Affinity maturation for an optimal balance between long-term immune coverage and short-term resource constraints — SI Appendix

Victor Chardès, Massimo Vergassola, Aleksandra M. Walczak,* and Thierry Mora*
Laboratoire de physique de l'École normale supérieure, CNRS, PSL University, Sorbonne Université, and Université de Paris, 75005 Paris, France

Appendix A: Mean-field de novo coverage

Here we show how the infection cost function defined in the main text,

$$I_t = \min\left[\phi, \left(\sum_{x\in P_{t-1}} c_{x,t} f(x,a_t)\right)^{-\alpha}\right], \tag{A1}$$

may be derived as the mean-field limit of a repertoire with memory and de novo compartments.

In addition to the evolving memory repertoire P_t already described in the main text, we define a de novo response made of random receptors $\mathcal{N}$, distributed uniformly with density ρ. Viruses may be recognized by either the memory or de novo clonotypes. The de novo coverage is defined as:

$$C_{\text{denovo}}(a_t) = \sum_{x\in\mathcal{N}} f(x,a_t), \tag{A2}$$

and the memory coverage as before:

$$C(a_t) = \sum_{x\in P_{t-1}} c_{x,t} f(x,a_t). \tag{A3}$$

(In this convention, each de novo clonotype has size one in arbitrary units.)

Depending on the values of these coverages, the system will choose to use either the de novo response, or an existing memory. In this decision, we factor in the fact that using the de novo response is more costly, which we account for using a prefactor $\beta < 1$. The cost is then defined as:

$$L_t = \max\left[\beta C_{\text{denovo}}(a_t), C(a_t)\right]^{-\alpha}. \tag{A4}$$

We can simplify this expression in the limit where de novo responding clonotypes are very numerous, but each offer weak coverage. In the limit of high density of de novo responding cells, $\rho\to\infty$, the coverage self-averages to its mean value:

$$C_{\text{denovo}} \approx \langle C_{\text{denovo}}\rangle = \rho\int d^dx\, f(x,a_t) = \rho U_d(q) r_0^d, \tag{A5}$$

with

$$U_d(q) = \int d^dy\, e^{-\|y\|^q} = S_d\int r^{d-1}dr\, e^{-r^q}, \tag{A6}$$

where we have done the change of variable $x = a_t + yr_0$, and where $S_d = 2\pi^{d/2}/\Gamma(d/2)$ is the surface area of the unit sphere.

Taking the $\rho\to\infty$ and $\beta\to 0$ limits, while keeping $\beta\rho$ finite, corresponds to a dense de novo response but where each de novo responding cell weakly covers the antigenic space. In this limit we recover the model of the main text

$$I_t = \min\left[\phi, C(a_t)^{-\alpha}\right], \tag{A7}$$

with $\phi = (\rho\beta U_d(q) r_0^d)^{-\alpha}$.

Appendix B: Transition from monoclonal to de novo phase at $\sigma_v = 0$

Here we derive an expression for the phase boundary between the monoclonal and polyclonal phases in the limit $\sigma_v = 0$, where the virus does not move.

In the special case where $\mu = 0$, clonotypes cannot multiply. At each time step, a number m_n of new clonotypes at created at $a_n = \text{const}$, distributed according to a Poisson law of mean $\bar{m}$. This number is added to existing clonotypes, of which a random fraction γ survives. If the previous number of clonotypes, M_n, is Poisson distributed with mean $\bar{M}_n$, the number of surviving ones M'_n is also Poisson distributed with mean $\gamma\bar{M}_n$ (since subsampling a Poisson-distributed number still gives a Poisson law). Then, the new number of clonotypes, $M_{n+1} = M'_n + m_n$, is also Poisson distributed, with the recurrence relation:

$$\bar{M}_{n+1} = \gamma\bar{M}_n + \bar{m}. \tag{B1}$$

At steady state, we have

$$\bar{M}_n \to \frac{\bar{m}}{1-\gamma}. \tag{B2}$$

Since all clonotypes are at $x = a_n$, the coverage is $C(a_n) = M$, so that the expected cost reads:

$$\mathcal{L} = \phi \exp\left(-\frac{\bar{m}}{1-\gamma}\right) + \sum_{M=1}^{+\infty} \frac{1}{M^\alpha} \exp\left(-\frac{\bar{m}}{1-\gamma}\right) \frac{1}{m!}\left(\frac{\bar{m}}{1-\gamma}\right)^M + \kappa\bar{m}. \tag{B3}$$

To find the transition from monoclonal to de novo response, $\bar{m} = 0$, we need to find the value of ϕ for which $\partial\mathcal{L}/\partial\bar{m}$ changes sign at $\bar{m} = 0$: if this derivative is positive, it is better to have $\bar{m} = 0$ (since the function is convex); if it is negative, there is benefit to be gained by increasing $\bar{m} > 0$. The condition:

$$\left.\frac{\partial\mathcal{L}}{\partial\bar{m}}\right|_{\bar{m}=0} = \kappa - \frac{\phi}{1-\gamma} + \frac{1}{1-\gamma} \tag{B4}$$

gives the transition point

$$\phi_c = 1 + \kappa(1-\gamma) \tag{B5}$$

For $\mu > 0$, we redefine M_n as the sum of all clonotype sizes, which is equal to the coverage, $C = M_n = \sum_{x\in P_n} c_{x,n}$. The recurrence relation is replaced by:

$$\bar{M}_{n+1} = \gamma(1+\mu)\bar{M}_n + \bar{m}. \tag{B6}$$

For $\gamma(1+\mu) > 1$, this number explodes, so that M is infinite, reducing the infection cost to 0 regardless of $\bar{m}$. The transition point is then

$$\phi_c = 0. \tag{B7}$$

For $\gamma(1+\mu) < 1$, $\bar{M}_n$ reaches a steady state value,

$$\bar{M}_n \to \frac{\bar{m}}{1-\gamma(1+\mu)}. \tag{B8}$$

Although M_n is not strictly distributed according to a Poisson law, it is still a good approximation, so that we can repeat the same argument as with $\mu = 0$,

$$\phi_c \approx 1 + \kappa(1-\gamma(1+\mu)). \tag{B9}$$

Appendix C: Analytical results in a solvable model

1. Model definition and main results

To gain insight into the transitions observed in the phase diagram of Fig. 2, we can make a series of simplifications and approximations about the model that allow for analytical progress. We assume a step function for the cross-reactivity function $f(x,a) = 1$ for $\|x-a\| \leq r_0$, and 0 otherwise, corresponding the limit $q = \infty$. Likewise, we assume

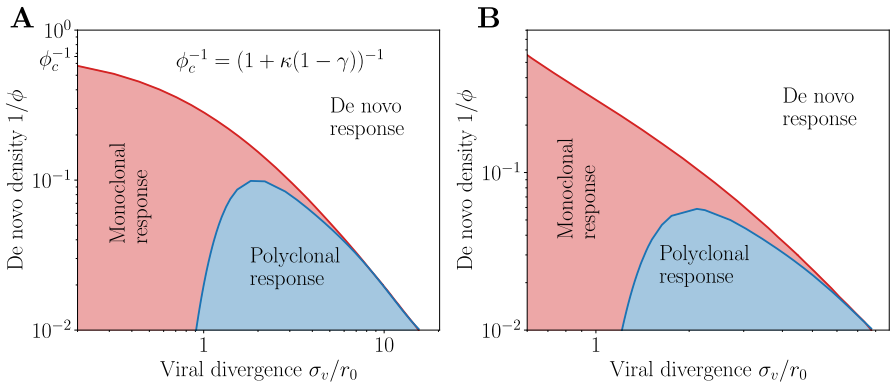

FIG. S1: **Phase diagram for various parameters.** **A.** Phase diagram for parameters $\mu = 0$, $\gamma = 0.85$ and $d = 2$. We observe the upper transition point $\phi_c = (1+\kappa(1-\gamma))^{-1}$ at $\sigma_v = 0$. **B.** Phase diagram for parameters $\mu = 0.5$, $\gamma = 0.85$ and $d = 3$. Since $\gamma(1+\mu) > 1$ the transition point $\phi_c = \infty$. In both **A** and **C** we observe that the phase diagram retains the same shape. In this panel $\alpha = 1$, $q = 2$ and $\kappa = 0.5/(1-\gamma)$.

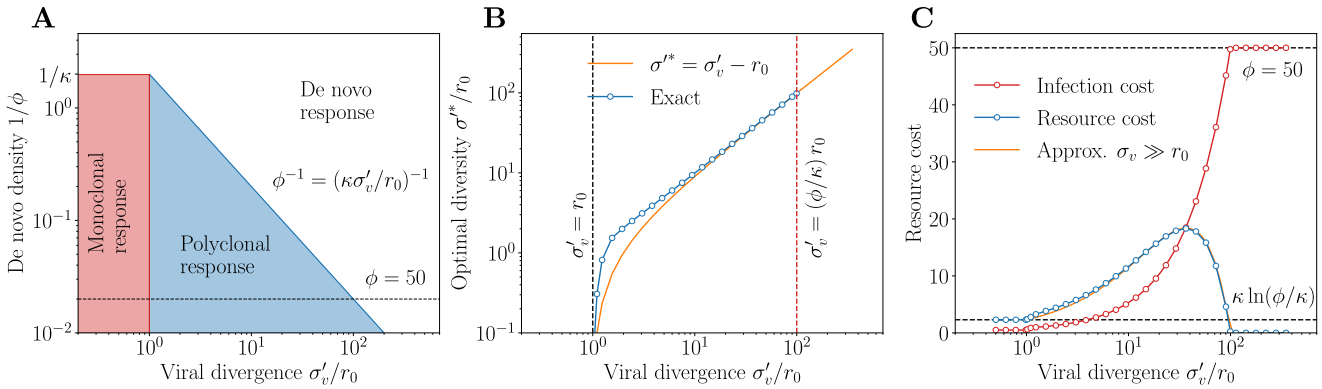

FIG. S2: **Analytical solution of a tractable model.** **A.** Exact phase diagram in $d = 1$ for the simplified model ($q = \infty$, $\gamma = 0$, and all-or-nothing infection cost). The boundary between monoclonal is given by $\sigma'_v = r_0$ and the boundary between polyclonal by $\phi^{-1} = (\kappa\sigma'_v/r_0)^{-d}$. **B.** Optimal memory diversity $\sigma'^* \approx \sigma'_v$ and **C.** optimal infection and plasticity costs for as a function of σ'_v for $\phi = 50$. σ' and σ'_v are rescaled versions of the diversity and divergence to match the variances of the original model.

a uniform distribution of viral antigenic mutations $a_{t+1} = a_t + \sigma'_v \eta'_{t+1}$, where η'_t is a random point of the d-dimensional unit ball, with $\sigma'_v = \sigma_v\sqrt{1+2/d}$ (so that the variance is the same as in the Gaussian case), and similarly for memory diversification, with new clonotypes drawn from a uniform distribution is a ball of radius $\sigma' = \sigma\sqrt{1+2/d}$. The infection cost is approximated by an all-or-nothing function, with $I_t = 0$ if there is any coverage $C(a_t) > 0$, and $I_t = \phi$ if $C(a_t) = 0$. We further assume $\gamma = 0$: all clonotypes are discarded at each time step, so that memory may only be used once.

In this simplified version of the model, the phase diagram and optimal parameters can be computed analytically. We first summarize the main results below. In the subsequent paragraphs, we provide detailed derivations in the case of arbitrary dimensions, and also provide additional exact formulas in the one-dimensional case.

One can show (see next paragraphs) that the transition from monoclonal to polyclonal response occurs exactly when the radius of the ball within which viral mutations occur reaches the cross-reactivity radius r_0:

$$\sigma'_v = r_0. \tag{C1}$$

Below this transition ($\sigma'_v < r_0$), the optimal strategy is to have no diversity at all and perfectly target the recognized antigen a_t, $\sigma'^* = 0$, as any memory cell at a_t will recognize the next infection. In this case the optimal mean number of memories $\bar{m}^* = \ln(\phi/\kappa)$ results in a trade off between the cost of new memories with the risk of not developing any memory at all by minimizing $\phi e^{-\bar{m}} + \kappa\bar{m}$. The transition from the monoclonal response to de novo phases is then given by $\phi = \kappa$, where $\bar{m}^* = 0$.

The polyclonal-to- de novo response transition may also be understood analytically. In the polyclonal response phase, the optimal strategy is, in either of the limits $\sigma'_v \gg r_0$ or $\bar{m}^* \ll 1$ (see next paragraphs):

$$\sigma'^* \approx \sigma'_v - r_0 \tag{C2}$$

$$\bar{m}^* \approx \frac{\sigma'^d_v}{r_0^d} \ln\left(\frac{\phi}{\kappa}\frac{r_0^d}{\sigma'^d_v}\right). \tag{C3}$$

In particular this result becomes exact at the transition from polyclonal to de novo response, where $\bar{m}^* = 0$. The transition is thus given by:

$$\phi^{-1} = \frac{r_0^d}{\kappa \sigma'^d_v}. \tag{C4}$$

The polyclonal response is outcompeted by the de novo one when the density of de novo responding cells (ϕ^{-1}) becomes larger than the probability density of new strains falling within the cross-reactivity radius ($r_0^d/{\sigma'_v}^d$), rescaled by the memory cost coefficient κ^{-1}.

Fig. S2 shows the resulting phase diagram, as well as the optimal diversity σ'^* and predicted costs for a fixed ϕ and $d = 1$. These predictions reproduce the main features of the full model, in particular the scaling of the immune diversity σ with σ_v (Fig. 2D vs. Fig. S2B) and the general shape of the optimal memory size $\bar{m}^*$ (Fig. 2E vs. Fig. S2C), which first increases as the virus becomes more divergent, to later drop to zero as memory becomes too costly to maintain and the system falls into the de novo phase.

2. General formulation

Define $P_{\text{hit}}(\sigma', r_0, r)$ as the probability that a random memory will recognize the next infection at distance r, i.e. the probability that a random point in the ball of radius σ' and a point at distance r from its center are at distance $\leq r_0$. The probability that none of m clonotypes recognize the virus, where m is drawn from a Poisson distribution of mean $\bar{m}$, reads:

$$P_{\text{miss}}(\sigma', r_0, r) = \sum_m e^{-\bar{m}} \frac{\bar{m}^m}{m!} (1 - P_{\text{hit}}(\sigma', r_0, r))^m = e^{-\bar{m} P_{\text{hit}}(\sigma', r_0, r)}. \tag{C5}$$

The best strategy maximizes this probability, averaged over the location of the next infection, over σ':

$$\bar{P}_{\text{miss}}(\sigma', \bar{m}, r_0, \sigma'_v) = \left\langle e^{-\bar{m} P_{\text{hit}}(\sigma', r_0, r)} \right\rangle_{\mathcal{B}(\sigma'_v)} = \frac{1}{\sigma'^d_v V_d} \int_0^{\sigma'_v} S_d r^{d-1} dr \, e^{-\bar{m} P_{\text{hit}}(\sigma', r_0, r)} \tag{C6}$$

where $\mathcal{B}_{\sigma'_v}$ is the ball of radius σ'_v, $V_d = \pi^{d/2}/\Gamma(d/2+1)$ is the volume of a unit ball, and $S_d = 2\pi^{d/2}/\Gamma(d/2)$ the area of the unit sphere.

Then the expected overall cost reads:

$$\mathcal{L} = \phi \bar{P}_{\text{miss}}(\sigma', \bar{m}, r_0, \sigma_v) + \kappa \bar{m} = \phi \left\langle e^{-\bar{m} P_{\text{hit}}(\sigma', r_0, r)} \right\rangle_{\mathcal{B}(\sigma'_v)} + \kappa \bar{m}. \tag{C7}$$

3. Exact location of the phase transitions, and approximate solution in the polyclonal phase

The location of the optimal σ' may be rigorously bounded from above and below. If $\sigma' < \sigma'_v - r_0$, then only part of the future positions of the virus are covered, so increasing σ' can bring no harm. Likewise, for $\sigma' > \sigma'_v + r_0$, memory covers parts of the antigenic space that have no chance of harboring the next virus, so that decreasing σ' is also always advantageous. Thus, the optimum σ'^* must satisfy:

$$\sigma'_v - r_0 \leq \sigma'^* \leq \sigma'_v + r_0. \tag{C8}$$

As already argued in the main text, when $\sigma'_v < r_0$, there is clearly no benefit to having $\sigma' > 0$, so the optimum is reached at $\sigma' = 0$. (C8) further shows that if $\sigma'_v > r_0$, then $\sigma'^* > 0$, so that a polyclonal phase is optimal. As a consequence, the transition from the monoclonal to polyclonal phases happens exactly at

$$\text{Monoclonal to polyclonal:} \quad \sigma'_v = r_0. \tag{C9}$$

In the monoclonal phase, memory always recognizes the next virus. The only risk of paying ϕ is when no memory is created, which happens with probability $e^{-\bar{m}}$, so that the cost reads:

$$\mathcal{L} = \phi e^{-\bar{m}} + \kappa \bar{m}. \tag{C10}$$

The optimal $\bar{m}^* = \ln(\phi/\kappa)$ cancels at the monoclonal-to- de novo response transition:

$$\text{Monoclonal to de novo response :} \quad \phi = \kappa. \tag{C11}$$

In the polyclonal phase, we could not find a general analytical solution, but there are two limits in which the solution may be calculated. The first limit is when $\sigma_v \gg r_0$. In that case, (C8) implies $\sigma'^* \approx \sigma'_v$, and

$$P_{\text{hit}}(\sigma'^*, r_0, r) \approx \frac{r_0^d}{\sigma_v'^d}, \tag{C12}$$

which doesn't depend on r. Then, minimizing

$$\mathcal{L} \approx \phi e^{-\bar{m}P_{\text{hit}}} + \kappa \bar{m} \tag{C13}$$

with respect to $\bar{m}$ yields:

$$\bar{m}^* = \frac{1}{P_{\text{hit}}} \ln\left(\frac{\phi}{\kappa} P_{\text{hit}}\right) = \frac{\sigma_v'^d}{r_0^d} \ln\left(\frac{\phi}{\kappa} \frac{r_0^d}{\sigma_v'^d}\right). \tag{C14}$$

The second limit in which things simplify is for small $\bar{m}$. Then the exponential in (C6) may be expanded at first order, yielding:

$$\mathcal{L} = \phi \left(1 - \bar{m} \langle P_{\text{hit}}(\sigma', r_0, r)\rangle_{\mathcal{B}(\sigma'_v)}\right) + \kappa \bar{m}, \tag{C15}$$

where $\langle\cdot\rangle_r$ is the mean of over the ball of radius σ'_v. Minimizing with respect to σ' is equivalent to maximizing $\langle P_{\text{hit}}(\sigma', r_0, r)\rangle_r$, which is the probability that a random point in the ball of radius σ' and a random point in the ball of radius σ'_v are separated by less than r_0. This probability is maximized for any $\sigma' \leq \sigma_v - r_0$, where it is equal to $(r_0/\sigma'_v)^d$. Increasing σ' beyond $\sigma_v - r_0$ can only lower the probability of recognition. Thus:

$$\min_{\sigma'} \mathcal{L} = \phi + \bar{m}\left(\kappa - \phi \frac{r_0^d}{\sigma_v'^d}\right). \tag{C16}$$

This gives us the condition for the transition from polyclonal to de novo response, where $\bar{m}^* = 0$. This happens when

$$\text{Polyclonal to de novo response :} \quad \phi = \kappa \frac{\sigma_v'^d}{r_0^d}. \tag{C17}$$

This condition gives us an exact expression for the location of the transition.

4. Exact solution in dimension 1

For $d = 1$, the cost $\mathcal{L}$ in (C7) may be calculated analytically, by using exact expressions of $P_{\text{hit}}(\sigma', r_0, r)$. When $\sigma'_v \leq r_0$, the optimal σ' is zero as explained in the main text. When $\sigma'_v > r_0$, we distinguish two cases: $r_0 < \sigma'_v \leq 2r_0$, and $\sigma'_v > 2r_0$.

a. Case $r_0 < \sigma'_v \leq 2r_0$. Since we know that the optimal σ' is between $\sigma'_v - r_0$ and $\sigma'_v + r_0$, we focus on that range. Then there are two subcases for σ'.

If $\sigma'_v - r_0 < \sigma' \leq r_0$, there are two contributions to the integral of $\bar{P}_{\text{miss}}$ over the position of the virus r. Either $r \leq r_0 - \sigma'$, then all memories recognize the virus, $P_{\text{hit}} = 1$; or $r_0 - \sigma' < r < \sigma'_v < r_0 + \sigma'$, in which case the recognition probability is given by the normalized intersection of two balls at distance r of radii σ' and r_0,

$$P_{\text{hit}} = \frac{\sigma' + r_0 - r}{2\sigma'}. \tag{C18}$$

Thus we obtain doing the integral over r in (C6):

$$\bar{P}_{\text{miss}}(\sigma', \bar{m}, r_0, \sigma'_v) = \frac{1}{\sigma'_v}\left[(r_0 - \sigma')e^{-\bar{m}} + \int_{r_0-\sigma'}^{\sigma'_v} \exp\left(-\bar{m}\frac{\sigma' + r_0 - r}{2\sigma'}\right) dr\right] \text{ if } \sigma'_v - r_0 < \sigma' \leq r_0. \tag{C19}$$

If $r_0 < \sigma' \leq \sigma'_v + r_0$, there are also two contributions. Either $r \leq \sigma' - r_0$, in which case there is no boundary effect, and the recognition probability is just $P_{\rm hit} = r_0/\sigma'_v$; or $\sigma' - r_0 < r \leq \sigma'_v \leq \sigma' + r_0$, in which case we have again (C18). Performing the integration in (C6) we obtain:

$$\bar{P}_{\rm miss}(\sigma', \bar{m}, r_0, \sigma'_v) = \frac{1}{\sigma'_v}\left[\exp\left(-\bar{m}\frac{r_0}{\sigma'}\right)(\sigma' - r_0) + \int_{\sigma'-r_0}^{\sigma'_v} \exp\left(-\bar{m}\frac{\sigma' + r_0 - r}{2\sigma'}\right) dr\right] \text{ if } r_0 < \sigma' \leq \sigma'_v + r_0. \quad \text{(C20)}$$

Numerical analysis shows that (C19) admits a minimum as a function of σ' in its interval of validity, $\sigma'_v - r_0 < \sigma'^* \leq r_0$, while (C20) is always increasing.

b. Case $\sigma'_v > 2r_0$. In this case, there is only a single subcase in the range of interest $\sigma'_v - r_0$ and $\sigma'_v + r_0$. This case is the same as the previous one considered, and the result is given by the same formula (C20). However, for $\sigma'_v > 2r_0$, this expression now admits a minimum $\sigma'_v - r_0 < \sigma'^* \leq \sigma'_v + r_0$.

We recover that in both cases (a and b), in the limit $\bar{m} \to 0$, this minimum is reached at $\sigma'_v - r_0$.

Appendix D: Population dynamics in sequential immunization

1. Clonotype growth and decay as a first-passage problem

We now want to study clonotype proliferation induced by a recall response. We focus on the limit of small mutation rates $\sigma_v \ll r_0$. Within this regime, the system is in the monoclonal phase with $\sigma^* = 0$. We can therefore focus on the case of a single clonotype at position $x = 0$ on the phenotypic space, and ask how successive challenges will modify its size. (Different initial conditions will only change the prefactor in front of the exponential modes in the distribution of first passage times, so the large time behavior of this probability distribution will be the same as discussed below.)

The clonotype has an initial size $c = 1$, and the virus drifts away from $x = 0$ with viral divergence σ_v. In the general model, cells have probability γ to survive from one challenge to the other. Proliferation is taken to be proportional to the cross reactivity radius, $\mu e^{-(r/r_0)^q}$. The population dynamics is thus given by the approximate recursion:

$$n_{t+1} \approx n_t \gamma \left[1 + \mu e^{-(r/r_0)^q}\right], \quad \text{(D1)}$$

where we have neglected birth-death noise. We can further simplify this equation to $n_{t+1} = \gamma(1 + \mu\Theta(r - r^*))n_t$, where $r^* = r_0 \ln(\gamma\mu/(1-\gamma))^{1/q}$ is defined as the radius at which the net fold-change factor crosses 1, i.e. when birth is exactly compensated by death. This means that, as long as the virus is within distance r^*, the clonotype grows with fold-change factor $\sim \Gamma$. As soon as it reaches r^*, and neglecting possible returns below r^* (which happen with probability 1 for $d \leq 2$, but with a frequency that does not affect the overall decay), it will decay with fold-change factor $\sim \gamma$. The problem is thus reduced to determining the first-passage time of the viral antigenic location at radius r^*.

We use a continuous approximation corresponding to a slowly evolving strain, $\sigma_v \ll r_0$:

$$a(0) = 0, \qquad da = \frac{\sigma_v}{\sqrt{d}} dW, \quad \text{(D2)}$$

where W is a Wiener process. The radius, given by $r(t) = |a(t)|$, behaves on average as:

$$\langle r(t)^2 \rangle = t\sigma_v^2. \quad \text{(D3)}$$

The time it takes for $r(t)$ to reach r^*, denoted by t^*, is approximately given by $\langle t^* \rangle \sim (r^*/\sigma_v)^2$.

The tail of the distribution for this first-passage time is dominated by rare events when the virus mutates less than expected between infections, leading to larger episodes of growth. We will show in the next two sections that the distribution of these exceptionally long t^* has an exponential tail:

$$P(t^* > t) \sim e^{-t/t_s}, \quad t_s \sim \langle t^* \rangle \sim \frac{r^{*2}}{\sigma_v^2}. \quad \text{(D4)}$$

This translates into a power-law tail for the peak clonotype abundance,

$$p(n^*) \sim \frac{1}{n^{*1+\beta}}, \quad \text{with } \beta \sim \frac{\sigma_v^2}{r^{*2} \ln \Gamma}. \quad \text{(D5)}$$

The same scaling holds for the distribution of all abundances, since the peak determines the rest of the trajectory.

Within the same simplified picture, the lifetime t_l of a clonotype is the sum of the time it takes to reach the peak, t^*, and the decay time until extinction, which is approximately $\ln(n^*)/\ln(1/\gamma)$:

$$t_l = t^* + \frac{\ln(n^*)}{\ln(1/\gamma)} = \left(1 + \frac{\ln \Gamma}{\ln(1/\gamma)}\right) t^*. \tag{D6}$$

Thus, t_l is proportional to t^*, and therefore also exponentially distributed:

$$p(t_l) \sim e^{-\lambda t_l}, \quad \lambda \sim \frac{\sigma_v^2}{r^{*2}} \left(1 + \frac{\ln \Gamma}{\ln(1/\gamma)}\right)^{-1}. \tag{D7}$$

Next we derive in detail the distributions of the first passage time of $r(t)$ to r^* to obtain Eq. D4.

2. First passage time in $d = 1$

The distribution of first passage time, $p(t)$, can be computed solving diffusion with a box of size $2r^*$ in $d = 1$ [1]:

$$p(t) = \sum_{n=0}^{+\infty} \frac{(2n+1)\sigma_v^2 \pi}{2r^{*2}} (-1)^n \exp\left(-\frac{(2n+1)^2 \pi^2 \sigma_v^2 t}{8r^{*2}}\right) \tag{D8}$$

The dominant term ($n = 0$) at long times gives an exponential decay:

$$p(t) \approx \frac{\sigma_v^2 \pi}{2r^{*2}} \exp\left(-\frac{\pi^2 \sigma_v^2 t}{8r^{*2}}\right). \tag{D9}$$

3. First passage in higher dimensions

We define $f(r, t)$ as the probability density that the virus has not yet reached r^* at time t, and is at radius r.

This probability density is solution to the diffusion equation with spherical symmetry and absorbing boundary conditions in arbitrary dimension $d > 1$:

$$\frac{\partial f}{\partial t} = \frac{\sigma_v^2}{2d}\left[\frac{\partial^2 f}{\partial r^2} + \frac{d-1}{r}\frac{\partial f}{\partial r}\right], \qquad f(r^*, t) = 0. \tag{D10}$$

Assuming separation of variables, $f(r, t) = T(t)R(r)$, we have:

$$\frac{T'(t)}{T(t)} = \frac{\sigma_v^2}{2d} \frac{R''(r) + \frac{d-1}{r}R'(r)}{R(r)} \equiv -\frac{\sigma_v^2}{2dr^{*2}}\lambda, \tag{D11}$$

where λ is to be determined later. This implies $T(t) = Ce^{-\lambda \frac{\sigma_v^2}{2dr^{*2}} t}$ where C is a constant. The radial part $R(r)$ is solution to:

$$R''(r) + \frac{d-1}{r}R'(r) = -\frac{\lambda}{r^{*2}}R(r). \tag{D12}$$

For $d = 1$ this equation reduces to a harmonic equation and we recover the above solution in 1D. Using the change of variable $R(r) = r^{1-d/2}g(r)$ we derive the following equation:

$$r^2 g''(r) + rg'(r) + \left(\lambda \frac{r^2}{r^{*2}} - \left(\frac{d}{2} - 1\right)^2\right) g(r) = 0. \tag{D13}$$

Changing the variable $x = \sqrt{\lambda} r/r^*$, the function $\tilde{g}(x) = g(xr^*/\sqrt{\lambda})$ is solution to the Bessel differential equation of order $d/2-1$. It can therefore be written as a superposition of a Bessel function of the first kind and a Bessel function of the second kind, both of order $d/2 - 1$. The Bessel function of the second kind having a singularity at $x = 0$, our solution is only given by the Bessel function of the first kind $g(xr^*/\sqrt{\lambda}) = BJ_{d/2-1}(x)$. The radial function R now reads:

$$R(r) = Br^{1-d/2} J_{d/2-1}(\sqrt{\lambda} r/r^*). \tag{D14}$$

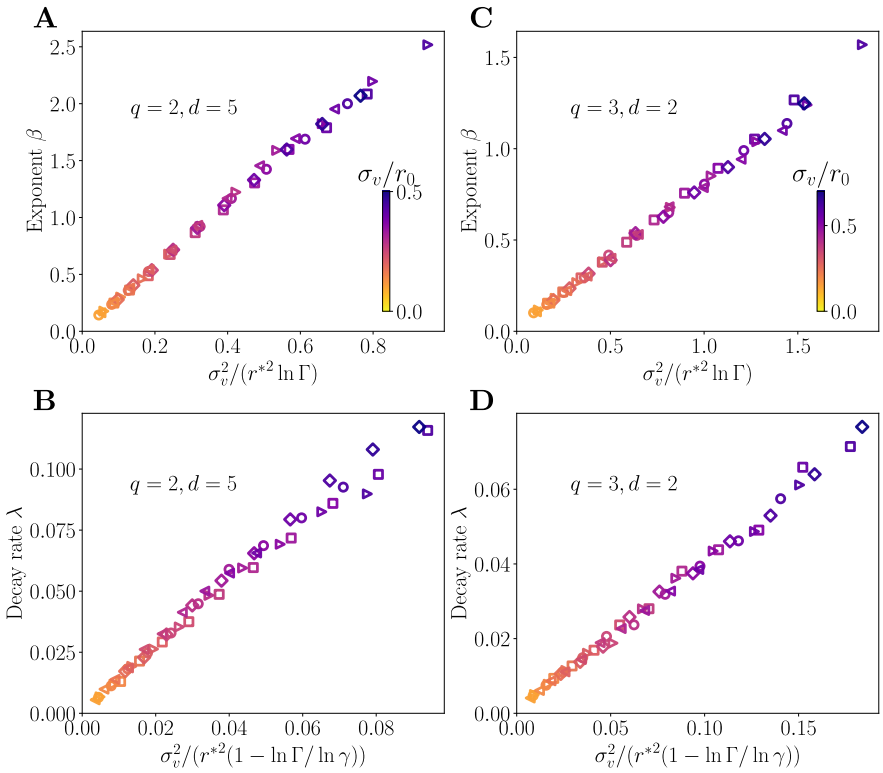

FIG. S3: **Scaling relations for various parameters. A-B.** Power law exponent β and lifetime decay rate λ in dimension $d = 5$ with a Gaussian cross-reactivity kernel with $q = 2$. **C-D.** Power law exponent β and lifetime decay rate λ in dimension $d = 2$ with a cross-reactivity kernel with $q = 3$. The different parameters used are $(\gamma = 0.82, \mu = 0.65)$ i.e. $\Gamma = 1.353$ (diamonds), $(\gamma = 0.8, \mu = 0.62)$ i.e. $\Gamma = 1.296$ (squares), $(\gamma = 0.85, \mu = 0.5)$ i.e. $\Gamma = 1.275$ (circles), $(\gamma = 0.87, \mu = 0.4)$ i.e. $\Gamma = 1.21$ (triangles $>$), $(\gamma = 0.9, \mu = 0.35)$ ie. $\Gamma = 1.21$ (triangles $<$). The strategy is optimized with $\phi = 100$, $\kappa = 0.5/(1-\gamma)$. We used $\alpha = 1$ throughout.

The absorbing boundary condition at $r = r^*$ gives us the condition $J_{d/2-1}(\sqrt{\lambda}) = 0$, which has an infinite number of solutions $j_{0,d/2-1}, ..., j_{n,d/2-1}, ...$, so that λ can take values

$$\lambda_n = j^2_{n,d/2-1}. \tag{D15}$$

The general solution to (D10) is given as a linear combination of all possible modes, with coefficients C_n determined from boundary conditions and the Dirac delta initial condition, $f(r, 0) = \delta(r)$:

$$f(r,t) = \sum_{n=0}^{+\infty} C_n r^{1-d/2} J_{d/2-1}\left(\frac{j_{n,d/2-1}}{r^*} r\right) \exp\left(-\frac{j^2_{n,d/2-1}}{2dr^{*2}} \sigma_v^2 t\right). \tag{D16}$$

The distribution of first passage times asymptotically follows the largest mode of this series, $n = 0$, so that:

$$p(t) \sim \exp\left(-\frac{j^2_{0,d/2-1}}{2dr^{*2}} \sigma_v^2 t\right). \tag{D17}$$

For instance for $d = 2, 3, 4$ we have $j_{0,0} \approx 2.40483$, $j_{0,1/2} = \pi$, $j_{0,1} \approx 3.83171$.

[1] Redner S (2001) *A Guide to First-Passage Processes* (Cambridge University Press, Cambridge).

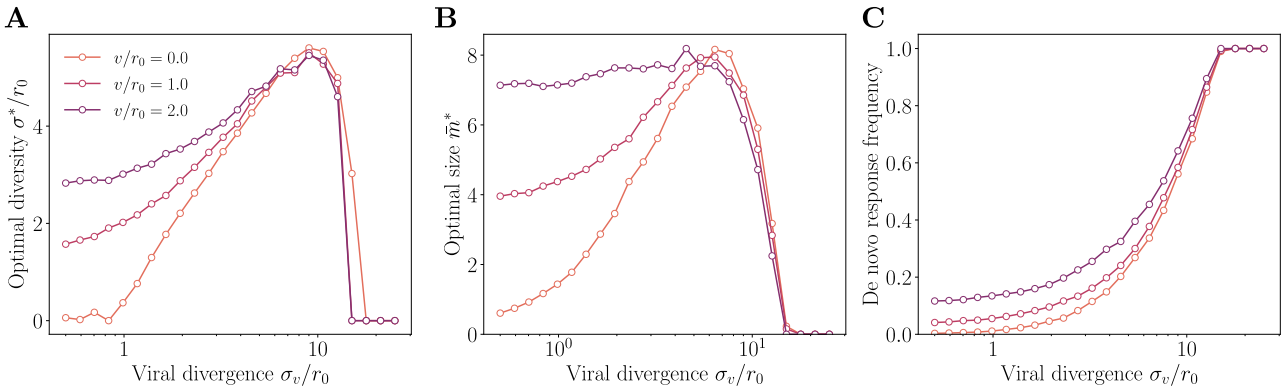

FIG. S4: **Optimal strategy in presence of drift. A.** Optimal diversity **B.** size and **C.** frequency of naive usage in response to an immunization challenge for different strain drift v/r_0. The strain follows a random walk with drift v in a fixed direction e_0: $a_{t+1} = a_t + ve_0 + \sigma_v \eta_t$. Parameters values: $\mu = 0.5$, $\gamma = 0.85$, $\kappa = 3.3$, $\alpha = 1$, $q = 2$, $d = 2$.