



Stochastic descriptors to study the fate and potential of naive T cell clonotypes in the periphery

J. R. Artalejo¹ · A. Gómez-Corral^{1,2} ·
M. López-García³ · C. Molina-París³

Received: 31 March 2015 / Revised: 20 April 2016 / Published online: 27 June 2016
© The Author(s) 2016. This article is published with open access at Springerlink.com

Abstract The population of naive T cells in the periphery is best described by determining both its T cell receptor diversity, or number of clonotypes, and the sizes of its clonal subsets. In this paper, we make use of a previously introduced mathematical model of naive T cell homeostasis, to study the fate and potential of naive T cell clonotypes in the periphery. This is achieved by the introduction of several new stochastic descriptors for a given naive T cell clonotype, such as its maximum clonal size, the time to reach this maximum, the number of proliferation events required to reach this maximum, the rate of contraction of the clonotype during its way to extinction, as well as the time to a given number of proliferation events. Our results show that two fates can be identified for the dynamics of the clonotype: extinction in the short-term if the clonotype experiences too hostile a peripheral environment, or establishment in the periphery in the long-term. In this second case the probability mass function for the maximum clonal size is bimodal, with one mode near one and the other mode far away from it. Our model also indicates that the fate of a recent thymic emigrant (RTE) during its journey in the periphery has a clear stochastic component, where the probability of extinction cannot be neglected, even in a friendly but competitive environment. On the other hand, a greater deterministic behaviour can be expected in the potential size of the clonotype seeded by the RTE in the long-term, once it escapes extinction.

✉ C. Molina-París
carmen@maths.leeds.ac.uk

¹ Department of Statistics and Operations Research, Faculty of Mathematics, Complutense University of Madrid, 28040 Madrid, Spain

² Instituto de Ciencias Matemáticas, CSIC-UAM-UC3M-UCM, Calle Nicolás Cabrera 13-15, Campus de Cantoblanco UAM, 28049 Madrid, Spain

³ Department of Applied Mathematics, School of Mathematics, University of Leeds, Leeds LS2 9JT, UK

Keywords T cell homeostasis · TCR diversity · Stochastic univariate birth and death process · Extinction · Stochastic descriptor · Competition · Cross-reactivity

Mathematics Subject Classification 60J28 Applications of continuous-time Markov processes on discrete state spaces · 92B05 General biology and biomathematics

1 Introduction

T cells are the set of lymphocytes characterised by the expression of a specialised receptor, called the T cell receptor (TCR). T cells can thus, be classified in “families” (or clonotypes) according to the molecular structure of the TCR they display on their membrane (Murphy et al. 2008). On average a T cell expresses 3×10^4 identical copies of a given TCR molecule. The number of T cells in the adaptive immune system of an adult human tends to a stationary “homeostatic” distribution (McLean et al. 1997; Freitas and Rocha 2000), specified by its size (the total number of T cells) and TCR diversity (the number of different clonotypes or TCR molecular structures). The size and diversity of the naive T cell pool (those T cells that have not taken part in a previous immune response) is essential to recognise a broad range of pathogens (Freitas and Rocha 1999, 2000; Tanchot and Rocha 1998). An adult mouse has a homeostatic population of 10^8 naive T cells (Mason 1998) and a TCR diversity estimated around 2×10^6 (Zarnitsyna et al. 2013), and a healthy adult human has a homeostatic population of 10^{11} naive T cells (Hazenberg et al. 2003) distributed in 2.5×10^7 different TCR specificities (Zarnitsyna et al. 2013; Johnson et al. 2014).

Naive T cells originate from T cell precursors (or thymocytes) that survive positive and negative selection in the thymus (Stritesky et al. 2012; Moran and Hogquist 2012; Bird 2009). Thus, naive T cells have not been activated by exposure to pathogen-derived peptide fragments or antigens. During their lifetime, naive T cells constantly circulate in the blood to visit lymph nodes, which we refer to as the “periphery” (Takada and Jameson 2009). The population of naive T cells, which have never taken part in an immune response, is maintained by continuous, yet slow, division in the periphery (or homeostatic proliferation) (Tanchot and Rocha 1998; Troy and Shen 2003; Takada and Jameson 2009). Antigen presenting cells provide the stimulatory signals that induce naive T cell division. The signals are delivered via the interaction between TCRs (on the surface of T cells) and the ligands (also referred to as self-peptides or self-antigens) expressed on the surface of antigen presenting cells. Although the number of naive T cells is approximately constant for much of an individual’s life, as one ages both the size and the TCR diversity of the naive pool decline (Goronzy et al. 2007). This loss of T cell receptor diversity is also observed for B cells (Gibson et al. 2009), and is associated with increased susceptibility to infection and poor health. In the long-term, during the ageing process, it is important to study both the probability of extinction and the mean time to extinction for different T cell clonotypes. This clearly requires a stochastic framework. Some of the homeostatic mechanisms which ensure that a maximum of “antigen space coverage” is achieved with the available numerical diversity of T cell clonotypes (Correia-Neves et al. 2001; Mahajan et al. 2005), have been analysed mathematically before by means of deterministic (De Boer and Perelson

1997) and stochastic (Stirk et al. 2008) models. Yet, the question of how the naive T cell population remains diverse is still not completely understood (De Boer et al. 2012).

During the homeostatic period the organisation of the peripheral naive T cell pool requires that an existing T cell must die if a new T cell is generated in the thymus or by peripheral cell division (Tanchot and Rocha 1998). It is then reasonable to pose a second challenge: to quantify the probability that a given TCR specificity, namely, a single recent thymic emigrant (RTE), can be established in the periphery, while competing with the pre-existing population of naive T cells. Recent thymic emigrants are those naive T cells that have just migrated out of the thymus, after surviving positive and negative selection, to become part of the peripheral T cell population (Fink 2013). Given that double positive thymocytes hardly divide in the thymic cortex and that single positive thymocytes divide at most once in the medulla of the thymus (Sinclair et al. 2013; Sritesky et al. 2013; Sawicka et al. 2014; Yates 2014), one expects a given TCR specificity to be introduced in the periphery by a single recent thymic emigrant. There is evidence to support the fact that RTEs do not have an intrinsic lifespan (Tanchot and Rocha 1998). Given the requirement of naive T cells for continuous TCR ligation to survive in the periphery, it is natural to hypothesise that the TCR cross-reactivity of a given RTE (how many self-peptides can provide a homeostatic stimulus to the TCR under consideration) (Sewell 2012), the number of different clonotypes (competitor clonotypes) that can bind those self-peptides, and the clonal size of each competitor, are key parameters that will determine the fate and potential in the periphery of a recent thymic emigrant.

The aim of this paper is to make use of a previously introduced mathematical model of naive T cell homeostasis (Stirk et al. 2008, 2010), and define new stochastic descriptors that allow us to provide answers to the previous questions. In Refs. Stirk et al. (2008, 2010), a continuous time birth and death Markov model that describes the time evolution of a peripheral naive T cell clonotype was introduced. In this reference, it was shown that extinction takes place with certainty for all parameter values of the model, and the average time to extinction was computed. However, the question of how the affinity and the cross-reactivity of a given T cell clonotype (or RTE) determines (1) its ability to become part of the peripheral naive T cell repertoire, (2) its maximum size if the clone is to get established in the naive peripheral pool, and (3) the timescale to get established in the periphery at its maximum size, was not considered. We study these questions by defining novel stochastic descriptors for a given TCR clonotype or RTE.

The paper is organised as follows: in Sect. 2 we introduce the stochastic birth and death process and define a number of stochastic descriptors (continuous and discrete) to characterise the maximum clonal size (Sect. 2.1) and the time to reach that size, the number of divisions to reach the maximum clonal size (Sect. 2.2), and the time to get to a given number of divisions (Sect. 2.3). Section 3 provides numerical results for the previously introduced descriptors, exploring the three parameter regimes described in Sect. 2. We conclude with a discussion of our results both from a mathematical and immunological perspective in Sect. 4.

2 Mathematical model

We make use of the mathematical model introduced in [Stirk et al. \(2008\)](#) to describe the time evolution of the number of cells in a given T cell clonotype, that is, the number of naive T cells that express identical T cell receptor molecules. The main assumption of the model is that T cells of a given clonotype compete for signals delivered by molecules (self-peptides also referred to as pMHC) expressed on antigen presenting cells (APCs). These signals induce one round of cell division, that is, a birth event. All cells of the given clonotype can also die and this is a death event in the model.

The underlying mathematical model is a birth and death process $\{X(t) : t \geq 0\}$ on the state space $\mathcal{X} = \mathbb{N} \cup \{0\}$, where the random variable $X(t)$ describes the number of cells of a given T cell clonotype as a function of time t . Its birth and death rates ([Fig. 1](#)) are specified for $i \in \mathbb{N} \cup \{0\}$ by [Stirk et al. \(2008\)](#):

$$\begin{aligned} \lambda_i &= \varphi e^{-\nu} \sum_{r=0}^{\infty} \frac{\nu^r}{r!} \frac{i}{r\langle n \rangle + i}, \\ \mu_i &= \mu i. \end{aligned} \tag{1}$$

Parameters $\nu \geq 0$, $\mu > 0$, $\varphi > 0$, and $\langle n \rangle \geq 1$ were introduced in [Stirk et al. \(2008\)](#), and they have the following meaning:

- μ is the per (naive) cell death rate for cells of the clonotype under consideration,
- φ is the per (naive) cell rate of homeostatic proliferation due to signals from the self-peptides that can bind to the TCR of the clonotype under consideration,
- ν is the number of other clonotypes (different from the one under consideration) that can compete for the same homeostatic proliferation signals, and
- $\langle n \rangle$ is the characteristic size of T cell clonotypes that compete for homeostatic proliferation signals with the clonotype of interest. We note that these competing clonotypes are not explicitly modelled and by assuming that they all have a characteristic number of naive T cells given by $\langle n \rangle$, the time evolution of the clonotype of interest can be reduced to a univariate Markov process (details about this approximation can be found in [Stirk et al. \(2008\)](#)).

It was shown in [Stirk et al. \(2008\)](#) that, for any value of the parameters, $0 \in \mathcal{X}$ is an absorbing state of the stochastic process, the time to extinction (or absorption) from any other state $i \in \mathcal{X}$ is finite with probability one, and its mean is also finite.

As introduced above, parameter φ is a measure both of the affinity and the cross-reactivity of the T cell receptor for self-pMHC molecules ([Sewell 2012](#)), in the sense that the rate of homeostatic proliferation depends not only on how well a TCR can



Fig. 1 Birth and death process (with absorption) for the time evolution of the number of cells in a given naive T cell clonotype

bind a given pMHC complex, but how many different pMHC complexes can provide survival signals to the T cell clonotype under consideration, characterised by its TCR molecule. Parameter ν is the number of competitors of the clonotype under consideration (Stirk et al. (2008)). Three regimes can be identified in parameter space that describe different immunological scenarios: the first one is the limit of no inter-clonal competition ($\nu \ll 1$), the second is the limit of large inter-clonal competition ($\nu \gg 1$), and the third one is the intermediate regime of competition ($\nu \approx 1$).

Hard niche clonotype In the special case $\nu \ll 1$, the population under study (the number of T cells that belong to a given TCR clonotype) does not compete with any other clonotypes, and thus the previous expressions for the birth and death rates simplify to (Stirk et al. (2008))

$$\begin{aligned}\lambda_i &= \varphi, \\ \mu_i &= \mu i.\end{aligned}$$

Soft niche clonotype The opposite limit, $\nu \gg 1$, represents a highly competitive environment, where the population of T cells under study competes with a large number of different clonotypes, and thus the expressions for the birth and death rates become (Stirk et al. (2008))

$$\begin{aligned}\lambda_i &= \frac{\varphi i}{\nu (n) + i}, \\ \mu_i &= \mu i.\end{aligned}$$

Intermediate niche clonotype In the case $\nu \approx 1$, the TCR clonotype will be referred to as an intermediate niche clonotype, and its birth and death rates will be given by Eq. (1).

These three regimes, hard, soft and intermediate niche clonotypes, will be explored in Sect. 3. Under these regimes, we introduce in this Section several stochastic descriptors which will allow us to study the dynamics of the clonotype under consideration. In particular, in Sect. 2.1 our interest is in the maximum clonal size reached by the clonotype and the time to reach this maximum size, for which we obtain analytical expressions for its probability mass function and its different order moments, respectively. We analyse the number of proliferation events to reach the maximum clonal size in Sect. 2.2, and in Sect. 2.3 we focus on the time to reach a given number of proliferation (or division) events, as a measure of the proliferative capacity (or potential) of the clonotype under study. Finally, the rate of contraction of the clonotype during its way to extinction can be studied by means of the time to contraction to a given clonal size, which is also analysed in Sect. 2.1.

2.1 Maximum clonal size and time to reach it

Our interest in this section is in the maximum number of cells belonging to the clonotype under consideration and the time needed to reach this maximum. To begin with, we define the random variables

$$\begin{aligned}
 X_i^{max} &= \max\{X(t) : t \geq 0 | X(0) = i\}, \\
 T_i^{max} &= \inf\{t : X(t) = X_i^{max}\},
 \end{aligned}$$

for $i \in \mathcal{X}$, which amount to the maximum clonal size attained by the clonotype, under the assumption that the initial size equals i , and the time to reach this maximum clonal size, respectively; note that $X_0^{max} = 0$ and $T_0^{max} = 0$, since 0 is an absorbing state in \mathcal{X} . In order to study both variables in parallel, we define the time to reach a total clonal size i_ρ , given that the initial clonal size equals i , as the auxiliary random variable

$$T_{i,i_\rho} = \inf\{t : X(t) = i_\rho | X(0) = i\},$$

for $i, i_\rho \in \mathcal{X}$, with $T_{i_\rho,i_\rho} = 0$. We point out here that these auxiliary random variables have their own immunological interest, since attaining a given clonal size i_ρ might indicate that the clonotype has become large enough to mount an immune response in a timely fashion, or it has grown too large and might lead to autoimmunity, so that a tolerant immune state has been lost. The analysis of T_{i,i_ρ} also allows the study of the random variables of interest X_i^{max} and T_i^{max} .

For a particular clonal size $i_\rho \in \mathcal{X}$, the random variable T_{i,i_ρ} might be analysed in a different manner depending on whether $i_\rho < i$ (time to *contraction* to a given clonal size i_ρ), or $i_\rho > i$ (time to *expansion* to a given clonal size i_ρ). We note that in the latter case, reaching i_ρ is not certain, so that T_{i,i_ρ} is a defective random variable, while reaching i_ρ is certain in the former case, since absorption at state 0 occurs with probability one (Stirk et al. 2008).

2.1.1 Time to contraction to a given clonal size $i_\rho < i$, given the initial clonal size i

For an initial clonal size i with $i > i_\rho$, the random variable T_{i,i_ρ} amounts to the time until absorption into i_ρ for a birth and death process defined on a single absorbing state i_ρ and the class of transient states, $\{i_\rho + 1, i_\rho + 2, \dots\}$, with birth rates $\{\lambda_j : j \geq i_\rho + 1\}$ and death rates $\{\mu_j : j \geq i_\rho + 1\}$. Since absorption of the underlying process $\{X(t) : t \geq 0\}$ occurs in a finite time almost surely (see Stirk et al. (2008)), an appeal to Karlin and McGregor (1957) allows us to describe the expected values of T_{i,i_ρ} from the equality

$$E[T_{i,i_\rho}^k] = k \sum_{n=i_\rho}^{i-1} \rho_n \sum_{j=n+1}^{\infty} \frac{E[T_{j,i_\rho}^{k-1}]}{\lambda_j \rho_j}, \quad k \geq 1,$$

where $\rho_{i_\rho} = 1$ and $\rho_n = \prod_{k=i_\rho+1}^n \lambda_k^{-1} \mu_k$, for $n \geq i_\rho + 1$; note that the expected values $E[T_{i_\rho,0}^k]$ corresponding to $i_\rho = 0$ are related to the total extinction of the clonotype. We refer the reader to Artalejo et al. (2012), where an analogous descriptor is analysed for the spread of an SIS epidemic among a population consisting of N individuals. In this case, the series $\sum_{j=n+1}^{\infty}$ becomes the finite sum $\sum_{j=n+1}^N$ and T_{i,i_ρ} can be seen as a phase-type random variable (see e.g., [Kulkarni (1996), Sect. 6.7] and [Latouche and Ramaswami (1999), Chapter 2]).

2.1.2 Time to expansion to a given clonal size $i_Q > i$, given the initial clonal size i , and analysis of X_i^{max} and T_i^{max}

For a fixed (and given) clonal size $i_Q \in \mathbb{N}$, we introduce the following notation:

$$\begin{aligned}
 v_{i,i_Q} &= \mathbb{P}(T_{i,i_Q} < \infty) = \mathbb{P}(X_i^{max} \geq i_Q), \quad 0 \leq i \leq i_Q, \\
 \phi_{i,i_Q}(s) &= E \left[e^{-s T_{i,i_Q}} 1_{\{T_{i,i_Q} < \infty\}} \right], \quad 0 \leq i \leq i_Q, \quad \Re(s) \geq 0, \\
 m_{i,i_Q}^{(k)} &= E \left[(T_{i,i_Q})^k 1_{\{T_{i,i_Q} < \infty\}} \right] = (-1)^k \frac{d^k}{ds^k} \phi_{i,i_Q}(s) \Big|_{s=0}, \quad 0 \leq i \leq i_Q, \quad k \geq 0,
 \end{aligned}
 \tag{2}$$

where $1_{\{T_{i,i_Q} < \infty\}}$ is a random variable that takes the values 1 if $T_{i,i_Q} < \infty$, and 0 otherwise. The quantities defined in Eq. (2) will allow us to analyse the random variables X_i^{max} and T_i^{max} , but the distribution of T_{i,i_Q} is defective, since $T_{i,i_Q} = \infty$ if the clonotype becomes extinct before reaching the size i_Q ; note that $1 - v_{i,i_Q}$ is the probability of such an extinction event, as

$$v_{i,i_Q} = \begin{cases} 0, & \text{if } i = 0, \\ 1, & \text{if } i = i_Q, \end{cases}$$

with $v_{i,i_Q} \in (0, 1)$ in the case $1 \leq i \leq i_Q - 1$.

By a first-step argument, the restricted Laplace-Stieltjes transforms satisfy the following set of linear equations:

$$\begin{aligned}
 \phi_{0,i_Q}(s) &= 0, \\
 \phi_{i,i_Q}(s) &= \frac{\mu_i}{s + \lambda_i + \mu_i} \phi_{i-1,i_Q}(s) + \frac{\lambda_i}{s + \lambda_i + \mu_i} \phi_{i+1,i_Q}(s), \quad 1 \leq i \leq i_Q - 1, \\
 \phi_{i_Q,i_Q}(s) &= 1.
 \end{aligned}
 \tag{3}$$

These equations can be rewritten in multiplicative form as

$$\beta_i \phi_{i-1,i_Q}(s) + \gamma_i \phi_{i,i_Q}(s) + \alpha_i \phi_{i+1,i_Q}(s) = \delta_i, \quad 1 \leq i \leq i_Q - 1,$$

with

$$\begin{aligned}
 \alpha_i &= \begin{cases} -\lambda_i, & \text{if } 1 \leq i \leq i_Q - 2, \\ 0, & \text{if } i = i_Q - 1, \end{cases} \\
 \beta_i &= \begin{cases} 0, & \text{if } i = 1, \\ -\mu_i, & \text{if } 2 \leq i \leq i_Q - 1, \end{cases} \\
 \gamma_i &= s + \lambda_i + \mu_i, \quad 1 \leq i \leq i_Q - 1, \\
 \delta_i &= \begin{cases} 0, & \text{if } 1 \leq i \leq i_Q - 2, \\ \lambda_{i_Q-1}, & \text{if } i = i_Q - 1. \end{cases}
 \end{aligned}$$

Then, by using forward elimination, we may obtain

$$G_i \phi_{i,i_\ell}(s) + \alpha_i \phi_{i+1,i_\ell}(s) = D_i, \quad 1 \leq i \leq i_\ell - 1,$$

with

$$G_i = \begin{cases} \gamma_1, & \text{if } i = 1, \\ \gamma_i - \frac{\beta_i \alpha_{i-1}}{G_{i-1}}, & \text{if } 2 \leq i \leq i_\ell - 1, \end{cases}$$

$$D_i = \begin{cases} 0, & \text{if } 1 \leq i \leq i_\ell - 2, \\ \lambda_{i_\ell-1}, & \text{if } i = i_\ell - 1. \end{cases}$$

In terms of the functions $g_i(s) = G_i - (s + \lambda_i)$, we may rewrite $\phi_{i,i_\ell}(s)$ as

$$\phi_{i,i_\ell}(s) = \frac{D_i - \alpha_i \phi_{i+1,i_\ell}(s)}{s + \lambda_i + g_i(s)},$$

since $g_i(s)$ satisfies

$$g_i(s) = \mu_i \frac{s + g_{i-1}(s)}{s + \lambda_{i-1} + g_{i-1}(s)}, \quad 2 \leq i \leq i_\ell - 1,$$

with $g_1(s) = \mu_1$. This implies that

$$\phi_{i,i_\ell}(s) = \prod_{k=i}^{i_\ell-1} \frac{\lambda_k}{s + g_k(s) + \lambda_k}, \quad 1 \leq i \leq i_\ell - 1, \tag{4}$$

from which it follows that $v_{i,i_\ell} \in (0, 1)$ for clonal sizes $1 \leq i \leq i_\ell - 1$, since $v_{i,i_\ell} = \phi_{i,i_\ell}(0)$.

In evaluating the *restricted* moments $m_{i,i_\ell}^{(k)}$, we first derive the probabilities v_{i,i_ℓ} as the values $\phi_{i,i_\ell}(0)$ from Eq. (4). In particular, it is seen that $v_{0,i_\ell} = 0$ and we can write

$$v_{i,i_\ell} = \left(\sum_{m=0}^{i_\ell-1} \zeta_m \right)^{-1} \sum_{k=0}^{i-1} \zeta_k, \quad 1 \leq i \leq i_\ell,$$

where $\zeta_0 = 1$ and $\zeta_i = \prod_{k=1}^i \lambda_k^{-1} \mu_k$, for $1 \leq i \leq i_\ell - 1$. Then, we have

$$m_{i,i_\ell}^{(0)} = v_{i,i_\ell}, \quad 0 \leq i \leq i_\ell,$$

$$m_{0,i_\ell}^{(k)} = 0, \quad k \geq 1,$$

$$m_{i_\ell,i_\ell}^{(k)} = 0, \quad k \geq 1.$$

Moments $m_{i,i_\varrho}^{(k)}$, for clonal sizes $1 \leq i \leq i_\varrho - 1$, are derived from Eq. (3). In fact, by taking derivatives in Eq. (3) we obtain

$$(\lambda_i + \mu_i) m_{i,i_\varrho}^{(k)} = \mu_i m_{i-1,i_\varrho}^{(k)} + \lambda_i m_{i+1,i_\varrho}^{(k)} + k m_{i,i_\varrho}^{(k-1)}, \quad 1 \leq i \leq i_\varrho - 1, \quad k \geq 1. \tag{5}$$

In order to solve Eq. (5), for a fixed value $k \geq 1$, we introduce some notation as follows:

$$\begin{aligned} x_{i,i_\varrho} &= m_{i,i_\varrho}^{(k)}, \\ \xi_{i,i_\varrho} &= k m_{i,i_\varrho}^{(k-1)}, \quad 0 \leq i \leq i_\varrho, \\ y_{i,i_\varrho} &= x_{i+1,i_\varrho} - x_{i,i_\varrho}, \quad 0 \leq i \leq i_\varrho - 1. \end{aligned}$$

Equation (5) is then equivalent to $\lambda_i y_{i,i_\varrho} + \xi_{i,i_\varrho} = \mu_i y_{i-1,i_\varrho}$, for $1 \leq i \leq i_\varrho - 1$, which implies that

$$y_{i,i_\varrho} = \zeta_i y_{0,i_\varrho} - \zeta_i b_i,$$

where $b_i = \sum_{j=1}^i \frac{\xi_{j,i_\varrho}}{\lambda_j \zeta_j}$. As a result, for $1 \leq i \leq i_\varrho - 1$, we have

$$x_{i+1,i_\varrho} = x_{1,i_\varrho} \sum_{j=0}^i \zeta_j - \sum_{j=1}^i \zeta_j b_j.$$

We also have $x_{i_\varrho,i_\varrho} = 0$ since $m_{i_\varrho,i_\varrho}^{(k)} = 0$, so that

$$\begin{aligned} x_{1,i_\varrho} &= \frac{\sum_{j=1}^{i_\varrho-1} \zeta_j b_j}{\sum_{j=0}^{i_\varrho-1} \zeta_j}, \\ x_{i,i_\varrho} &= \frac{\sum_{j=i}^{i_\varrho-1} \zeta_j \sum_{m=0}^{i-1} \zeta_m \sum_{k=m+1}^j \frac{\xi_{k,i_\varrho}}{\lambda_k \zeta_k}}{\sum_{j=0}^{i_\varrho-1} \zeta_j}, \quad 2 \leq i \leq i_\varrho - 1, \\ x_{i_\varrho,i_\varrho} &= 0. \end{aligned} \tag{6}$$

Equation (6) is a recursive procedure that allows us to compute the k th order moments, $m_{i,i_\varrho}^{(k)}$, in terms of previously computed moments, $m_{i,i_\varrho}^{(k-1)}$, since $x_{i,i_\varrho} = m_{i,i_\varrho}^{(k)}$ and $\xi_{i,i_\varrho} = k m_{i,i_\varrho}^{(k-1)}$. In the special cases $k = 0$ and 1, it is readily seen that

$$\lim_{i_\varrho \rightarrow \infty} v_{i,i_\varrho} = \begin{cases} 0, & \text{if } 0 \leq i \leq i_\varrho - 1, \\ 1, & \text{if } i = i_\varrho, \end{cases}$$

and the asymptotic value $\lim_{i_\varrho \rightarrow \infty} m_{i,i_\varrho}^{(1)}$ is always finite, with $\lim_{i_\varrho \rightarrow \infty} m_{i,i_\varrho}^{(1)} = 0$ for initial clonal sizes $i \in \{0, i_\varrho\}$, since $\sum_{m=0}^\infty \zeta_m = \infty$, for any value of the parameter ν .

We finally focus on the random variables X_i^{max} and T_i^{max} . Note that the distribution of X_i^{max} is readily derived from the values v_{i,i_ϱ} , since $v_{i,i_\varrho} = \mathbb{P}(X_i^{max} \geq i_\varrho)$, and the k th order moment, $m_i^{max,(k)}$, of the time T_i^{max} to reach the maximum clonal size can be evaluated as

$$\begin{aligned} m_i^{max,(k)} &= \sum_{i_\varrho=i}^{\infty} E \left[(T_i^{max})^k \mid X_i^{max} = i_\varrho \right] \mathbb{P}(X_i^{max} = i_\varrho) \\ &= \sum_{i_\varrho=i+1}^{\infty} m_{i,i_\varrho}^{(k)} \left(1 - \frac{v_{i,i_\varrho+1}}{v_{i,i_\varrho}} \right). \end{aligned}$$

For practical or computational purposes, the above series should be replaced by the finite sum

$$\sum_{i_\varrho=i+1}^{K_q} m_{i,i_\varrho}^{(k)} \left(1 - \frac{v_{i,i_\varrho+1}}{v_{i,i_\varrho}} \right), \tag{7}$$

where K_q can be selected as the $(100q)$ th percentile of X_i^{max} for a probability $q \in (0, 1)$ close enough to 1. Although an analytical study of $\sum_{i_\varrho=K_q+1}^{\infty} m_{i,i_\varrho}^{(k)} (1 - v_{i,i_\varrho}^{-1} v_{i,i_\varrho+1})$ does not seem to be feasible, we note that by using the finite sum in Eq. (7) we ensure that the probability mass accumulated by X_i^{max} is greater than q , whence the probability $1 - q$ can be interpreted as a global error measure. Furthermore, this type of truncation procedure has been efficiently used for epidemics (Almaraz et al. 2016) and competition processes (Gómez-Corral and López García 2011, 2012).

2.2 Number of proliferation events to reach the maximum clonal size

We are now interested in the number N_i^{max} of one-step transitions $i' \rightarrow i' + 1$ of \mathcal{X} occurring in the random interval $[0, T_i^{max}]$, which allows us to record the number of proliferation events to reach the maximum clonal size, for $i \in \mathcal{X}$. This descriptor can be seen as a discrete version of the random variable T_i^{max} in Sect. 2.1, and its analysis can be carried out by means of the number of proliferation events to reach a total clonal size, which is defined as the number N_{i,i_ϱ} of one-step transitions $i' \rightarrow i' + 1$ of \mathcal{X} to register $X(t) = i_\varrho$ for the first time.

We first introduce the following notation:

$$\begin{aligned} \psi_{i,i_\varrho}(z) &= E \left[z^{N_{i,i_\varrho}} 1_{\{N_{i,i_\varrho} < \infty\}} \right], \quad |z| \leq 1, \\ u_{i,i_\varrho} &= \mathbb{P}(N_{i,i_\varrho} < \infty), \\ \tilde{m}_{i,i_\varrho}^{(k)} &= E \left[N_{i,i_\varrho} (N_{i,i_\varrho} - 1) \cdots (N_{i,i_\varrho} - k + 1) 1_{\{N_{i,i_\varrho} < \infty\}} \right], \quad k \geq 0, \end{aligned}$$

for initial clonal sizes $1 \leq i \leq i_\varrho$. It can be easily verified that $\psi_{0,i_\varrho}(z) = u_{0,i_\varrho} = \tilde{m}_{0,i_\varrho}^{(k)} = 0$, $\psi_{i_\varrho,i_\varrho}(z) = u_{i_\varrho,i_\varrho} = 1$ and $\tilde{m}_{i_\varrho,i_\varrho}^{(k)} = 0$, for $k \geq 1$. For clonal sizes $1 \leq i \leq i_\varrho - 1$, it is seen that $u_{i,i_\varrho} = v_{i,i_\varrho}$, since $\psi_{i,i_\varrho}(1) = \phi_{i,i_\varrho}(0)$ and, consequently, the probabilities u_{i,i_ϱ} can be computed from Eq. (4), and they amount to the probabilities of reaching the clonal size i_ϱ before extinction, given that the initial clonal size is i .

By taking derivatives on the equality

$$\psi_{i,i_\varrho}(z) = \frac{\mu_i}{\lambda_i + \mu_i} \psi_{i-1,i_\varrho}(z) + \frac{\lambda_i}{\lambda_i + \mu_i} z \psi_{i+1,i_\varrho}(z), \quad 1 \leq i \leq i_\varrho - 1,$$

we obtain

$$(\lambda_i + \mu_i) \tilde{m}_{i,i_\varrho}^{(k)} = \mu_i \tilde{m}_{i-1,i_\varrho}^{(k)} + \lambda_i \tilde{m}_{i+1,i_\varrho}^{(k)} + \lambda_i k \tilde{m}_{i+1,i_\varrho}^{(k-1)}, \quad 1 \leq i \leq i_\varrho - 1, \quad k \geq 1, \tag{8}$$

which is similar to Eq. (5) with the term $k m_{i,i_\varrho}^{(k-1)}$ replaced by $\lambda_i k \tilde{m}_{i+1,i_\varrho}^{(k-1)}$. This implies that, similarly to the solution of Eq. (6), the solution of Eq. (8) has the form

$$\tilde{m}_{i,i_\varrho}^{(k)} = \frac{\sum_{j=i}^{i_\varrho-1} \zeta_j \sum_{m=0}^{i-1} \zeta_m \sum_{n=m+1}^j \frac{k \tilde{m}_{n+1,i_\varrho}^{(k-1)}}{\zeta_n}}{\sum_{j=0}^{i_\varrho-1} \zeta_j}, \quad 1 \leq i \leq i_\varrho - 1, \quad k \geq 1. \tag{9}$$

Similarly to Sect. 2.1.2, it can be seen that $\lim_{i_\varrho \rightarrow \infty} \tilde{m}_{i,i_\varrho}^{(1)}$ is always finite, with $\lim_{i_\varrho \rightarrow \infty} \tilde{m}_{i,i_\varrho}^{(1)} = 0$ for initial clonal sizes $i \in \{0, i_\varrho\}$.

In the case of the random variable N_{i,i_ϱ} , not only every k th order factorial moment can be obtained from Eq. (9), but also its distribution. We can write

$$\tilde{x}_{i,i_\varrho}^j = \mathbb{P}(N_{i,i_\varrho} = j), \quad 0 \leq i \leq i_\varrho, \quad j \geq i_\varrho - i,$$

with $\tilde{x}_{0,i_\varrho}^j = 0$ for $j \geq 0$, $\tilde{x}_{i_\varrho,i_\varrho}^j = 1$ if $j = 0$, and 0 if $j \geq 1$, $\tilde{x}_{i,i_\varrho}^j = 0$ for $1 \leq i \leq i_\varrho - 1$ and $0 \leq j \leq i_\varrho - i - 1$, and $\tilde{x}_{i,i_\varrho}^0 = 0$ for $0 \leq i \leq i_\varrho - 1$. For $j \geq 1$, we have

$$\begin{aligned} \tilde{x}_{0,i_\varrho}^j &= 0, \\ \tilde{x}_{i,i_\varrho}^j &= \frac{\mu_i}{\lambda_i + \mu_i} \tilde{x}_{i-1,i_\varrho}^j + \frac{\lambda_i}{\lambda_i + \mu_i} \tilde{x}_{i+1,i_\varrho}^{j-1}, \quad 1 \leq i \leq i_\varrho - 1, \\ \tilde{x}_{i_\varrho,i_\varrho}^j &= 0, \end{aligned}$$

so that a recursion on j and i can be obtained in order to compute the probability mass function of N_{i,i_ϱ} . The probability mass function obtained in this way is consistent with the expressions obtained in Eq. (9) for the factorial moments, but the proof is omitted here.

Finally, the factorial moments of the random variable N_i^{max} can be approximated in a similar way as described for Eq. (7). In particular, if we denote $\tilde{m}_i^{max,(k)} = E[N_i^{max}(N_i^{max} - 1) \cdots (N_i^{max} - k + 1)]$, we have

$$\begin{aligned} \tilde{m}_i^{max,(k)} &= \sum_{i_\varrho=i+1}^{\infty} \tilde{m}_{i,i_\varrho}^{(k)} \left(1 - \frac{v_{i,i_\varrho+1}}{v_{i,i_\varrho}}\right) \\ &\approx \sum_{i_\varrho=i+1}^{K_q} \tilde{m}_{i,i_\varrho}^{(k)} \left(1 - \frac{v_{i,i_\varrho+1}}{v_{i,i_\varrho}}\right). \end{aligned} \tag{10}$$

2.3 Time to reach a given number of proliferation events

In this section, we fix an initial clonal size i' and a number $D \geq 0$ of proliferation events, and we study the time $T_{i'}^D$ to reach a total number D of proliferation events (one-step transitions $i'' \rightarrow i'' + 1$ of \mathcal{X}), with $T_{i'}^0 = 0$. In order to study this descriptor, we consider the augmented process $\{(X(t), D(t)) : t \geq 0\}$, where $D(t)$ denotes the number of division events up to time t , with $(X(0), D(0)) = (i', 0)$. For the initial clonal size i' , the augmented process starts from state $(i', 0)$ and is defined on the finite state space

$$\mathcal{X}_0 \cup \mathcal{X}_D(i') \cup \mathcal{X}_T(i'),$$

where

$$\begin{aligned} \mathcal{X}_0 &= \{(0, d) : 0 \leq d \leq D - 1\}, \\ \mathcal{X}_D(i') &= \{(i, D) : 2 \leq i \leq i' + D\}, \\ \mathcal{X}_T(i') &= \{(i, d) : 1 \leq i \leq i' + d, 0 \leq d \leq D - 1\}. \end{aligned}$$

States in $\mathcal{X}_0 \cup \mathcal{X}_D(i')$ are absorbing states. In particular, states in \mathcal{X}_0 represent the extinction of the clonotype when the number D of proliferation events has not been reached, while states in $\mathcal{X}_D(i')$ reflect that the number D of proliferation events has been reached. States in $\mathcal{X}_T(i')$ are transient states. Figure 2 illustrates the dynamics of the augmented process in the case $D = 3$.

In a more general setting, we define the random variable

$$T_{(i,d)}^D = \inf\{t : (X(t), D(t)) \in \mathcal{X}_D(i') | (X(0), D(0)) = (i, d)\},$$

for states $(i, d) \in \mathcal{X}_0 \cup \mathcal{X}_D(i') \cup \mathcal{X}_T(i')$. Note that the descriptor $T_{i'}^D$ is equivalent to the random variable $T_{(i',0)}^D$, and states in \mathcal{X}_0 and $\mathcal{X}_D(i')$ lead to the values $T_{(0,d)}^D = \infty$ and $T_{(i,D)}^D = 0$, respectively. For states in $\mathcal{X}_T(i')$, the random variable $T_{(i,d)}^D$ is defective, and $\mathbb{P}(T_{(i,d)}^D = \infty) > 0$ is the probability of reaching \mathcal{X}_0 before getting to $\mathcal{X}_D(i')$.

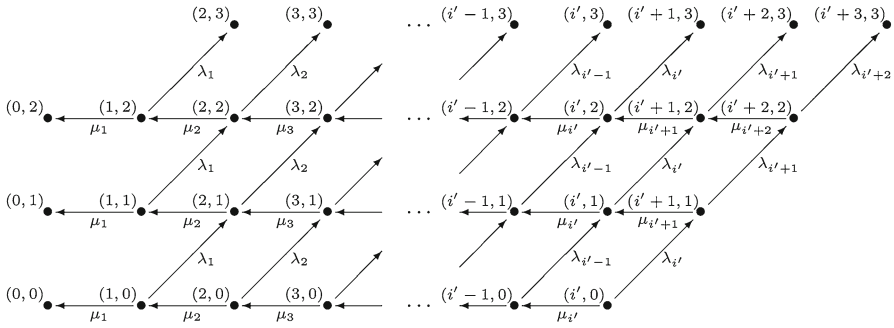


Fig. 2 Transitions between augmented states in the case $D = 3$

For states $(i, d) \in \mathcal{X}_0 \cup \mathcal{X}_D(i') \cup \mathcal{X}_T(i')$, we introduce the notation

$$\begin{aligned}
 w_{(i,d)}^D &= \mathbb{P}(T_{(i,d)}^D < \infty), \\
 \Phi_{(i,d)}^D(s) &= E \left[e^{-sT_{(i,d)}^D} 1_{\{T_{(i,d)}^D < \infty\}} \right], \quad \Re(s) \geq 0, \\
 \hat{m}_{(i,d)}^{D,(k)} &= E \left[\left(T_{(i,d)}^D \right)^k 1_{\{T_{(i,d)}^D < \infty\}} \right], \quad k \geq 0,
 \end{aligned}$$

and we observe that

$$w_{(i,d)}^D = \begin{cases} 0, & \text{if } (i, d) \in \mathcal{X}_0, \\ \in (0, 1), & \text{if } (i, d) \in \mathcal{X}_T(i'), \\ 1, & \text{if } (i, d) \in \mathcal{X}_D(i'). \end{cases}$$

As a result, the Laplace-Stieltjes transforms $\Phi_{(i,d)}^D(s)$ verify the boundary conditions $\Phi_{(0,d)}^D(s) = 0$ if $(0, d) \in \mathcal{X}_0$, and $\Phi_{(i,D)}^D(s) = 1$ if $(i, D) \in \mathcal{X}_D(i')$. For states $(i, d) \in \mathcal{X}_T(i')$, we have

$$\Phi_{(i,d)}^D(s) = \frac{\mu_i}{s + \lambda_i + \mu_i} \Phi_{(i-1,d)}^D(s) + \frac{\lambda_i}{s + \lambda_i + \mu_i} \Phi_{(i+1,d+1)}^D(s), \quad (11)$$

which yields a finite system of linear equations that can be solved in a recursive manner (Algorithm 1).

The boundary conditions for the moments $\hat{m}_{(i,d)}^{D,(k)}$ are given by

$$\begin{aligned}
 \hat{m}_{(i,d)}^{D,(0)} &= w_{(i,d)}^D, \quad (i, d) \in \mathcal{X}_0 \cup \mathcal{X}_D(i') \cup \mathcal{X}_T(i'), \\
 \hat{m}_{(i,d)}^{D,(k)} &= 0, \quad (i, d) \in \mathcal{X}_0 \cup \mathcal{X}_D(i'),
 \end{aligned}$$

Algorithm 1 Computation of the Laplace-Stieltjes transforms $\Phi_{(i,d)}^D(s)$ for states $(i, d) \in \mathcal{X}_0 \cup \mathcal{X}_D(i') \cup \mathcal{X}_T(i')$.

```

Step 1:  $d := D$ ;
        $i := 1$ ;
       while  $i < i' + d$ , do
            $i := i + 1$ ;
            $\Phi_{(i,d)}^D(s) := 1$ ;
       enddo.
Step 2: While  $d > 1$ , do
        $d := d - 1$ ;
        $i := 0$ ;
        $\Phi_{(i,d)}^D(s) := 0$ ;
       while  $i < i' + d$ , do
            $i := i + 1$ ;
            $\Phi_{(i,d)}^D(s) := \frac{\mu_i \Phi_{(i-1,d)}^D(s) + \lambda_i \Phi_{(i+1,d+1)}^D(s)}{s + \lambda_i + \mu_i}$ ;
       enddo;
   enddo.

```

where the probabilities $w_{(i,d)}^D$, for states $(i, d) \in \mathcal{X}_T(i')$, are derived from Algorithm 1 by selecting $s = 0$. If we take derivatives in Eq. (11), we may write down

$$\hat{m}_{(i,d)}^{D,(k)} = \frac{k}{\lambda_i + \mu_i} \hat{m}_{(i,d)}^{D,(k-1)} + \frac{\mu_i}{\lambda_i + \mu_i} \hat{m}_{(i-1,d)}^{D,(k)} + \frac{\lambda_i}{\lambda_i + \mu_i} \hat{m}_{(i+1,d+1)}^{D,(k)}, \quad (12)$$

for states $(i, d) \in \mathcal{X}^T(i')$. Eq. (12) can be solved (Algorithm 2) by adapting our arguments in Algorithm 1.

3 Numerical results

In this Section we carry out a set of numerical experiments in order to analyse the potential of a recent thymic emigrant (of a given clonotype) in the periphery to expand to different clonal sizes, or to proliferate for a given number of divisions, as well as the random times of these events, and the rate of contraction of the clonotype under consideration. We study these random variables under several competition and signalling environments, which are specified by the choice of parameters ν , $\langle n \rangle$ and φ . We point out here that parameters φ and ν should be seen as intrinsically related to the particular TCR expressed by the T cells within the clonotype under consideration, since this TCR determines the number of different self-peptides that the T cell can interact with. On the other hand, $\langle n \rangle$ is not directly related to the TCR and should be considered an environmental parameter, since it is the characteristic size of the competing clonotypes. We set the per cell death rate $\mu = 1$, so that the time unit in the process under study is the mean lifetime of a T cell.

We consider throughout this Section homeostatic signalling rates that have been selected for a wide range of parameter values, in view of some preliminary numerical experiments. In particular, our results regarding the hard, the intermediate and the soft

Algorithm 2 Computation of the moments $\hat{m}_{(i,d)}^{D,(k)}$ for states $(i, d) \in \mathcal{X}_0 \cup \mathcal{X}_D(i') \cup \mathcal{X}_T(i')$.

```

Step 1:  $k' := 0$ ;
 $d := D$ ;
 $i := 1$ ;
while  $i < i' + d$ , do
     $i := i + 1$ ;
     $\hat{m}_{(i,d)}^{D,(k')} := 1$ ;
enddo;
while  $d > 1$ , do
     $d := d - 1$ ;
     $i := 0$ ;
     $\hat{m}_{(i,d)}^{D,(k')} := 0$ ;
    while  $i < i' + d$ , do
         $i := i + 1$ ;
         $\hat{m}_{(i,d)}^{D,(k')} := \frac{\mu_i \hat{m}_{(i-1,d)}^{D,(k')} + \lambda_i \hat{m}_{(i+1,d+1)}^{D,(k')}}{\lambda_i + \mu_i}$ ;
    enddo;
enddo.
Step 2: While  $k' < k$ , do
     $k' := k' + 1$ ;
     $d := D$ ;
     $i := 1$ ;
    while  $i < i' + d$ , do
         $i := i + 1$ ;
         $\hat{m}_{(i,d)}^{D,(k')} := 0$ ;
    enddo;
    while  $d > 1$ , do
         $d := d - 1$ ;
         $i := 0$ ;
         $\hat{m}_{(i,d)}^{D,(k')} := 0$ ;
        while  $i < i' + d$ , do
             $i := i + 1$ ;
             $\hat{m}_{(i,d)}^{D,(k')} := \frac{k \hat{m}_{(i,d)}^{D,(k'-1)} + \mu_i \hat{m}_{(i-1,d)}^{D,(k')} + \lambda_i \hat{m}_{(i+1,d+1)}^{D,(k')}}{\lambda_i + \mu_i}$ ;
        enddo;
    enddo;
enddo;
    
```

niche cases amount to values of the signalling rate φ , the number ν of competitors, and the characteristic size $\langle n \rangle$ of these competitors ranging from *friendly* scenarios, where the T cell clonotype under consideration can become established in the periphery, to *hostile* environments, where this establishment is not possible at all. In particular, we analyse:

- (a) The hard niche case, which corresponds to a clonotype (or RTE) with no competitors ($\nu = 0$), in Table 1, with numerical results based on the stochastic descriptors X_i^{max} , T_i^{max} and N_i^{max} , for various small-to-moderate homeostatic signalling rates, $\varphi \in \{0.1, 1.0, 5.0, 10.0, 20.0, 30.0, 40.0, 50.0, 75.0, 100.0\}$. The hard niche case is also analysed in Fig. 3, where the probability mass function of X_i^{max} is plotted for various homeostatic signalling rates, $\varphi \in \{0.1, 1.0, 5.0, 10.0\}$,

Table 1 Expected values $E[X_i^{max}]$, $E[T_i^{max}]$ and $E[N_i^{max}]$ as a function of φ for the hard niche case (i.e., $v = 0$), for parameters $\mu = 1.0$ and $i = 1$

φ	$E[X_i^{max}]$	$E[T_i^{max}]$	$E[N_i^{max}]$	q	0.25	0.5	0.75	0.99
0.1	1.08143	0.07834	0.08617	K_q	1	1	1	2
				$\mathbb{P}(X_i^{max} \geq K_q)$	0.99526	0.99526	0.99526	0.08617
				$E[T_i, K_q X_i^{max} \geq K_q]$	0	0	0	0.90909
				$\sigma[T_i, K_q X_i^{max} \geq K_q]$	0	0	0	0.90909
				$E[N_i, K_q X_i^{max} \geq K_q]$	0	0	0	1
				$\sigma[N_i, K_q X_i^{max} \geq K_q]$	0	0	0	0
1.0	1.84694	0.52477	1.07984	K_q	1	2	3	5
				$\mathbb{P}(X_i^{max} \geq K_q)$	0.99351	0.49351	0.24351	0.02292
				$E[T_i, K_q X_i^{max} \geq K_q]$	0	0.5	1.25	2.79412
				$\sigma[T_i, K_q X_i^{max} \geq K_q]$	0	0.5	1.03078	2.03962
				$E[N_i, K_q X_i^{max} \geq K_q]$	0	1	2.5	6.29412
				$\sigma[N_i, K_q X_i^{max} \geq K_q]$	0	0	0.86603	2.66205
5.0	8.40494	12.72341	67.15374	K_q	4	10	12	16
				$\mathbb{P}(X_i^{max} \geq K_q)$	0.74812	0.54931	0.32916	0.00999
				$E[T_i, K_q X_i^{max} \geq K_q]$	0.72892	9.43564	21.56767	39.98014
				$\sigma[T_i, K_q X_i^{max} \geq K_q]$	0.52755	8.08455	19.77093	37.53902
				$E[N_i, K_q X_i^{max} \geq K_q]$	4.02410	49.43688	113.44516	214.03020
				$\sigma[N_i, K_q X_i^{max} \geq K_q]$	1.43975	39.26485	98.40159	191.6886
10.0	21.50413	927.95677	9.29×10^3	K_q	22	24	26	29
				$\mathbb{P}(X_i^{max} \geq K_q)$	0.80227	0.61532	0.25744	0.01306
				$E[T_i, K_q X_i^{max} \geq K_q]$	209.05940	736.27200	1745.83800	2435.43200
				$\sigma[T_i, K_q X_i^{max} \geq K_q]$	206.35540	733.34640	1742.73800	2432.12500

Table 1 continued

φ	$E[X_i^{max}]$	$E[T_i^{max}]$	$E[N_i^{max}]$	q	0.25	0.5	0.75	0.99
20.0	48.33950	1.01×10^7	2.02×10^8	$E[N_i, K_q X_i^{max} \geq K_q]$	2.09×10^3	7.37×10^3	1.75×10^4	2.44×10^4
				$\sigma[N_i, K_q X_i^{max} \geq K_q]$	2.06×10^3	7.33×10^3	1.74×10^4	2.43×10^4
				K_q	50	51	53	56
				$\mathbb{P}(X_i^{max} \geq K_q)$	0.79967	0.65553	0.24416	0.01118
				$E[T_i, K_q X_i^{max} \geq K_q]$	3.81×10^6	7.71×10^6	1.88×10^7	2.51×10^7
				$\sigma[T_i, K_q X_i^{max} \geq K_q]$	3.81×10^6	7.71×10^6	1.88×10^7	2.51×10^7
30.0	75.17354	1.48×10^{11}	4.44×10^{12}	$E[N_i, K_q X_i^{max} \geq K_q]$	7.62×10^7	1.54×10^8	3.77×10^8	5.03×10^8
				$\sigma[N_i, K_q X_i^{max} \geq K_q]$	7.62×10^7	1.54×10^8	3.77×10^8	5.03×10^8
				K_q	77	78	80	83
				$\mathbb{P}(X_i^{max} \geq K_q)$	0.82173	0.67261	0.24499	0.01096
				$E[T_i, K_q X_i^{max} \geq K_q]$	5.25×10^{10}	1.10×10^{11}	2.73×10^{11}	3.62×10^{11}
				$\sigma[T_i, K_q X_i^{max} \geq K_q]$	5.25×10^{10}	1.10×10^{11}	2.73×10^{11}	3.62×10^{11}
40.0	102.01704	2.44×10^{15}	9.75×10^{16}	$E[N_i, K_q X_i^{max} \geq K_q]$	1.57×10^{12}	3.28×10^{12}	8.19×10^{12}	1.09×10^{13}
				$\sigma[N_i, K_q X_i^{max} \geq K_q]$	1.57×10^{12}	3.28×10^{12}	8.19×10^{12}	1.09×10^{13}
				K_q	104	105	107	110
				$\mathbb{P}(X_i^{max} \geq K_q)$	0.83761	0.69029	0.25417	0.01138
				$E[T_i, K_q X_i^{max} \geq K_q]$	8.06×10^{14}	1.72×10^{15}	4.42×10^{15}	5.93×10^{15}
				$\sigma[T_i, K_q X_i^{max} \geq K_q]$	8.06×10^{14}	1.72×10^{15}	4.42×10^{15}	5.93×10^{15}
40.0	102.01704	2.44×10^{15}	9.75×10^{16}	$E[N_i, K_q X_i^{max} \geq K_q]$	3.22×10^{16}	6.88×10^{16}	1.77×10^{17}	2.37×10^{17}
				$\sigma[N_i, K_q X_i^{max} \geq K_q]$	3.22×10^{16}	6.88×10^{16}	1.77×10^{17}	2.37×10^{17}

Table 1 continued

φ	$E[X_i^{max}]$	$E[T_i^{max}]$	$E[N_i^{max}]$	q	0.25	0.5	0.75	0.99
50.0	128.84185	4.28×10^{19}	2.14×10^{21}	K_q	131	132	134	137
				$\tilde{\mathbb{P}}(X_i^{max} \geq K_q)$	0.85166	0.70967	0.26879	0.01220
				$E[T_i, K_q X_i^{max} \geq K_q]$	1.30×10^{19}	2.84×10^{19}	7.60×10^{19}	1.04×10^{20}
				$\sigma[T_i, K_q X_i^{max} \geq K_q]$	1.30×10^{19}	2.84×10^{19}	7.60×10^{19}	1.04×10^{20}
				$E[N_i, K_q X_i^{max} \geq K_q]$	6.52×10^{20}	1.42×10^{21}	3.80×10^{21}	5.19×10^{21}
				$\sigma[N_i, K_q X_i^{max} \geq K_q]$	6.52×10^{20}	1.42×10^{21}	3.80×10^{21}	5.19×10^{21}
75.0	196.43319	2.06×10^{30}	1.55×10^{32}	K_q	199	200	201	205
				$\tilde{\mathbb{P}}(X_i^{max} \geq K_q)$	0.83576	0.67186	0.444024	0.00960
				$E[T_i, K_q X_i^{max} \geq K_q]$	7.42×10^{29}	1.58×10^{30}	2.77×10^{30}	4.97×10^{30}
				$\sigma[T_i, K_q X_i^{max} \geq K_q]$	7.42×10^{29}	1.58×10^{30}	2.77×10^{30}	4.97×10^{30}
				$E[N_i, K_q X_i^{max} \geq K_q]$	5.57×10^{31}	1.19×10^{32}	2.07×10^{32}	3.73×10^{32}
				$\sigma[N_i, K_q X_i^{max} \geq K_q]$	5.57×10^{31}	1.19×10^{32}	2.07×10^{32}	3.73×10^{32}
100.0	264.12728	1.01×10^{41}	1.11×10^{43}	K_q	267	268	269	273
				$\tilde{\mathbb{P}}(X_i^{max} \geq K_q)$	0.81774	0.63708	0.39927	0.00800
				$E[T_i, K_q X_i^{max} \geq K_q]$	4.59×10^{40}	9.55×10^{40}	1.61×10^{41}	2.68×10^{41}
				$\sigma[T_i, K_q X_i^{max} \geq K_q]$	4.59×10^{40}	9.55×10^{40}	1.61×10^{41}	2.68×10^{41}
				$E[N_i, K_q X_i^{max} \geq K_q]$	4.59×10^{42}	9.55×10^{42}	1.61×10^{43}	2.68×10^{43}
				$\sigma[N_i, K_q X_i^{max} \geq K_q]$	4.59×10^{42}	9.55×10^{42}	1.61×10^{43}	2.68×10^{43}

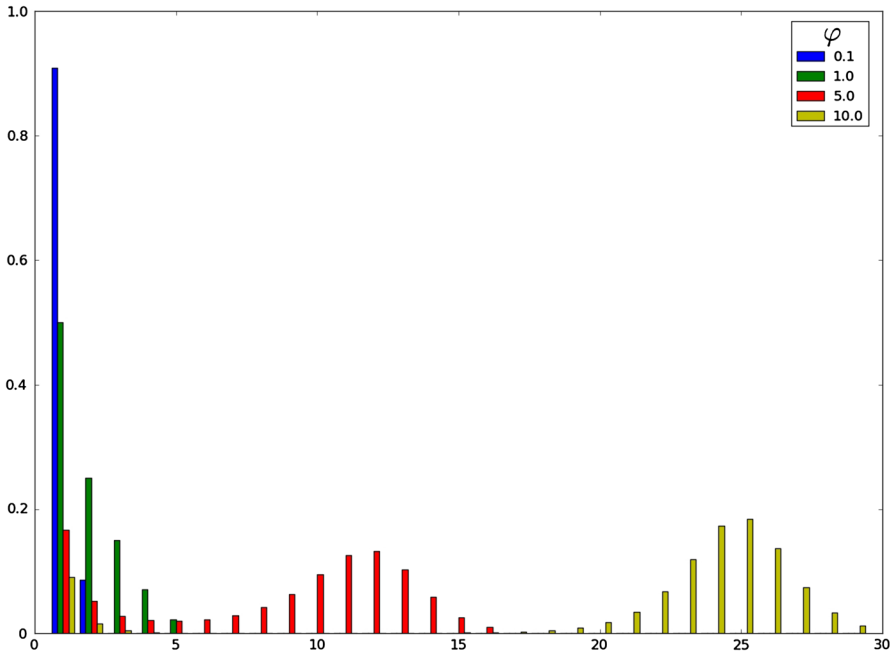


Fig. 3 Probability mass function of X_i^{max} for different values of φ in the hard niche case (*i.e.*, $\nu = 0$), for $\mu = 1.0$ and the initial clonal size $i = 1$

- (b) The intermediate niche case in Table 2 and Fig. 4, which corresponds to moderate average numbers of competitors, $\nu \in \{1, 10, 50\}$, and moderate homeostatic signalling rates, $\varphi = 50$,
- (c) The soft niche case in Table 3 and Fig. 5, which reflects large average numbers of competitors, $\nu \in \{500, 1000\}$, and high homeostatic signalling rates, $\varphi = 500$.

In Tables 1, 2 and 3, the mean values $E[X_i^{max}]$, $E[T_i^{max}]$ and $E[N_i^{max}]$ are computed by considering only clonal sizes up to the 99th percentile $K_{0.99}$ of X_i^{max} , and by means of the truncating procedures given by Eqs. (7) and (10). However, for parameter values $(\nu, \langle n \rangle) \in \{(500, 100), (1000, 50), (1000, 100)\}$ in Table 3, values of $E[X_i^{max}]$, $E[T_i^{max}]$ and $E[N_i^{max}]$ are computed by considering only clonal sizes up to the percentile $K_{0.999}$ of X_i^{max} , since $K_{0.99} = 1$ in these cases which could lead to misleading results. In this way, the quantities obtained are approximations to the *true* ones, and we focus on the particular case in which the clonotype under study has been recently released to the periphery (RTE), so that $i = 1$. A direct comparison of our results with results obtained by Gillespie simulations suggests a significant level of precision for the truncation procedures proposed by Eqs. (7) and (10). For example, values $E[X_i^{max}] = 8.40494$, $E[T_i^{max}] = 12.72341$ and $E[N_i^{max}] = 67.15374$ reported in Table 1 for $\varphi = 5.0$ have Gillespie (10^6 simulations) counterparts $E[X_i^{max}] = 8.49793$, $E[T_i^{max}] = 12.91816$ and $E[N_i^{max}] = 68.21512$, leading to relative error values, $1 - \frac{\text{truncated value}}{\text{simulated value}} = 0.0109, 0.0150$ and 0.0155 , respectively.

Table 2 Expected values $E[X_i^{max}]$, $E[T_i^{max}]$ and $E[N_i^{max}]$ for different choices of $(v, (n))$ in the intermediate niche case for the parameters $(\mu, \varphi) = (1.0, 50.0)$ and the initial clonal size $i = 1$

v	(n)	$E[X_i^{max}]$	$E[T_i^{max}]$	$E[N_i^{max}]$	q	0.25	0.5	0.75	0.99
1	1	123.25013	1.57×10^{18}	7.68×10^{19}	K_q	127	128	129	133
					$\mathbb{P}(X_i^{max} \geq K_q)$	0.78258	0.60668	0.38305	0.00847
					$E[T_i, K_q X_i^{max} \geq K_q]$	7.22×10^{17}	1.43×10^{18}	2.33×10^{18}	3.83×10^{18}
					$\sigma[T_i, K_q X_i^{max} \geq K_q]$	7.22×10^{17}	1.43×10^{18}	2.33×10^{18}	3.83×10^{18}
					$E[N_i, K_q X_i^{max} \geq K_q]$	3.54×10^{19}	7.00×10^{19}	1.14×10^{20}	1.88×10^{20}
					$\sigma[N_i, K_q X_i^{max} \geq K_q]$	3.54×10^{19}	7.00×10^{19}	1.14×10^{20}	1.88×10^{20}
					K_q	71	73	75	79
50	68.61117	2.16×10^8	5.92×10^9	$\mathbb{P}(X_i^{max} \geq K_q)$	0.83010	0.60581	0.27023	0.00942	
				$E[T_i, K_q X_i^{max} \geq K_q]$	5.97×10^7	1.86×10^8	3.75×10^8	5.22×10^8	
				$\sigma[T_i, K_q X_i^{max} \geq K_q]$	5.97×10^7	1.86×10^8	3.75×10^8	5.22×10^8	
				$E[N_i, K_q X_i^{max} \geq K_q]$	1.64×10^9	5.10×10^9	1.03×10^{10}	1.43×10^{10}	
				$\sigma[N_i, K_q X_i^{max} \geq K_q]$	1.64×10^9	5.10×10^9	1.03×10^{10}	1.43×10^{10}	
				K_q	60	62	63	67	
				100	57.61605	2.24×10^7	5.15×10^8	$\mathbb{P}(X_i^{max} \geq K_q)$	0.78631
$E[T_i, K_q X_i^{max} \geq K_q]$	8.71×10^6	2.54×10^7	3.59×10^7					5.44×10^7	
$\sigma[T_i, K_q X_i^{max} \geq K_q]$	8.71×10^6	2.54×10^7	3.59×10^7					5.44×10^7	
$E[N_i, K_q X_i^{max} \geq K_q]$	2.01×10^8	5.85×10^8	8.27×10^8					1.25×10^9	
$\sigma[N_i, K_q X_i^{max} \geq K_q]$	2.01×10^8	5.85×10^8	8.27×10^8					1.25×10^9	
K_q	92	95	97					102	
10	75.87474	4.57×10^9	1.82×10^{11}					$\mathbb{P}(X_i^{max} \geq K_q)$	0.75703
				$E[T_i, K_q X_i^{max} \geq K_q]$	5.38×10^8	3.50×10^9	8.10×10^9	1.31×10^{10}	

Table 2 continued

ν	$\langle n \rangle$	$E[X_i^{max}]$	$E[T_i^{max}]$	$E[N_i^{max}]$	q	0.25	0.5	0.75	0.99
50					$\sigma[T_i, K_q X_i^{max} \geq K_q]$	5.38×10^8	3.50×10^9	8.10×10^9	1.31×10^{10}
					$E[N_i, K_q X_i^{max} \geq K_q]$	2.15×10^{10}	1.40×10^{11}	3.23×10^{11}	5.24×10^{11}
					$\sigma[N_i, K_q X_i^{max} \geq K_q]$	2.15×10^{10}	1.40×10^{11}	3.23×10^{11}	5.24×10^{11}
					K_q	1	1	1	3
					$\mathbb{P}(X_i^{max} \geq K_q)$	0.99869	0.99869	0.99869	0.01028
					$E[T_i, K_q X_i^{max} \geq K_q]$	0	0	0	1.48326
					$\sigma[T_i, K_q X_i^{max} \geq K_q]$	0	0	0	1.14561
					$E[N_i, K_q X_i^{max} \geq K_q]$	0	0	0	2.10206
					$\sigma[N_i, K_q X_i^{max} \geq K_q]$	0	0	0	0.33538
					K_q	1	1	1	2
100					$\mathbb{P}(X_i^{max} \geq K_q)$	0.99682	0.99682	0.99682	0.05227
					$E[T_i, K_q X_i^{max} \geq K_q]$	0	0	0	0.94455
					$\sigma[T_i, K_q X_i^{max} \geq K_q]$	0	0	0	0.94455
					$E[N_i, K_q X_i^{max} \geq K_q]$	0	0	0	1
					$\sigma[N_i, K_q X_i^{max} \geq K_q]$	0	0	0	0
					K_q	1	2	3	20
					$\mathbb{P}(X_i^{max} \geq K_q)$	0.99229	0.49229	0.32338	0.00242
					$E[T_i, K_q X_i^{max} \geq K_q]$	0	0.5	1.00676	10.60944
					$\sigma[T_i, K_q X_i^{max} \geq K_q]$	0	0.5	0.82202	6.66161
					$E[N_i, K_q X_i^{max} \geq K_q]$	0	1	2.33784	77.95166
50	1	2.99647	1.06499	4.53753	$\sigma[N_i, K_q X_i^{max} \geq K_q]$	0	0	0.67229	43.81203

Table 2 continued

ν	$\langle n \rangle$	$E[X_i^{max}]$	$E[T_i^{max}]$	$E[N_i^{max}]$	q		0.25	0.5	0.75	0.99
50		1.01919	0.01920	0.01959	K_q	$\mathbb{P}(X_i^{max} \geq K_q)$	1	1	1	2
						$E[T_i, K_q X_i^{max} \geq K_q]$	0.99959	0.99959	0.99959	0.01959
						$\sigma[T_i, K_q X_i^{max} \geq K_q]$	0	0	0	0.98000
						$E[N_i, K_q X_i^{max} \geq K_q]$	0	0	0	0.98000
						$\sigma[N_i, K_q X_i^{max} \geq K_q]$	0	0	0	1
100		1.00990	0.00990	0.01000	K_q	$\mathbb{P}(X_i^{max} \geq K_q)$	1	1	1	2
						$E[T_i, K_q X_i^{max} \geq K_q]$	0.99990	0.99990	0.99990	0.01000
						$\sigma[T_i, K_q X_i^{max} \geq K_q]$	0	0	0	0.98990
						$E[N_i, K_q X_i^{max} \geq K_q]$	0	0	0	0.98990
						$\sigma[N_i, K_q X_i^{max} \geq K_q]$	0	0	0	1

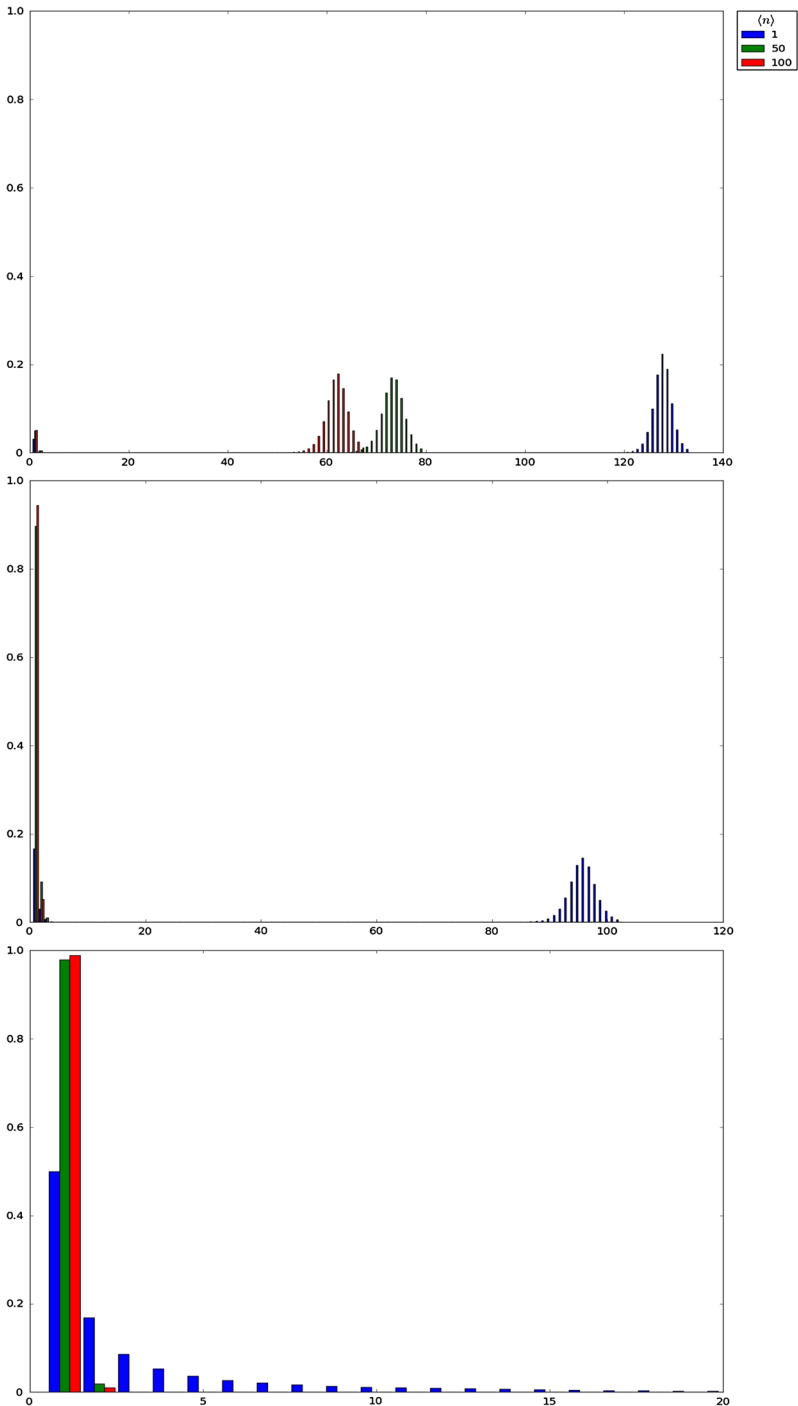


Fig. 4 Probability mass function of X_i^{max} as a function $\langle n \rangle$, in the intermediate niche case for (from top to bottom) $\nu \in \{1, 10, 50\}$, the parameters $(\mu, \varphi) = (1.0, 50.0)$ and the initial clonal size $i = 1$

Table 3 Expected values $E[X_i^{max}]$, $E[T_i^{max}]$ and $E[N_i^{max}]$ for different choices of $(v, (n))$ in the soft niche case, for the parameters $(\mu, \varphi) = (1.0, 500.0)$ and the initial clonal size $i = 1$

v	(n)	$E[X_i^{max}]$	$E[T_i^{max}]$	$E[N_i^{max}]$	q	0.25	0.5	0.75	0.99
500	1	3.64779	1.36144	9.36164	K_q	1	1	3	44
					$\tilde{P}(X_i^{max} \geq K_q)$	0.99048	0.99048	0.32293	0.00066
					$E[T_i, K_q X_i^{max} \geq K_q]$	0	0	1.00166	23.19127
					$\sigma[T_i, K_q X_i^{max} \geq K_q]$	0	0	0.81786	14.13329
					$E[N_i, K_q X_i^{max} \geq K_q]$	0	0	2.33378	363.97348
					$\sigma[N_i, K_q X_i^{max} \geq K_q]$	0	0	0.66722	222.16995
50	1	1.01882	0.01884	0.01922	K_q	1	1	1	2
					$\tilde{P}(X_i^{max} \geq K_q)$	0.99961	0.99961	0.99961	0.01922
					$E[T_i, K_q X_i^{max} \geq K_q]$	0	0	0	0.98039
					$\sigma[T_i, K_q X_i^{max} \geq K_q]$	0	0	0	0.98039
					$E[N_i, K_q X_i^{max} \geq K_q]$	0	0	0	1
					$\sigma[N_i, K_q X_i^{max} \geq K_q]$	0	0	0	0
100	1	1.00970	0.00971	0.00980	K_q	1	1	1	1
					$\tilde{P}(X_i^{max} \geq K_q)$	0.99010	0.99010	0.99010	0.99010
					$E[T_i, K_q X_i^{max} \geq K_q]$	0	0	0	0
					$\sigma[T_i, K_q X_i^{max} \geq K_q]$	0	0	0	0
					$E[N_i, K_q X_i^{max} \geq K_q]$	0	0	0	0
					$\sigma[N_i, K_q X_i^{max} \geq K_q]$	0	0	0	0
1000	1	1.54331	0.34871	0.64267	K_q	1	1	2	6
					$\tilde{P}(X_i^{max} \geq K_q)$	0.99225	0.99225	0.32536	0.00795
					$E[T_i, K_q X_i^{max} \geq K_q]$	0	0	0.66689	2.77159

Table 3 continued

ν	$\langle n \rangle$	$E[X_i^{max}]$	$E[T_i^{max}]$	$E[N_i^{max}]$	q	0.25	0.5	0.75	0.99
50				0.00980	$\sigma[T_i, K_q X_i^{max} \geq K_q]$	0	0	0.66689	1.77604
					$E[N_i, K_q X_i^{max} \geq K_q]$	0	0	1	7.28252
					$\sigma[N_i, K_q X_i^{max} \geq K_q]$	0	0	0	2.52032
					K_q	1	1	1	1
					$\tilde{\mathbb{P}}(X_i^{max} \geq K_q)$	0.99010	0.99010	0.99010	0.99010
					$E[T_i, K_q X_i^{max} \geq K_q]$	0	0	0	0
100			0.00495	$\sigma[T_i, K_q X_i^{max} \geq K_q]$	0	0	0	0	0
				$E[N_i, K_q X_i^{max} \geq K_q]$	0	0	0	0	0
				$\sigma[N_i, K_q X_i^{max} \geq K_q]$	0	0	0	0	0
				K_q	1	1	1	1	
				$\tilde{\mathbb{P}}(X_i^{max} \geq K_q)$	0.99503	0.99503	0.99503	0.99503	
				$E[T_i, K_q X_i^{max} \geq K_q]$	0	0	0	0	
				$\sigma[T_i, K_q X_i^{max} \geq K_q]$	0	0	0	0	0
				$E[N_i, K_q X_i^{max} \geq K_q]$	0	0	0	0	0
				$\sigma[N_i, K_q X_i^{max} \geq K_q]$	0	0	0	0	0
				K_q	1	1	1	1	
				$\tilde{\mathbb{P}}(X_i^{max} \geq K_q)$	0.99503	0.99503	0.99503	0.99503	
				$E[T_i, K_q X_i^{max} \geq K_q]$	0	0	0	0	

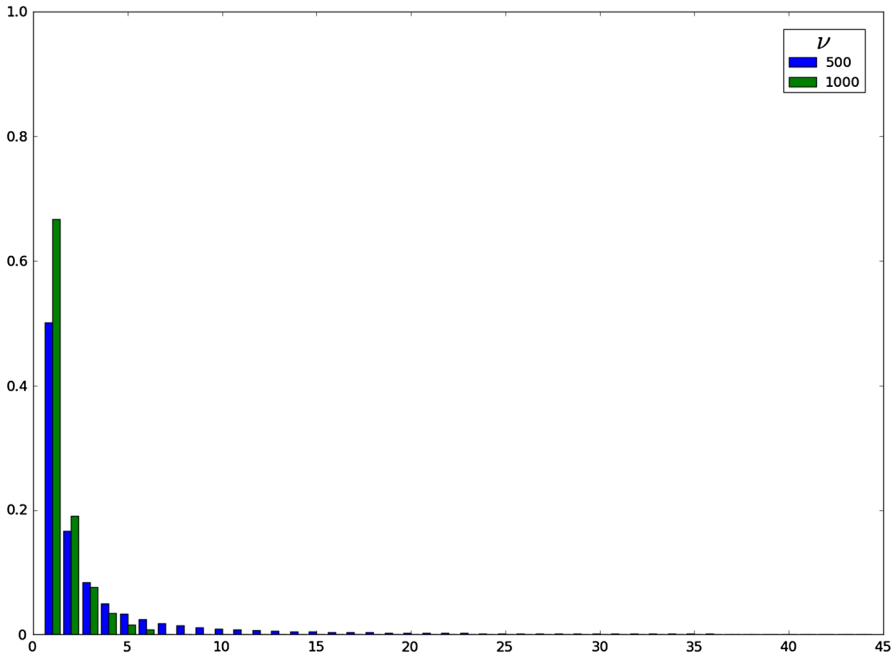


Fig. 5 Probability mass function of X_i^{max} in the soft niche case, for $\nu \in \{500, 1000\}$, the parameters $(\mu, \varphi) = (1.0, 500.0)$, $(n) = 1$ and the initial clonal size $i = 1$

The distribution of X_i^{max} is analysed in Tables 1, 2 and 3 by computing its $(100q)$ th percentiles K_q for $q \in \{0.25, 0.5, 0.75, 0.99\}$, where the percentile K_q is the first value $x \geq 1$, such that $\mathbb{P}(X_i^{max} \leq x) \geq q$. The probability $\mathbb{P}(X_i^{max} \geq K_q)$ of reaching the clonal sizes represented by K_q can be exactly obtained from Eq. (4). However, these probabilities are replaced in Tables 1, 2 and 3 by their truncated versions $\tilde{\mathbb{P}}(X_i^{max} \geq K_q) = \mathbb{P}(K_{0.99} \geq X_i^{max} \geq K_q)$. We have done so as these truncated values result in good approximations to the true probabilities, but with the advantage that, for the particular cases yielding $K_q = 1$ (for example, $\varphi \in \{0.1, 1.0\}$ and $q = 0.25$ in Table 1), they provide the total probability mass of X_i^{max} considered in the truncations given by Eqs. (7) and (10).

The time and the number of proliferation events to reach the maximum clonal size can be analysed in greater depth by studying the conditional values $E[T_{i,K_q} | X_i^{max} \geq K_q]$, $\sigma[T_{i,K_q} | X_i^{max} \geq K_q]$, $E[N_{i,K_q} | X_i^{max} \geq K_q]$ and $\sigma[N_{i,K_q} | X_i^{max} \geq K_q]$, where $\sigma[X]$ denotes the standard deviation of the random variable X . We note here that we consider conditional values due to the fact that the random variables T_{i,K_q} and N_{i,K_q} are defective; that is, we have $T_{i,K_q} = N_{i,K_q} = \infty$ if the clonal size K_q is not reached, which occurs with non-zero probability. These values provide information about the time and the number of proliferation events to reach different clonal sizes (K_q with $q \in \{0.25, 0.5, 0.75, 0.99\}$), under the assumption that those clonal sizes have been, in fact, reached.

When focusing on the hard niche case (Table 1; Fig. 3), the scenarios that allow the establishment of the RTE in the periphery are given by values of $E[X_i^{max}]$ and

$E[T_i^{max}]$ significantly larger than 1, which correspond to larger values of the homeostatic signalling rate φ . It is clear that the mean value, $E[X_i^{max}]$, has a clearly increasing behaviour with respect to the signalling rate φ . Moreover, results corresponding to those hostile environments within the hard niche case, corresponding to low signal rates, represent the pressure against the survival and expansion of the clonotype in the short-term. In those cases ($\varphi \in \{0.1, 1.0, 5.0\}$), the most representative sample path is the one which amounts to the immediate extinction of the clonotype from the initial clonal size $i = 1$, which gives $X_i^{max} = 1$, or the occurrence of a few proliferation events (yielding a maximum size of the clonotype near 1), and the subsequent extinction of the clonotype. On the other hand, in those cases where the clonotype survives in the mid-term (higher signalling rates), we observe a high concentration of the probability mass function of X_i^{max} around some mode, which is reflected in the concentrated values obtained for the percentiles $K_{0.25}$, $K_{0.5}$, $K_{0.75}$ and $K_{0.99}$. In particular, the probability mass function of X_i^{max} is unimodal when the clonotype becomes extinct in the short-term with high probability, and bimodal when the clonotype survives and expands in the mid-term, with moderate probabilities.

The bimodal shape of X_i^{max} in friendly, yet competitive peripheral environments, within the hard niche case, becomes more apparent when analysing results plotted in Fig. 3. In particular, we plot in Fig. 3 the probability mass function of X_i^{max} for the hard niche case for signalling rates $\varphi \in \{0.1, 1.0, 5.0, 10.0\}$. The clonotype under consideration has higher potential for expansion for larger values of the homeostatic signalling rate φ . However, in these cases, the probability mass function of X_i^{max} shows a bimodal shape, with a mode equal to 1 and the other taking different values depending on the particular parameters. This shape, together with the results of Table 1, should be interpreted as a binary outcome scenario: in these cases, the clonotype has moderate or high probabilities of reaching its potential clonal size represented by the second mode, which will only happen if it escapes the extinction at the beginning of its lifetime; on the other hand, there exists, with some significant probability, the chance of the clonotype becoming extinct in the short-term (during its first transitions), which yields the first mode equal to 1. We point out here that this also means that (for example, see the case $\varphi = 10.0$ in Fig. 3), once the clonotype escapes extinction in the short-term, the probabilities of reaching different clonal sizes, such as $i_q \in \{5, 10, 15\}$, are practically equal, since they represent the probability of reaching clonal sizes around the second mode. Finally, when the rate of signal is not enough (see the case $\varphi = 0.1$ in Fig. 3), the probability mass function of X_i^{max} is unimodal around 1. This represents the case in which the clonotype will become immediately extinct (or extinct in the short-term after reaching small sizes around 1) from its initial clonal size 1.

The increase of $E[X_i^{max}]$ associated with increasing values of the homeostatic signalling rate in Table 1 leads also to increasing mean values for $E[T_i^{max}]$, $E[N_i^{max}]$, $E[T_{i,K_q}|X_i^{max} \geq K_q]$ and $E[N_{i,K_q}|X_i^{max} \geq K_q]$. This is easily explained by noting that, in these cases, the clonotype has a greater potential to reach bigger clonal sizes, so that the time and number of proliferation events to reach those sizes should also be larger. The variability of these random variables increases, as well, which is observed by analysing the values $\sigma[T_{i,K_q}|X_i^{max} \geq K_q]$ and $\sigma[N_{i,K_q}|X_i^{max} \geq K_q]$. Moreover, the coefficient of variance $CV(X) = \sigma[X]/E[X]$ of these random variables seems to slowly increase to one, when increasing φ . In any case, the large values of $E[T_i^{max}]$ and

$E[N_i^{max}]$ imply that, although in friendly conditions a substantial clonotype maximum size will be reached with high probability, there exists a significant impediment against this event, represented by the large number of proliferation and cell death events taking place until this maximum size is reached.

The intermediate niche case is analysed in Table 2 and Fig. 4. In this case, friendly environments can be identified with small values of ν ($\nu = 1$ and $\langle n \rangle \in \{1, 50, 100\}$ in Table 2), or with moderate values of ν together with small values of $\langle n \rangle$ ($\nu = 10$ and $\langle n \rangle = 1$ in Table 2), while hostile scenarios correspond to greater numbers of competitors and larger clonal sizes for them. Again, the clonotype (or RTE) can become established in the periphery under friendly enough environments, reaching a maximum clonal size near $E[X_i^{max}]$, represented by the larger mode of the distribution of X_i^{max} (given by the value around which its percentiles are concentrated). Under too hostile scenarios, the clonotype will not become established and its maximum clonal size remains near 1. Again, this binary outcome scenario can be properly identified by analysing the results in Fig. 4, where the probability mass function of X_i^{max} is plotted for different parameter regimes for the intermediate niche case. It is worth noting that even in the case in which the most probable outcome is the establishment of the clonotype (e.g., $\nu = 1$ and $\langle n \rangle = 1$ in Fig. 4), there exists a non-negligible probability of extinction in the short-term. This probability is identified with the probability mass of X_i^{max} around the value 1.

We note here that there exist several scenarios where a certain expected behaviour can be identified. For example, in the particular case $(\nu, \langle n \rangle) = (50, 50)$ in Table 2, we obtain the following percentiles of X_i^{max} : $(K_{0.25}, K_{0.5}, K_{0.75}, K_{0.99}) = (1, 1, 1, 2)$, which represent a high concentration of the probability mass function around the value 1. That is, there is a high probability of the clonotype becoming immediately extinct from its initial clonal size $i = 1$. In particular, for percentiles $K_q = 1$, the corresponding time T_{i,K_q} to reach the clonal size K_q is equal to 0, since the initial clonal size is already $i = 1$, and the standard deviation equals 0. Similar comments can be made for N_{i,K_q} . The number of proliferation events to reach $K_q = 2$ (given that it is reached) is equal to 1, with probability 1, and then its standard deviation is 0. It should be noted that the distribution of the random time $T_{i,2}$, under the assumption that $T_{i,2} < \infty$, can be seen as the distribution of the minimum between two independent and exponentially distributed random variables X and Y with means λ_1^{-1} and μ_1^{-1} in the case $X < Y$. As a result, $T_{i,2}$ is exponentially distributed with mean $(\lambda_1 + \mu_1)^{-1}$, and $E[T_{i,2}|X_i^{max} \geq 2] = \sigma[T_{i,2}|X_i^{max} \geq 2] = 1/(\lambda_1 + \mu_1) = 0.98000$. Moreover, we note that one should carefully interpret the value of $E[X_i^{max}]$ in Tables 1, 2 and 3, since these values have been obtained making use of the truncation procedure described in Eqs. (7) and (10).

Finally, the soft niche case representing a hostile environment given by a large number of competitors and big clonal sizes for them, is analysed in Table 3 and Fig. 5. Results in Table 3 represent scenarios where the most probable outcome is the short-term extinction of the clonotype (values of $E[X_i^{max}]$, $E[T_i^{max}]$ and $E[N_i^{max}]$ near 1, percentiles K_q accumulated around 1). This outcome can be identified by analysing the probability mass function of X_i^{max} for these cases, which is concentrated around 1 (see Fig. 5).

In Figs. 6, 7, 8 and 9 and Table 4, we focus on the auxiliary descriptors T_{i,i_ϱ} , with $i_\varrho < i$, and T_i^D , which have not been analysed previously. In particular, we consider in Figs. 6, 7, and 8 the intermediate niche case and we analyse the time T_{i,i_ϱ} to contraction to a given clonal size $i_\varrho < i$. The mean time, $E[T_{i,i_\varrho}]$, and the standard deviation, $\sigma[T_{i,i_\varrho}]$, are computed for different values of i ($i = 10, 100$ and 500 for Figs. 6, 7 and 8, respectively), i_ϱ , and parameters ν, φ and $\langle n \rangle$. The results in these figures show relatively little dependence of this time to contraction on the competition parameters ($\nu, \langle n \rangle$). The initial i and final i_ϱ clonal sizes, and the signalling rate φ seem to have a higher impact on the behaviour of this random variable, since larger values of φ are related to a higher resistance of the clonotype against contraction.

In Table 4 and Fig. 9, the time T_i^D to reach a particular number of proliferation events, D , is studied for the soft niche case. In particular, the probabilities $\mathbb{P}(T_i^D < \infty)$

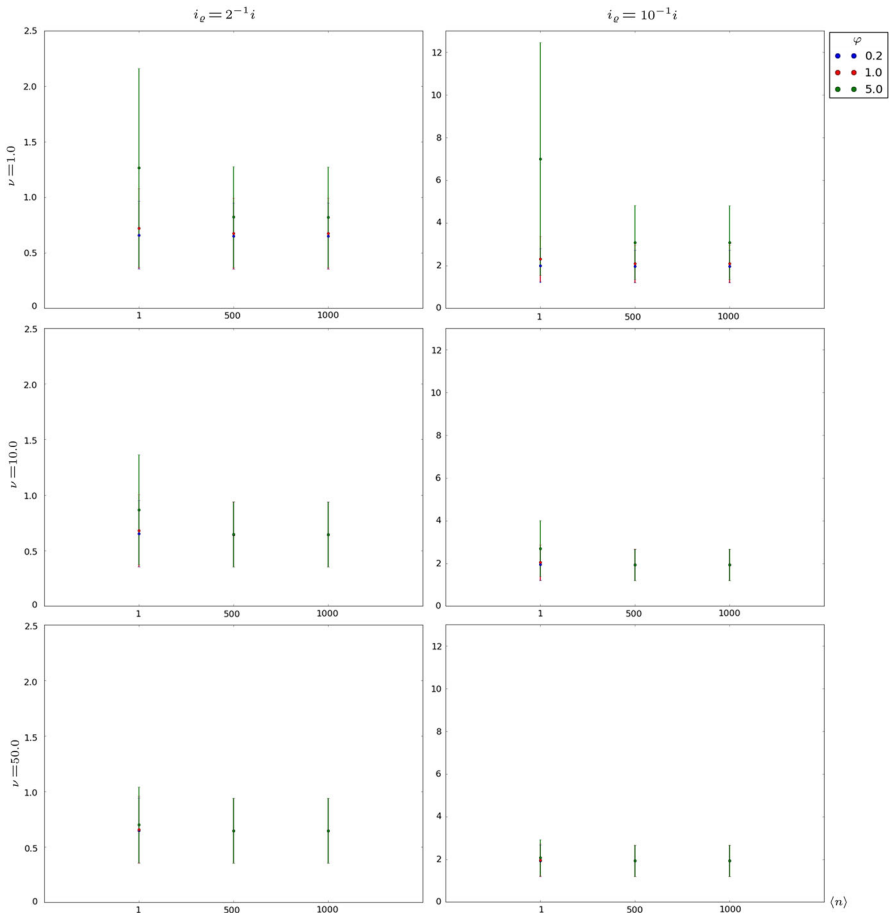


Fig. 6 Expected value $E[T_{i,i_\varrho}]$ plus and minus standard deviation $\sigma[T_{i,i_\varrho}]$ for $i = 10$, in the intermediate niche case, for $\mu = 1.0$ and various choices of i_ϱ, ν, φ and $\langle n \rangle$

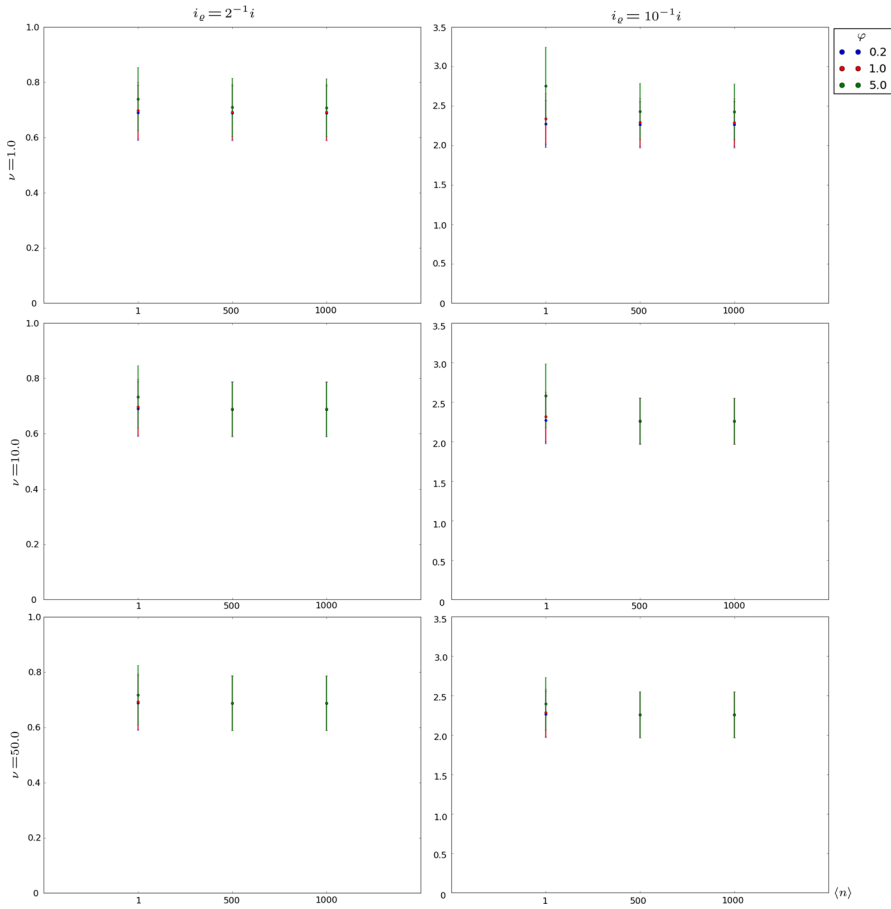


Fig. 7 Expected value $E[T_{i,i_0}]$ plus and minus standard deviation $\sigma[T_{i,i_0}]$ for $i = 100$, in the intermediate niche case, for $\mu = 1.0$ and various choices of i_0, ν, φ and $\langle n \rangle$

are given in Table 4, and the conditional mean values $E[T_i^D | T_i^D < \infty]$ and conditional standard deviations $\sigma[T_i^D | T_i^D < \infty]$ are shown in Fig. 9 for different values of i, D and $\langle n \rangle$. In contrast to the time to contraction, the results of Table 4 and Fig. 9 indicate that T_i^D (which is a measure of the proliferation capacity of the clonotype) seems to be highly sensitive to $\langle n \rangle$, the characteristic number of naive T cells in a competitor clonotype. When the competition environment is too hostile, which is represented by large values of $\langle n \rangle$, the potential of the clonotype in the periphery to reach a certain number, D , of proliferation events is notably reduced, and this is reflected by the decrease of the probabilities $\mathbb{P}(T_i^D < \infty)$. At the same time, means and standard deviations in Fig. 9 increase in those scenarios. On the other hand, for small values of $\langle n \rangle$, there exist parameter regimes where the target number of divisions, D , is easily reached (see, for example, the cases associated with the initial clonal size $i = 100$ in Table 4).

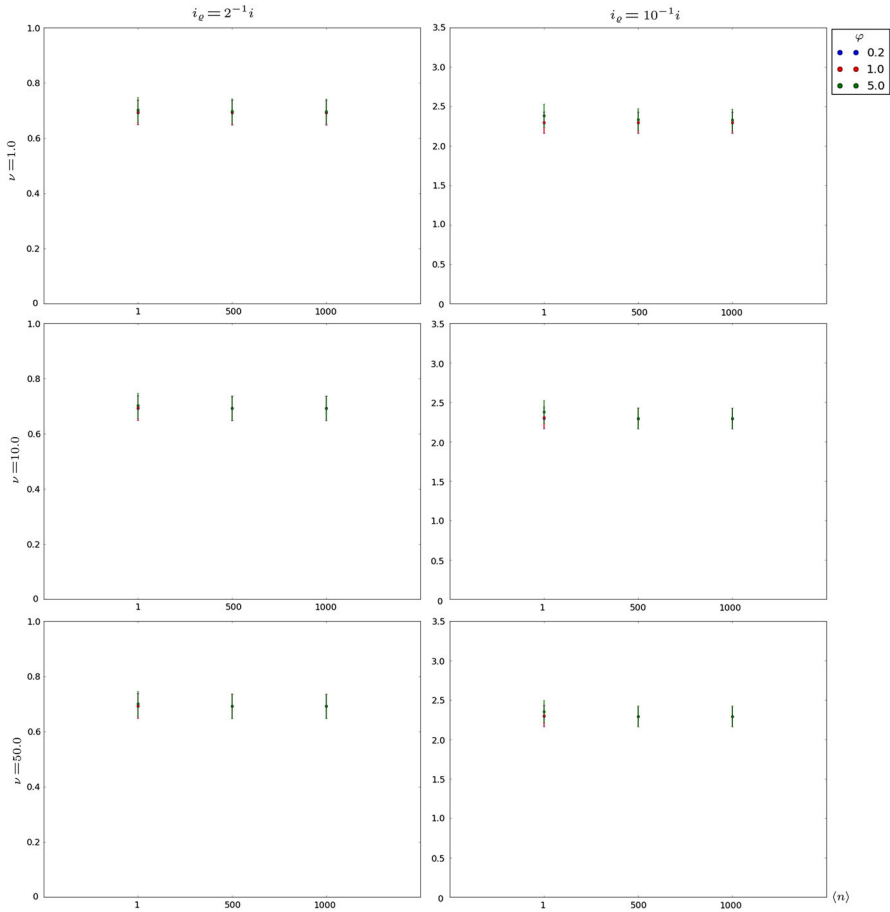


Fig. 8 Expected value $E[T_{i,i_0}]$ plus and minus standard deviation $\sigma[T_{i,i_0}]$ for $i = 500$, in the intermediate niche case, for $\mu = 1.0$ and various choices of i_0 , ν , φ and $\langle n \rangle$

4 Discussion

In this paper we have analysed the fate of a given naive T cell clonotype or a recent thymic emigrant (with initial clonal size $i = 1$), when competing with pre-established peripheral T cell clonotypes for homeostatic proliferation signals. The dynamics of a given clonotype has been based on a mathematical stochastic model originally developed in [Stirk et al. \(2008\)](#). Here, our objective is to study the fate and potential of naive T cell clonotypes in the periphery by analysing, for a given clonotype, its maximum clonal size, the time to reach this maximum, the number of proliferation events required to reach this maximum, the rate of contraction of the clonotype during its way to extinction, as well as the time to a given number of proliferation events.

We have introduced in Sect. 2 several stochastic descriptors in terms of defective and non-defective random variables. The use of Laplace-Stieltjes transforms, probability

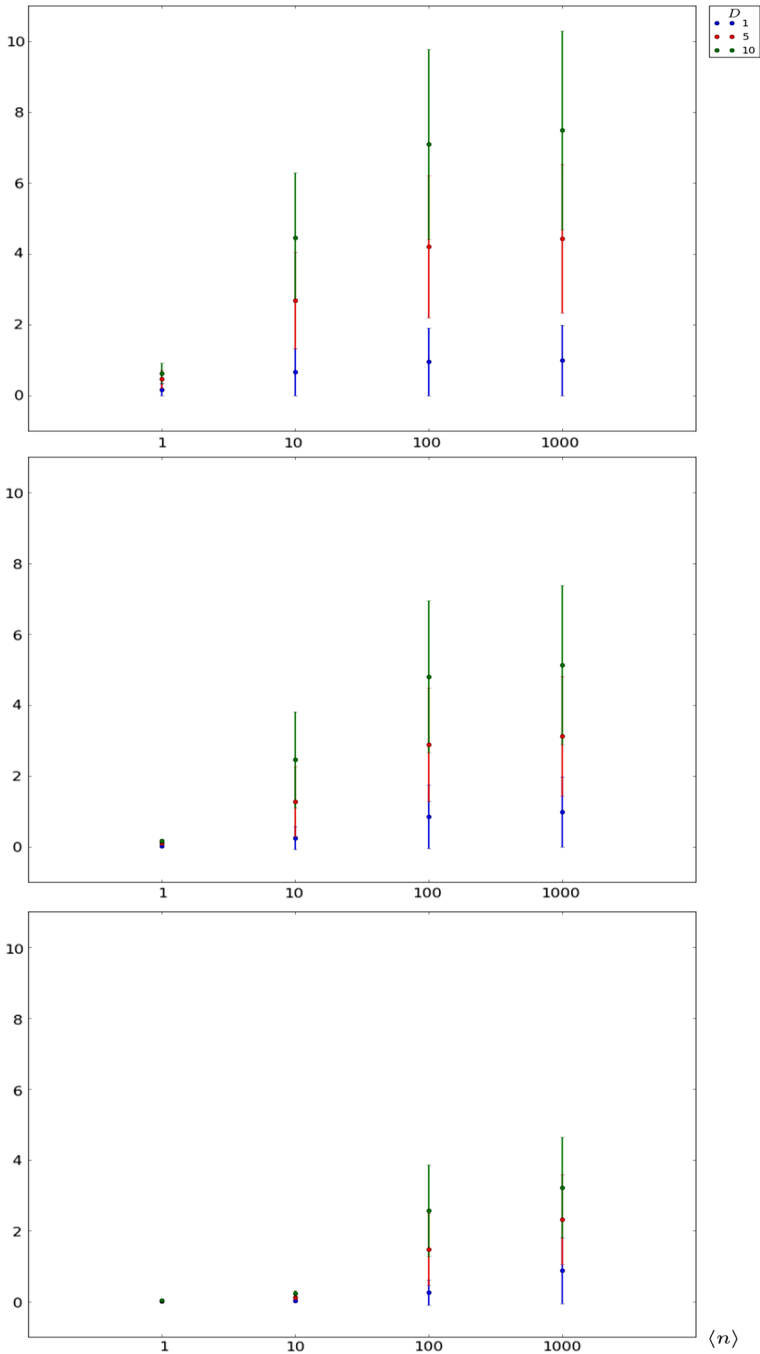


Fig. 9 Conditional expected value $E[T_i^D | T_i^D < \infty]$ plus and minus conditional standard deviation $\sigma[T_i^D | T_i^D < \infty]$ for (from *top* to *bottom*) $i \in \{1, 10, 100\}$, in the soft niche case, and for various choices of the number D of proliferation events, and the parameters $\mu = 1.0$, $\varphi = 1000.0$ and $\nu = 200.0$

Table 4 Probability $\mathbb{P}(T_i^D < \infty)$ as a function of the initial clonal size i in the soft niche case, for various choices of the number D of proliferation events, and the parameters $\mu = 1.0$, $\varphi = 1000.0$ and $\nu = 200.0$

i	D	$\langle n \rangle$	1	10	100	1000
1	1		0.83263	0.33322	0.04761	0.00497
	5		0.79944	0.05986	$<10^{-5}$	$<10^{-9}$
	10		0.79876	0.01653	$<10^{-9}$	$<10^{-18}$
10	1		0.99999	0.98249	0.38600	0.04865
	5		0.99999	0.74816	0.00113	$<10^{-7}$
	10		0.99999	0.41791	$<10^{-6}$	$<10^{-16}$
100	1	1	1	1	0.99230	0.39263
	5	1	1	0.99999	0.58648	0.00022
	10	1	1	0.99999	0.05467	$<10^{-9}$

generating functions, auxiliary random variables and absorbing augmented Markov processes, together with first-step analysis, allow us to obtain, in some cases, the distribution of the random variables of interest, or to exactly compute or approximate the different order moments of these variables.

In Sect. 3 several numerical experiments have been carried out in order to study the descriptors previously defined, under different environmental conditions for the clonotype under consideration. In particular, when a recent thymic emigrant reaches the periphery, it will compete with a small or large number of different clonotypes depending on the TCR it expresses (which in turn determines its cross-reactivity). The parameter that encodes the number of competitors is ν in the model. The characteristic clonal size of these competitors, $\langle n \rangle$, as well as the total rate of stimulatory signal provided by the self-peptides that the given clonotype can recognise, φ , also affect the dynamics of the clonotype under study, so that three different regimes are defined (hard, soft or intermediate niche). For these three different scenarios, numerical results have been provided for our descriptors, which allow us to follow the survival and extinction dynamics of the clonotype or RTE. Our main results can be summarised as follows:

- Two fates can be identified for the dynamics of the clonotype: if the clonotype under study does not receive enough stimulatory signal, or if it faces too hostile a competitive environment (encoded by a large number of competitors and/or large clonal sizes of these competitors), the survival probability of the clonotype in the mid-term is low. This fate can be seen by analysing the results in Fig. 4 for $\nu = 50$, or Fig. 5. If the clonotype faces moderate competitive environments, the probability of the clonotype escaping extinction and surviving in the long-term is significant, and it can be approximated in terms of the probability mass function of X_i^{max} .
- In this second case, the bimodal structure of the probability mass function of X_i^{max} (where one mode is 1 or near 1, and the other mode is far away from it) reflects a binary outcome scenario:
 - One possibility is the extinction of the clonotype in the short-term, with non-negligible probability. The probability of this event can be identified with the sum of the probabilities around 1 in the probability mass function of X_i^{max} .

- If such an event does not occur, the clonotype escapes extinction and it reaches its maximum size, which will be a value near the second mode of the probability mass function of X_i^{max} .

For example, in Fig. 4 for $\nu = 1$ and $\langle n \rangle = 1$, the probability mass around 1 represents the probability of the clonotype becoming extinct after a few proliferation and death events (then, the maximum clonal size will be 1 or a value near 1). The long term survival probability of the clonotype can be approximately identified with $\mathbb{P}(X_i^{max} \geq 6)$. We note here that the probability mass for states between 6 and 110 is almost null, which reflects the fact that once the clonotype escapes extinction, it will reach maximum clonal sizes above 110 almost surely. Moreover, the high concentration of the probability mass function around the second mode (which can also be identified by analysing the percentiles of X_i^{max} in Tables 1, 2 and 3) should be interpreted as the fact that the set of parameters $(\varphi, \nu, \langle n \rangle)$ (and, thus, the molecular properties of the corresponding TCR and the clonal sizes of competing T cell clonotypes) highly determines the maximum size of a clonotype in the periphery.

Recent experimental work suggests that whether a recent thymic emigrant survives or not upon its arrival in the periphery, depends on the clonal sizes of the pre-existing and competing naive T cell clonotypes: under lymphopenic conditions, that is, small values of $\langle n \rangle$, RTEs seem to be able to proliferate and become established in the periphery, but when the peripheral naive T cell pool is full, RTEs struggle to get incorporated into the population of recirculating lymphocytes (Fink and Hendricks 2011; Houston et al. 2011; Berkley and Fink 2014). This empirical observation is in line with our results: the fate of a recent thymic emigrant during its initial journey in the periphery has a clear stochastic component, where the probability of extinction cannot be neglected, even under favourable conditions for the RTE, and which can be studied in terms of the mass probability function of X_i^{max} . On the other hand, a greater deterministic component can be identified in the potential size of the clonotype seeded by the RTE in the long-term, once it escapes extinction (which is more likely for small values of $\langle n \rangle$), which is reflected in the concentration of the mass probability function of X_i^{max} around the second mode.

Other interesting results that have been derived from our study of the stochastic descriptors defined here are:

1. There seems to exist some threshold value for the parameter φ when analysing the time and the number of proliferation events to reach the maximum clonal size. This threshold behaviour can be observed, for example, in Table 1 for parameters $\varphi \in \{10.0, 20.0\}$ and $\nu = 0$ (hard niche case). It can be observed that, for these cases, small changes in $E[X_1^{max}]$ are, however, associated with large changes in $E[T_1^{max}]$ and $E[N_1^{max}]$. These results suggest that different clonotypes with similar maximum clonal sizes can display very different behaviours as to how this maximum size is achieved (average time and number of divisions), depending on the value of φ , the homeostatic proliferation signalling rate.
2. Figures 6, 7 and 8 allow us to show that the rate of contraction of the given clonotype on its way to extinction seems to depend much more on φ , the homeostatic

proliferation signalling rate, than on the competition parameters (ν , $\langle n \rangle$). At the same time, the competition parameters do significantly affect the proliferation rate of the clonotype, which is analysed in Table 4 and Fig. 9, in terms of the time to reach a particular number of proliferation events.

Acknowledgements This work was supported by the Government of Spain (Ministry of Economy and Competitiveness), Project MTM-2011-23864 (Jesús Rodríguez Artalejo, Antonio Gómez-Corral, Martín López-García), Project FIS2009-12964-C05-03 (Carmen Molina-París) and Project FIS2013-47949-C2-2-P (Carmen Molina-París), the Leverhulme Trust, Research Grant RPG-2012-772 (Martín López-García, Carmen Molina-París), BBSRC (Research Development Fellowship BB/G023395/1, Carmen Molina-París), FP7 PITN-GA-2012-317040 (Carmen Molina-París), and the Severo Ochoa Excellence Program, Project SEV-2015-0554 (Antonio Gómez-Corral).

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- Almaraz E, Gómez-Corral A, Rodríguez-Bernal MT (2016) On the time to reach a critical number of infections in epidemic models with infective and susceptible immigrants. *BioSystems* 144:68–77
- Artalejo J, Economou A, López-Herrero MJ (2012) Stochastic epidemic models revisited: analysis of some continuous performance measures. *J Biol Dyn* 6:189–211
- Berkley A, Fink P (2014) Cutting edge: CD8+ recent thymic emigrants exhibit increased responses to low-affinity ligands and improved access to peripheral sites of inflammation. *J Immunol* 193:3262–3266
- Bird L (2009) T-cell development: thymocytes run the “gauntlet”. *Nat Rev Immunol* 9:2
- Correia-Neves M, Waltzinger C, Mathis D, Benoist C (2001) The shaping of the T cell repertoire. *Immunity* 14:21–32
- De Boer R, Perelson A (1997) Competitive control of the self-renewing T cell repertoire. *Int Immunol* 9:779–790
- De Boer R, Perelson A, Ribeiro R (2012) Modelling deuterium labelling of lymphocytes with temporal and/or kinetic heterogeneity. *J R Soc Interf* 9:2191–2200
- Fink P (2013) The biology of recent thymic emigrants. *Ann Rev Immunol* 31:31–50
- Fink P, Hendricks D (2011) Post-thymic maturation: young T cells assert their individuality. *Nat Rev Immunol* 11:544–549
- Freitas A, Rocha B (1999) Peripheral T cell survival. *Curr Opin Immunol* 11:152–156
- Freitas A, Rocha B (2000) Population biology of lymphocytes: the flight for survival. *Annu Rev Immunol* 18:83–111
- Gibson K, Wu Y, Barnett Y, Duggan O, Vaughan R, Kondeatis E, Nilsson BO, Wikby A, Kipling D, Dunn-Walters D (2009) B cell diversity decreases in old age and is correlated with poor health status. *Aging Cell* 8:18–25
- Gómez-Corral A, López García M (2012) Extinction times and size of the surviving species in a two-species competition process. *J Math Biol* 64:255–289
- Gómez-Corral A, López García M (2012) On the number of births and deaths during an extinction cycle, and the survival of a certain individual in a competition process. *Comput Math Appl* 64:236–259
- Goronzy JJ, Lee W-W, Weyand CM (2007) Aging and T-cell diversity. *Exp Gerontol* 42:400–406
- Hazenber M, Borghans J, De Boer R, Miedema F (2003) Thymic output: a bad TREC record. *Nat Immunol* 4:97–99
- Houston E, Higdon L, Fink P (2011) Recent thymic emigrants are preferentially incorporated only into the depleted T-cell pool. *Proc Natl Acad Sci* 108:5366–5371
- Johnson P, Goronzy J, Antia R (2014) A population biological approach to understanding the maintenance and loss of the T-cell repertoire during aging. *Immunology* 142:167–175
- Karlin S, McGregor J (1957) The classification of birth and death processes. *Trans Am Math Soc* 86:366–400
- Kulkarni V (1996) Modeling and analysis of stochastic systems. CRC Press, Boca Raton

- Latouche G, Ramaswami V (1999) Introduction to matrix analytic methods in stochastic modeling. Society for industrial and applied mathematics, ASA-SIAM series on statistics and applied probability
- Mahajan V, Leskov I, Chen J (2005) Homeostasis of T cell diversity. *Cell Molec Immunol* 2:1–10
- Mason D (1998) A very high level of crossreactivity is an essential feature of the T-cell receptor. *Immunol Today* 19:395–404
- McLean AR, Rosado MM, Agenes F, Vasconcellos R, Freitas AA (1997) Resource competition as a mechanism for B cell homeostasis. *Proc Natl Acad Sci* 94:5792–5797
- Moran A, Hogquist K (2012) T-cell receptor affinity in thymic development. *Immunology* 135:261–267
- Murphy K, Travers P, Walport M, Janeway C (2008) *Janeway's immunobiology*. Garland Science Publications
- Sawicka M, Sritesky G, Reynolds J, Abourashchi N, Lythe G, Molina-París C, Hogquist K (2014) From pre-DP, post-DP, SP4, and SP8 thymocyte cell counts to a dynamical model of cortical and medullary selection. *Front Immunol* 5(article 19):1–14
- Sewell A (2012) Why must T cells be cross-reactive? *Nat Rev Immunol* 12:669–677
- Sinclair C, Bains I, Yates AJ, Seddon B (2013) Asymmetric thymocyte death underlies the CD4: CD8 T-cell ratio in the adaptive immune system. *Proc Natl Acad Sci* 110:E2905–E2914
- Stirk E, Lythe G, van den Berg H, Hurst G, Molina-París C (2010) The limiting conditional probability distribution in a stochastic model of T cell repertoire maintenance. *Math Biosci* 224:74–86
- Stirk E, Molina-París C, van den Berg H (2008) Stochastic niche structure and diversity maintenance in the T cell repertoire. *J Theor Biol* 255:237–249
- Sritesky GL, Jameson SC, Hogquist KA (2012) Selection of self-reactive T cells in the thymus. *Ann Rev Immunol* 30:95–114
- Sritesky G, Xing Y, Erickson J, Kalekar L, Wang X, Mueller D, Jameson S, Hogquist K (2013) Murine thymic selection quantified using a unique method to capture deleted T cells. *Proc Natl Acad Sci* 110:4679–4684
- Takada K, Jameson S (2009) Naive T cell homeostasis: from awareness of space to a sense of place. *Nat Rev Immunol* 9:823–832
- Tanchot C, Rocha B (1998) The organization of mature T-cell pools. *Immunol Today* 19:575–579
- Troy AE, Shen H (2003) Cutting edge: homeostatic proliferation of peripheral T lymphocytes is regulated by clonal competition. *J Immunol* 170:672–676
- Yates A (2014) Theories and quantification of thymic selection. *Front Immunol* 5(article 13):1–15
- Zarnitsyna V, Evavold B, Schoettle L, Blattman J, Antia R (2013) Estimating the diversity, completeness, and cross-reactivity of the T cell repertoire. *Front Immunol* 4(article 485):1–11