Fluctuating fitness shapes the clone-size distribution of immune repertoires

Jonathan Desponds[a], Thierry Mora[b,1], and Aleksandra M. Walczak[a]

[a]Laboratoire de Physique Théorique, CNRS, Université Pierre et Marie Curie (UPMC) and École Normale Supérieure, 75005 Paris, France; and [b]Laboratoire de Physique Statistique, CNRS, UPMC and École Normale Supérieure, 75005 Paris, France

Edited by José N. Onuchic, Rice University, Houston, TX, and approved November 11, 2015 (received for review July 2, 2015)

The adaptive immune system relies on the diversity of receptors expressed on the surface of B- and T cells to protect the organism from a vast amount of pathogenic threats. The proliferation and degradation dynamics of different cell types (B cells, T cells, naive, memory) is governed by a variety of antigenic and environmental signals, yet the observed clone sizes follow a universal power-law distribution. Guided by this reproducibility we propose effective models of somatic evolution where cell fate depends on an effective fitness. This fitness is determined by growth factors acting either on clones of cells with the same receptor responding to specific antigens, or directly on single cells with no regard for clones. We identify fluctuations in the fitness acting specifically on clones as the essential ingredient leading to the observed distributions. Combining our models with experiments, we characterize the scale of fluctuations in antigenic environments and we provide tools to identify the relevant growth signals in different tissues and organisms. Our results generalize to any evolving population in a fluctuating environment.

immune repertoire | population dynamics | fluctuating fitness | lymphocyte receptor | repertoire sequencing

Antigen-specific receptors expressed on the membrane of B- and T cells (B-cell receptors, BCRs and T-cell receptors, TCRs) recognize pathogens and initiate an adaptive immune response (1). An efficient response relies on the large diversity of receptors that is maintained from a source of newly generated cells, each expressing a unique receptor. These progenitor cells later divide or die, and their offspring make up clones of cells that share a common receptor. The sizes of clones vary, as they depend on the particular history of cell divisions and deaths in the clone. The clone-size distribution thus bears signatures of the challenges faced by the adaptive system. Understanding the form of the clone-size distribution in healthy individuals is an important step in characterizing the antigenic recognition process and the functioning of the adaptive immune system. It also presents an important starting point for describing statistical deviations seen in individuals with compromised immune responses.

High-throughput sequencing experiments in different cell types and species (2–9) have allowed for the quantification of clone sizes and their distributions (2, 9–11). Previous population dynamics approaches to repertoire evolution have taken great care in precisely modeling these processes for each compartment of the population, through the various mechanisms by which cells grow, die, communicate, and change phenotype (12–17). However, one of the most striking properties of repertoire statistics revealed by high-throughput sequencing is the observation of power laws in clone-size distributions (Fig. 1 *A* and *B*), which holds true for various species (human, mice, zebrafish), cell type (B- and T cells), and subsets (naive and memory, CD4 and CD8), and seems to be insensitive to these context-dependent details. It remains unclear, however, what universal features of these dynamics lead to the observed power-law distributions. Here we identify the key biological parameters of the repertoire dynamics that govern its behavior.

The wide range and types of interactions that influence a B- or T-cell fate happen in a complex, dynamical environment with inhomogeneous spatial distributions. They are difficult to measure in vivo, making their quantitative characterization elusive. Motivated by the universality of the observed clone-size distribution, we describe the effective interaction between the immune cells and their environment as a stochastic process governed by only a few relevant parameters. All cells proliferate and die depending on the strength of antigenic and cytokine signals they receive from the environment, which together determine their net growth rate (Fig. 1*C*). This effective fitness that fluctuates in time is central to our description. We find that its general properties determine the form of the clone-size distribution. We distinguish two broad classes of models, according to whether these fitness fluctuations are clone-specific (mediated by their specific BCR or TCR) or cell-specific (mediated by phenotypic fluctuations such as the number of cytokine receptors). We identify the models that are compatible with the experimentally observed distributions of clone sizes. These distributions do not depend on the detailed mechanisms of cell signaling and growth, but rather emerge as a result of self-organization, with no need for fine-tuned interactions. Performing a series of validated approximations, we find a simple algebraic relationship constraining the different timescales of the problem by the experimentally observed exponent of the clone-size distribution. This result allows for testable predictions and estimates of the rates that govern the diversity of a clonal distribution.

Results

Clone Dynamics in a Fluctuating Antigenic Landscape. The fate of the cells of the adaptive immune system depends on a variety of clone-specific stimulations. The recognition of pathogens triggers large events of fast clone proliferation followed by a relative decay, with some cells being stored as memory cells to fend off future infections. Naive cells, which have not yet recognized an antigen, do not usually undergo such extreme events of proliferation and

Significance

Receptors on the surface of lymphocytes specifically recognize foreign pathogens. The diversity of these receptors sets the range of infections that can be detected and fought off. Recent experiments show that, despite the many differences between these receptors in different cell types and species, their distribution of diversity is a strikingly reproducible power law. By introducing effective models of repertoire dynamics that include environmental and antigenic fluctuations affecting lymphocyte growth or "fitness," we show that a temporally fluctuating fitness is responsible for the observed heavy tail distribution. These models are general and describe the dynamics of various cell types in different species. They allow for the classification of the functionally relevant repertoire dynamics from the features of the experimental distributions.

Author contributions: J.D., T.M., and A.M.W. designed research, performed research, contributed new reagents/analytic tools, analyzed data, and wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Freely available online through the PNAS open access option.

[1]To whom correspondence should be addressed. Email: tmora@lps.ens.fr.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1512977112/-/DCSupplemental.

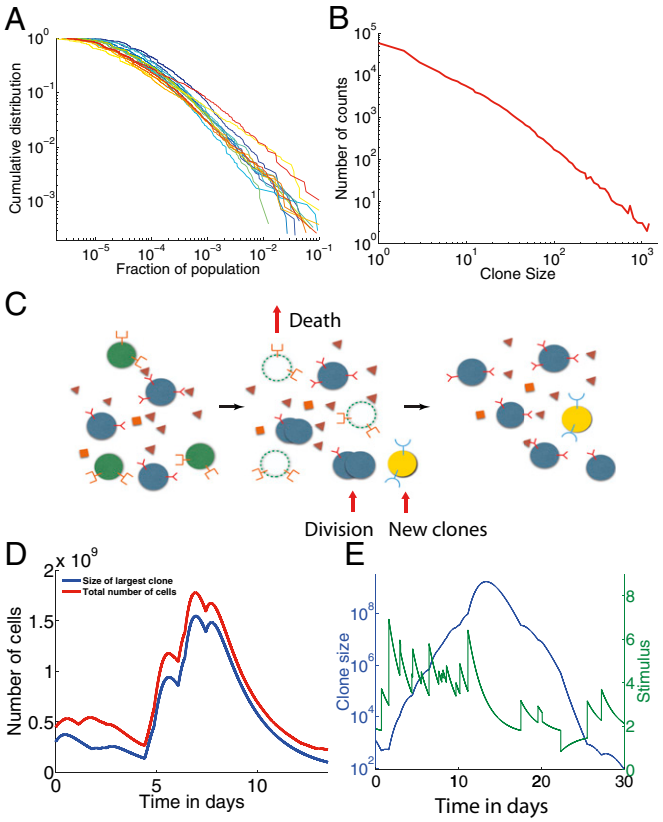

Fig. 1. Experimental clone-size distributions have heavy tails. (*A*) B-cell zebrafish experimental cumulative clone-size distribution for 14 fish as a function of the fraction of the population occupied by that clone from data in Weinstein et al. (2). (*B*) Clone-size distribution for murine T cells from Zarnitsyna et al. (11) (data plotted as presented in original paper). (*C*) The dynamics of adaptive immune cells include specific interactions with antigens that promote division and prevent cell death. New cells are introduced from the thymus or bone marrow output on average balance each other to create a steady-state population. (*D* and *E*) Example trajectories from simulations of the immune cell population dynamics in Eq. **1**. The total number of cells (*D*) shows large variations after an exceptional event of a large pathogenic invasion. One or a few cells that react to that specific antigen grow up to a macroscopic portion of the total population, and then decrease back to normal sizes after the invasion. A typical clone-size trajectory along with its pathogenic stimulation $\sum_j K_{ij} a_j(t)$ shows the coupling between clone growth and variations of the antigenic environment (*E*). Parameters used: $s_C = 2{,}000$ day^{-1}, $C_0 = 2$, $s_A = 1.96 \cdot 10^7$ day^{-1}, $a_{j,0} = a_0 = 1$, $\lambda = 2$ day^{-1}, $p = 10^{-7}$, $\nu = 0.98$ day^{-1}, and $\mu = 1.18$ day^{-1}.

death, but their survival relies on short binding events (called "tickling") to antigens that are natural to the organism (self-proteins) (18, 19). Because receptors are conserved throughout the whole clone (with the exception of B-cell hypermutations), clones that are better at recognizing self-antigens and pathogens will on average grow to larger populations than bad binders. By analogy to Darwinian evolution, they are "fitter" in their local, time-varying environment.

We first present a general model for clonal dynamics that accounts for the characteristics common to all cell types, following previous work by de Boer, Perelson, and collaborators (14, 20, 21). We later explore the effect of specific features such as hypermutations, memory/naive compartmentalization, and thymic output decay on the clone-size distribution.

We denote by $a_j(t)$ the overall concentration of an antigen j as a function of time. We assume that after its introduction at a random time t_j, this concentration decays exponentially with a characteristic lifetime of antigens λ^{-1}, $a_j(t) = a_{j,0} e^{-\lambda(t-t_j)}$ as pathogens are cleared out of the organism, either passively or through the action of the immune response. Lymphocyte receptors are specific to certain antigens, but this specificity is degenerate, a phenomenon referred to as cross-reactivity or polyspecificity. The extent to which a lymphocyte expressing receptor i interacts with antigen j (foreign or self) is encoded in the cross-reactivity function K_{ij}, which is zero if i and j do not interact, or a positive number drawn from a distribution to be specified, if they do. In general, interactions between lymphocytes and antigens effectively promote growth and suppress cell death, but for simplicity we can assume that the effect is restricted to the division rate. In a linear approximation, this influence is proportional to $\sum_j K_{ij} a_j(t)$, i.e., the combined effect of all antigens j for which clone i is specific. This leads to the following dynamics for the evolution of the size C_i of clone i (Fig. 1*C*):

$$\frac{dC_i}{dt} = \left(\nu + \sum_j K_{ij} a_j(t) - \mu\right) C_i + B\xi_i(t), \qquad [1]$$

where ν and μ are the basal division and death rates, respectively, and where $B\xi_i(t)$ is a birth–death noise of intensity $B^2 = (\nu + \sum_j K_{ij} a_j(t) + \mu) C_i$, with $\xi_i(t)$ a unit Gaussian white noise (see *SI Appendix*, section A for details about birth–death noise).

New clones, with a small typical initial size C_0, are constantly produced and released into the periphery with rate s_C (Fig. 1*C*). For example, a number on the order of $s_C = 10^8$ new T cells is output by the thymus daily in humans (22). Because the total number of T cells is on the order of 10^{11}, this means that the net effect of cell death and proliferation results in a negative average growth rate of 10^{-3} days^{-1} in homeostatic conditions (22). Because the probability of rearranging the exact same receptor independently is very low ($<10^{-10}$) (23), we assume that each new clone is unique and comes with its own set of cross-reactivity coefficients K_{ij}. Assuming a rate s_A of new antigens, the average net growth rate in Eq. **1** is $f_0 = \nu + \langle a_{j,0} \rangle \langle K \rangle s_A \lambda^{-1} - \mu < 0$, and the stationary number of clones should fluctuate around $N_C \approx s_C |f_0|^{-1}$ clones. This is just an average, and treating each clone independently may lead to large variations in the total number of cells (i.e., the sum of sizes of all clones). To maintain a constant population size, clones compete with each other for specific resources (pathogens or self-antigens) and homeostatic control can be maintained by a global resource such as Interleukin 7 or Interleukin 2. Here we do not model this homeostatic control explicitly, but instead assume that the division and death rates ν, μ are tuned to achieve a given repertoire size. We verified that adding an explicit homeostatic control did not affect our results (*SI Appendix*, Fig. S2 and *SI Appendix*, section B).

We simulated the dynamics of a population of clones interacting with a large population of antigens. Each antigen interacts with each present clone with probability $p = 10^{-7}$, and with strength K_{ij} drawn from a Gaussian distribution of mean 1 and variance 1 (truncated to positive values). Although it has been argued that the breadth of cross-reactivity and affinity to self-antigens are correlated (24, 25), here for simplicity we draw them independently, as we do not expect this correlation to qualitatively affect the results. A typical trajectory of the antigenic stimulation undergone by a given clone, $\sum_j K_{ij} a_j$, is shown in Fig. 1*E* (green curve), and shows how clone growth tracks the variations of the antigenic environment. When the stimulation is particularly strong, the model recapitulates the typical behavior experimentally observed at the population level following a pathogenic invasion (26, 27), as illustrated in Fig. 1*D*: The population of a clone explodes (red curve), driving the growth of the total population (blue curve), while taking over a large fraction of the carrying capacity of the system, and then decays back as the infection is cleared.

On average, the effects of division and death almost balance each other, with a slight bias toward death because of the turnover imposed by thymic or bone marrow output. However, at a given time, a clone that has high affinity for several present antigens will undergo a transient but rapid growth, whereas most other clones will decay slowly toward extinction. In other words, locally in time, the antigenic environment creates a unique "fitness" for each clone. Because growth is exponential in time,

these differential fitnesses can lead to very large differences in clone sizes, even if variability in antigen concentrations or affinities is nominally small. We thus expect to observe large tails in the distribution of clone size. Fig. 2*A* shows the cumulative probability distribution function (CDF) of clone sizes obtained at steady state (blue curve) showing a clear power-law behavior for large clones, spanning several decades.

The exponent of the power law is independent of the introduction size of clones (Fig. 2*A*, *Inset*) and the specifics of the randomness in the environment (exponential decay, random number of partners, random interaction strength) as long as its first and second moment are kept fixed (*SI Appendix*, Fig. S3 and *SI Appendix*, section C).

Simplified Models and the Origin of the Power Law. To understand the power-law behavior observed in the simulations, and its robustness to various parameters and sources of stochasticity, we decompose the overall fitness of a clone at a given time (its instantaneous growth rate) into a constant, clone-independent part equal to its average $f_0 < 0$, and a clone-specific fluctuating part of zero mean, denoted by $f_i(t)$. This leads to rewriting Eq. **1** as

$$\frac{dC_i}{dt} = [f_0 + f_i(t)]C_i(t) + B\xi_i(t), \qquad [2]$$

with $B^2 \approx (|f_0| + 2\mu)C_i$.

The function $f_i(t)$ encodes the fluctuations of the environment as experienced by clone i. Because antigens can be recognized by several receptors, these fluctuations may be correlated between clones. Assuming that these correlations are weak, $\langle f_i(t) f_j(t')\rangle \approx 0$, amounts to treating each clone independently of each other, and thus to reducing the problem to the single clone level. The stochastic process giving rise to $f_i(t)$ is a sum of Poisson-distributed exponentially decaying spikes. This process is not easily amenable to analytical treatment, but we can replace it with a simpler stochastic process with the same temporal autocorrelation function. This autocorrelation is given by $\langle f_i(t) f_i(t')\rangle = A^2 e^{-\lambda|t-t'|}$, with the antigenic noise strength $A^2 = s_A p a_0^2 \langle K^2\rangle \lambda^{-1}$, and where we recall that λ^{-1} is the characteristic lifetime of antigens. The simplest process with the same autocorrelation function is given by an overdamped spring in a thermal bath, or Ornstein–Uhlenbeck process,

$$\frac{df_i}{dt} = -\lambda f_i + \sqrt{2}\gamma\eta_i(t), \qquad [3]$$

with $\eta_i(t)$ a Gaussian white noise of intensity 1 and $\gamma = A\sqrt{\lambda}$ quantifies the strength of variability of the antigenic environment (*SI Appendix*, section D). This is also the process of maximum entropy or caliber (28) with that autocorrelation function (*SI Appendix*, section E and ref. 29).

The effect of the birth–death noise $B\xi_i(t)$ is negligible compared with the fitness variations for large clones and it has no effect on the tail (*SI Appendix*, Fig. S5 and *SI Appendix*, section F). It can thus be ignored when looking at the tail of the distribution and its power-law exponent, but it will play an important role for defining the range over which the power law is satisfied.

The population dynamics described by Eqs. **2** and **3** can be reformulated in terms of a Fokker–Planck equation for the joint abundance ρ of clones of a given log size $x = \log C$ and a given fitness f:

$$\frac{\partial\rho(x,f,t)}{\partial t} = -(f_0 + f)\frac{\partial\rho}{\partial x} + \lambda\frac{\partial(f\rho)}{\partial f} + \gamma^2\frac{\partial^2\rho}{\partial f^2} + s(x,f), \qquad [4]$$

where the source term $s(x,f)$ describes new clones arriving at rate s_C with size C_0 and normally distributed fitnesses of variance $\langle f^2\rangle = \gamma^2/\lambda$. This Fokker–Planck equation can be solved numerically with finite element methods with an absorbing boundary condition at $x = 0$ to account for clone extinction. The solution, represented by the black curve in Fig. 2*A*, matches closely that of the full simulated population dynamics (in blue). The power-law behavior is apparent above a transition point that depends on the distribution of introduction sizes of new clones and the parameters of the model (see below). Intuitively, the microscopic details of the noise are not expected to matter when considering long timescales, as a consequence of the central limit theorem. However, the long tails of the distribution of clone sizes involve rare events and belong to the regime of large deviations, for which these microscopic details may be important. Therefore, the agreement between the process described by the overdamped spring and the exponentially decaying, Poisson-distributed antigens is not guaranteed, and in fact does not hold in all parameter regimes (*SI Appendix*, Fig. S8).

We can further simplify the properties of the noise by assuming that its autocorrelation time is small compared with other timescales. This leads to taking the limit $\gamma, \lambda \to \infty$ while keeping their ratio $\sigma = \gamma/\lambda$ constant, so that $f_i(t)$ is just a Gaussian white noise with $\langle f_i(t) f_i(t')\rangle = 2\sigma^2\delta(t-t')$ (*SI Appendix*, section F and *SI Appendix*, Fig. S4). The corresponding Fokker–Planck equation now reads

$$\partial_t\rho(x,t) = -f_0\partial_x\rho(x,t) + \sigma^2\partial_x^2\rho(x,t) + s(x), \qquad [5]$$

with $s(x) = s_C\delta(x - \log(C_0))$. This equation can be solved analytically at steady state, and the resulting clone-size distribution is, for $C > C_0$,

$$\rho(C) = \frac{s_C}{\alpha\sigma^2}\frac{1}{C^{\alpha+1}}, \qquad [6]$$

with $\alpha = |f_0|/\sigma^2 = \lambda|f_0|/A^2$ (details in *SI Appendix*, section F). The full solution, represented in Fig. 2*A* in red, captures well the long-tail behavior of the clone-size distribution despite ignoring the temporal correlations of the noise, and approaches the solution of the colored-noise model (Eq. **3**) as $\lambda, \gamma \to \infty$, as expected (Fig. 2*A*).

The power-law behavior and its exponent depend on the noise intensity, but are otherwise insensitive to the precise details of the microscopic noise, including its temporal properties. Fat tails (small α) are expected when the average cell lifetime is long (small $|f_0|$) and when the antigenic noise is high (large σ or A). The explicit expression for the exponent of the power law $1 + \alpha$ as a function of the biological parameters can be used to infer the antigenic noise strength A^2 directly from data. The typical net clone decay rate $|f_0| \approx 10^{-3}$ can be estimated from thymic output and repertoire size, as discussed earlier. The characteristic lifetime of antigens λ^{-1} is harder to estimate, as it corresponds to the turnover time of the antigens that the body is exposed to, but is probably on the order of days or a few weeks, $\lambda \approx 0.1$ day^{-1}. We estimated $\alpha = 1 \pm 0.2$ from the zebrafish data of Fig. 1*A* (2, 10) using canonical methods of power-law exponent extraction (30) (see *SI Appendix*, section G for details), and also found a similar value in human T cells (31). The resulting estimate, $A = 10^{-2}$ day^{-1}, is rather striking, as it implies that fluctuations in the net clone growth rate, A, are much larger than its average f_0.

Whereas the distribution always exhibits a power law for large clones, this behavior does not extend to clones of arbitrarily small sizes, where the details of the noise and how new clones are introduced matter. We define a power-law cutoff C^* as the smallest clone size for which the cumulative distribution function differs from its best power-law fit by less than 10%. Using numerical solutions to the Fokker–Planck equation associated with the colored-noise model, we can draw a map of C^* as a function of the parameters of the system. In Fig. 2 *B* and *C* we show how C^* varies as a function of the introduction size for different values of the dimensionless parameter related to the effective strength of antigen fluctuations relative to their characteristic lifetime at fixed power-law exponents. In principle, one can use this dependency to infer effective parameters from data. In practice, when dealing with data it is more convenient to consider the value of the cumulative distribution at C^*, rather than C^* itself. For example, fixing $C_0 = 4$ and fitting the curve of Fig. 1*A* with our simplified model using λ as an adjustable parameter, we obtain $\lambda \approx 0.14$ day^{-1} (*SI Appendix*, section G),

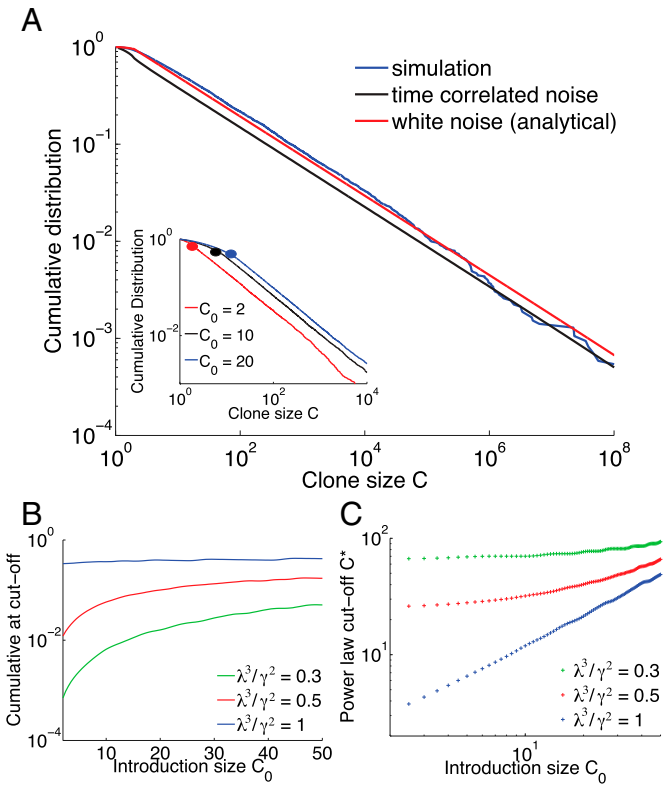

Fig. 2. Clone-size distributions for populations with fluctuating antigenic, clone-specific fitness. (*A*) Comparison of simulations and simplified models of clone dynamics. Blue curve: cumulative distribution of clone sizes obtained from the simulation of Eq. **1**. Black curve: a simplified, numerically solvable model of random clone-specific growth, also predicts a power-law behavior. Red curve: analytical solution for the Gaussian white-noise model, Eq. **4**. Parameters used: $\nu = 0.98$ day^{-1}, $\mu = 1.18$ day^{-1}, $\lambda = 2$ day^{-1}, $s_C = 2{,}000$ day^{-1}, $C_0 = 2$, and $s_A = 1.96 \cdot 10^7$ day^{-1}. (*Inset*) The exponent is independent of the initial clone size. Results from simulation with different values of the introduction clone size. The cutoff value of the power-law behavior, represented here as a dot, is strongly dependent on the value of C_0. Parameters are $\nu = 0.2$ day^{-1}, $\mu = 0.4$ day^{-1}, $\lambda = 2$ day^{-1}, $\gamma = 1$ day$^{-3/2}$, and $s_C = 5{,}000$. (*B*) Value of the CDF at the point of the power-law cutoff as a function of the introduction clone size C_0 for different values of a dimensionless parameter related to the effective strength of antigen fluctuations relative to their characteristic lifetime λ^3/γ^2 for a fixed power-law exponent α. We use the CDF because it is robust, invariant under multiplicative rescaling of the clone sizes. This way we do not need to correct directly for PCR multiplication or sampling. Parameters for *B* and *C* are $\nu = 4.491$ days^{-1}, $\mu = 5.489$ days^{-1}, and $\alpha = -0.998$. (*C*) Power-law cutoff as a function of the introduction clone size.

which corresponds to a characteristic lifetime of antigens of around a week. Although this estimate must be taken with care, because of possible PCR amplification biases plaguing the small clone size end of the distribution, the procedure described here can be applied generally to any future repertoire sequencing dataset for which reliable sequence counts are available.

A Model of Fluctuating Phenotypic Fitness. So far, we have assumed that fitness fluctuations are identical for all members of a same clone. However, the division and death of lymphocytes do not only depend on signaling through their TCR or BCR. For example, cytokines are also growth inducers and homeostatic agents (32, 33), and the ability to bind to cytokines depends on single-cell properties such as the number of cytokine receptors on the membrane of a given cell, independent of their BCR or TCR. Other stochastic single-cell factors may affect cell division and death. These signals and factors are cell-specific, as opposed to the clone-specific properties related to BCR or TCR binding. Together, they define a global phenotypic state of the cell that determines its time-varying fitness, independent of the clone and its TCR or BCR. This does not mean that these phenotypic fitness fluctuations are independent across the cells belonging to the same clone. Cells within a clone share a common ancestry, and may have inherited some phenotypic properties of their common ancestors, making their fitnesses effectively correlated with each other. However, this phenotypic memory gets lost over time, unlike fitness effects mediated by antigen-specific receptors.

We account for these phenotypic fitness fluctuations by a function $f_c(t)$ quantifying how much the fitness of an individual cell c differs from the average fitness f_0. This fitness difference is assumed to be partially heritable, which we model by

$$\frac{df_c}{dt} = -\lambda_c f_c(t) + \sqrt{2}\gamma_c \eta_c(t), \qquad [7]$$

where λ_c^{-1} is the heritability, or the typical time over which the fitness-determining trait is inherited, γ_c quantifies the variability of the fitness trait, and $\eta_c(t)$ is a cell-specific Gaussian white noise of power 1. Despite its formal equivalence with Eq. **3**, it is important to note that here the fitness dynamics occurs at the level of the single cell (and its offspring) instead of the entire clone. The dynamics of the fitness $f_i(t)$ of a given clone i can be approximated from Eq. **7** by averaging the fitnesses $f_c(t)$ of cells in that clone, yielding

$$\frac{dC_i}{dt} = [f_0 + f_i(t)]C_i(t) + \sqrt{(\nu+\mu)C_i(t)}\xi_i(t), \qquad [8]$$

$$\frac{df_i}{dt} = -\lambda_c f_i(t) + \frac{1}{\sqrt{C_i(t)}}\sqrt{2}\gamma_c \eta_i(t), \qquad [9]$$

where $\eta_i(t)$ and $\xi_i(t)$ are clone-specific white noise of intensity 1, and ν and μ are the average birth and death rates, respectively, so that $f_0 = \nu - \mu$ (details in *SI Appendix*, section I). The difference with Eq. **3** is the $1/\sqrt{C_i(t)}$ prefactor in the fitness noise $\eta_i(t)$, which stems from the averaging of that noise over all cells in the clone, by virtue of the law of large numbers. Because of this prefactor, the fitness noise is now of the same order of magnitude as the birth–death noise, which must now be fully taken into account. Taking Eqs. **8** and **9** at the population level gives a Fokker–Planck equation with a source term accounting for the import of new clones. We verify the numerical steady-state Fokker–Planck solution against Gillespie simulations (*SI Appendix*, Fig. S6; see *SI Appendix*, section H for details).

Fig. 3 *A* and *B* shows the distribution of clone sizes for different values of the phenotypic relaxation rate λ_c and environment amplitude γ_c. These distributions vary from a sharp exponential drop in the case of low heritability (large λ_c) to heavier tails in the case of long conserved cell states (small λ_c). To quantify the extent to which these distributions can be described as heavy-tailed, we fit them to a power law with exponential cutoff, $\rho(C) \propto C^{-1-\alpha} e^{-C/C_m}$, where C_m is the value below which the distribution could be interpreted as an (imperfect) power law. Fig. 3*C* shows a strong dependency of this cutoff with the phenotypic memory λ_c^{-1}. The longer the phenotypic memory λ_c^{-1}, the more clone-specific the fitness looks, and the more the distribution can be mistaken for a power law in a finite-size experimental distribution. Larger birth–death noise also extends the range of validity of the power law. As a result, and despite the absence of a true power-law behavior, these models of fluctuating phenotypic fitnesses cannot be discarded based on current experimental data.

The model can be solved exactly at the two extremes of the heritability parameter λ_c. In the limit of infinite heritability ($\lambda_c \to 0$) the system is governed by selective sweeps. The clone with the largest fitness completely dominates the population, until it is replaced by a better one, giving rise to a trivial clone-size distribution. In the opposite limit, when heritability goes to 0 ($\lambda_c \to +\infty$), the Fokker–Planck equation can be solved analytically (*SI Appendix*, sections I and J), yielding an exact power law with exponential cutoff, $\rho(C) \propto C^{-1-\alpha} e^{-C/C_m}$, with $\alpha = -[1 + (\mu+\nu)\lambda_c^2/2\gamma_c^2]^{-1}$ and

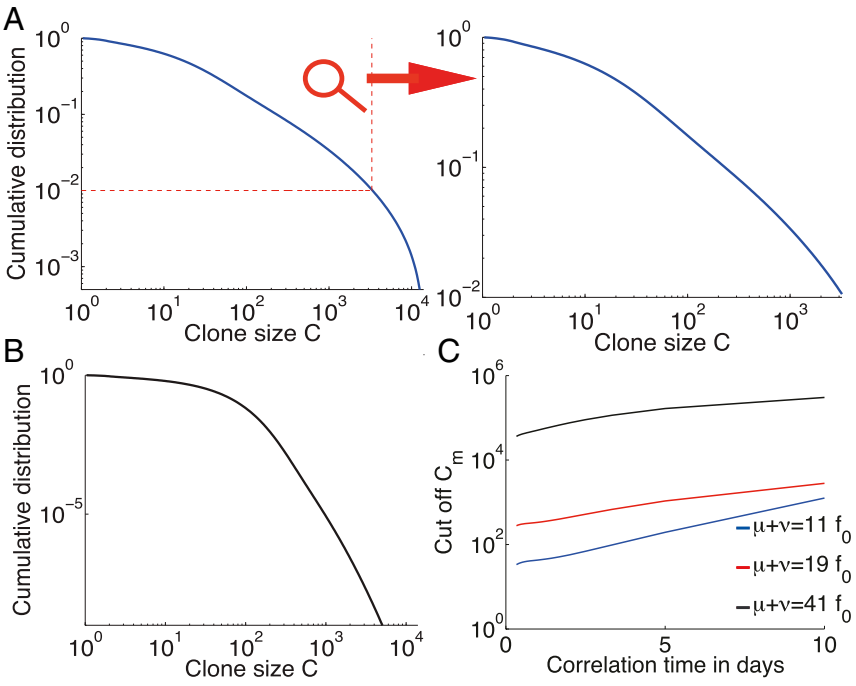

Fig. 3. Clone-size distributions for populations with a cell-specific fluctuating phenotypic fitness. (*A*) Cumulative distribution of clone sizes for moderate phenotypic heritability (λ_c^{-1}). The distribution is power-law–like for small clone values and drops above a cutoff around 0.01 of clone-size probability. An experiment that does not sequence the repertoire deeply enough could report a power-law behavior (see zoom). Parameters are $\nu = 0.17$ days^{-1}, $\mu = 0.3$ day^{-1}, $\lambda_c = 0.4$ days^{-1}, and $\gamma_c = 0.5$ days$^{-3/2}$. $C_0 = 2$ for all three graphs. (*B*) An example of a distribution of clone sizes from a cell-specific model with very low environmental noise, close to the pure birth–death limit. The distribution is flat ($\alpha = 0$) and then drops exponentially. It does not resemble experimental data. Parameters are $\nu = 0.1$ days^{-1}, $\mu = 0.3$ days^{-1}, $\lambda_c = 2$ days^{-1}, and $\gamma_c = 5$ days$^{-3/2}$. (*C*) Value of the cumulative distribution at the exponential cutoff as a function of the speed of environment variations λ_c, for different birth–death noise levels. Parameters are $f_0 = -0.998$ days^{-1} and $f_0\lambda_c^2/\gamma_c^2 = 0.998$.

$C_m = (\mu - \nu)^{-1}[(\mu + d\nu)/2 + \gamma_c^2/\lambda_c^2]$. The numerical solution of Fig. 3*B* is close to this limit. Note that even with a negligible exponential cutoff, the predicted $\alpha < 0$ contradicts experimental observations.

Discussion

The model introduced in this paper describes the stochastic nature of the immune dynamics with a minimal number of parameters, helping interpret the different regimes. These parameters are effective in the sense that they integrate different levels of signaling, pathways, and mechanisms, focusing on the long timescales of clone dynamics. We assumed that they are general enough that different cell types (B- and T cells) or subsets (naive or memory) can be described by the same dynamical equations despite their differences. How do refined models including these differences affect our results?

Naive and memory cells differ in their turnover rate, i.e., their death rate, memory cells being renewed at a pace 10 times faster than naive ones (34). In our model, this difference is reflected in a higher birth–death noise for memory cells. We have shown that this noise had no effect on the tail of the clone-size distribution for clone-specific fitness (*SI Appendix*, Fig. S5), whereas it was important for the case of a cell-specific fitness, where birth–death noise contributed to the distribution to the same extent as fitness fluctuations. However, some repertoire datasets mix both naive and memory sets, and one could wonder whether our results hold for such mixtures. To examine this question, we simulated a simple two-compartment model where naive cells get irreversibly converted into memory cells when their stimulation is above a certain threshold (see *SI Appendix*, section K for details). We found that when fitness was clone-specific, the clone-size distribution of the mixture and that of memory cells alone still follow a power law, whereas that of naive cells only does so when conversion to memory upon stimulation is partial (*SI Appendix*, Fig. S12). Repeating the same analysis for the cell-specific fitness model, we found that clone-size distributions for each phenotype differed according to their respective birth–death noises, with a longer tail for memory cells as expected from their higher turnover rate.

The main difference between B- and T cells ignored by our model is that BCRs accumulate hypermutations upon proliferation. We studied this effect by allowing proliferating clones to spawn new clones with slightly modified affinities to antigens (*SI Appendix*, section L). The resulting clone-size distribution still follows a power law (*SI Appendix*, Fig. S13), although with a slightly smaller exponent due to increased stochasticity.

Another simplifying assumption of our model is that the dynamics reaches a steady state. This may be challenged by the decay of the thymic output s_C with age. To estimate the importance of this effect, we simulated the model of a clone-specific fitness with an exponentially decaying source term, combined with a decreasing $|f_0|$ chosen to keep the population constant on average (*SI Appendix*, section M). The clone-size distributions at different points in time, shown in *SI Appendix*, Fig. S14, still follow a power law. Interestingly, the exponent α is predicted to decrease with age, consistent with $\alpha \propto |f_0|$.

We showed that the relevant sources of stochasticity for the shape of the clone-size distributions fall into two main categories, depending on how cell fate is affected by the environment. Either the stochastic elements of clone growth act in a clone-specific way, through their receptor (BCR or TCR), leading to power-law distributions with exponent ≥1, or in a cell-specific way, e.g., through their variable level of sensitivity to cytokines (and more generally through any phenotypic trait affecting cell fitness), leading to exponentially decaying distributions with a power-law prefactor. These two types of signals (clone-specific and cell-specific) are important for the somatic evolution of the immune system (21, 32, 33, 35–37) and our analysis shows that the shape of the clone-size distribution is informative of their relative importance to the repertoire dynamics. It provides a first theoretical setting and an initial systematic classification for modeling immune repertoire dynamics. Our method applied to high-throughput sequencing data can be used to quantify how much each type of signal contributes to the overall dynamics, and what is the driving force for the different cell subsets. For example, although it is reasonable to speculate that clone-specific signals should dominate for memory cells (through antigen recognition), and cell-specific selection for naive cells (through cytokine-mediated homeostatic division), the relative importance of these signals for both cell types is yet to be precisely quantified, and may vary across species. A clear power law over several decades would strongly hint at dynamics dominated by interactions with antigens, whereas a faster decaying distribution would favor a scenario where individual cell fitness fluctuations dominate. Applying these methods to data from memory cells can give orders of magnitude for the

division and half-life of memory lymphocytes, as well as the typical number of cells C_0 from a clone that are stored as memory following an infection.

The application of our method to data from the first immune repertoire survey [BCRs in zebrafish (2)] suggests that clone-specific noise dominates in that case, allowing us to infer a relation between the dynamical parameters of the model from the observed power-law exponent ≈ 2. However, there are a few issues with applying our method directly to data in the current state of the experiments. First, the counts (i.e., how many cells have the same receptor sequence and belong to the same clone) from many high-throughput repertoire sequencing experiments are imperfect because of PCR bias and sampling problems. New methods using single-molecule barcoding have been developed for RNA sequencing (8, 38, 39), but they do not solve the problem entirely, as the number of expressed mRNA molecules may not faithfully represent the cell numbers because of possible expression bias. In addition, most studies (with the exception of ref. 40) have been sequencing only one of the two chains of lymphocyte receptors, which is insufficient to determine clone identity unambiguously. As methods improve, however, our model can be applied to future data to distinguish different sources of fitness stochasticity and to put reliable constraints on biological parameters. Studying clone-size distributions in healthy individuals allows us to characterize signatures of normally functioning immune systems. By comparing them to the same properties in individuals suffering from immune diseases or cancer, our approach could be used to identify sources of anomalies.

Thanks to its generality, our model is also relevant beyond its immunological context, and follows previous attempts to explain power laws in other fields (41–43). The dynamics described here corresponds to a generalization of the neutral model of population genetics (44) where thymic or bone marrow outputs are now reinterpreted as new mutations or speciations, and where we have added a genotypic or phenotypic fitness noise (receptor or cell-specific noise, respectively). It was recently shown that such genotypic fitness noise strongly affects the fixation probability and time in a population of two alleles (45, 46). Note that, because new thymic or bone marrow clones are unrelated to existing clones, there are no lineage histories, in contrast with previous theoretical work on evolving populations in fluctuating fitness landscapes (47–49). Our main result (Eq. **6**) shows how fitness noise can cause the clone-size distribution (called "frequency spectrum" in the context of population genetics) to follow a power law with an arbitrary exponent >1 in a population of fixed size, whereas the classical neutral model gives a power law of exponent 1 with an exponential cutoff (as shown in our exact solution with $\gamma_c = 0$). Our results can be used to explain complex allele frequency spectra using fluctuating fitness landscapes.

ACKNOWLEDGMENTS. This work was supported in part by Grant ERCStG 306312.

1. Janeway C (2005) *Immunobiology* (Garland Science, New York).
2. Weinstein JA, Jiang N, White RA, 3rd, Fisher DS, Quake SR (2009) High-throughput sequencing of the zebrafish antibody repertoire. *Science* 324(5928):807–810.
3. Ndifon W, et al. (2012) Chromatin conformation governs T-cell receptor $J\beta$ gene segment usage. *Proc Natl Acad Sci USA* 109(39):15865–15870.
4. Thomas N, et al. (2014) Tracking global changes induced in the CD4 T cell receptor repertoire by immunization with a complex antigen using short stretches of CDR3 protein sequence. *Bioinformatics* 30(22):3181–3188.
5. Larimore K, McCormick MW, Robins HS, Greenberg PD (2012) Shaping of human germline IgH repertoires revealed by deep sequencing. *J Immunol* 189(6):3221–3230.
6. Sherwood AM, et al. (2011) Deep sequencing of the human TCR and TCR repertoires suggests that TCR rearranges after and T cell commitment. *Sci Transl Med* 3(90):90ra61.
7. Robins HS, et al. (2009) Comprehensive assessment of T-cell receptor beta-chain diversity in alphabeta T cells. *Blood* 114(19):4099–4107.
8. Zvyagin IV, et al. (2014) Distinctive properties of identical twins' TCR repertoires revealed by high-throughput sequencing. *Proc Natl Acad Sci USA* 111(16):5980–5985.
9. Warren RL, et al. (2011) Exhaustive T-cell repertoire sequencing of human peripheral blood samples reveals signatures of antigen selection and a directly measured repertoire size of at least 1 million clonotypes. *Genome Res* 21(5):790–797.
10. Mora T, Walczak AM, Bialek W, Callan CG, Jr (2010) Maximum entropy models for antibody diversity. *Proc Natl Acad Sci USA* 107(12):5405–5410.
11. Zarnitsyna VI, Evavold BD, Schoettle LN, Blattman JN, Antia R (2013) Estimating the diversity, completeness, and cross-reactivity of the T cell repertoire. *Front Immunol* 4:485.
12. Stirk ER, Lythe G, van den Berg HA, Molina-París C (2010) Stochastic competitive exclusion in the maintenance of the naïve T cell repertoire. *J Theor Biol* 265(3):396–410.
13. Stirk ER, Molina-París C, van den Berg HA (2008) Stochastic niche structure and diversity maintenance in the T cell repertoire. *J Theor Biol* 255(2):237–249.
14. de Boer RJ, Freitas AA, Perelson AS (2001) Resource competition determines selection of B cell repertoires. *J Theor Biol* 212(3):333–343.
15. Almeida ARM, et al. (2012) Quorum-sensing in CD4(+) T cell homeostasis: A hypothesis and a model. *Front Immunol* 3:125.
16. Hapuarachchi T, Lewis J, Callard RE (2013) A mechanistic model for naive CD4 T cell homeostasis in healthy adults and children. *Front Immunol* 4:366.
17. Reynolds J, Coles M, Lythe G, Molina-París C (2012) Deterministic and stochastic naïve T cell population dynamics: Symmetric and asymmetric cell division. *Dyn Syst* 27:75–103.
18. Troy AE, Shen H (2003) Cutting edge: Homeostatic proliferation of peripheral T lymphocytes is regulated by clonal competition. *J Immunol* 170:672–676.
19. Mak T, Saunders M (2006) *The Immune Response: Basic and Clinical Principles* (Elsevier/Academic, San Diego), Vol 1.
20. de Boer RJ, Perelson AS (1994) T cell repertoires and competitive exclusion. *J Theor Biol* 169(4):375–390.
21. Freitas AA, Rosado MM, Viale AC, Grandien A (1995) The role of cellular competition in B cell survival and selection of B cell repertoires. *Eur J Immunol* 25(6):1729–1738.
22. Bains I, Antia R, Callard R, Yates AJ (2009) Quantifying the development of the peripheral naive CD4+ T-cell pool in humans. *Blood* 113(22):5480–5487.
23. Murugan A, Mora T, Walczak AM, Callan CG, Jr (2012) Statistical inference of the generation probability of T-cell receptors from sequence repertoires. *Proc Natl Acad Sci USA* 109(40):16161–16166.
24. Kosmrlj A, Jha AK, Huseby ES, Kardar M, Chakraborty AK (2008) How the thymus designs antigen-specific and self-tolerant T cell receptor sequences. *Proc Natl Acad Sci USA* 105(43):16671–16676.
25. Kosmrlj A, et al. (2010) Effects of thymic selection of the T-cell repertoire on HLA class I-associated control of HIV infection. *Nature* 465(7296):350–354.
26. Murali-Krishna K, et al. (1998) Counting antigen-specific CD8 T cells: A reevaluation of bystander activation during viral infection. *Immunity* 8(2):177–187.
27. Kaech SM, Wherry EJ, Ahmed R (2002) Effector and memory T-cell differentiation: implications for vaccine development. *Nat Rev Immunol* 2(4):251–262.
28. Pressé S, Ghosh K, Lee J, Dill KA (2013) Principles of maximum entropy and maximum caliber in statistical physics. *Rev Mod Phys* 85:1115–1141.
29. Cavagna A, et al. (2014) Dynamical maximum entropy approach to flocking. *Phys Rev E Stat Nonlin Soft Matter Phys* 89(4):042707.
30. Clauset A, Shalizi CR, Newman MJ (2009) Power-law distributions in empirical data. *SIAM Rev* 51:661–703.
31. Bolkhovskaya OV, Zorin DY, Ivanchenko MV (2014) Assessing T cell clonal size distribution: A non-parametric approach. *PLoS One* 9(9):e108658.
32. Schluns KS, Kieper WC, Jameson SC, Lefrançois L (2000) Interleukin-7 mediates the homeostasis of naïve and memory CD8 T cells in vivo. *Nat Immunol* 1(5):426–432.
33. Tan JT, et al. (2001) IL-7 is critical for homeostatic proliferation and survival of naive T cells. *Proc Natl Acad Sci USA* 98(15):8732–8737.
34. de Boer RJ, Perelson AS (2013) Quantifying T lymphocyte turnover. *J Theor Biol* 327:45–87.
35. Seddon B, Zamoyska R (2002) TCR signals mediated by Src family kinases are essential for the survival of naive T cells. *J Immunol* 169(6):2997–3005.
36. Tanchot C, Lemonnier FA, Pérarnau B, Freitas AA, Rocha B (1997) Differential requirements for survival and proliferation of CD8 naïve or memory T cells. *Science* 276(5321):2057–2062.
37. Nesić D, Vukmanović S (1998) MHC class I is required for peripheral accumulation of CD8+ thymic emigrants. *J Immunol* 160(8):3705–3712.
38. Best K, Oakes T, Heather JM, Taylor JS, Chain B (2014) Sequence and primer independent stochastic heterogeneity in PCR amplification efficiency revealed by single molecule barcoding. bioRxiv. Available at dx.doi.org/10.1101/011411.
39. Vollmers C, Sit RV, Weinstein JA, Dekker CL, Quake SR (2013) Genetic measurement of memory B-cell recall using antibody repertoire sequencing. *Proc Natl Acad Sci USA* 110(33):13463–13468.
40. DeKosky BJ, et al. (2015) In-depth determination and analysis of the human paired heavy- and light-chain antibody repertoire. *Nat Med* 21(1):86–91.
41. Sornette D, Cont R (1997) Convergent multiplicative processes repelled from zero: Power laws and truncated power laws. *Journal de Physique I, EDP Sciences* 7(3):431–434.
42. Marsili M, Maslov S, Zhang YC (1998) Dynamical optimization theory of a diversified portfolio. *Physica A* 253:9.
43. Mitzenmacher M (2004) A brief history of generative models for power law and lognormal distributions. *Internet Math* 1:226–251.
44. Kimura M (1983) *The Neutral Theory of Molecular Evolution* (University Press, New York).
45. Cvijović I, Good BH, Jerison ER, Desai MM (2015) The *Fate* of a *Mutation* in a *F*luctuating *E*nvironment preprint.
46. Melbinger A, Vergassola M (2015) The impact of environmental fluctuations on evolutionary fitness functions. *Scientific Reports* 5:15211.
47. Leibler S, Kussell E (2010) Individual histories and selection in heterogeneous populations. *Proc Natl Acad Sci USA* 107(29):13183–13188.
48. Mustonen V, Lässig M (2010) Fitness flux and ubiquity of adaptive evolution. *Proc Natl Acad Sci USA* 107(9):4248–4253.
49. Rivoire O, Leibler S (2014) A model for the generation and transmission of variations in evolution. *Proc Natl Acad Sci USA* 111(19):E1940–E1949.

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:

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

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

$$P(n)\mu n = P(n+1)\nu(n+1) \tag{A2}$$

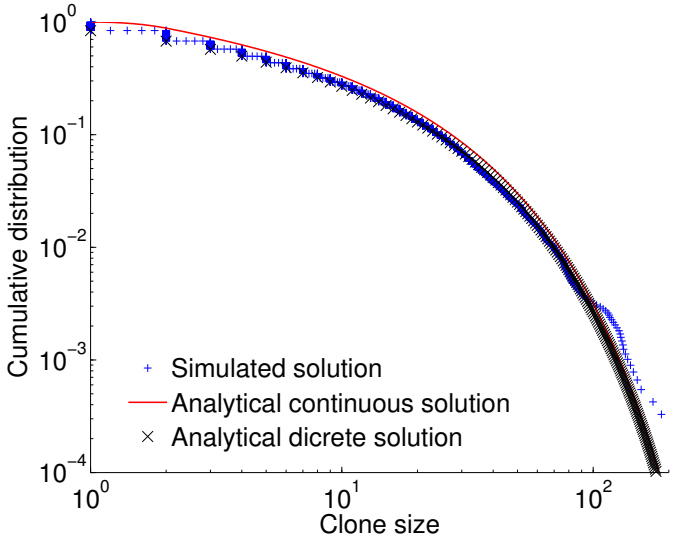

FIG. S1: We compare results from a full Gillespie simulation (blue crosses) of a system with only birth-death dynamics with analytical prediction for a discrete system (black crosses, Eq. A3) and a continuous system (red curve, Eq. A12). The prediction with discrete variables is more accurate for small 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

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

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

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

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

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

and the corresponding Fokker-Planck equation reads

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

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

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

where K is an integration constant. Defining

$$C_m = (\mu+\nu)/(2|f_0|) \tag{A8}$$

for $x < x_0$ we obtain

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

and for $x > x_0$

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

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

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

or in terms of the clone size

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

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.

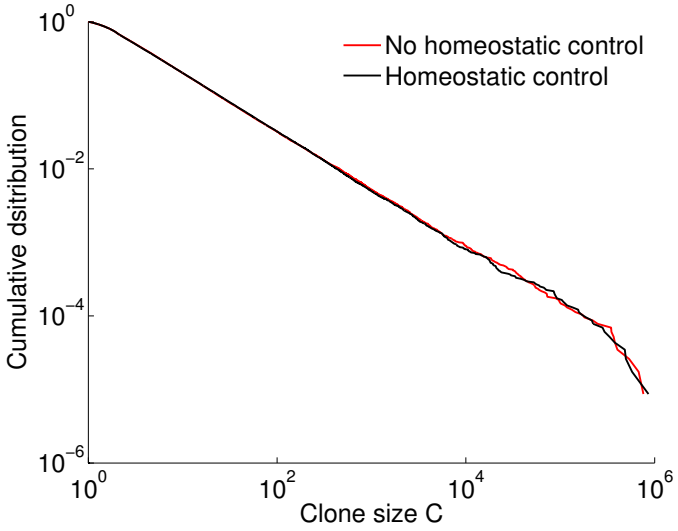

FIG. S2: Adding an explicit homeostatic control term does not affect the clone size distribution compared to tuning the degradation and death rates to obtain a given repertoire size as is done in the main text. Comparison of the clone size distribution with an explicit homeostatic control term given 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.

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 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

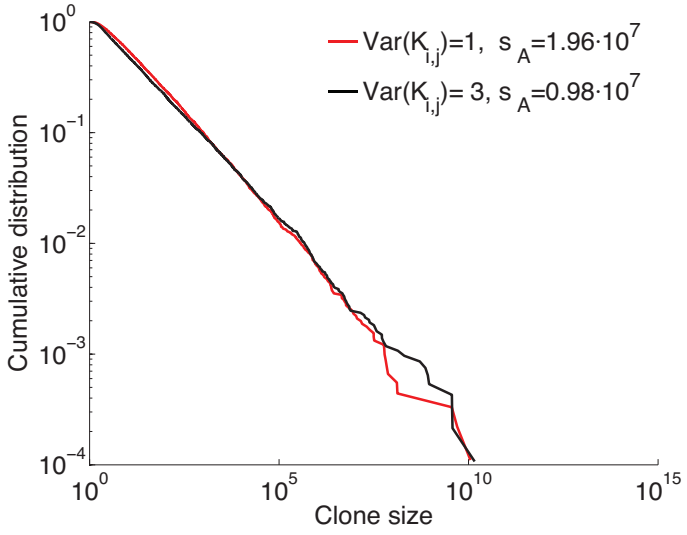

FIG. S3: Repartitioning the sources of stochasticity between the number of new antigens per time unit or the variability of binding probabilities does not influence the clone size distributions. We compare simulations of the full system dynamics defined by Eq. 1 of the main text with two sets of values $s_{\rm A}$ 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 $A^2 = s_{\rm A} p a_0^2 \langle K^2\rangle \lambda^{-1}$. The parameters were taken to be (as in Fig. 1) $s_{\rm C} = 2000$ 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 $s_{\rm A} = 1.96 \cdot 10^7$ while for the black curve the variance of the entries of K_{ij} is 3, so that $\langle K^2\rangle = 4$, and $s_{\rm A} = 0.98 \cdot 10^7$.

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). Setting $x = \log C$, 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

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

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

$$\partial_t C_i = f_0 C_i + \sqrt{2}\sigma C_i\eta_i + \sqrt{(\nu+\mu)C_i(t)}\xi_i, \quad \text{(F1)}$$

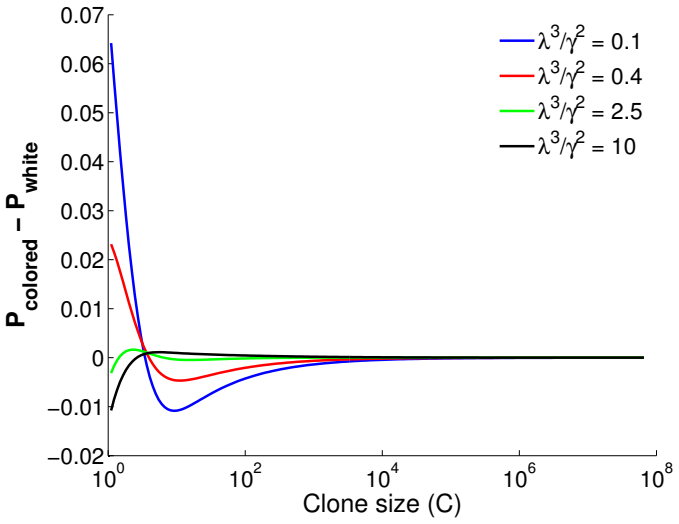

FIG. S4: Comparison between clone size distribution obtained as solutions of the time-correlated and time-uncorrelated noise models (without birth death noise). As the values of the dimensionless parameter related to the effective strength of antigen fluctuations relative to their characteristic lifetime 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.

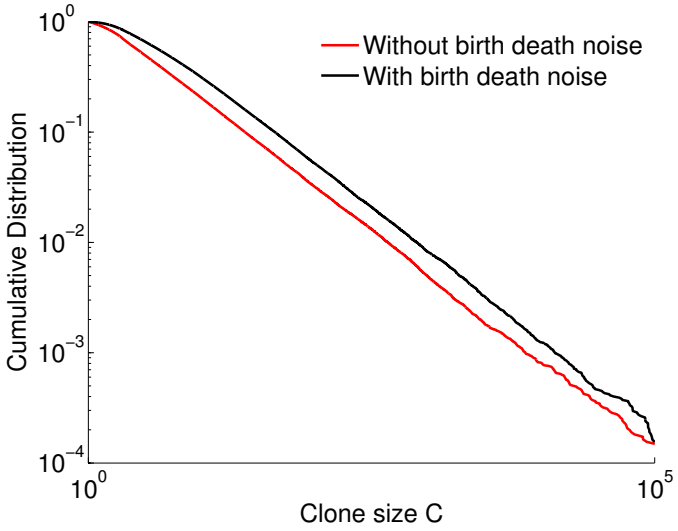

FIG. S5: We compare simulations of the Langevin dynamics with time correlated antigenic noise with birth-death noise (black line) to the same dynamics without the birth-death noise (red line). All other parameters are kept fixed.We find similar values of the power law exponents but different small clone 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 $\langle \eta_i(t)\eta_i(t')\rangle = \delta(t-t')$ follows the Stratanovich convention and ξ_i is as before. The equation for the logarithm of the clone size $x = \log C$ is

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

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

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

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

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

with the corresponding Fokker Planck equation

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

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

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

where we have defined

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

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

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

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

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

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 $x = \log C$,

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

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

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

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

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

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

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

and the solution ρ_1 to the homogeneous equation

$$f_0\rho_1 = \sigma^2\rho_1' + e^{-x}\frac{\mu+\nu}{2}\rho_1' \tag{F14}$$

of solution:

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

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

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

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

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

or in terms of the clone size

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

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

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

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:

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

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.

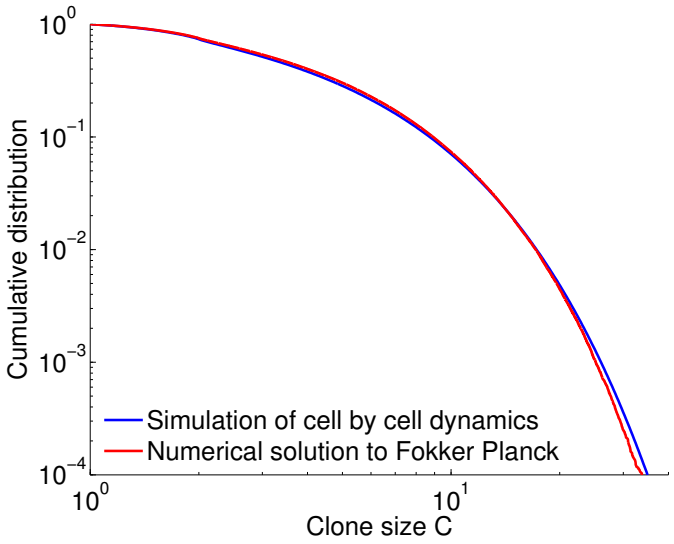

FIG. S6: Comparison of the Fokker-Planck solution (red line) and explicit Gillespie simulations of the dynamics (blue line) for the cell specific fitness model discussed in the "A model of fluctuating phenotypic fitness" Results section of the main text, show good agreement allowing us to use the population 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$^{-3/2}$.

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 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:

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

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

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

where f_i is the average fitness in clone i

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

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

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

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

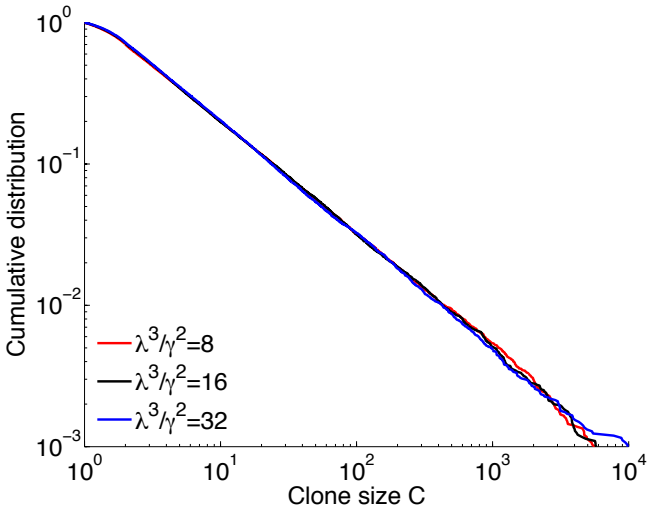

FIG. S7: Varying the dimensionless parameter related to the effective strength of antigen fluctuations relative to their characteristic lifetime λ^3/γ^2 does not affect the exponent of the power law if the ratio between exponential decay λ and standard deviation of the variation γ is kept constant. For 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.

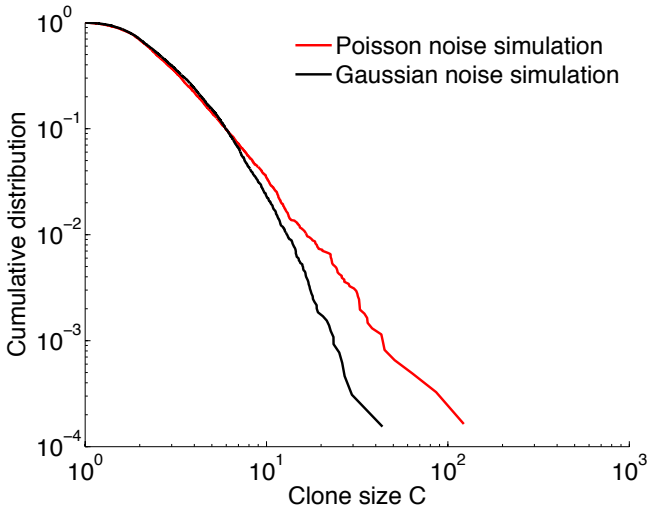

FIG. S8: Large deviations can influence the effect of Poisson noise on the simulated clone size distributions and create a discrepancy between Poisson noise (red line) and the Gaussian approximations (black line) we assume in the main text. The discrepancy is most apparent for small clones. We simulated 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.

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 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:

$$\frac{dC_i}{dt}=f_0C_i+\sqrt{2C_i}\frac{\gamma_c}{\lambda_c}\eta_i+\frac{\gamma_c^2}{\lambda_c^2}+\sqrt{(\mu+\nu)C_i}\xi_i \tag{J1}$$

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:

$$\begin{cases}\partial_t C_i=f_0C_i+fC_i,\\ \partial_t f_i=-\lambda_c f_i+\sqrt{\frac{2}{C_i}}\gamma_c\eta_i,\end{cases} \tag{J2}$$

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

$$\begin{aligned}\rho(x)=e^{-e^x/C_m+x}\Big[KEi(e^x/C_m)-KEi(C_m^{-1}) \qquad (J7)\\ -\frac{s_C\lambda_c^2}{\gamma_c^2}Ei(\frac{e^x}{C_m})+\frac{s_C\lambda_c^2}{\gamma_c^2}Ei(\frac{e^{x_0}}{C_m})\Big], \qquad (J8)\end{aligned}$$

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

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

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.

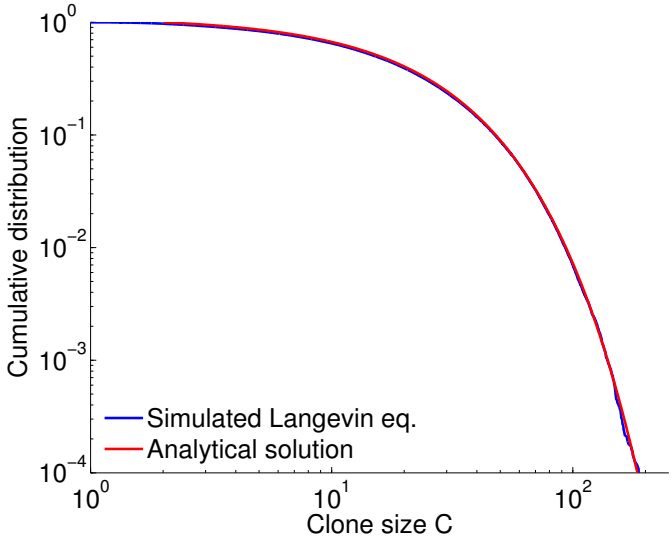

FIG. S9: The result of a simulation of the Langevin equation of the white noise cell-specific fitness model (blue line) compared to the analytical prediction of Eq. J11 (red line) show very good agreement. The parameters are $\nu = 0.2$ day^{-1}, $\mu = 0.4$ day^{-1}, $C_0 = 2$, $\lambda_c = 4$ day^{-1} and $\gamma_c = 8$ day$^{-3/2}$.

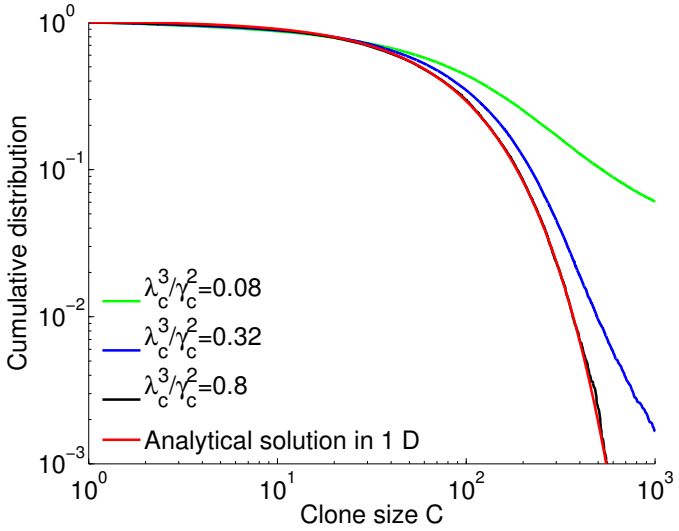

FIG. S10: Convergence of the cell-specific fitness models (Eq. I6) without birth-death noise to Eq. J11 in the limit of no heritability ($\lambda_c \to \infty$). For all four curves $\alpha = 0.2$. Parameters used: $\nu = 0.2$ day^{-1}, $\mu = 0.25$ day^{-1}, $C_0 = 2$ and 1000 new clones introduced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

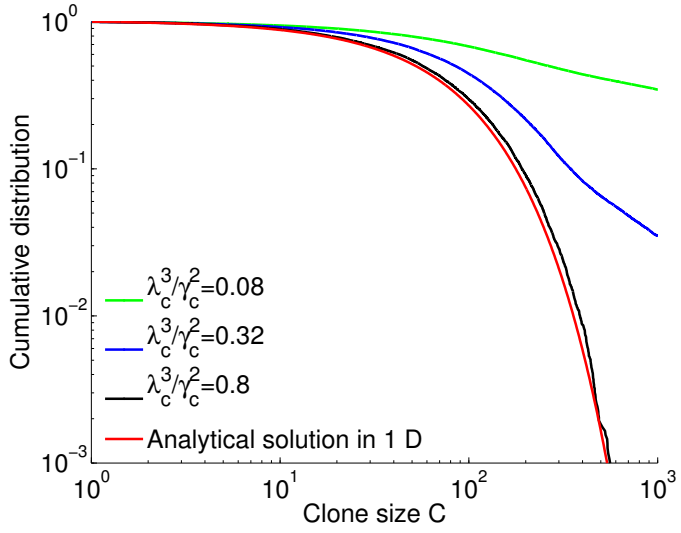

FIG. S11: Convergence of the cell-specific models (Eq. I6) with birth-death noise to the analytical result of Eq. J15 (red line). Keeping constant α while $\lambda_c \to \infty$ and $\gamma_c \to \infty$ we recover the solution of Eq. J15. Parameters are the same as in Fig. S10

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 $0 \leq 1+\alpha \leq 1$ 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

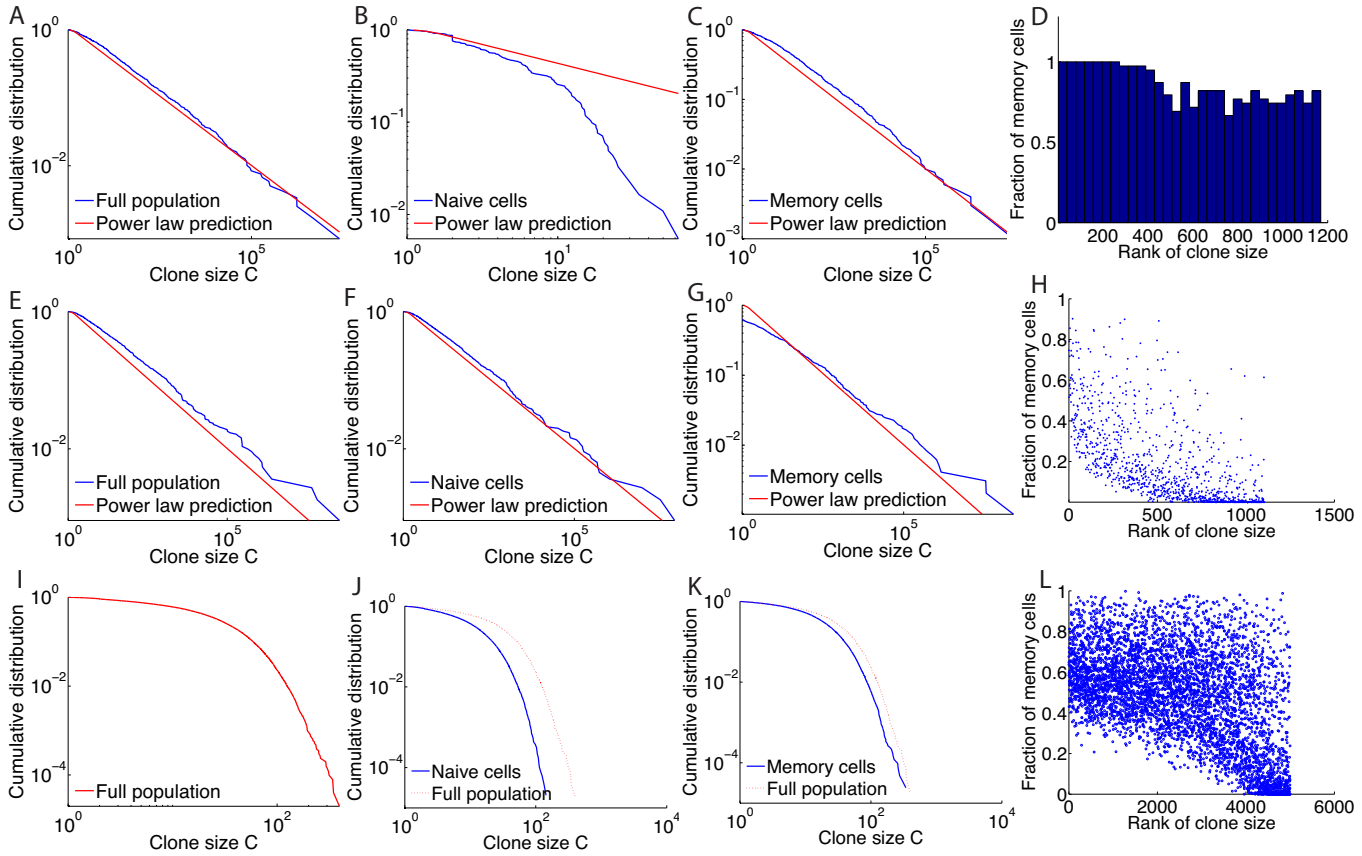

FIG. S12: Simulation results for clone and cell specific model with two cell compartments for naive and memory. Panels A to D are results from clone-specific fitness model with a switching rate θ from naive to memory taken to be infinite (the whole clone switches instantly to memory when above a fitness threshold) and fitness threshold $f_{\rm mem} = 1$ 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 $f_{\rm mem} = 1$ day^{-1}. For both clone-specific simulations the parameters are: $s_{\rm C} = 200$ day^{-1} , $C_0 = 2$, $s_{\rm A} = 1.96 \cdot 10^7$ 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 $f_{\rm mem} = 0.5$. The other parameters are: $s_{\rm C} = 10^4$ 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 $f_{\rm mem}$ 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).

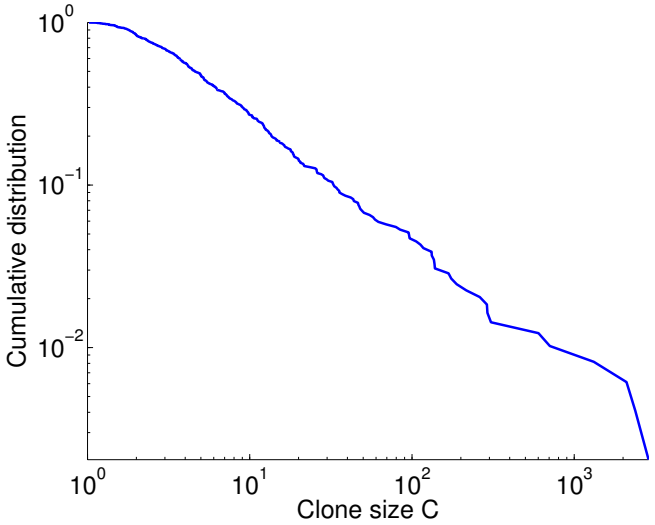

FIG. S13: We show the clone size distribution that results from simulating a model of clone-specific fitness with somatic hypermuations as described in Appendix L and Eq. L1. The distribution exhibits clear power law behavior. Hypermutation parameters are: $f_{\rm hyp} = 4$ days^{-1}, $r_{\rm hyp} = 0.01$ days^{-1}, $r_{\rm del} = 0.01$, $p_{\rm hyp} = 0.5$, $\psi = 0.7$ and $\sigma_{hyp} = 0.05$. Other parameters are: $s_{\rm C} = 200$ day^{-1} , $C_0 = 2$, $s_{\rm A} = 1.5 \cdot 10^7$ day^{-1}, $\langle a_{j,0} \rangle = 1$, 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.

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 $f_i > f_{\rm hyp}$, offspring with hypermutated receptors are being produced with rate $r_{\rm hyp}$. A large fraction $r_{\rm del}$ of those are assumed to have acquired deleterious mutations and are removed from the pool. The rest (fraction $1-r_{\rm del}$) 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:

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

where ψ is a parameter controling the heritability of the values of the K entries, and $p_{\rm hyp}$ 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 binding affinities with probability $(1 - p_{\rm hyp})p$ (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-

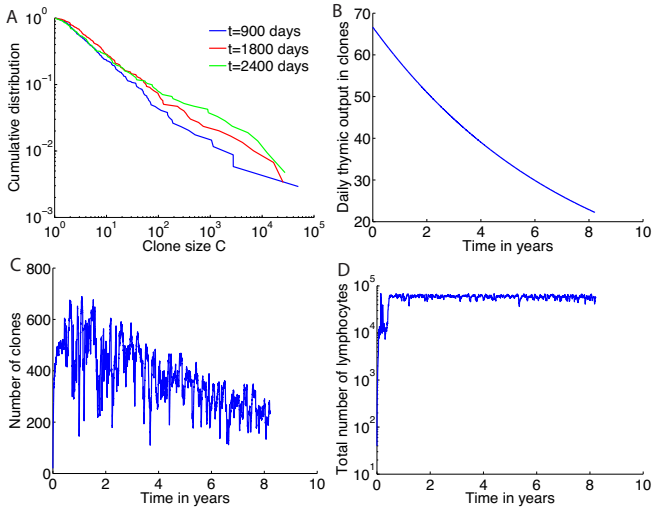

FIG. S14: Results of a simulation of a model of clone-specific fitness with a decaying source term and balancing decrease of $|f_0|$ to keep the population size constant. A. The clone size distributions at different time points maintains a power law behavior with an exponent α that decreases with time. 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}, $\langle a_{j,0}\rangle = 1$, $\mathrm{Var}(a_{j,0}) = 1$, $\lambda_c = 2$ 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

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

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 :

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

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.