Fluctuating fitness shapes the clone size distribution of immune repertoires: Supplementary information

Jonathan Desponds, Thierry Mora, Aleksandra M. Walczak

Appendix A: Simple birth-death process with no fitness fluctuations, and its continuous limit

In this Appendix we derive the steady-state clone size distribution for a system that does not experience any environmental stimulation or noise, but is governed by a birth death process. We will show that the small number fluctuations arising from the discrete nature of birth and death are not sufficient to explain the observed distributions. We also show that our choice of a continuous birth death process is equivalent to its discrete version.

The multiplicative birth–death process corresponds to the following discrete dynamics:

$$\begin{cases} P(n \to n+1) = \mu n dt \\ P(n \to n-1) = \nu n dt, \end{cases} \quad (A1)$$

where μ is the division rate, ν the death rate. We assume that the population of cells of size n is maintained out of equilibrium by a source of new cells. The steady state solution for cell numbers above the value of the source satisfies detailed balance

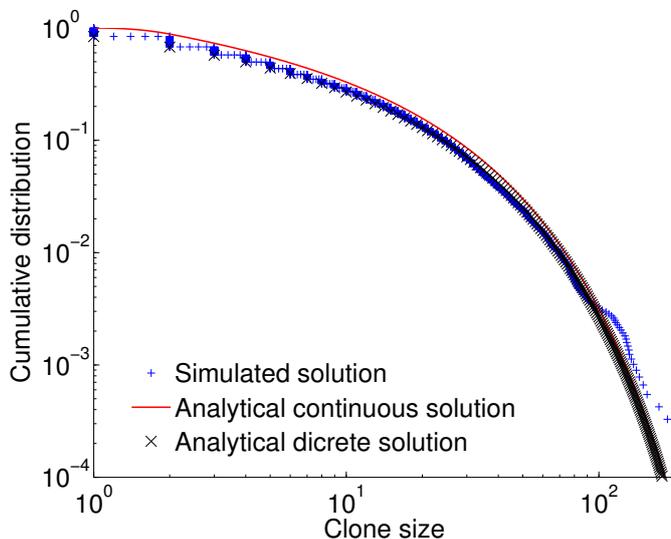$$P(n)\mu n = P(n+1)\nu(n+1) \quad (A2)$$

FIG. S1: We compare results from a full Gillespie simulation (blue crosses) of a system with only birth-death dynamics with analytical prediction for a discrete system (black crosses, Eq. A3) and a continuous system (red curve, Eq. A12). The prediction with discrete variables is more accurate for small clones but the behaviour of all systems is the same for large populations. The parameters are $\nu = 1.45$ day^{-1}, $\mu = 1.5$ day^{-1}, $C_0 = 2$ and we introduce 2000 new clones per day.

and, assuming the death rate is larger than the birth rate, takes the form

$$P(n) \sim \frac{K}{n} e^{-n \log \nu / \mu}. \quad (A3)$$

The continuous counterpart of this discrete stochastic process corresponds to the following linear-noise approximation:

$$\partial_t C_i = f_0 C_i + \sqrt{(\mu + \nu) C_i} \xi, \quad (A4)$$

where $\langle \xi_i(t) \xi_i(t') \rangle = \delta(t - t')$ and $f_0 = \mu - \nu < 0$ (and we use the Îto convention). In terms of $x = \log C$ the Langevin equation is

$$\partial_t x = f_0 + \sqrt{\mu + \nu} e^{-x/2} \xi - e^{-x} \frac{(\mu + \nu)}{2}, \quad (A5)$$

and the corresponding Fokker-Planck equation reads

$$\partial_t \rho = \partial_x(-f_0 \rho) + \partial_x^2 \left(\frac{\mu + \nu}{2} e^{-x} \rho \right) + \partial_x \left(e^{-x} \rho \frac{\mu + \nu}{2} \right) + s(x), \quad (A6)$$

where $s(x)$ is the distribution of sizes of newly arriving clones. At steady state, we find

$$K - s_C \theta(x - x_0) = -f_0 \rho + \frac{\mu + \nu}{2} e^{-x} \rho', \quad (A7)$$

where K is an integration constant. Defining

$$C_m = (\mu + \nu)/(2|f_0|) \quad (A8)$$

for $x < x_0$ we obtain

$$\rho(x) = e^{-e^x/C_m} K \int_0^x e^x e^{e^x/C_m} = K C_m(1 - e^{-(e^x - 1)/C_m}) \quad (A9)$$

and for $x > x_0$

$$\rho(x) = e^{-e^x/C_m} C_m \Big[K e^{e^x/C_m} - K e^{1/C_m} \quad (A10)$$
$$- \frac{s_C}{|f_0| C_m} e^{e^x/C_m} + \frac{s_C}{|f_0| C_m} e^{e^{x_0}/C_m} \Big]$$

To ensure convergence we set $K = s_C/(|f_0| C_m)$ and the steady solution of the Fokker-Planck equation is

$$\rho(x) = \begin{cases} \frac{s_C}{|f_0|}(1 - e^{-(e^x - 1)/C_m}) \text{, if } x < x_0 \\ \frac{s_C}{|f_0|}(e^{e^{x_0}/C_m} - e^{C_m^{-1}}) e^{-e^x/C_m} \text{, if } x > x_0 \end{cases} \quad (A11)$$

or in terms of the clone size

$$\rho(C) = \begin{cases} \frac{1}{C}(1 - e^{-(C-1)/C_m}) \text{, if } C < C_0 \\ (e^{C_0/C_m} - e^{C_m^{-1}}) \frac{e^{-C/C_m}}{C} \text{, if } C > C_0 \end{cases} \quad (A12)$$

This result is exactly equivalent to that of Eq. A3 when $\nu - \mu = |f_0| \ll \mu, \nu$. The accuracy of the approximation is verified in Fig. S1. Even for very large exponential cutoff values, C_m, the apparent exponent is $\alpha = 0$, corresponding to a flat cumulative distribution. This distribution is inconsistent with experiments, regardless of sequencing depth and we conclude that pure birth-death noise is not sufficient to explain the observed distributions.

Appendix B: Effects of explicit global homeostasis

In the simulations of clone dynamics in a fluctuating environment presented in the "Clone dynamics in a fluctuating antigenic landscape" Results section of the main text, we did not explicitly include a homeostatic control term, but tuned the division and death rates to achieve a given repertoire size. Here we add an explicit homeostatic term to the growth and degradation terms in the Langevin simulations described by Eq. 1 of the main text

$$-h \left[\frac{\sum_i C_i}{N} \right]^r , \tag{B1}$$

where N is a carrying capacity, h is the homeostatic constant multiplicator and r is the exponent of homeostatic response that described the sharpness of the response when approaching then carrying capacity limit. Comparing in Fig. S2 the resulting clone size distribution obtained with the explicit homeostatic term to the distribution from the simulations in the main text, we see that the explicit homeostatic term does not have an effect on the form of the distribution. It does have an effect on the trajectory of certain clones, and in particular on the response of the system to a very large invasion, making it an important feature of the dynamics of the immune system. However, as shown by the results in Fig. S2 its net effect on the clone size distribution can be taken into account by tuning division and death. When considering specific trajectories in the mean field approximation homeostatic control will add a systematic negative drift to the clonal population and can be accounted for by an additional contribution to f_0.

Appendix C: Details of noise partition do not influence the clone size distribution function

In the simulation of the dynamics of receptors experiencing a clone-specific fitness presented in the "Clone dynamics in a fluctuating antigenic landscape" Results section of the main text we distributed the noise between the different random distributions: the poisson distributed number of new antigens (s_A), the variance of the initial concentrations ($a_{j,0}$) and the variance of the binding probability (the values of K_{ij}). We made specific choices for this reparation by picking specific parameters of the random processes. Here we show that these specific
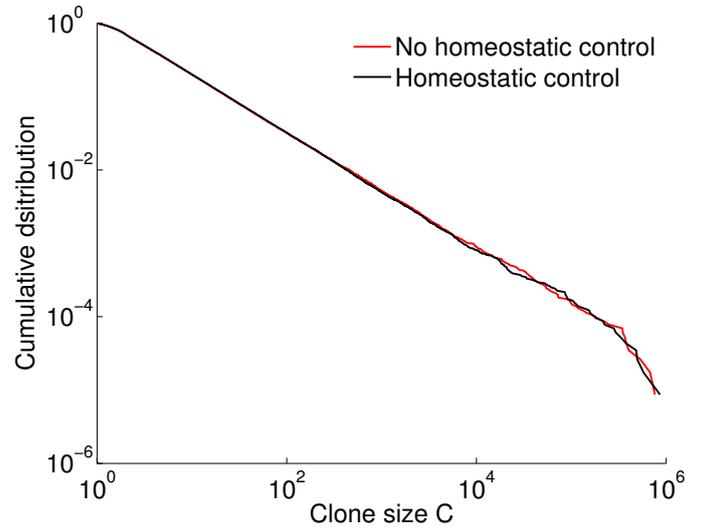

FIG. S2: Adding an explicit homeostatic control term does not affect the clone size distribution compared to tuning the degradation and death rates to obtain a given repertoire size as is done in the main text. Comparison of the clone size distribution with an explicit homeostatic control term given by Eq. B1 (black line) to the distribution presented in the main text (red line). We simulate the Langevin equation for a division rate $\nu = 0.2$ days^{-1}, death rate $\mu = 0.4$ days^{-1}, introduction size $C_0 = 2$, environmental correlation time of $\lambda^{-1} = 0.5$ days and an amplitude of variations of the environment $A = 1.41$ days^{-1} without any homeostatic control for the red curve and with carrying capacity $N = 4 \cdot 10^{10}$ ($h = 1$) and a homeostatic exponent $r = 3$ for the black curve.

choices of repartitioning the contributions to the noise do not influence the clone size distributions. Fig. S3 compares clone size distributions obtained with different values of the poisson distributed number of newly arriving antigen N_a and the variance of the Gaussian distributed binding probabilities K_{ij}, reproducing the same distributions in both cases.

Appendix D: Model of temporally correlated clone-specific fitness fluctuations

In the "Simplified models and the origin of the power law" Results section of the main text we make a series of approximations to effectively describe the dynamics of immune cells: we first approximate the antigenic environment by a random process with time correlated (colored) noise and we later neglect these temporal correlations. In this section and Appendix F we give the details that lead to the specific forms of the effective equations. In this Appendix we derive the Fokker-Planck equations for the time correlated noise model. In Appendix F we will consider the limit of an infinitely quickly changing environment.

The Langevin equations describing the dynamics of cells experiencing clone specific fitness fluctuations with
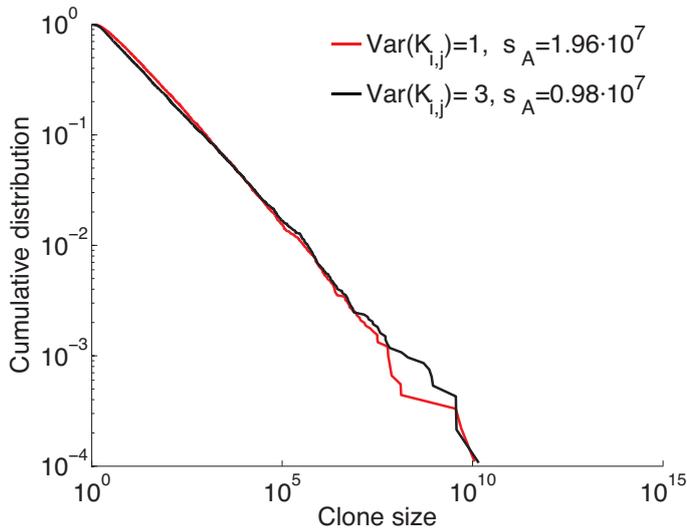
FIG. S3: Repartitioning the sources of stochasticity between the number of new antigens per time unit or the variability of binding probabilities does not influence the clone size distributions. We compare simulations of the full system dynamics defined by Eq. 1 of the main text with two sets of values s_A of the poisson distributed number of newly arriving antigen N_a and the variance of the Gaussian distributed binding probabilities K_{ij} that give the same total environmental noise $A^2 = s_A p a_0^2 \langle K^2 \rangle \lambda^{-1}$. The parameters were taken to be (as in Fig. 1) $s_C = 2000$ day^{-1}, $C_0 = 2$, day^{-1}, $a_{j,0} = a_0 = 1$, $\lambda = 2$ day^{-1}, $p = 10^{-7}$, $\nu = 0.98$ day^{-1}, $\mu = 1.18$ day^{-1}. For the red curve the variance of the entries of K_{ij} is 1, so that $\langle K^2 \rangle = 2$ and $s_A = 1.96 \cdot 10^7$ while for the black curve the variance of the entries of K_{ij} is 3, so that $\langle K^2 \rangle = 4$, and $s_A = 0.98 \cdot 10^7$.

a finite correlation time are

$$\frac{dC_i}{dt} = [f_0 + f_i(t)]C_i(t) + \sqrt{(\nu + \mu)C_i(t)}\xi_i(t), \quad \text{(D1)}$$

$$\frac{df_i}{dt} = -\lambda f_i(t) + \sqrt{2}\gamma \eta_i(t), \quad \text{(D2)}$$

where $\langle \xi_i(t)\xi_i(t') \rangle = \delta(t - t')$ represents birth death noise in the linear-noise approximation (with the Îto convention) and $\langle \eta_i(t)\eta_i(t') \rangle = \delta(t - t')$ is the noise of antigenic environment. The autocorrelation function of this Ornstein-Uhlenbeck process is

$$\langle f_i(t)f_i(t') \rangle = e^{-\lambda(t+t')}\left(\langle f_i(0)^2 \rangle - \frac{\gamma^2}{\lambda} \right) + \frac{\gamma^2}{\lambda}e^{-\lambda|t-t'|}. \quad \text{(D3)}$$

We pick the steady-state value of the initial fitness distribution to cancel the first in Eq. D3, $\langle f_i(0)^2 \rangle = \gamma^2/\lambda$ and obtain

$$\langle f_i(t)f_i(t') \rangle = \frac{\gamma^2}{\lambda}e^{-\lambda|t-t'|}, \quad \text{(D4)}$$

(conditioned on the integral of the net growth rate $f + f_0$ being positive so that the clone does not go extinct). Set-

ting $x = \log C$, we obtain a new set of Langevin equations

$$\partial_t x_i = f_0 + f_i + \sqrt{\mu + \nu}e^{-x_i/2}\xi_i - e^{-x_i}\frac{(\mu + \nu)}{2}, \quad \text{(D5)}$$

$$\frac{df_i}{dt} = -\lambda f_i + \sqrt{2}\gamma \eta_i, \quad \text{(D6)}$$

where the birth-death noise is now treated in the Îto convention. The corresponding Fokker-Planck equation for the distribution of fitness and clone size at time t, $\rho(x, f, t)$, verifies

$$\partial_t \rho = \partial_x(-f_0\rho) + \partial_f(\lambda f \rho) + \partial_f^2(\gamma^2\rho) + \quad \text{(D7)}$$
$$\partial_x^2\left(\frac{\mu+\nu}{2}e^{-x}\rho \right) + \partial_x\left(e^{-x}\rho\frac{\mu+\nu}{2} \right)$$
$$+ s(x, f),$$

where $s(x, f)$ is the source of new clones. We solve this equation numerically using finite element methods to obtain clone size distributions for the clone-specific fitness model.

Appendix E: The Ornstein Uhlenbeck process and maximum entropy

In this Appendix we show that the maximum entropy or maximum caliber process with autocorrelation function $\langle x(t)x(t + s) \rangle = A^2 e^{-\lambda|s|}$ corresponds to the Ornstein-Uhlenbeck process. We consider this continuous maximum entropy process as the continuous limit of a simpler maximum entropy system in discrete time. Burg's maximum entropy theorem [1] states that the maximum entropy process in discrete time that constrains $\langle X_n(t)^2 \rangle = A^2$ and $\langle X_n(t)X_{n+1}(t) \rangle = A^2 e^{-\lambda\tau}$ corresponds to the following Markovian dynamics:

$$X_{n+1} = e^{-\lambda\tau}X_n + \sqrt{1 - e^{-2\lambda\tau}}A\eta, \quad \text{(E1)}$$

where η is Gaussian white noise. In the limit of $\tau \to 0$ we recover the constrained autocorrelation function in the vicinity of $s = 0^+$: $\langle x(t)^2 \rangle = A^2$, $(d/ds)\langle x(t)x(t + s) \rangle|_{s=0^+} = -\lambda A^2$, and Eq. E1 converges to an Ornstein-Uhlenbeck process.

Appendix F: Model solution for white-noise clone-specific fitness fluctuations

In the limit of infinitely quickly fluctuating environments, $\gamma \to +\infty$ and $\lambda \to +\infty$ while keeping their ratio $\sigma = \gamma/\lambda$ constant, the autocorrelation of the fitness noise approaches a Dirac delta function, and the fluctuating part of the growth rate $f_i(t)$ converges to Gaussian white noise, $\langle f_i(t)f_i(t') \rangle = 2\sigma^2\delta(t - t')$. Effectively the immune cell dynamics are now described by a one dimensional Langevin equation for the clone size

$$\partial_t C_i = f_0 C_i + \sqrt{2}\sigma C_i \eta_i + \sqrt{(\nu + \mu)C_i(t)}\xi_i, \quad \text{(F1)}$$
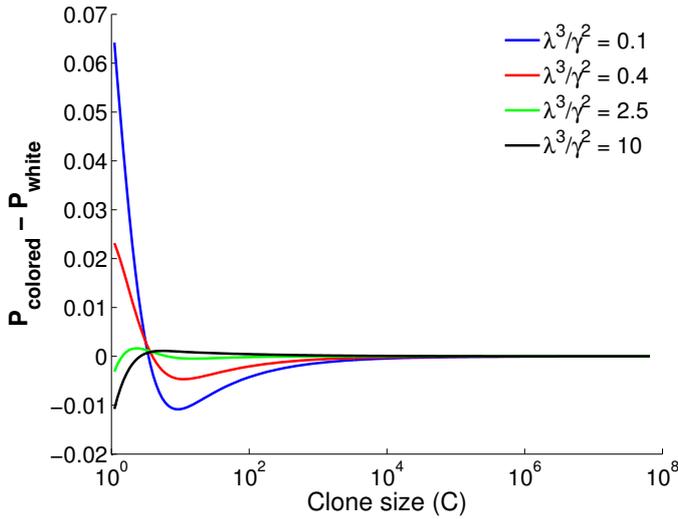
FIG. S4: Comparison between clone size distribution obtained as solutions of the time-correlated and time-uncorrelated noise models (without birth death noise). As the values of the dimensionless parameter related to the effective strength of antigen fluctuations relative to their characteristic lifetime λ^3/γ^2 grow the time correlated noise prediction converges to the exact power-law solution of the white-noise model. The cut-off value of the power law decreases with λ^3/γ^2. All simulations performed at a constant value of $\alpha = |f_0|\lambda^2/\gamma^2$ set to 0.5. The value of f_0 is kept fixed to -0.5 days^{-1} for all solutions.

where $\langle\eta_i(t)\eta_i(t')\rangle = \delta(t-t')$ follows the Stratanovich convention and ξ_i is as before. The equation for the logarithm of the clone size $x = \log C$ is

$$\partial_t x_i = f_0 + \sqrt{2}\sigma\eta_i + \sqrt{\mu+\nu}e^{-x_i/2}\xi_i - e^{-x_i}\frac{(\mu+\nu)}{2}. \quad (F2)$$

We explicitly checked that the numerical solution to the clone specific fitness model in Eqs. D1 and D2 converged to the dynamics described by Eq. F1, as demonstrated in Fig. S4.

We now solve this equation analytically, starting with the case of no birth-death noise: Eq. F1 simplifies to

$$\partial_t C_i = f_0 C_i + \sqrt{2}\sigma C_i\eta_i \quad (F3)$$

The equation for $x = \log C$ (using the Stratanovich convention) is

$$\partial_t x_i = f_0 + \sqrt{2}\sigma\eta_i, \quad (F4)$$

with the corresponding Fokker Planck equation

$$\partial_t \rho(x,t) = \partial_x(-f_0\rho) + \frac{1}{2}\partial_x[2\sigma^2\partial_x\rho] + s(x), \quad (F5)$$

where $s(x)$ is the source term describing the size of newly introduced clones. Assuming a constant initial clone size, $s(x) = s_C\delta(x-x_0)$, the steady state solution is

$$\rho(x) = e^{-\alpha x}\frac{1}{\alpha}\left[Ke^{\alpha x} - K - s_C\sigma^2 e^{\alpha x} + s_C\sigma^2 e^{x_0}\right], \quad (F6)$$
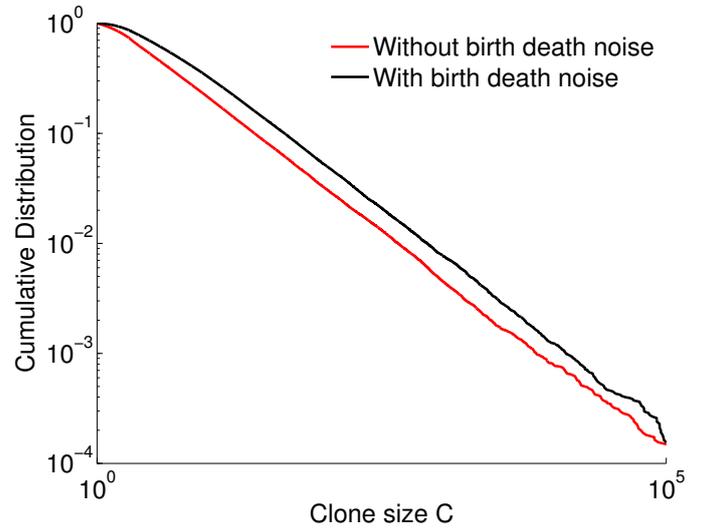

FIG. S5: We compare simulations of the Langevin dynamics with time correlated antigenic noise with birth-death noise (black line) to the same dynamics without the birth-death noise (red line). All other parameters are kept fixed. We find similar values of the power law exponents but different small clone behaviours. The parameters are $\nu = 0.2$ day^{-1}, $\mu = 0.4$ day^{-1} (for red curve simply $f_0 = -0.2$ day^{-1}), $C_0 = 2$, $\lambda = 2$ day^{-1} and $\gamma = 1$ day$^{-3/2}$

where we have defined

$$\alpha = |f_0|/\sigma^2, \quad (F7)$$

and K is an integration constant. Imposing that ρ vanishes at infinity sets $K = s_C\sigma^2$ and the final form of the steady state clone size distribution is

$$\rho(x) = \begin{cases} \frac{s_C}{|f_0|}\left(1 - e^{-\alpha x}\right) & \text{if } x < x_0 \\ \frac{s_C}{|f_0|}e^{-\alpha x}\left(e^{x_0} - 1\right) & \text{if } x > x_0, \end{cases} \quad (F8)$$

or in terms of clone size $C = e^x$,

$$\rho(C) = \begin{cases} \frac{s_C}{|f_0|C}\left(1 - \frac{1}{C^\alpha}\right) & \text{if } C < C_0 \\ \frac{s_C}{|f_0|}\frac{1}{C^{\alpha+1}}\left(\frac{1}{C^{x_0}} - 1\right) & \text{if } C > C_0. \end{cases} \quad (F9)$$

In all simulations and solutions we find that for large clones, the model of temporally correlated fitness fluctuations behaves as the its white noise limit. This behaviour can be explained by the fact that large clones need a long time to become large. At these long timescales, the characteristic time of noise correlation is negligible and the noise may be approximated as white. For this reason, the exponent α of the power law computed assuming a white noise for the fitness fluctuations is still valid even when that noise is actually correlated in time.

Next, we re-introduce the birth-death noise and solve the general equation. The Langevin equation for $x = \log C$,

$$\partial_t x = f_0 + \sqrt{2}\sigma\eta + \sqrt{\mu+\nu}e^{-x/2}\xi - e^{-x}\frac{(\mu+\nu)}{2} \quad (F10)$$

results in the Fokker-Planck equation for the distribution of clone sizes

$$\partial_t \rho = \partial_x(-f_0\rho) + \frac{1}{2}\partial_x[2\sigma^2\partial_x\rho] + \partial_x^2\left(\frac{\mu+\nu}{2}e^{-x}\rho\right)$$
$$+\partial_x\left(e^{-x}\rho\frac{\mu+\nu}{2}\right) + s(x). \quad \text{(F11)}$$

Assuming that the initial size is constant, the steady state solution is given by the solution of the inhomogeneous linear equation:

$$K - s_C\theta(x-x_0) = -f_0\rho + \sigma^2\rho' + e^{-x}\frac{\mu+\nu}{2}\rho'. \quad \text{(F12)}$$

The full solution is the sum $\rho = \rho_0 + \rho_1$ of the particular solution,

$$\rho_0(x) = \begin{cases} \frac{K}{|f_0|} & \text{for } x < x_0, \\ \frac{K-s_C}{|f_0|} & \text{for } x > x_0, \end{cases} \quad \text{(F13)}$$

and the solution ρ_1 to the homogeneous equation

$$f_0\rho_1 = \sigma^2\rho_1' + e^{-x}\frac{\mu+\nu}{2}\rho_1' \quad \text{(F14)}$$

of solution:

$$\rho_1(x) = K'\left[\frac{e^x + \frac{(\mu+\nu)}{2\sigma^2}}{1 + \frac{(\mu+\nu)}{2\sigma^2}}\right]^{-\alpha}, \quad \text{(F15)}$$

with $\alpha = |f_0|/\sigma^2$. Therefore, for $x > x_0$

$$\rho(x) = K'\left[\frac{e^x + \frac{(\mu+\nu)}{2\sigma^2}}{1 + \frac{(\mu+\nu)}{2\sigma^2}}\right]^{-\alpha} + \frac{K-s}{|f_0|} \quad \text{(F16)}$$

we set $K = s$ for convergence and obtain the steady state clone size distribution for large x

$$\rho(x) = \left[e^x + \frac{\mu+\nu}{2\sigma^2}\right]^{-\alpha}, \quad \text{(F17)}$$

or in terms of the clone size

$$\rho(C) = \frac{1}{C\left(C + \frac{\mu+\nu}{2\sigma^2}\right)^\alpha}. \quad \text{(F18)}$$

We see that the white noise solution with birth–death noise has the same large clone power law behaviour as without birth–death noise. Fig. S5 illustrates how birth death noise in the clone-specific fitness models with time correlated noise also does not affect the power law exponent but only the cut off of the power law.

Appendix G: Data analysis

In the main text we report values of the power law exponents and power law cut off values obtained from the high throughput sequencing repertoire study of clone size distributions of zebrafish B-cell heavy chain receptors of Weinstein et al. [2]. We extracted the power law exponent and the best fit for the starting point of the power law, defined as its lower bound cutoff, from the discrete clone size distributions plotted in Fig. 1 of the main text using the methods discussed by Clauset and Newman [3]. Specifically, for each point of the cumulative clone size distribution we compute an estimate of the power law exponent with that point as cutoff (i.e the best fit of the power law including only the values of the distribution above that point) using

$$\alpha(C_{\min}) = 1 + n\left[\sum_{i=1}^n \log\left(\frac{C_i}{C_{\min}}\right)\right], \quad \text{(G1)}$$

where $C_{\min}$ is the cut off and n is the number of points with y-axis values above $C_{\min}$. For each of these cut-off values we compute the Kolmogorov-Smirnov distance between the data and the estimated power law distribution:

$$d(C_{\min}) = \max_{C>C_{\min}}|F_d(C) - F_e(C;C_{\min})| \quad \text{(G2)}$$

where the maximum is taken over all values above the cut off $C_{\min}$, F_d is the cumulative distribution function (CDF) of the data and $F_e(C;C_{\min})$ is the CDF of the estimated power law distribution with $C_{\min}$ as a cutoff, using Eq. G1. The the cut off is taken to be the minimum of this distance over all possible cut off values and the exponent is the exponent found for this value.

The obtained power law parameters are presented in Table I. The power law exponent gives reproducible values for different individuals and agrees with values of the same exponent obtained from human data [4]. We note that the power law exponent of the cumulative distribution function is α for a power law distribution with exponent $1 + \alpha$. As discussed in detail in the main text, the reliability of the cutoff estimate C^* is sensitive to experimental precision of capturing the rare clones. In the presented dataset the reads were not barcoded and the counts had to be renormalized by a known PCR amplification factor. Therefore, these normalized counts could not to used as normal counts, making the definition of a cut-off clone size problematic. To overcome this problem, we estimate the power law cut-off from the value of the cumulative distribution function at the cut-off clone size (instead of the cut-off clone size itself). That value is invariant under rescaling of absolute clone size values, unlike C^*.

We notice that the steady state solution is invariant under a full rescaling of time in the equations of the dynamics. This means that the system can be described by two dimensionless parameters, $\alpha = f_0\lambda^2/\gamma^2$ and λ^3/γ^2, and the introduction size C_0. Fitting α to data and assuming value for C_0, we can compare the value of the power law cut-off in data and in simulations to fit the remaining dimensionless parameter, λ^3/γ^2. Estimating f_0 based on thymic output we can predict the order of magnitude of λ and γ.

Fish	$1+\alpha$	C^*	$\log(1-\mathrm{CDF}(C^*))$
A	2.0591	32.6445	- 3.1389
B	2.0214	10.7231	-1.8644
C	2.0708	16.7386	-2.4655
D	2.0670	14.9313	-2.1492
E	2.0529	8.2685	-1.8332
F	2.0006	5.8972	-1.6161
G	1.9867	52.2909	-2.7329
H	2.2242	32.1719	-2.6877
I	2.0835	18.4385	-2.2757
J	1.6907	44.4885	-2.2877
K	1.7641	3.6030	-0.9907
L	1.9417	18.5298	-2.2730
M	1.9901	18.5531	-2.2031
N	1.8877	108.4732	-2.7984

TABLE I: Fit of the power law exponent of the clone size distribution $1+\alpha$ and power law cut-off value C^* for zebrafish B-cell heavy chain D segment data from Weinstein et al [2] presented in Fig. 1. The fit for 14 fish (named A to N) shows a similar fit of the power law exponent.

Appendix H: Cell specific simulations

In the "A model of fluctuating phenotypic fitness" Results section of the main text, we present results of Fokker-Planck simulations for the cells dynamics. Here we verify that the stochastic dynamics of cells subject to a fluctuating cell-specific fitness are well approximated at the population level by a Fokker-Planck equation with a source term accounting for the import of new clones by comparing its numerical steady-state solution obtained by a finite elements method to explicit Gillespie simulations. We simulated the dynamics of clones using a Gillespie algorithm where cell division and death are accounted for explicitly and depend linearly on a fitness $f_c(t)$ fluctuating according to Eq. 7. The death rate is kept constant (above the average birth rate) and the fluctuations of the fitness only affect the birth rate (with the constraint that the birth rate is always positive). The agreement between the results of this detailed simulation and the Fokker-Planck solution, shown in Fig. S6, validates the linear-noise approximation for the birth-death noise as well as the averaging argument leading to Eq. 8 and 9. This allows us to rely on the Fokker-Planck solution to explore parameter space.

Appendix I: Model of cell-specific fitness fluctuations, and its limit of no heritability

The cell specific fitness model described in the "A model of fluctuating phenotypic fitness" Results section of the main text arises as a description of a population where each cell experiences its own growth fluctuations
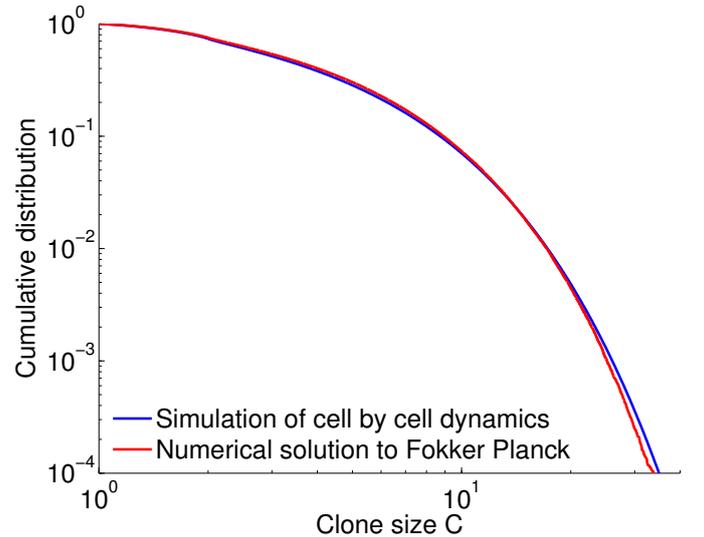

FIG. S6: Comparison of the Fokker-Planck solution (red line) and explicit Gillespie simulations of the dynamics (blue line) for the cell specific fitness model discussed in the "A model of fluctuating phenotypic fitness" Results section of the main text, show good agreement allowing us to use the population level Fokker-Planck solution to explore parameter space. Parameters were taken to be $\nu = 0.5$ day^{-1}, $\mu = 0.8$ day^{-1}, $C_0 = 2$, $\lambda_c = 4$ days^{-1} and $\gamma_c = 4$ day$^{-3/2}$.

but cells deriving from the same lineage remain correlated. In this Appendix we derive the equations that describe the dynamics of clones in this system.

Each cell c experiences a time-correlated multiplicative noise from environmental growth factors. For cells j in a given cell lineage (or clone) i, each individual cell's fitness follows the stochastic dynamics:

$$\partial_t f_c(t) = -\lambda_c f_c + \sqrt{2}\gamma_c \eta_c \quad (I1)$$

where $\langle \eta_c(t)\eta_c(t')\rangle = \delta(t-t')$. Averaging over all cells in the clone, we obtain

$$\begin{cases} \partial_t C_i = f_0 C_i + f_i C_i + \sqrt{(\mu+\nu)C_i}\,\xi_i \\ \partial_t f_i = -\lambda_c f_i + \sqrt{\dfrac{2}{C_i}}\gamma_c \eta_i, \end{cases} \quad (I2)$$

where f_i is the average fitness in clone i

$$f_i(t) = \frac{1}{C_i}\sum_{c\in i} f_c(t), \quad (I3)$$

and where we have added a birth-death noise term $\sqrt{(\mu+\nu)C_i}\,\xi_i$. We use the Îto convention for the birth-death noise, $\langle \xi_i(t)\xi_i(t')\rangle = \delta(t-t')$ and the Stratanovich one for the environmental noise $\langle \eta_i(t)\eta_i(t')\rangle = \delta(t-t')$. The equivalent equations for $x = \log C$ are

$$\partial_t x_i = f_0 + f_i + \sqrt{\mu+\nu}\,e^{-x_i/2}\xi - e^{-x_i}\frac{\mu+\nu}{2} \quad (I4)$$

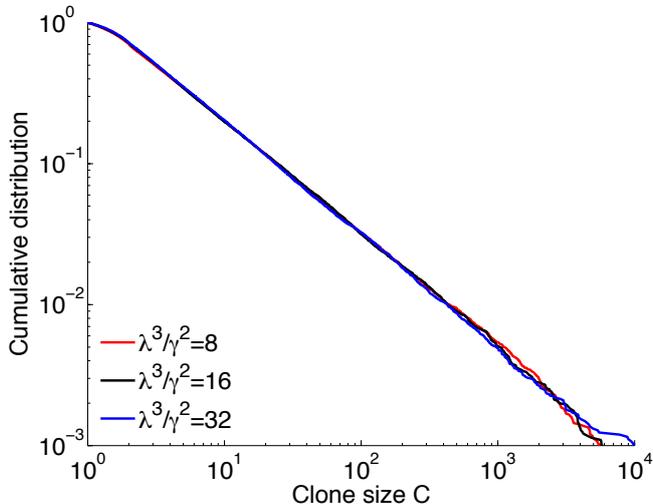$$\partial_t f_i = -\lambda_c f_i + \sqrt{2}e^{-x_i/2}\gamma_c \eta_i \quad (I5)$$

FIG. S7: Varying the dimensionless parameter related to the effective strength of antigen fluctuations relative to their characteristic lifetime λ^3/γ^2 does not affect the exponent of the power law if the ratio between exponential decay λ and standard deviation of the variation γ is kept constant. For all three curves the exponent is $\alpha = 0.8$ and $\nu = 0.5$ days^{-1}, $\mu = 0.8$ days^{-1}, $C_0 = 2$ while λ and γ vary.

and the Fokker-Planck equation is

$$
\begin{aligned}
\partial_t \rho(t,x,f) = &- (f_0 + f)\partial_x \rho + \lambda_c \partial_f(f\rho) + e^{-x}\gamma_c \partial_f^2 \rho \\
&+ \frac{\mu + \nu}{2}\partial_x(e^{-x}\rho) + \frac{\mu + \nu}{2}\partial_x^2(e^{-x}\rho) \\
&+ s(x,f),
\end{aligned}
\tag{I6}
$$

where $s(x,f)$ is the joint distribution of size and fitness or newly arriving clones (from thymic or bone marrow output). This is the full Fokker-Planck equation that is solved numerically in the main text using the finite elements method.

Because of the $1/\sqrt{C_i}$ prefactor in front of the noise term, we could expect fitness fluctuations to behave like a birth-death noise in the limit of low heritability ($\lambda_c \to \infty$). In the remainder of this Appendix we show that this is not the case, and we show how to take the limit of no heritability properly.

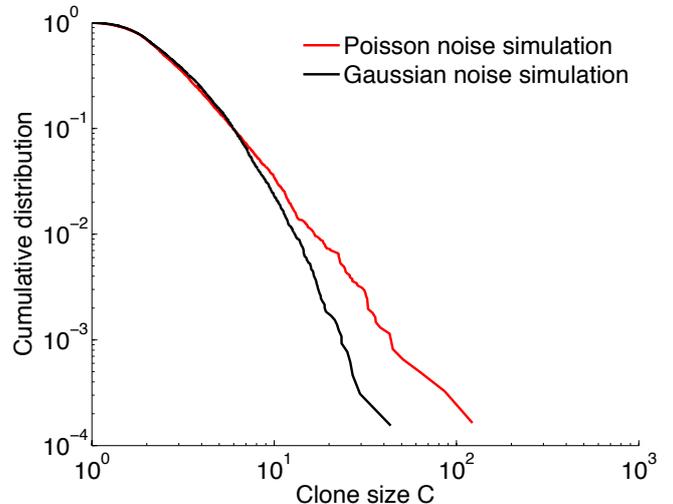Consider the limit of $\lambda_c \to \infty$ and $\gamma_c \to \infty$, keeping the

FIG. S8: Large deviations can influence the effect of Poisson noise on the simulated clone size distributions and create a discrepancy between Poisson noise (red line) and the Gaussian approximations (black line) we assume in the main text. The discrepancy is most apparent for small clones. We simulated the Langevin dynamics of the Gaussian model with $\nu = 0.5$ day^{-1}, $\mu = 1$ day^{-1}, $C_0 = 2$, $\lambda = 3$ day^{-1} and $\gamma = 1$ day$^{-3/2}$ and the same dynamics with Poisson noise and $\nu = 0.5$ day^{-1}, $\mu = 1$ day^{-1}, $C_0 = 2$, $\lambda = 3$ day^{-1} and $s_A = 10^7$ day^{-1}. In both cases we introduce $s_C = 2000$ new clones per day.

ratio γ_c/λ_c constant, so that f does not become infinitesimally small. The equation for the environmental stimulation f in $x = \log C$ space is given by (in Stratanovich convention)

$$
\partial_t f = -\lambda_c f + \sqrt{2}\gamma_c e^{-x/2}\eta.
\tag{I7}
$$

Direct integration gives

$$
f(t) = \sqrt{2}\gamma_c \int_0^t e^{-\lambda_c u} e^{-x(t-u)/2}\eta(t-u)du
\tag{I8}
$$

and we divide the integral into two sub-integrals for $k > 0$

$$
\begin{aligned}
f(t) = &\sqrt{2}\gamma_c \int_{k/\lambda_c}^t e^{-\lambda_c u} e^{-x(t-u)/2}\eta(t-u)du \\
&+ \sqrt{2}\gamma_c \int_0^{k/\lambda_c} e^{-\lambda_c u} e^{-x(t-u)/2}\eta(t-u)du.
\end{aligned}
\tag{I9}
$$

With infinite precision, for any value of t, we set the integral of η to be bounded and obtain the first integral is with probability $1 - \epsilon$ smaller in norm than

$$\sqrt{2}\gamma_c\sqrt{t}K(\epsilon)e^{-k}, \tag{I10}$$

where $K(\epsilon)$ is a constant to control the variations of the integral of ξ with probability ϵ (the time factor for the control of the integral is in the $\sqrt{t}$).

The second sub-integral is

$$\begin{aligned} &\sqrt{2}\gamma_c\int_0^{k/\lambda_c} e^{-\lambda_c u}e^{-x(t-u)/2}\eta(t-u)du \\ &\approx e^{-x(t^-)/2}\eta(t)\sqrt{2}\frac{\gamma_c}{\lambda_c}(1-e^{-k}). \end{aligned} \tag{I11}$$

We choose $k = \sqrt{\lambda_c}$ and in the limit of $\lambda_c \to \infty$ and $\gamma_c \to \infty$ keeping $\gamma_c/\lambda_c = $ const we obtain the final form of environmental fluctuations

$$f(t) \longrightarrow \sqrt{2\frac{\gamma_c}{\lambda_c}}e^{-x(t^-)}\eta(t), \tag{I12}$$

where t^- means the left-hand limit. $f(t)$ depends only on the past, which means that in $x = \log C$ space the noise is similar to a birth-death noise in the Îto convention. Yet in terms of clone sizes C additional Îto terms make the effect of environmental fluctuations different from classical birth-death dynamics.

Appendix J: Model solutions for cell-specific fitness fluctuations in the limit of no heritability

In this Appendix we solve the model of cell-specific fitness fluctuations in the limit where trait heritability is low. In this limit, the dynamics is described by a model with an instantaneous random fitness that is uncorrelated for cells in the same clone. The resulting Langevin equation reads:

$$\frac{dC_i}{dt} = f_0 C_i + \sqrt{2C_i}\frac{\gamma_c}{\lambda_c}\eta_i + \frac{\gamma_c^2}{\lambda_c^2} + \sqrt{(\mu+\nu)C_i}\xi_i \tag{J1}$$

where all noise is treated in the Îto convention, and where the extra term γ_c^2/λ_c^2 comes the converting back the low-heritability limit of the fitness fluctuations, given by Eq. I12, into $C = e^x$ space. We note that although the fitness and birth-death noise have very similar forms, the birth-death noise is self-generated and intrinsic, while the fitness noise is environmental and extrinsic. This small difference greatly affects the steady-state clone size distribution.

To see this, we first consider the case of no birth-death noise. In the cell-specific fitness model consider the following equations with the Stratanovich rule:

$$\begin{cases} \partial_t C_i = f_0 C_i + f C_i, \\ \partial_t f_i = -\lambda_c f_i + \sqrt{\frac{2}{C_i}}\gamma_c\eta_i, \end{cases} \tag{J2}$$

and its equivalent for $x = \log(C)$

$$\begin{cases} \partial_t x_i = f_0 + f_i, \\ \partial_t f_i = -\lambda_c f_i + e^{-x_i/2}\gamma_c\eta_i \end{cases} \tag{J3}$$

In Appendix I we have shown that in the limit of $\lambda_c \to \infty$ and $\gamma_c \to \infty$, the system reduces to the one dimensional equation

$$\partial_t x_i = f_0 + e^{-x_i/2}\sqrt{2}\frac{\gamma_c}{\lambda_c}\eta_i \tag{J4}$$

with the Îto rule for the white noise η_i. The corresponding Fokker-Planck equation is

$$\partial_t \rho = \partial_x(-f_0\rho) + \frac{1}{2}\partial_x^2\left[\frac{2\gamma_c^2}{\lambda_c^2}e^{-x}\rho\right] + s(x). \tag{J5}$$

Assuming a deterministic introduction size $s(x) = s_C\delta(x - x_0)$, at steady-sate we get

$$K - s_C\theta(x - x_0) = -f_0\rho + e^{-x}\frac{\gamma_c^2}{\lambda_c^2}\rho' - \frac{\gamma_c^2}{\lambda_c^2}\rho e^{-x}, \tag{J6}$$

which for $x > x_0$ is solved by

$$\rho(x) = e^{-e^x/C_m+x}\Big[KEi(e^x/C_m) - KEi(C_m^{-1}) \tag{J7}$$

$$-\frac{s_C\lambda_c^2}{\gamma_c^2}Ei(\frac{e^x}{C_m}) + \frac{s_C\lambda_c^2}{\gamma_c^2}Ei(\frac{e^{x_0}}{C_m})\Big], \tag{J8}$$

where K is an integration constant, Ei is the exponential integral function and

$$C_m = \frac{\gamma_c^2}{|f_0|\lambda_c^2}. \tag{J9}$$

The divergence of Ei at infinity sets $K = s_C\lambda_c^2/(\gamma_c^2)$ and the clone size distribution is

$$\rho(x) = \begin{cases} \left(Ei(e^x/C_m) - Ei(C_m^{-1})\right)e^{-e^x/C_m+x} & \text{for } x < x_0 \\ \left(Ei(e^{x_0}/C_m) - Ei(C_m^{-1})\right)e^{-e^x C_m+x} & \text{for } x > x_0 \end{cases} \tag{J10}$$

or in terms of $x = \log C$

$$\rho(C) = \begin{cases} e^{-C/C_m}\left(Ei(C/C_m) - Ei(C_m^{-1})\right) & \text{for } C < C_0 \\ e^{-C/C_m}\left(Ei(e^{x_0}/C_m) - Ei(C_m^{-1})\right) & \text{for } C > C_0 \end{cases} \tag{J11}$$

The validity of this solution is checked in Fig. S9 and the convergence of the full solution of Eq. I6 (with no birth-death noise) to the analytical solution in the limit of no heritability ($\lambda_c \to \infty$) is show in Fig. S10.

For comparison, in a pure birth-death process (no fitness fluctuations) the clone-size distribution is, for C large enough, $\rho(C) \sim e^{-C/C_m}/C$ where $C_m = (\mu+\nu)/(2(\mu-\nu))$, as shown in Appendix A. These two solutions both have an exponential cutoff, but have very different power-law exponents, corresponding to $\alpha = 0$ and $\alpha = -1$, respectively.
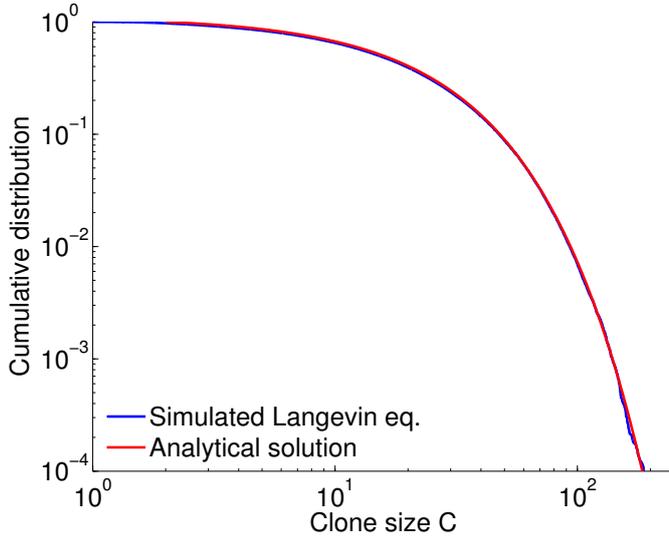
FIG. S9: The result of a simulation of the Langevin equation of the white noise cell-specific fitness model (blue line) compared to the analytical prediction of Eq. J11 (red line) show very good agreement. The parameters are $\nu = 0.2$ day^{-1}, $\mu = 0.4$ day^{-1}, $C_0 = 2$, $\lambda_c = 4$ day^{-1} and $\gamma_c = 8$ day$^{-3/2}$.
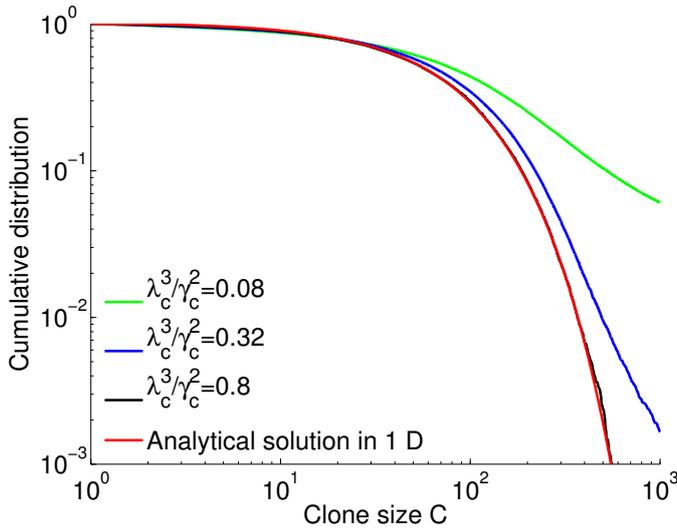

FIG. S10: Convergence of the cell-specific fitness models (Eq. I6) without birth-death noise to Eq. J11 in the limit of no heritability ($\lambda_c \to \infty$). For all four curves $\alpha = 0.2$. Parameters used: $\nu = 0.2$ day^{-1}, $\mu = 0.25$ day^{-1}, $C_0 = 2$ and 1000 new clones introduced each day.

We now add the birth-death noise, *i.e.* consider both types of noise, still in the limit of no heritability. The corresponding Langevin equation reads:

$$\partial_t x_i = f_0 + \sqrt{\mu + \nu}e^{-x_i/2}\xi - e^{-x_i}\frac{\mu + \nu}{2} + e^{-x_i/2}\frac{\sqrt{2}\gamma_c}{\lambda_c}\eta \tag{J12}$$

where all noise is in the Îto convention. Integrating the Fokker Planck associated to this equation gives at steady
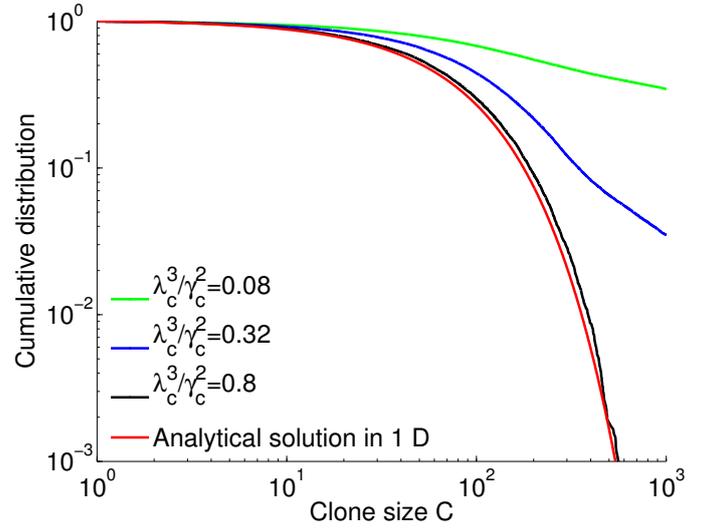

FIG. S11: Convergence of the cell-specific models (Eq. I6) with birth-death noise to the analytical result of Eq. J15 (red line). Keeping constant α while $\lambda_c \to \infty$ and $\gamma_c \to \infty$ we recover the solution of Eq. J15. Parameters are the same as in Fig. S10

state condition

$$K - s_C\theta(x - x_0) = -f_0\rho + \left[\frac{\mu + \nu}{2} + \frac{\gamma_c^2}{\lambda_c^2}\right]e^{-x}\rho' - \frac{\gamma_c^2}{\lambda_c^2}e^{-x}\rho. \tag{J13}$$

In order for ρ to be well defined we set $K = s_C$. For $x > x_0$ the equation is homogeneous and solved by separation of variables:

$$\frac{d\rho}{\rho}e^{-x}\left[\frac{\mu + \nu}{2} + \frac{\gamma_c^2}{\lambda_c^2}\right] = \left(f_0 + \frac{\gamma_c^2}{\lambda_c^2}e^{-x}\right)\rho, \tag{J14}$$

and gives the solution:

$$\rho(C) = \frac{Ke^{-C/C_m}}{C^{1+\alpha}}, \tag{J15}$$

with

$$\alpha = -\left(1 + \frac{(\mu + \nu)\lambda_c^2}{2\gamma_c^2}\right)^{-1}, \tag{J16}$$

which is a power-law with an exponent $0 \le 1 + \alpha \le 1$ and an exponential cutoff

$$C_m = (\mu - \nu)^{-1}\left(\frac{\mu + \nu}{2} + \frac{\gamma_c^2}{\lambda_c^2}\right). \tag{J17}$$

The convergence of the solution of the full system, Eq. I6, to this solution is checked in Fig. S11.

Appendix K: Dynamics of naive and memory cells

In this section we present our results on the division of the population between naive and memory cells and its
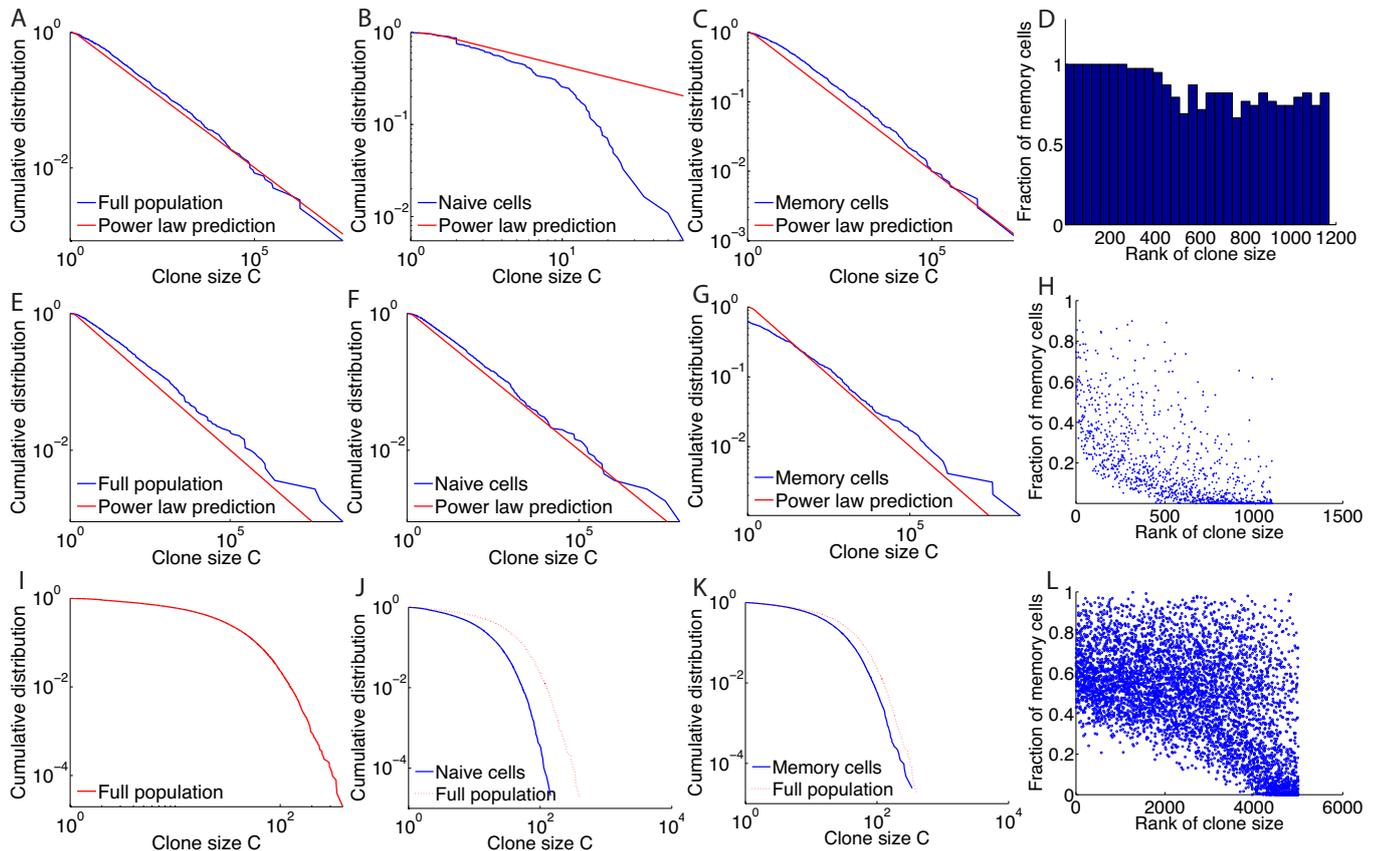
FIG. S12: Simulation results for clone and cell specific model with two cell compartments for naive and memory. Panels A to D are results from clone-specific fitness model with a switching rate θ from naive to memory taken to be infinite (the whole clone switches instantly to memory when above a fitness threshold) and fitness threshold $f_{\mathrm{mem}} = 1$ day^{-1}. Panels E to H are results for a model with clone-specific fitness with a finite switching rate $\theta = 0.05$ days^{-1} and fitness threshold $f_{\mathrm{mem}} = 1$ day^{-1}. For both clone-specific simulations the parameters are: $s_{\mathrm{C}} = 200$ day^{-1}, $C_0 = 2$, $s_{\mathrm{A}} = 1.96 \cdot 10^7$ day^{-1}, $\langle a_{j,0} \rangle = 1$, Var$(a_{j,0}) = 1$, $\lambda = 2$ day^{-1}, $p = 10^{-7}$, $\nu = 0.98$ day^{-1}, $\mu = 1.18$ day^{-1}. Panels I to L are results from simulations of a model with cell-specific fitness with a switching rate $\theta = 0.25$ and threshold $f_{\mathrm{mem}} = 0.5$. The other parameters are: $s_{\mathrm{C}} = 10^4$ day^{-1}, $C_0 = 2$, $\lambda_c = 2$ day^{-1}, $\gamma_c = 4$ day$^{-3/2}$, $\nu = 0.5$ day^{-1}, $\mu = 0.7$ day^{-1}. Panels A, E and I show the clone size distribution of the whole population adding memory and naive contributions to each clone and the power law prediction from the white noise model for clone-specific fitness. Panels B, F and J show the clone size distributions of the naive pool of cells compared to the white noise prediction for the clone-specific fitness (B, F) and the full population distribution for the cell-specific dynamics (J). Panels C, G and K show the clone size distributions of the memory pools (same comparisons as for naive). Panels D, H and L show the fraction of memory cells in clones as a function of their rank (biggest clones have smallest ranks) as a histogram for an infinite switching rate (because clones are either all naive or all memory) and as scatter plots for the two other types of dynamics.

impact on the distribution of clone sizes. In our simulations and analysis so far we have always considered the system to be uniform, because most of the data available at this time is not sorted into naive and memory/effector cells and because the main difference between naive and memory cells (higher stimulation of memory cells by binding events) is already included in our models.

In principle, memory and naive cells could have a completely different set of parameters. None of the values of these parameters are known with high, accuracy although it emerges from all studies that memory cells have a higher turnover rate (or death rate μ) than naive cells.

However, our estimate of f_0 (which is the average division rate minus the death rate) cannot be performed for separate groups of naive and memory cells without knowledge of their total population and the rate of conversion from naive to memory cells. For these reasons we keep the same effective f_0 for the whole population.

We model the immune system with two pools of cells: naive and memory/effector for both the clone-specific and cell-specific fitness models. Clones from the naive pool with fitness over a given threshold f_{mem} turn irreversibly into memory cells at a certain rate θ per day. In both cases the two pools have the same dynamics but mem-

ory cells have a higher turnover: the death rate μ and the basal birth rate ν are higher in the memory pool but their difference f_0 is unchanged. This means the birth-death noise is higher in the memory pool. We find that in the clone-specific fitness model it does not affect the power-law exponent of the clone-size distribution, but it does affect strongly the distribution (and more specifically the cutoff value C_m) in the cell-specific fitness model, as birth-death noise is of the same order of magnitude as the environmental noise (Fig. S12).

In the clone-specific fitness model, we find that the distribution still displays power-law behavior with the expected exponent (Fig. S12A and E). For very high rates of conversion from naive to memory we see that naive cell distributions drop exponentially above a threshold, as all high fitness clones are completely converted into memory (Fig. S12B). For lower rates of conversion both memory and naive pools have heavy tails and the memory pool has a higher power law cutoff for small values (Fig. S12F and G). For the cell-specific fitness model we find that the memory pool can have significantly heavier tails (as its dynamics is much faster) and a higher cutoff C_m (a power-law like behavior in a wider range) than the naive pool (Fig. S12A-B-C). In all cases we recover that naive clones are smaller than memory clones, or in other words large clones are mostly made up of memory cells (Fig. S12D-H-L).

Appendix L: Effects of hypermutations

In this section we show that including the effect of somatic hypermutations in the clone-specific fitness dynamics does not change the power law behavior of the distribution. We model the somatic hypermutations by replacing a small fraction of the offspring of the fastest expanding clones by new clones with binding affinities close to the ones of their parents. For each clone such that $f_i > f_{\text{hyp}}$, offspring with hypermutated receptors are being produced with rate r_{hyp}. A large fraction r_{del} of those are assumed to have acquired deleterious mutations and are removed from the pool. The rest (fraction $1 - r_{\text{del}}$) form new clones of size 1 (in our definition, which differs from the usual convention for B cells, a clone is a subset of cells with the exact same receptor sequence). The interaction matrix $K_{i',j}$ of each new, hypermutated clone i' is formed from the interaction matrix $K_{i,j}$ of its progenitor i by changing each non-zero entry of $K_{i,j}$ to:

$$K_{i',j} = \begin{cases} 0 & \text{with probability } 1 - p_{\text{hyp}} \\ \psi K_{i,j} + (1 - \psi) + \sigma_{\text{hyp}}\zeta & \text{otherwise,} \end{cases} \quad (\text{L1})$$

where ψ is a parameter controling the heritability of the values of the K entries, and p_{hyp} the probability that the specificity to a given antigen is passed on to the hypermutated offspring; ζ is a Gaussian variable of mean 0 and variance 1. To compensate the loss of specificity, zero entries of $K_{i,j}$ are assigned new, non-zero values of
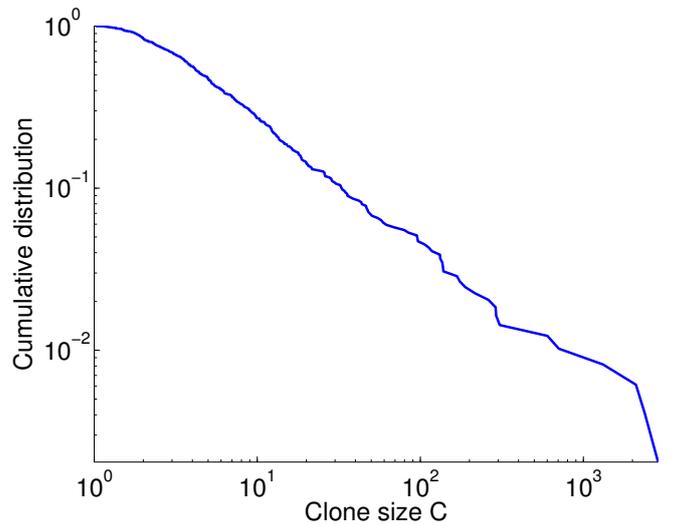

FIG. S13: We show the clone size distribution that results from simulating a model of clone-specific fitness with somatic hypermuations as described in Appendix L and Eq. L1. The distribution exhibits clear power law behavior. Hypermutation parameters are: $f_{\text{hyp}} = 4$ days^{-1}, $r_{\text{hyp}} = 0.01$ days^{-1}, $r_{\text{del}} = 0.01$, $p_{\text{hyp}} = 0.5$, $\psi = 0.7$ and $\sigma_{hyp} = 0.05$. Other parameters are: $s_C = 200$ day^{-1}, $C_0 = 2$, $s_A = 1.5 \cdot 10^7$ day^{-1}, $\langle a_{j,0} \rangle = 1$, $\text{Var}(a_{j,0}) = 1$, $\lambda = 2$ day^{-1}, $p = 10^{-3}$, $\nu = 0.75$ day^{-1}, $\mu = 1.15$ day^{-1}. Non zero $K_{i,j}$ entries from thymic output have mean 1 and standard deviation 0.3.

binding affinities with probability $(1 - p_{\text{hyp}})p$ (where we recall that p is the probability for a given clone to be specific to a given antigen), so that the number of non-zero values of K remains the same on average. The value of these new binding affinities are drawn completely at random, as before (no inheritance).

A small part of the hypermutated clones branch out and undergo affinity maturation, meaning that they are selected generation after generation. Their fitness increases until the environment varies enough for their branch to be obsolete and decay back to low fitnesses. The effect of hypermutations on the distribution depends on the ratio between the speed at which hypermutated lineages drift in fitness space and the time scale for variations of the environment (λ^{-1}).

Somatic hypermutations add a source of stochasticity in fitness and increase the number of large clones. Accordingly, simulations of the model with hypermutations (see Fig. S13) show that the clone size distribution still exhibits power law behavior, but with a lower exponent (heavier tails) due to the extra stochasticity induced by hypermutations.

Appendix M: Time dependent source terms and aging

In this section we investigate the effect of a decaying thymic output on the distribution of clones for the anti-
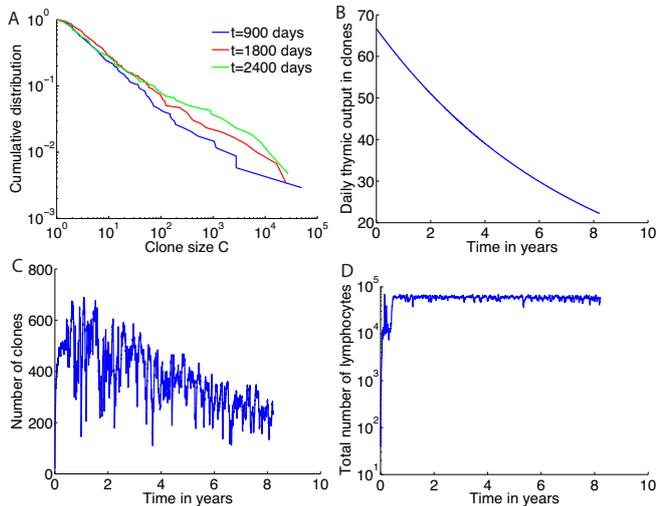
FIG. S14: Results of a simulation of a model of clone-specific fitness with a decaying source term and balancing decrease of $|f_0|$ to keep the population size constant. A. The clone size distributions at different time points maintains a power law behavior with an exponent α that decreases with time. B. Decay of the thymic output with time. C. Total number of clones is decreasing with time. D. Total number of cells is maintained by tuning the rate f_0. Parameters used are: source decay timescale $tau = 8.3$ yr, $s_{C,0} = 200$ day^{-1}, $C_0 = 2$, $s_A = 1.5 \cdot 10^7$ day^{-1}, $\langle a_{j,0} \rangle = 1$, Var$(a_{j,0}) = 1$, $\lambda_c = 2$ day^{-1}, $p = 10^{-7}$, $\nu + \mu = 1.9$ day^{-1}, $f_0 = -0.4$ day^{-1} at time $t = 0$.

gen recognition based model. In all our simulations we assume that the source of new clones (thymic output) produces a number of clones that is on average constant with time. It is an approximation since in humans or in mice thymic output is high at birth and during growth and slowly decreases during adult life. This decrease is very slow compared to the time scales involved in this analysis [5] and so within the time frames considered it

can be considered constant. In this section we look at the effect of this decrease over long time scales.

We model the decrease of thymic output with an exponentially decaying (with time) source term. In real organisms, homeostatic control ensures that the total number of cells in the body is conserved during this reduction of thymic output. We do not model this homeostatic control explicitly, but rather tune the difference between birth and death rates f_0 to keep the total population constant on average, which we showed was equivalent (see Fig. S2). Simple averaging of the dynamics shows that

$$\frac{d\langle N \rangle}{dt} = f_0 N + n_C \langle f_i C_i \rangle + s_C \qquad (M1)$$

where n_C is the number of clones in the system and N is the total number of cells. Since our source term is a function of time, to have on average a constant total population size we need to define :

$$f_0(t) = -\frac{n_C(t)\langle f_i C_i \rangle + s_C(t)}{N}. \qquad (M2)$$

We show the results of a simulation in Fig. S14 with $s_C = s_{C,0} e^{-t/\tau}$, $\tau = 8.3$ yr. We recover results known in humans and get predictions for the behavior of the exponent of the power law at different ages. We find that, with the decrease of thymic output, the number of clones is decreasing (Fig. S14C), meaning that clones become on average fitter (i.e. better at recognizing antigens), but at the expense of repertoire diversity. Keeping the population constant (Fig. S14D) slowly decreases the decaying rate of clones $|f_0|$ and so is expected to decrease the exponent, which behaves as $\alpha = \lambda |f_0|/A^2$. Accordingly, simulations show a clear power-law behavior in the clone-size distribution (Fig. S14A), with the tail of the distribution becoming heavier with age. We thus expect older organisms with lower thymic output to have a larger tail in their clone-size distribution. We predict thymectomy to lead to distributions with very fat tails.

[1] Cover TM, Thomas JA (1991) *Elements of Information Theory* (Wiley-Interscience, New York, NY, USA).

[2] Weinstein JA, Jiang N, White RA, Fisher DS, Quake SR (2009) High-throughput sequencing of the zebrafish antibody repertoire. *Science* 324:807–810.

[3] Clauset A, Shalizi CR, Newman MJ (2009) Power-law distributions in empirical data. *SIAM Rev* 51:661–703.

[4] Bolkhovskaya OV, Zorin DY, Ivanchenko MV (2014) Assessing T Cell Clonal Size Distribution: A Non-Parametric Approach. *PLoS ONE* 9:e108658.

[5] Bains I, Antia R, Callard R, Yates AJ (2009) Quantifying the development of the peripheral naive CD4 + T-cell pool in humans. *Blood* 113:5480–5487.