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

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. 1C). 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

---

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 with novel, unique receptors. Division, death, and thymic 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. 2A 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. 2A, *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. 2A, 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. 2A 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. 2A).

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. 1A (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. 1A with our simplified model using λ as an adjustable parameter, we obtain $\lambda \approx 0.14$ day^{-1} (*SI Appendix*, section G),
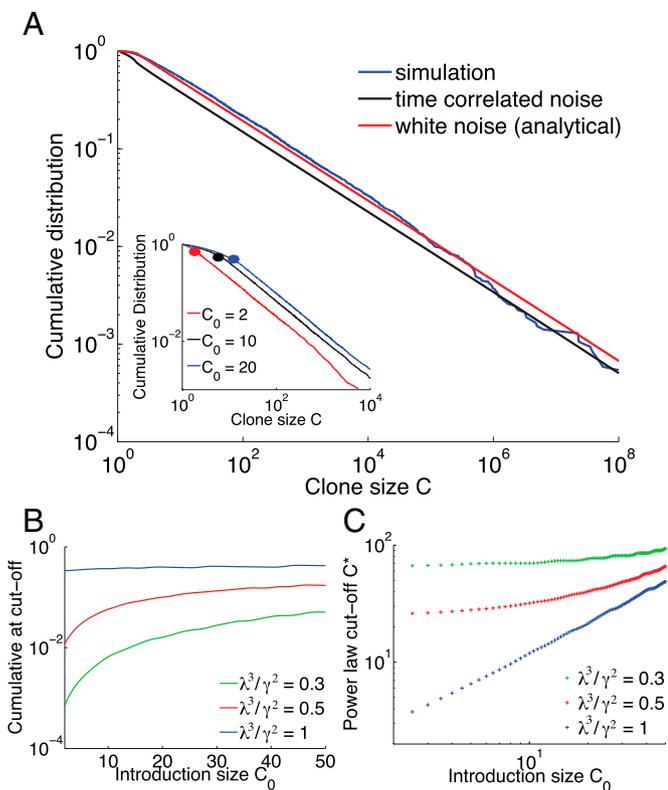
Fig. 2. Clone-size distributions for populations with fluctuating antigenic, clone-specific fitness. (*A*) Comparison of simulations and simplified models of clone dynamics. Blue curve: cumulative distribution of clone sizes obtained from the simulation of Eq. **1**. Black curve: a simplified, numerically 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
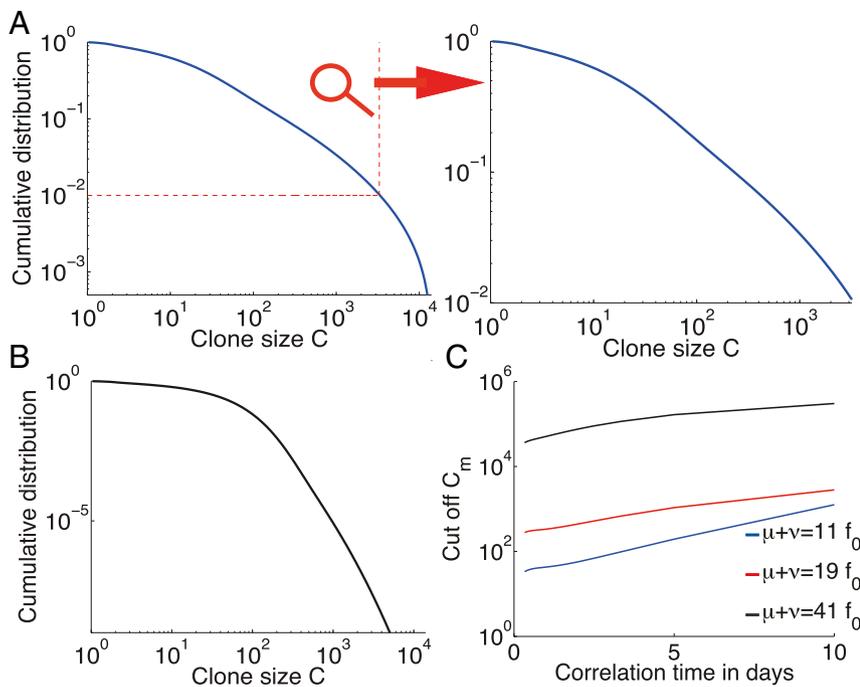
Fig. 3. Clone-size distributions for populations with a cell-specific fluctuating phenotypic fitness. (*A*) Cumulative distribution of clone sizes for moderate phenotypic heritability (λ_c^{-1}). The distribution is power-law–like for small clone values and drops above a cutoff around 0.01 of clone-size 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

Desponds et al.

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.

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 Jβ 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 TCRα and TCRβ repertoires suggests that TCRβ 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 *Fluctuating Environment* 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.

PHYSICS

EVOLUTION