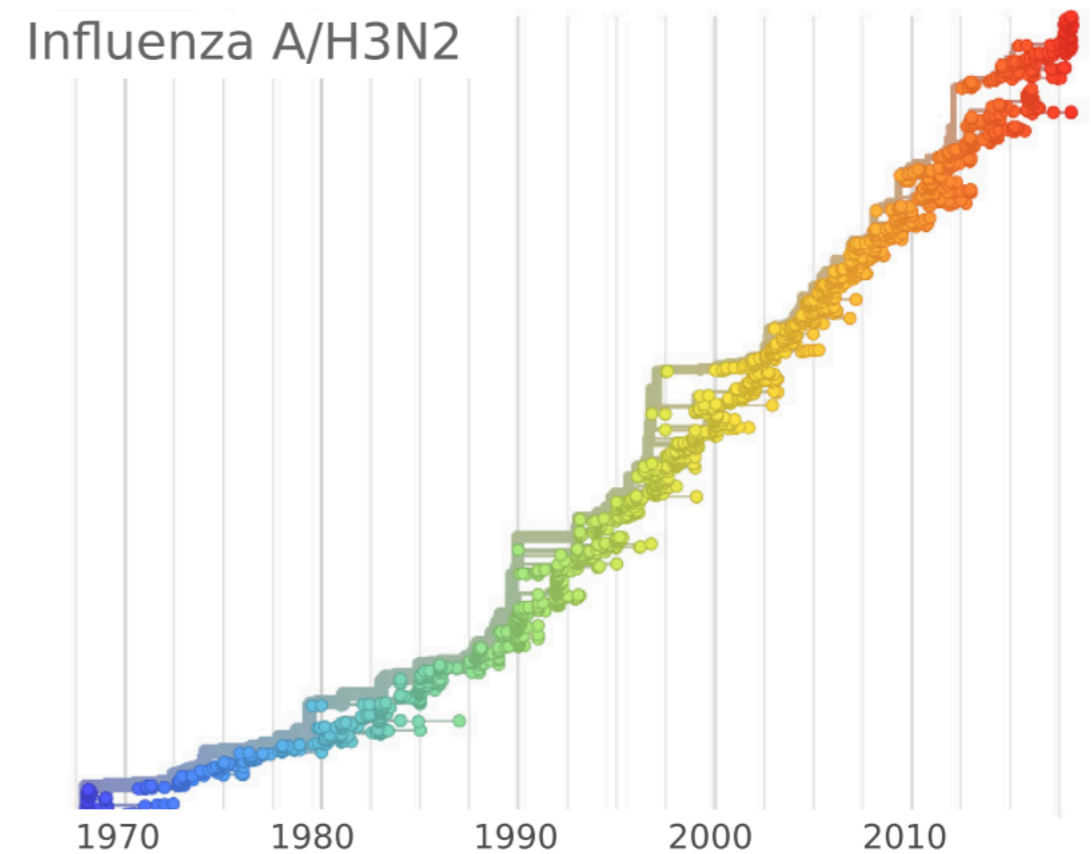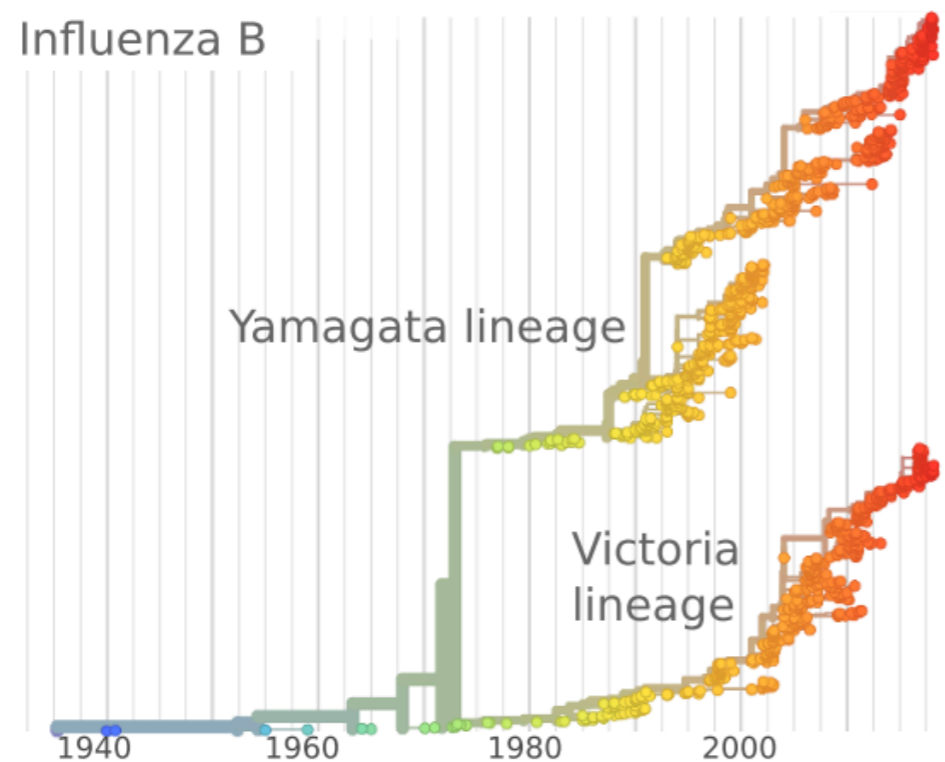
Viral evolution

- evolution of influenza H3N2 looks linear

- one trunk driving evolution

- low diversity, strong selection

Influenza A/H3N2

Yan, Neher and Shraiman 2019

Influenza B

Yamagata lineage

Victoria lineage

- sometimes, splits happen

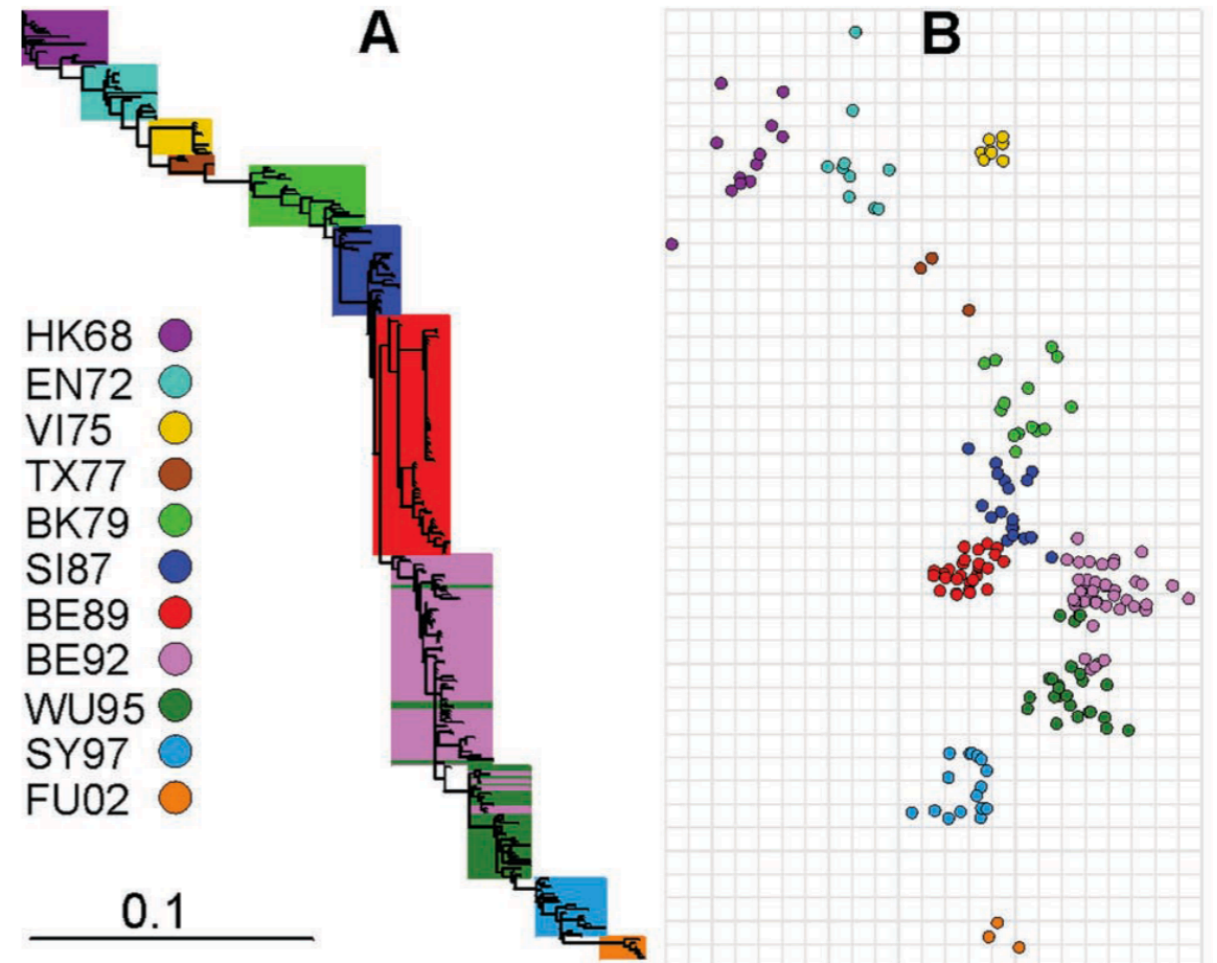*Yan, Neher and Shraiman 2019*
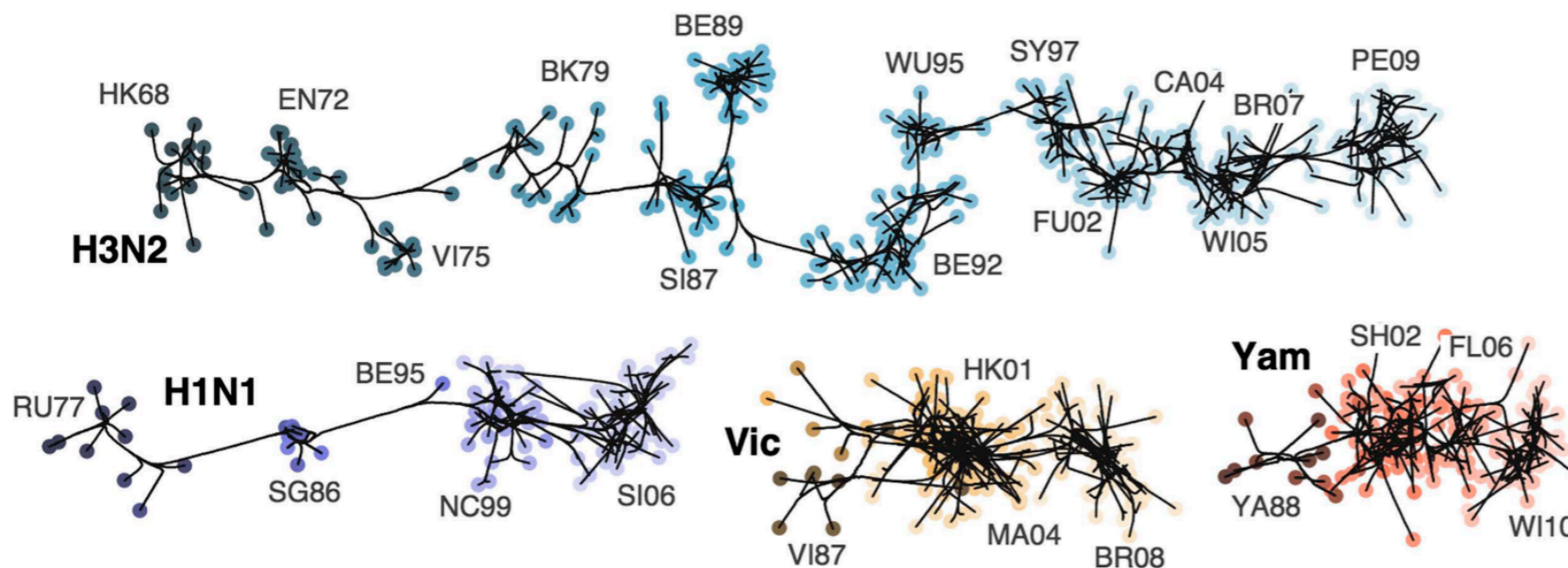
Antigenic map

- **antigenic distance** =
  similarity of response to
  antibody sera
  (hemmaglutination inhibition)

- often projected onto 2D map
  (dim. reduction)

HK68
EN72
VI75
TX77
BK79
SI87
BE89
BE92
WU95
SY97
FU02

0.1

A

B

Smith et al 2004

H3N2

HK68   EN72   BK79   BE89   WU95   SY97   CA04   PE09
                                                         BR07
VI75   SI87   BE92   FU02   WI05

H1N1   BE95
RU77   Vic   HK01   Yam   SH02 FL06
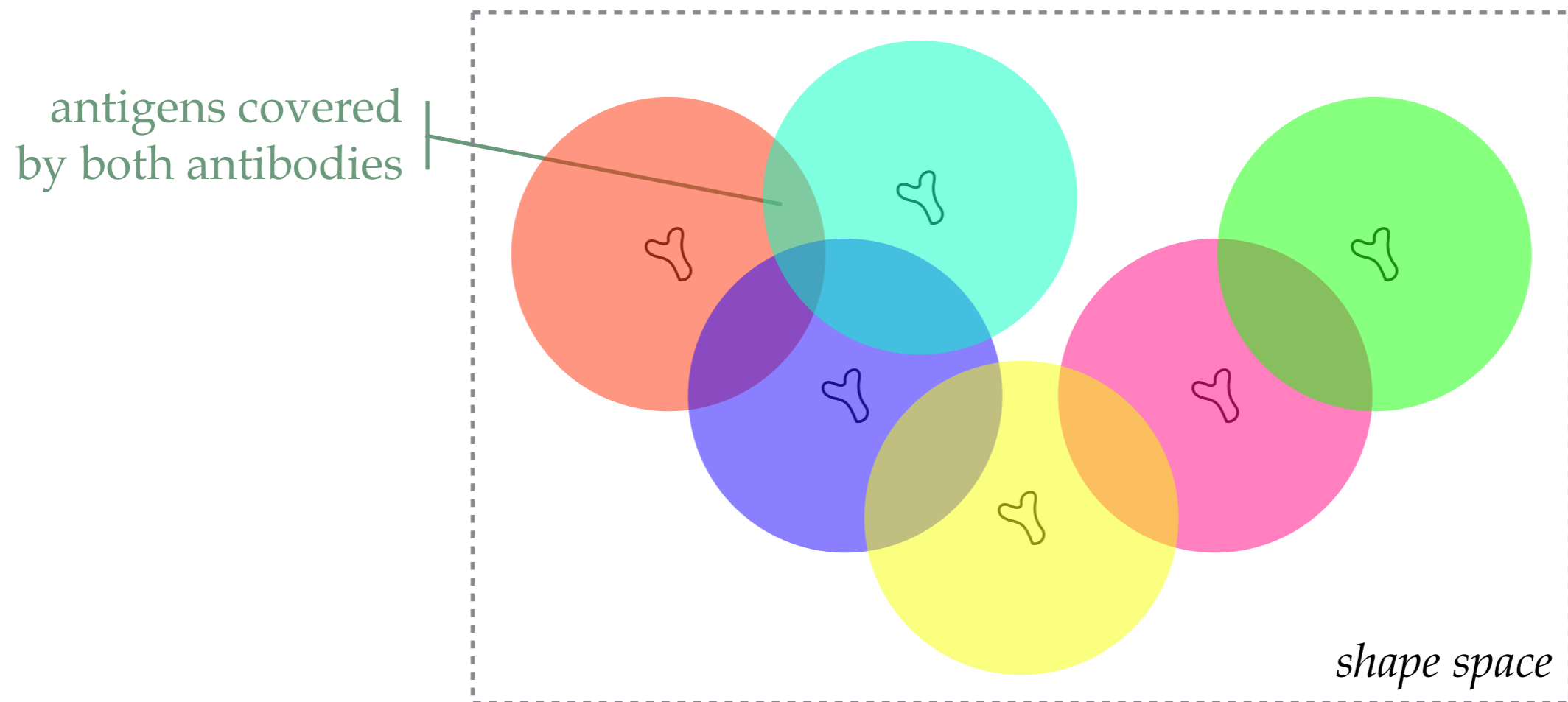SG86   NC99   SI06   VI87   MA04   BR08   YA88   WI10

Bedford et al 2014

Immune "shape space"

- shape space: space of antibodies covering all possible specificities *(Segel and Perelson 1989)*

- antibodies have cross-reactivities

antigens covered
by both antibodies

shape space

- low-dimensional: dimensions 5 to 8 for *all antigens* *(Smith et al 1997)*

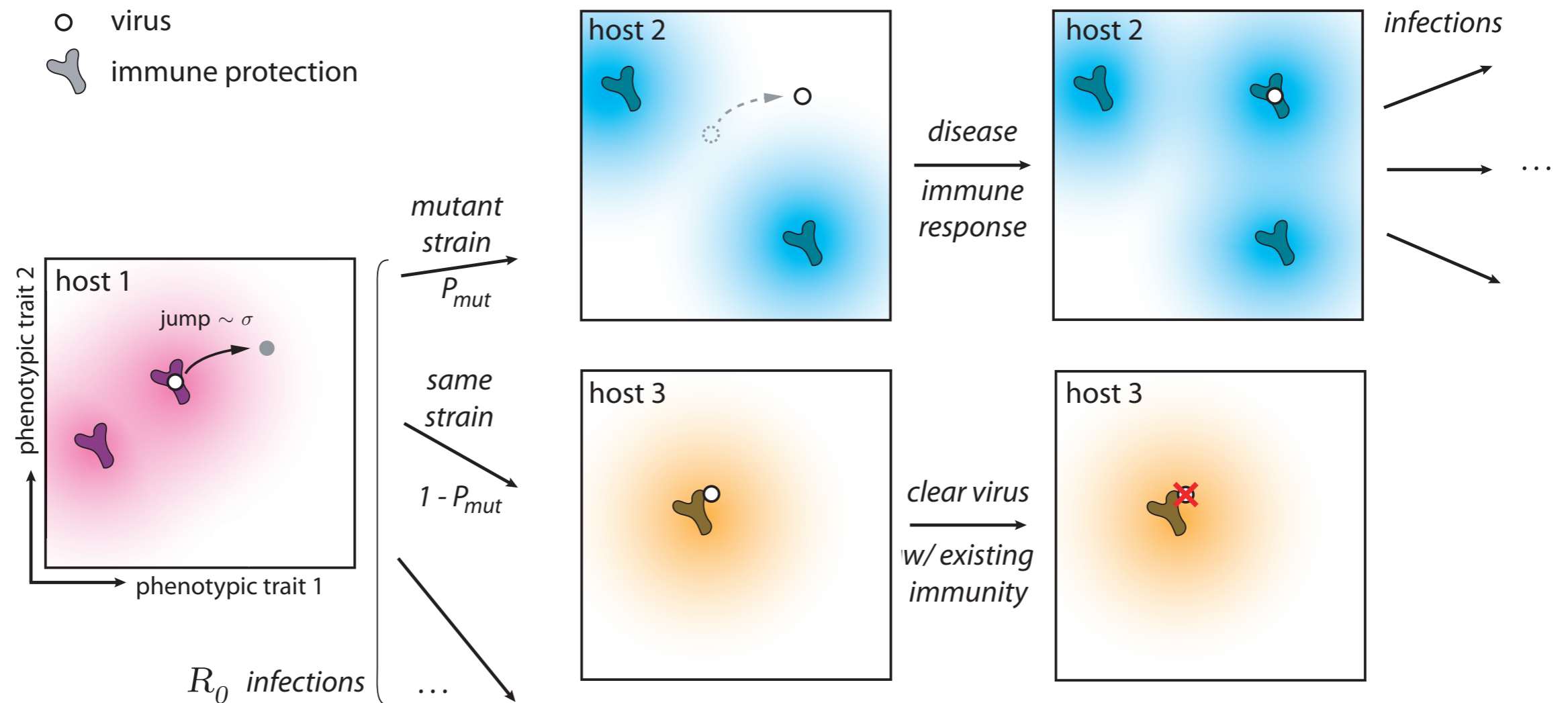 even lower for given family of targets e.g. HA

Co-evolution model

- antibodies and antigens live in same space
- see also

Bedford Rambault Pascual 2012
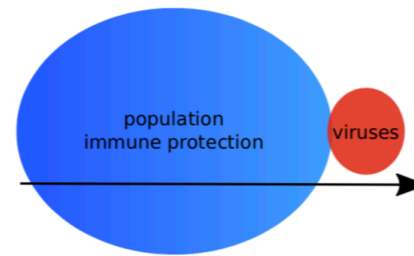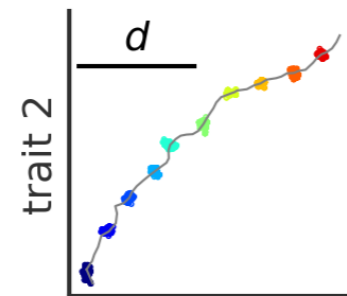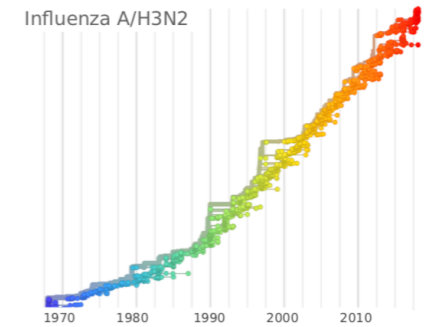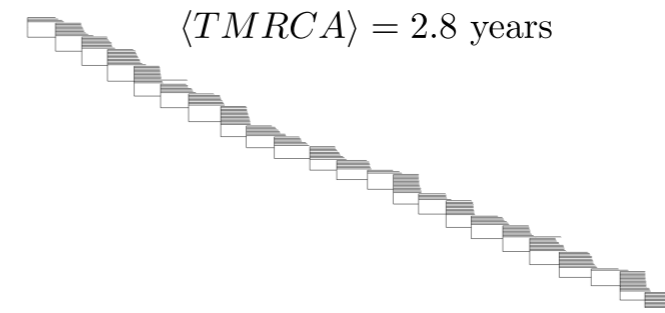*Rouzine and Rozhnova 2018*
Yan, Neher and Shraiman 2019

○ virus

immune protection

phenotypic trait 2

host 1

$\sim \sigma$

phenotypic trait 1

mutant strain P_{mut}

same strain $1 - P_{mut}$

R_0 infections

host 2

host 2

disease immune response

infections

...

host 3

host 3

clear virus w/ existing immunity

4 evolutionary trends

Ballistic

20   40

years
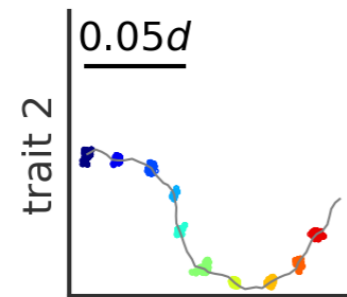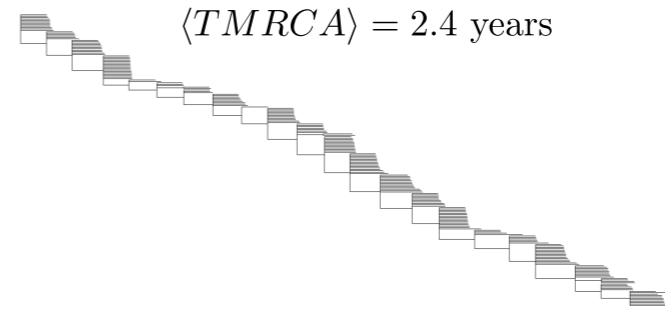
d

trait 2

population
immune protection

viruses

$\langle TMRCA \rangle = 2.8$ years

Influenza A/H3N2

1970 1980 1990 2000 2010

$0.05d$

trait 2

$\langle TMRCA \rangle = 2.4$ years

$0.5d$

trait 2
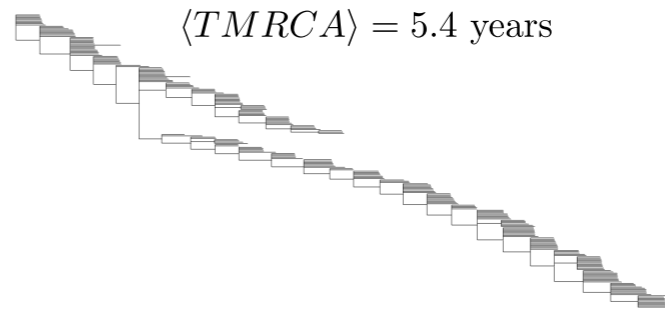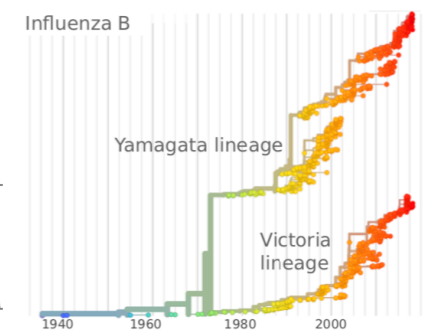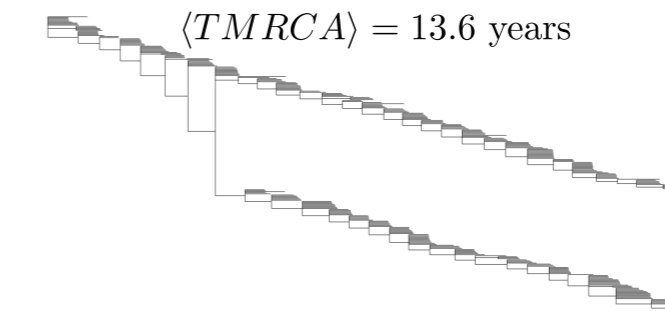
$\langle TMRCA \rangle = 5.4$ years

$0.05d$

trait 2

$\langle TMRCA \rangle = 13.6$ years

Influenza B

Yamagata lineage

Victoria lineage

1940 1960 1980 2000

trait 1

Marchi Lässig Mora Walczak 2019

average number of lineages

mutation rate (per day)

10^{-1}

10

10^{-4}

20   40

years

d

trait 2

0.05d

trait 2

4

3

2

1

mutation jump size

20   40

years

d

trait 2

TMRCA (yrs)

40

20

0.05d

trait 2

trait 1

TMRCA (yrs)

40

20

0

25
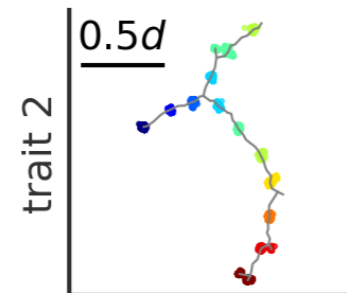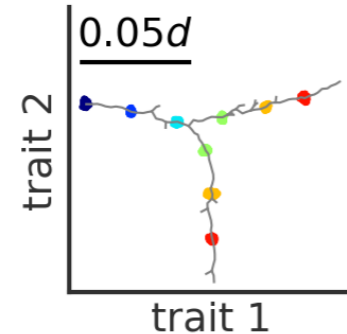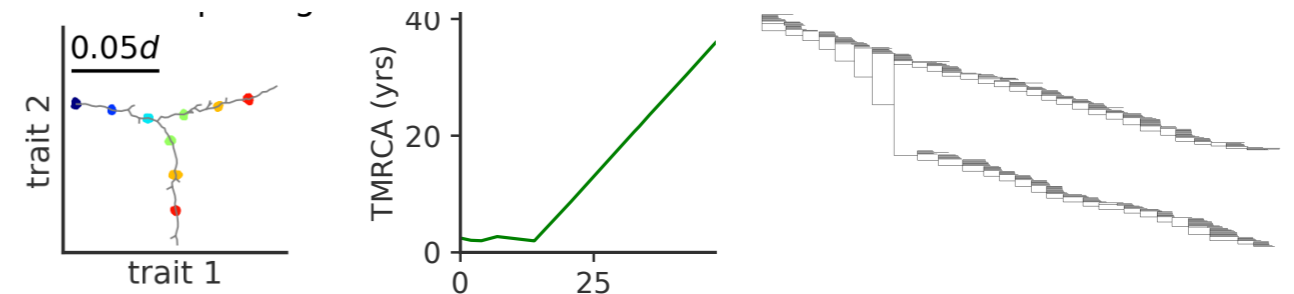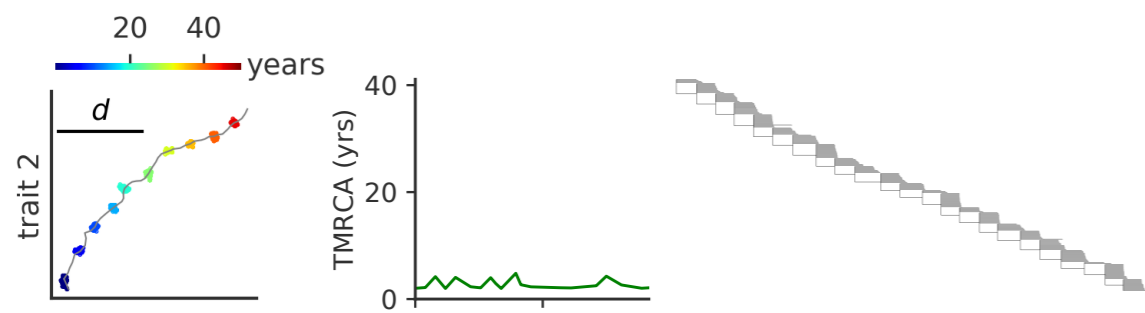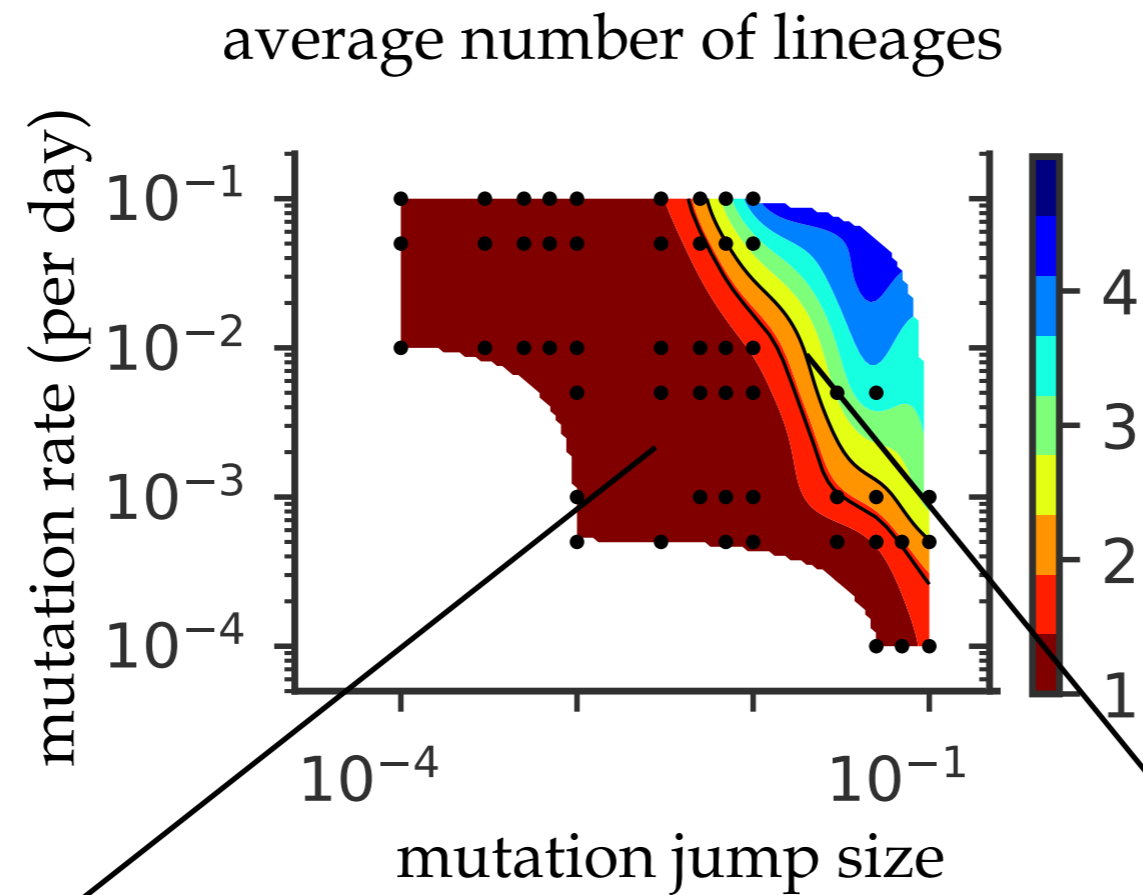
e rate $= 5 \cdot 10^{-4}$)

Coarse grained model

virus n immune memory h

viral fitness

D

viral population size
= number of infected hosts

$$N(t) = \int dx\, n(x, t) \neq \text{const}$$

mutations

$$\partial_t n(\mathbf{x}, t) = f(\mathbf{x}, t) n(\mathbf{x}, t) + D\partial_{\mathbf{x}}^2 n + \sqrt{n(\mathbf{x}, t)} \eta(\mathbf{x}, t)$$
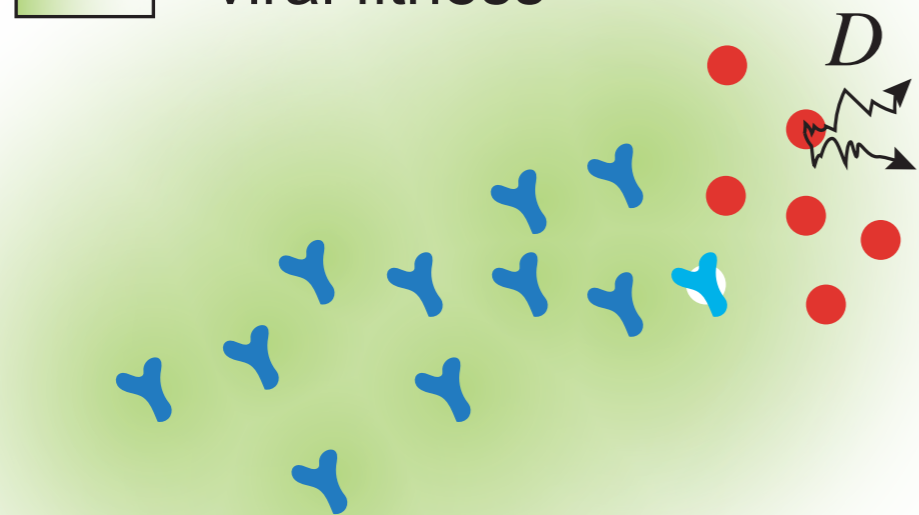
growth

demographic
noise

fitness

number of injectable hosts

$$f(\mathbf{x}, t) = \ln[R_0(1 - c(\mathbf{x}, t))^M]$$

susceptibility of random
host to strain x

immunity
acquisition

$$\partial_t h(\mathbf{x}, t) = \frac{1}{N_h}\left[n(\mathbf{x}, t) - N(t)\frac{h(\mathbf{x}, t)}{M} \right]$$
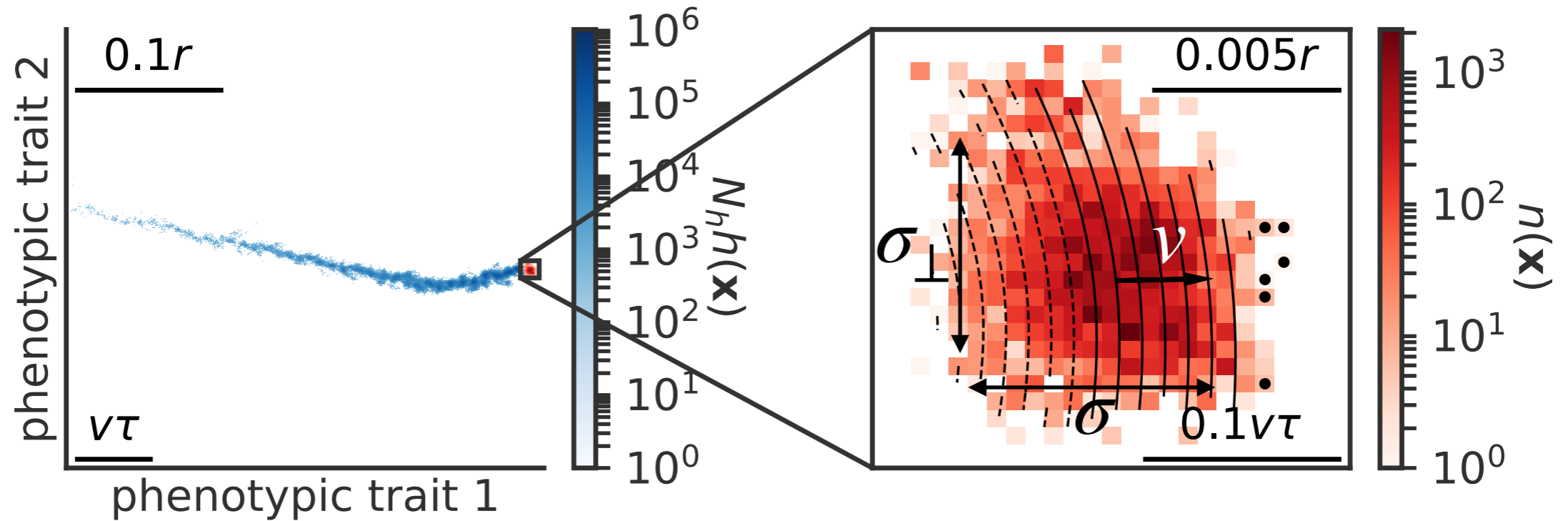
compensatory
memory loss

immune coverage

$$c(\mathbf{x}, t) = \frac{1}{M}\int d\mathbf{x}'\, h(\mathbf{x}', t) e^{-|\mathbf{x} - \mathbf{x}'|}$$

host population
size

$$N_h = \text{const}$$

\# memories carried
by each host

$$\tau = \frac{M N_h}{N(t)}$$

immune
turnover time

$$M = \int dx\, h(x, t) = \text{const}$$

time - units of infection cycles

Antigenic wave

- simulations: same behaviour as agent-based model

- distribution of fitness

Tsimring Levin Kessler 1996
Rouzine et al 2003
Desai Fisher 2007
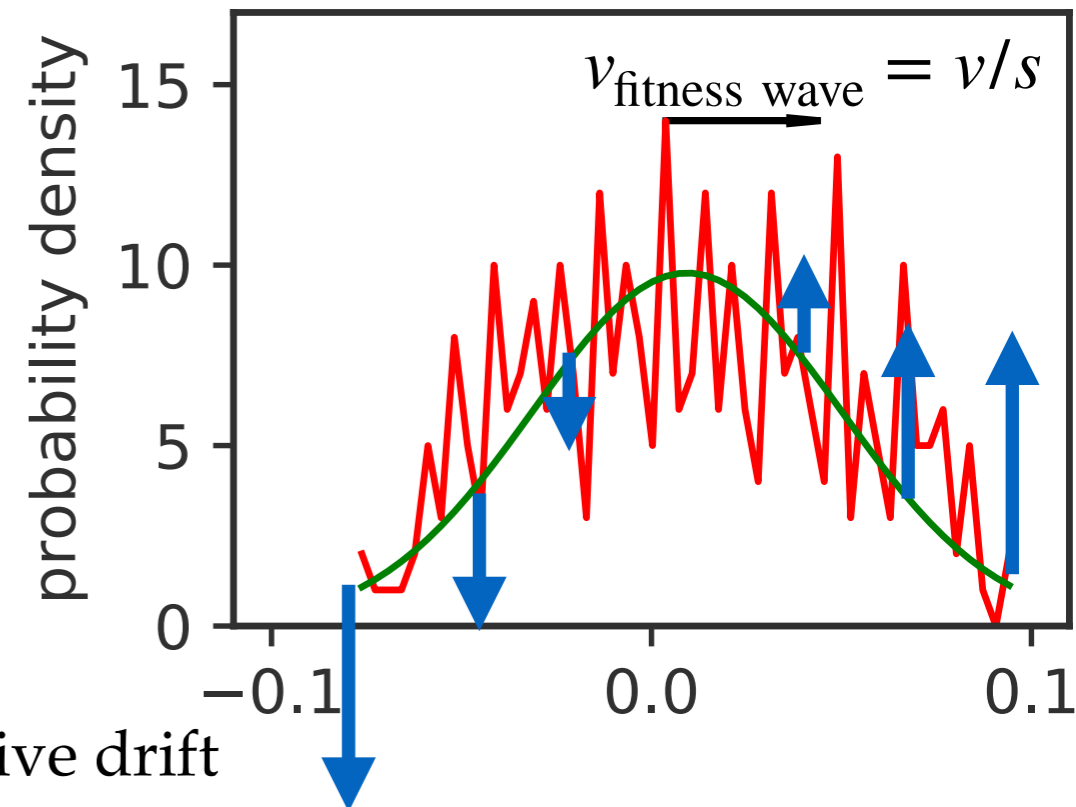
- link to fitness wave theory ?

  - \>1 dimension $\rightarrow D_{\text{fitness wave}} = Ds^2$

$$n(\mathbf{x}, t) = n_1(x_1 - vt)\phi(x_2, \dots, x_d)$$

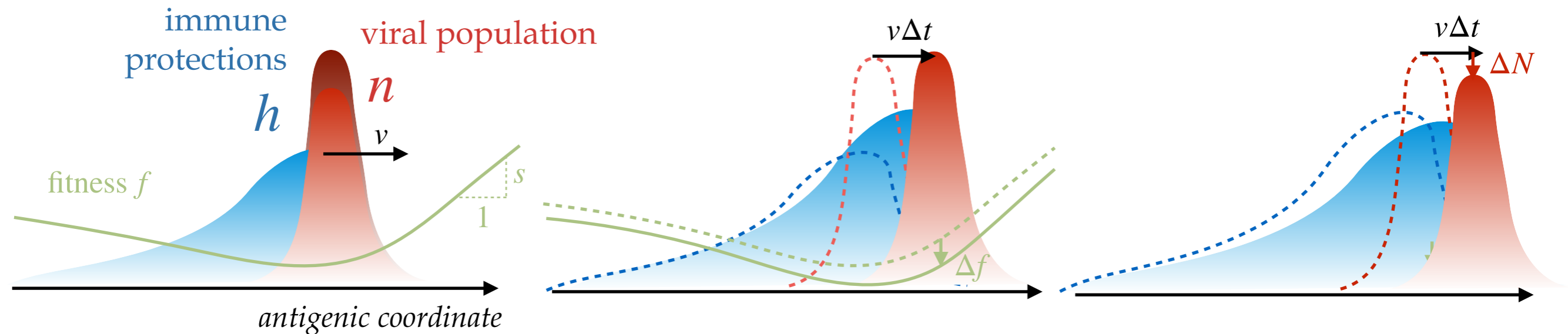 - ~ constant fitness gradient $f \approx f_0 + s(x_1 - vt)$

  - ∥ mutations isotropic (beneficial & deleterious)

  - ⊥ mutations neutral

  - non-constant population size → continuous negative drift

Population control - main wave direction

- non-constant population size → continuous negative drift
- population size regulated by immune pressure

$$f \approx f_0 + s\,(x_1 - vt\,)$$

Rouzine and Rozhnova 2018
Yan, Neher and Shraiman 2019

immune protections h viral population n fitness f v s 1   *antigenic coordinate* $v\Delta t$ Δf $v\Delta t$ ΔN

$$\partial_t h(\mathbf{x}, t) = \frac{1}{N_h}\left[n(\mathbf{x}, t) - N(t)\frac{h(\mathbf{x}, t)}{M}\right] \quad \rightarrow \quad \tau = \frac{MN_h}{N(t)} \quad \text{immune turnover time} \quad \rightarrow \quad \text{N stable}$$

population size set by $f_0 = 0$ →
$$f(\mathbf{x}, t) = \ln\left[R_0\left(1 - \frac{e^{-u}}{1+v\tau}\right)^M\right] \approx f_0 + su$$

$$\left(R_0^{1/M} - 1\right) = MN_h v/N \quad \rightarrow \quad N/v = MN_h\left(R_0^{1/M} - 1\right)$$

&   selection coefficient $s = |\nabla f| = M\left(R_0^{1/M} - 1\right) \quad \rightarrow \quad \frac{N}{N_h} = vs \quad$ incidence rate

Population control - main wave direction

- non-constant population size → continuous negative drift

$$f \approx f_0 + s\,(x_1 - vt\,)$$

Rouzine and Rozhnova 2018
Yan, Neher and Shraiman 2019

- population size regulated by immune pressure

population size $\quad \dfrac{N}{N_h} = vs \quad$ & $\quad$ selection coefficient $\quad s = |\nabla f| = M\left(R_0^{1/M} - 1\right)$

$$D_{\text{fitness wave}} = Ds^2$$

- speed regulated by population size $\quad$ *(Neher and Hallatschek 2013)*

$$v_{\text{fitness wave}} = v/s$$

$$v \approx D^{2/3}s^{1/3}\left[24\ln(N(Ds^2)^{1/3})\right]^{1/3}$$

→ N & v in terms of model parameters
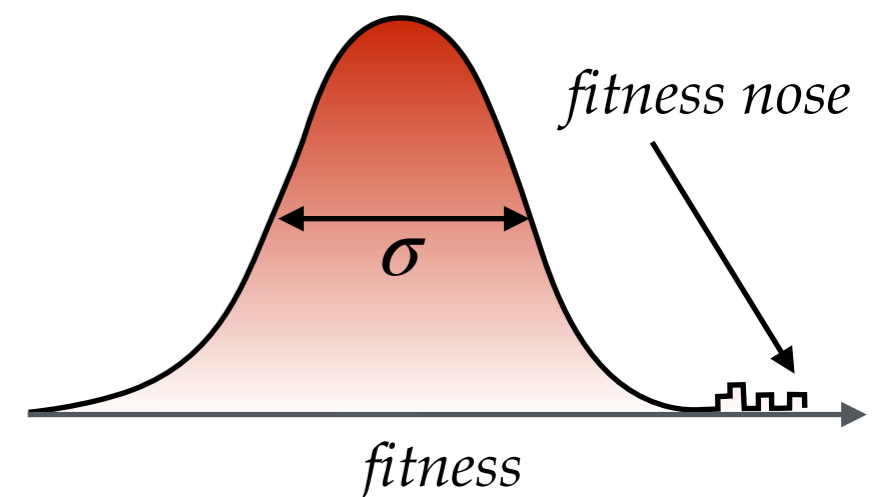
Shape of the wave

- comparison to simulations

$$N = N_h v s$$

$$v \approx D^{2/3} s^{1/3} \left[24 \ln(N_h v s (D s^2)^{1/3}) \right]^{1/3}$$
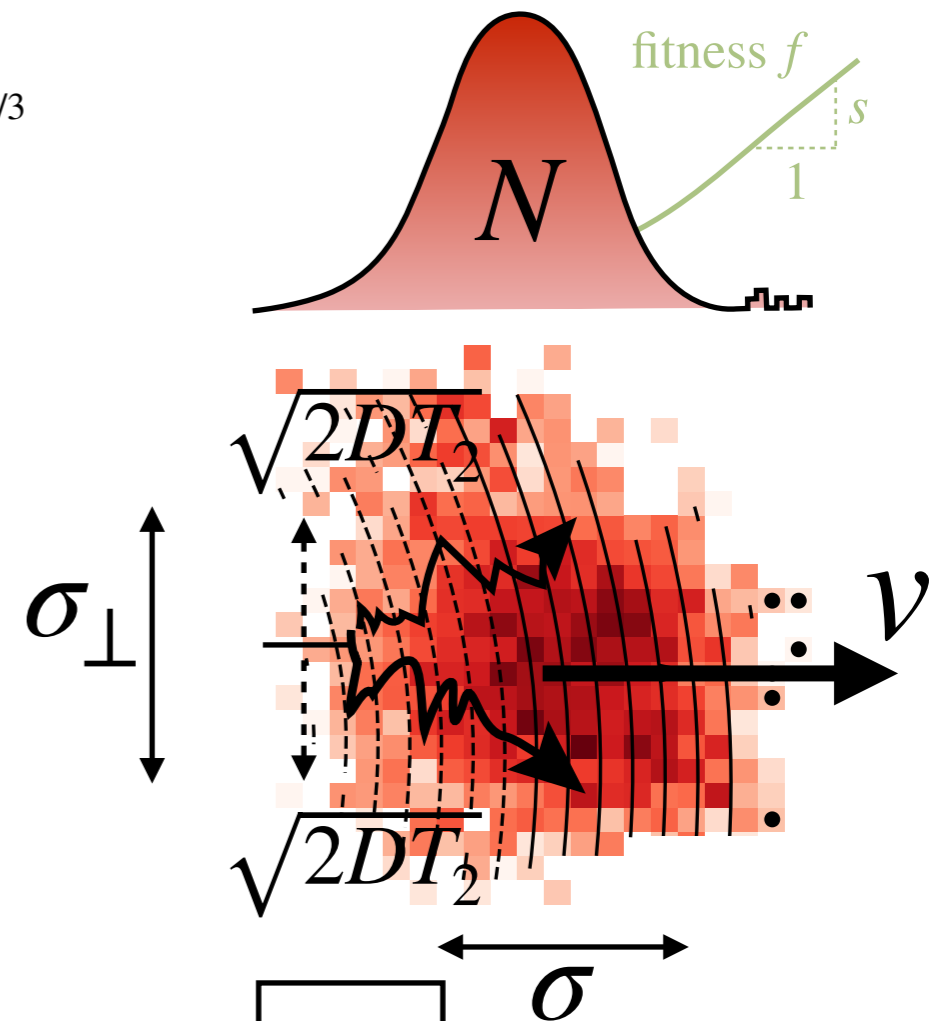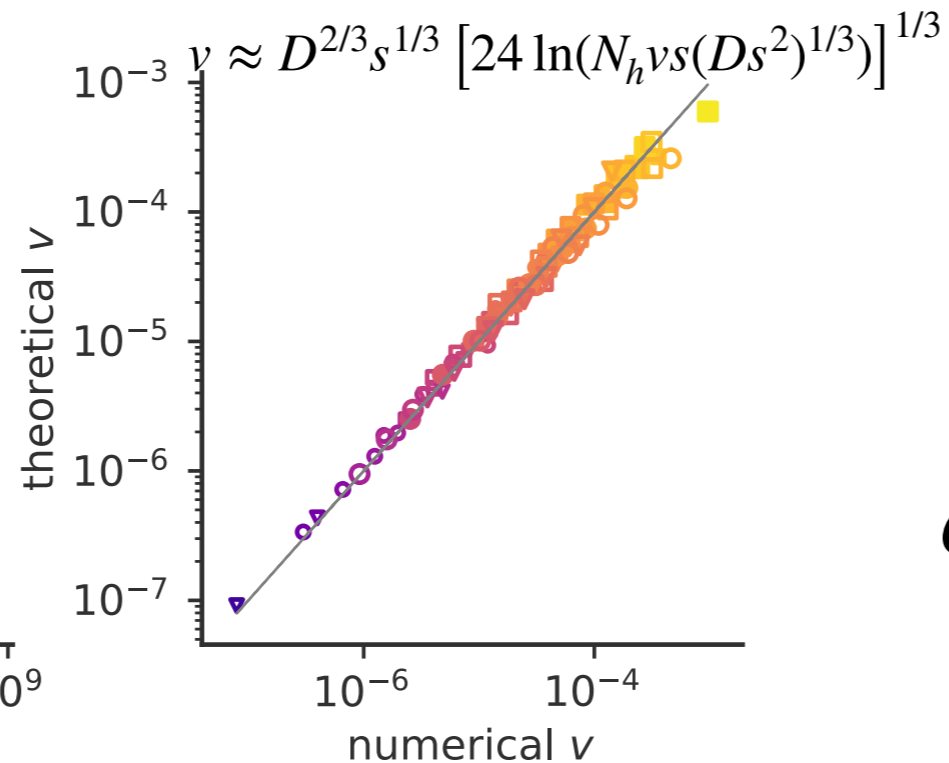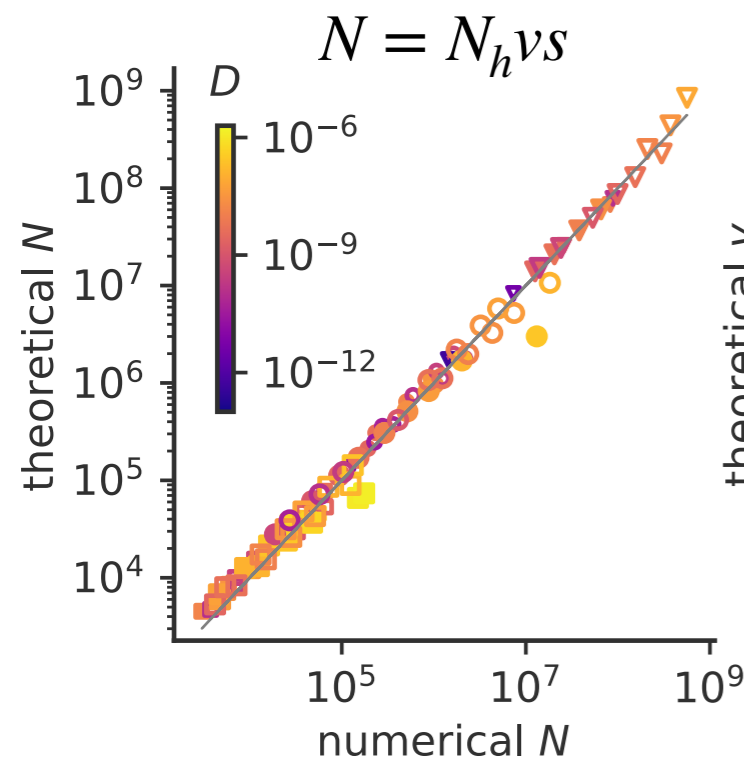
Fisher's theorem

$$\partial_t f = \mathrm{var}(f) = \cancel{s^2} \sigma^2$$

$$sv = s^2 \sigma^2$$

fitness f

$$N$$

$$\sqrt{2DT_2}$$

$$\sigma_\perp$$

$$v$$

$$\sqrt{2DT_2}$$

$$\sigma$$

$$T_2$$

$$\langle T_2 \rangle = \frac{1.66 \sigma^2}{2D}$$

(Neher and Hallatschek 2013)

$$2\sigma_\perp^2 = 2D \times 2T_2$$
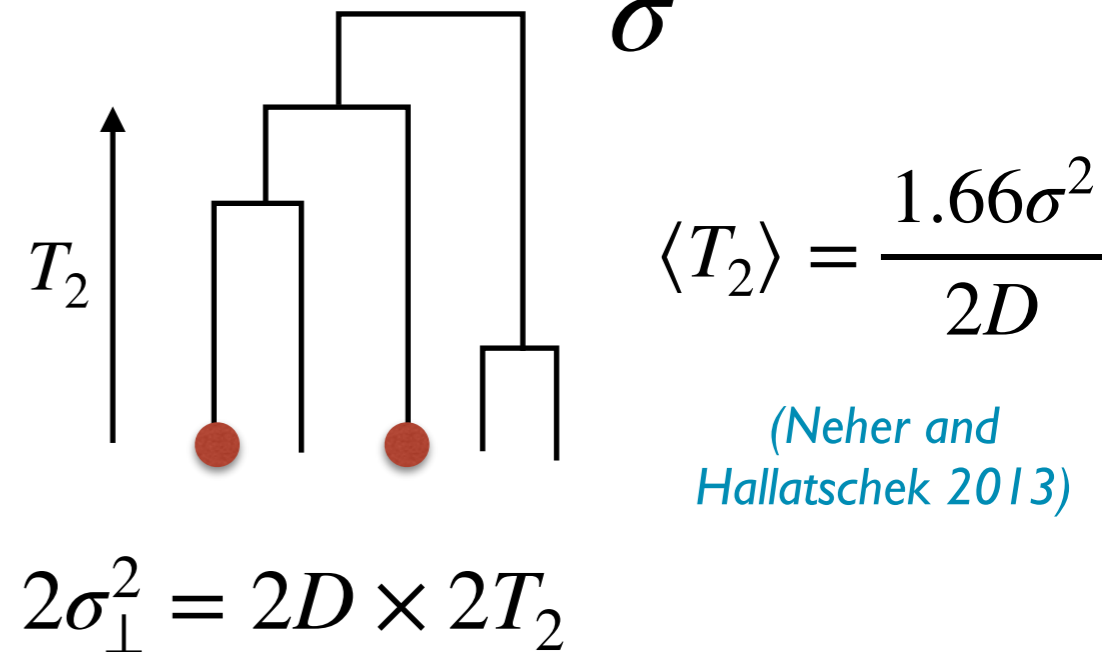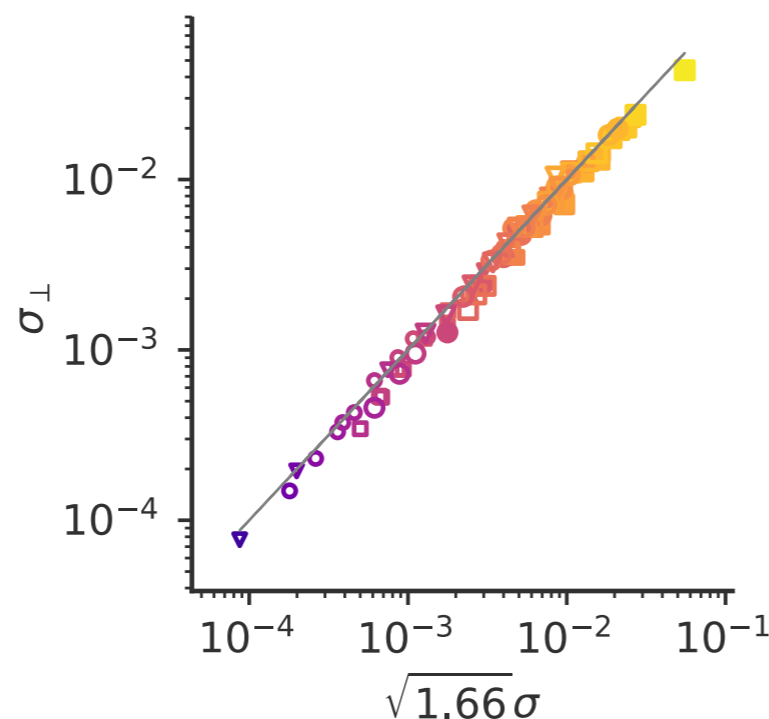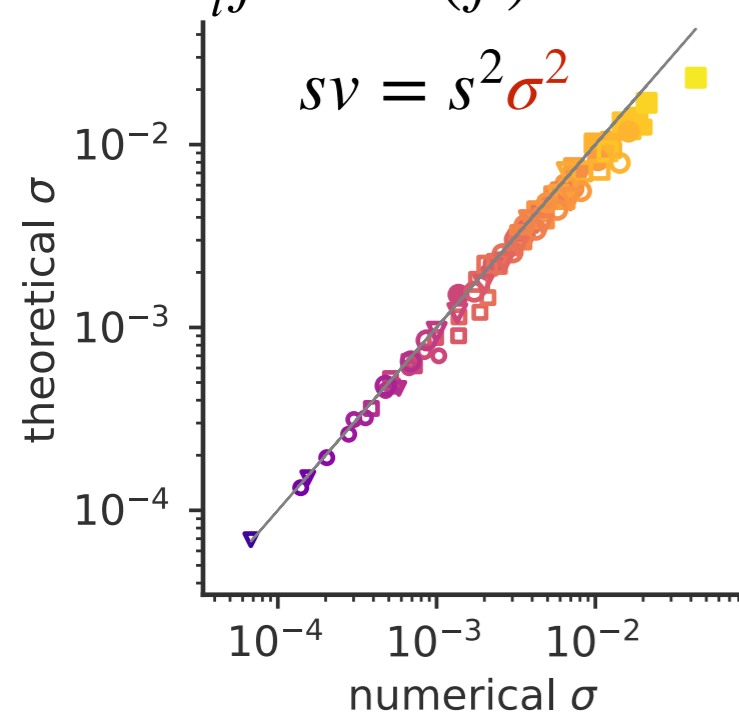
$\rightarrow$ full charaterisation of the wave

Position of the wave: effective EOM

- motion driven by fitness gradient

$$\frac{d\vec{x}}{dt} = \vec{v} = v(N)\frac{\vec{\nabla} f}{|\vec{\nabla} c|} + \text{genetic drift}$$

inertia

angular diffusion

- fitness $\iff$ immune coverage

$$c(\mathbf{x}, t) = \int_{-\infty}^{t} \frac{dt'}{\tau} e^{-\frac{t-t'}{\tau} - |\mathbf{x} - \mathbf{x}(t')|}$$

$$\tau = \frac{MN_h}{N} \sim [\text{incidence rate}]^{-1}$$
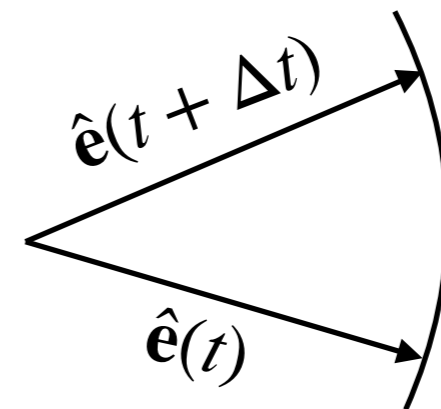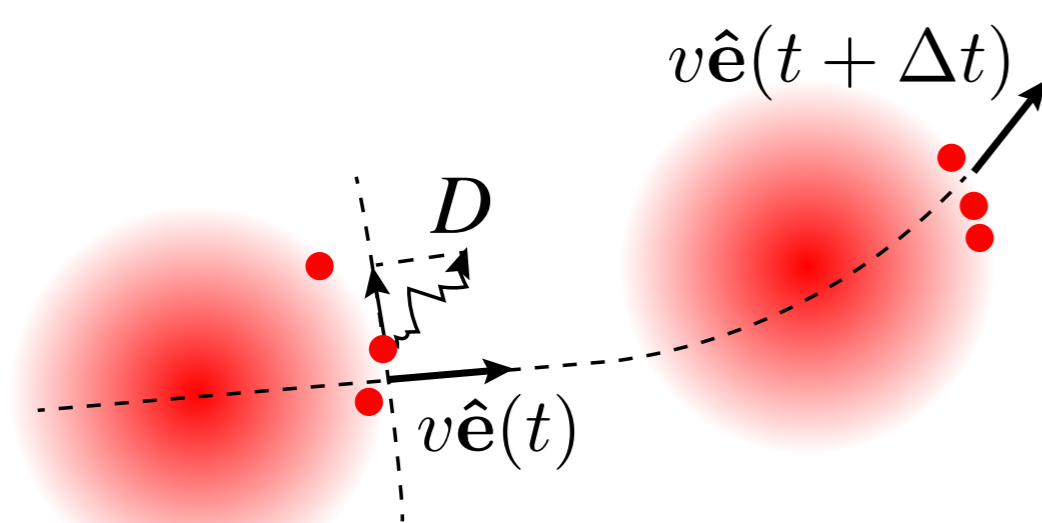
- drift in perpendicular direction reorients wave

$$\hat{\mathbf{e}}(t) \sim \partial_t \mathbf{x} / |\partial_t \mathbf{x}|$$

$v\hat{\mathbf{e}}(t + \Delta t)$

D

$v\hat{\mathbf{e}}(t)$
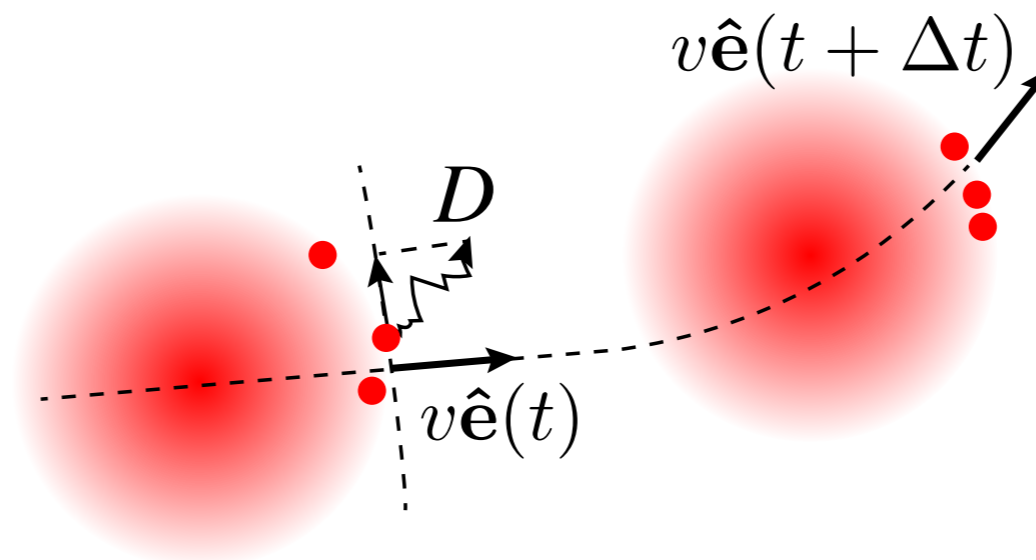
$$\langle \hat{\mathbf{e}}(t) \cdot \hat{\mathbf{e}}(t + \Delta t) \rangle = e^{-\Delta t / t_{\text{persist}}}$$

Angular diffusion and persistance time

$$\langle \hat{\mathbf{e}}(t) \cdot \hat{\mathbf{e}}(t + \Delta t) \rangle = e^{-\Delta t/t_{\text{persist}}}$$

effectiv   **f orientatio**

$$t_{\text{persist}} = \frac{1}{4D(1 + (v\tau)^{-1})^2} = \qquad (R_0^{1/M} -$$

$$= \frac{1}{4D} R_0^{-2/M}$$

time for a **single virus**
to escape immunity

$= \sqrt{8D}$

$v \to$ small

- large R_0 or small M $\to$ speeds up wave but small t_{persist}

$10r$

$20v\tau$

r

$0.5v\tau$

$v\hat{\mathbf{e}}(t + \Delta t)$

$\delta \mathbf{x}_\perp$

$v\hat{\mathbf{e}}(t)$

$\to$ for fixed N_h epidemic waves move faster and
change course more often (less predictable)

$\langle \hat{\mathbf{e}}(t) \cdot \hat{\mathbf{e}}(t + \Delta t) \rangle = e^{-\Delta t/t_{\text{persist}}}$

t_{persist}

theory

simulations

Angular diffusion and persistance time
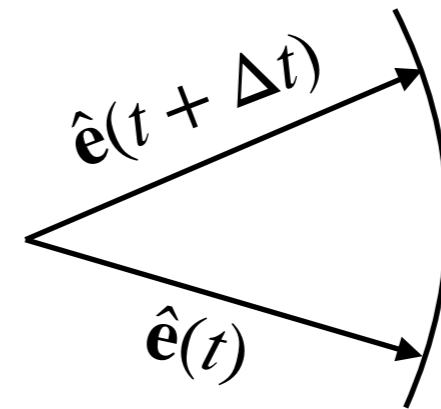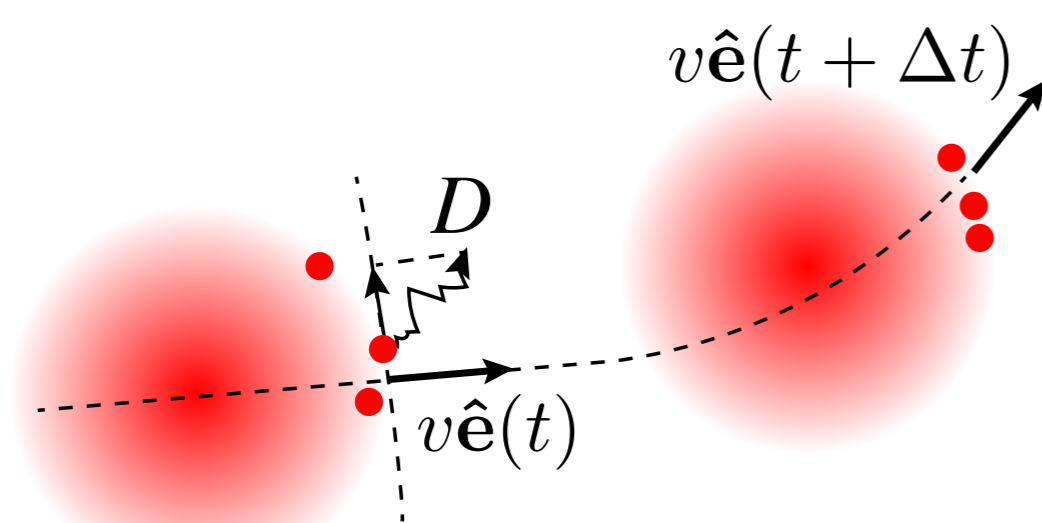
at large times

$$\partial_t \hat{\mathbf{e}} = \sqrt{8D} \left(1 + (v\tau)^{-1}\right) \boldsymbol{\xi}_\perp(t)$$

$$\langle \hat{\mathbf{e}}(t) \cdot \hat{\mathbf{e}}(t + \Delta t) \rangle = e^{-\Delta t/t_{\text{persist}}}$$

effective diffusion in space of orientations

$$t_{\text{persist}} = \frac{1}{4D(1 + (v\tau)^{-1})^2} =$$

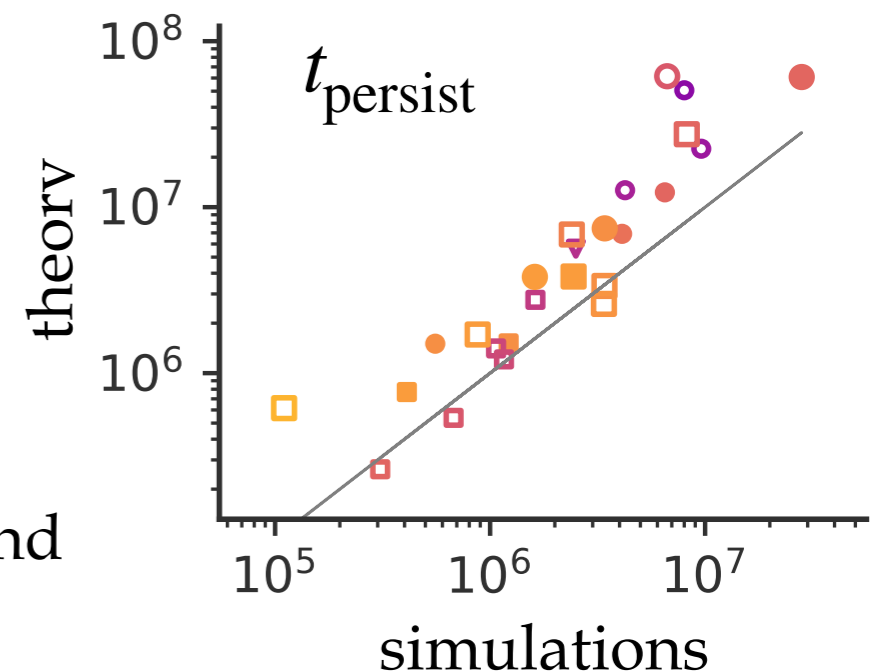$$= \frac{1}{4D} R_0^{-2/M} \gg \frac{1}{v} \sim \frac{1}{D^{2/3-\epsilon}} = t_{\text{escape}} \gg t_{\text{coalescence}} \sim \frac{1}{D^{1/3-\epsilon}}$$

time for a **single virus**
to escape immunity

*driven by neutral
mutations (slow)* $\perp$

time for the **viral
population** to escape
immunity

*driven by selected
mutations (fast)* $||$

← account for immune
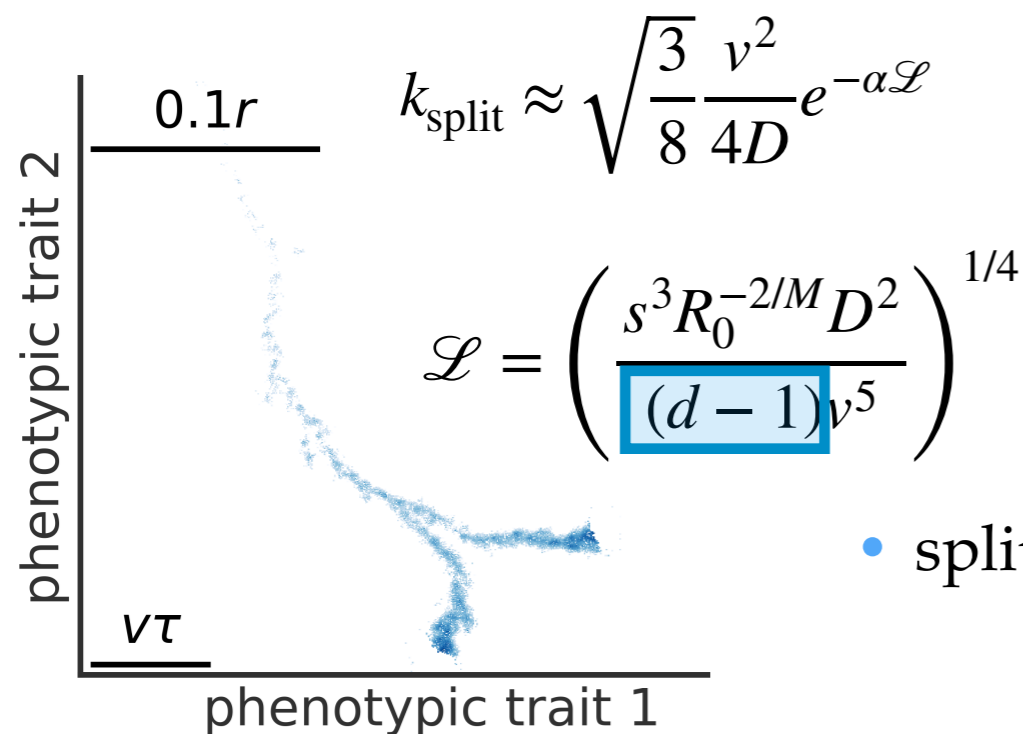system memory

Predictability and splitting

- can one predict where the virus is going?

- $t_{\text{predict}} : \Delta \mathbf{x}_\perp \sim \text{cross} - \text{reactivity range}$

$$t_{\text{predict}} \sim [8(d-1)D/3]^{-1/3} T^{2/3}$$

- higher dimensions $\rightarrow$ more possibilities for deviation from course $\rightarrow$ less predictable

$t \gg T$

$$\partial_t^2 \mathbf{x}_\perp = \frac{\sqrt{8D}}{T} \xi_\perp(t)$$

$$\Delta \mathbf{x}_\perp \sim \frac{8(d-1)D}{3T^2} t^3$$

- d small: $t_{\text{persist}} > t_{\text{predict}} > t_{\text{escape}}$

- d large: $t_{\text{predict}} \ll t_{\text{escape}}$

- wave split: when 2 sublineages reach $\Delta \mathbf{x}_\perp \sim 1$ (cross-reactivity range)

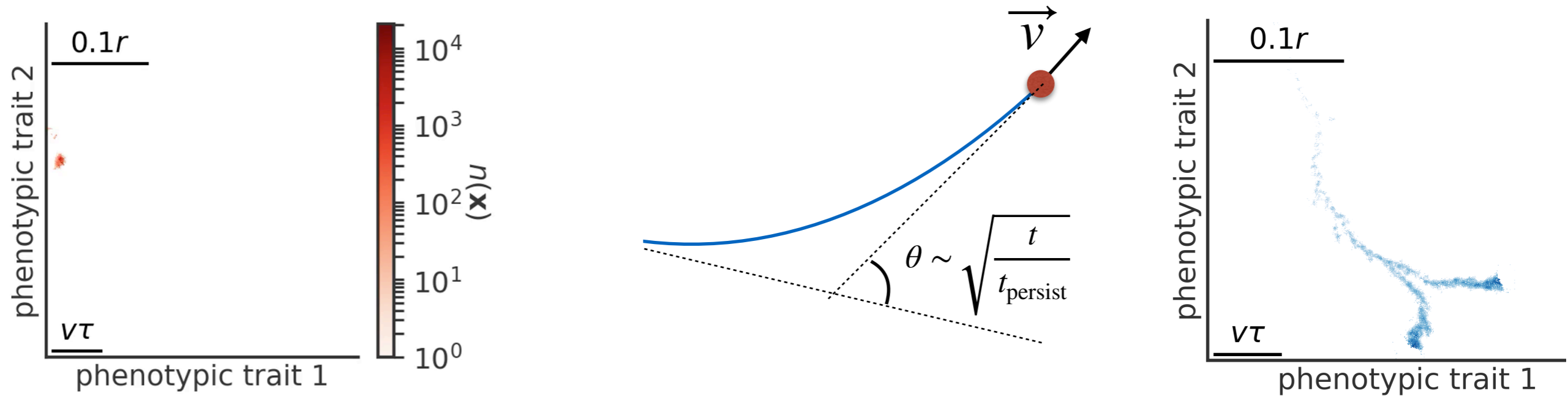$$k_{\text{split}} \approx \sqrt{\frac{3}{8}} \frac{v^2}{4D} e^{-\alpha \mathscr{L}}$$

$$\mathscr{L} = \left(\frac{s^3 R_0^{-2/M} D^2}{(d-1) v^5} \right)^{1/4}$$

0.1r

phenotypic trait 2

$v\tau$

phenotypic trait 1

$\kappa_{\text{split}}/((3/2)^{1/2} v^2/8D)$

$\mathscr{L}$
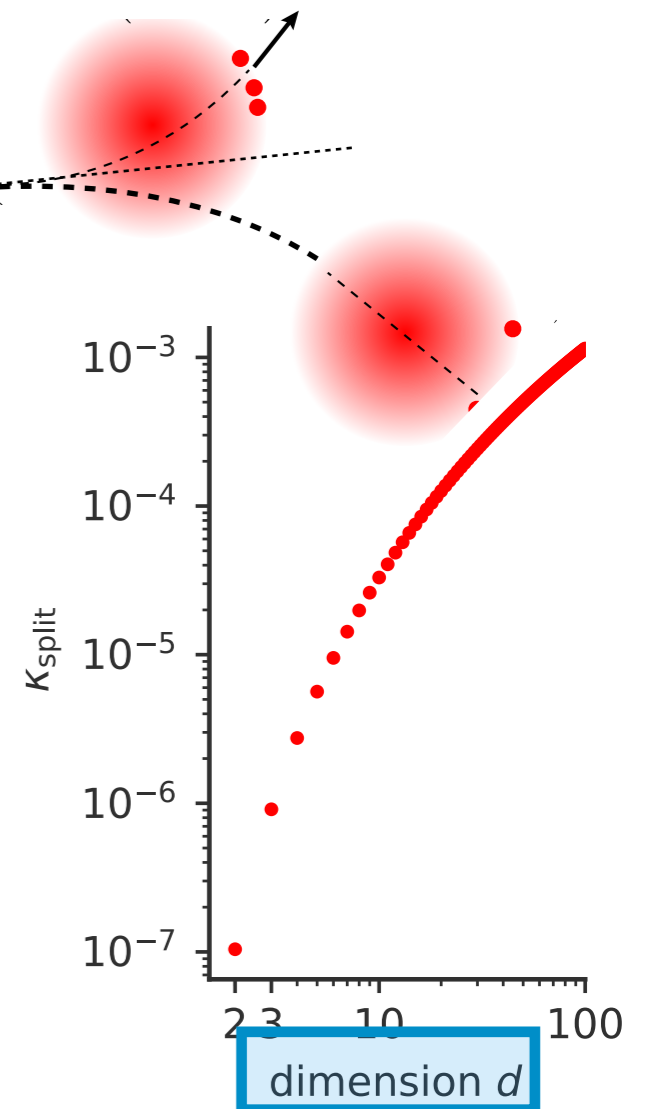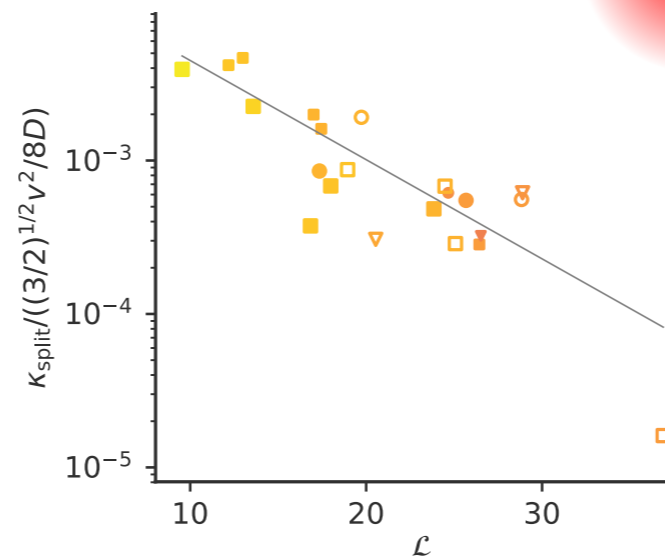
κ_{split}

dimension d

- splitting rate grows with dimension
  $\rightarrow$ departure from canalised evolution easier

- splitting less likely than deflections

(see Yan, Neher and Shraiman 2019 for infinite dim.)

analytical theory of co-evolution in phenotypic space

$$v \approx D^{2/3} s^{1/3} \left[24 \ln(N_h v D^{1/3} s^{5/3}) \right]^{1/3} \gg D$$

$$t_{\text{persist}} = \frac{1}{4D} R_0^{-2/M}$$

$$k_{\text{split}} \approx \sqrt{\frac{3}{8}} \frac{v^2}{4D} e^{-\alpha \mathscr{L}}$$

$$\mathscr{L} = \left(\frac{s^3 R_0^{-2/M} D^2}{(d-1) v^5} \right)^{1/4}$$

effect of immune system

$$t_{\text{split}} \gg t_{\text{persist}} \gg t_{\text{escape}} \gg t_{\text{coalescence}}$$

$$t_{\text{predict}}$$