Multi-type Galton-Watson processes with affinity-dependent selection applied to antibody affinity maturation

Irene Balelli, Vuk Milisic, Gilles Wainrib

Noname manuscript No.
(will be inserted by the editor)

Multi-type Galton-Watson processes with affinity-dependent selection applied to antibody affinity maturation

Irene Balelli · Vuk Milišić · Gilles Wainrib

Received: date / Accepted: date

Abstract We analyze the interactions between division, mutation and selection in a simplified evolutionary model, assuming that the population observed can be classified into fitness levels. The construction of our mathematical framework is motivated by the modeling of antibody affinity maturation of B-cells in Germinal Centers during an immune response. This is a key process in adaptive immunity leading to the production of high affinity antibodies against a presented antigen. Our aim is to understand how the different biological parameters affect the system's functionality. We identify the existence of an optimal value of the selection rate, able to maximize the number of selected B-cells for a given generation.

Keywords Multi-type Galton-Watson process · Germinal center reaction · Affinity-dependent selection · Evolutionary landscapes

Mathematics Subject Classification (2010) 60J80 · 60J85 · 60J85

Contents

I. Balelli · V. Milišić
Université Paris 13, Sorbonne Paris Cité, LAGA, CNRS (UMR 7539), laboratoire d'excellence Inflamex. F-93430 - Villetaneuse - France.
E-mail: balelli@math.univ-paris13.fr · milisic@math.univ-paris13.fr

I. Balelli
ISPED, Centre INSERM U1219, and INRIA - Statistics in System Biology and Translational Medicine Team. F-33000 - Bordeaux - France. E-mail: irene.balelli@inserm.fr

G. Wainrib
Ecole Normale Supérieure, Département d'Informatique.
45 rue d'Ulm, 75005 - Paris - France. E-mail: gilles.wainrib@ens.fr

G. Wainrib
Owkin, Inc. E-mail: gilles.wainrib@owkin.com

1 Introduction

Antibody Affinity Maturation (AAM) takes place in Germinal Centers (GCs), specialized micro-environnements which form in the peripheral lymphoid organs upon infection or immunization [37,10]. GCs are seeded by ten to hundreds distinct B-cells [34], activated after the encounter with an antigen, which initially undergo a phase of intense proliferation [10]. Then, AAM is achieved thanks to multiple rounds of division, Somatic Hypermutation (SHM) of the B-cell receptor proteins, and subsequent selection of B-cells with improved ability of antigen-binding [20]. B-cells which successfully complete the GC reaction output as memory B-cells or plasma cells [38,10]. Indirect evidence suggests that only B-cells exceeding a certain threshold of antigen-affinity differentiate into plasma cells [30]. The efficiency of GCs is assured by the contribution of other immune molecules, for instance Follicular Dendritic Cells (FDCs) and follicular helper T-cells (Tfh). Nowadays the key dynamics of GCs are well characterized [20,10,13,34]. Despite this there are still mechanisms which remain unclear, such as the dynamics of clonal competition of B-cells, hence how the selection acts. In recent years a number of mathematical models of the GC reaction has appeared to investigate these questions, such as [22,40], where agent-based models are developed and analyzed through extensive numerical simulations, or [45] where the authors establish a coarse-grained model, looking for optimal values of *e.g.* the selection strength and the initial B-cell fitness maximizing the affinity improvement.

Our aim in this paper is to contribute to the mathematical foundations of adaptive immunity by introducing and studying a simplified evolutionary model inspired by AAM, including division, mutation, affinity-dependent selection and death. We focus on interactions between these mechanisms, identify and analyze the parameters which mostly influence the system functionality, through a rigorous mathematical analysis. This research is motivated by important biotechnological applications. Indeed, the fundamental understanding of the evolutionary mechanisms involved in AAM have been inspiring many methods for the synthetic production of specific antibodies for drugs, vaccines or cancer immunotherapy [2,19,32]. This production process involves

the selection of high affinity peptides and requires smart methods to generate an appropriate diversity [9]. Beyond biomedical motivations, the study of this learning process has also given rise in recent years to a new class of bio-inspired algorithms [7,27,35], mainly addressed to solve optimization and learning problems.

We consider a model in which B-cells are classified into $N+1$ affinity classes with respect to a presented antigen, N being an integer big enough to opportunely describe the possible fitness levels of a B-cell with respect to a specific antigen [41,43]. A B-cell is able to increase its fitness thanks to SHMs of its receptors: only about 20% of all mutations are estimated to be affinity-affecting mutations [31,33]. By conveniently defining a transition probability matrix, we can characterize the probability that a B-cell belonging to a given affinity class passes to another one by mutating its receptors thanks to SHMs. Therefore we define a selection mechanism which acts on B-cells differently depending on their fitness. We mainly focus on a model of *positive and negative selection* in which B-cells submitted to selection either die or exit the GC as output cells, according to the strength of their affinity with the antigen. Hence, in this case, no recycling mechanism is taken into account. Nevertheless the framework we set is very easy to manipulate: we can define and study other kinds of affinity-dependent selection mechanisms, and eventually include recycling mechanisms, which have been demonstrated to play an important role in AAM [39]. We demonstrate that independently from the transition probability matrix defining the mutational mechanism and the affinity threshold chosen for positive selection, the optimal selection rate maximizing the number of output cells for the t^{th} generation is $1/t$, $t \in \mathbb{N}$ (Proposition 6).

From a mathematical point of view, we study a class of multi-types Galton-Watson (GW) processes (*e.g.* [14,3]) in which, by considering dead and selected B-cells as two distinct types, we are able to formalize the evolution of a population submitted to an affinity-dependent selection mechanism. To our knowledge, the problem of affinity-dependent selection in GW processes has not been deeply investigated so far.

In Section 2 we define the main model analyzed in this paper. We give as well some definitions that we will use in next sections. Section 3 contains the main mathematical results. A convenient use of a multi-type GW process allows to study the evolution of both GC and output cells over time. We determine the optimal value of the selection rate which maximizes the expected number of selected B-cells at any given maturation cycle in Section 3.3. We conclude Section 3 with some numerical simulations. In Section 4 we define two possible variants of the model described in previous sections, and provide some mathematical results and numerical simulations as well. This evidences how the mathematical tools used in Section 3 easily apply to define other affinity-dependent selection models. Finally, in Section 5 we discuss our modeling assumptions and give possible extensions and limitations of our mathematical

model. In order to facilitate the reading of the paper, some technical mathematical demonstrations, as well as some classical results about Galton-Watson theory are reported in the Appendix for interested readers.

2 Main definitions and modeling assumptions

This section provides the mathematical framework of this article. Let us suppose that given an antigen target cell $\overline{\mathbf{x}}$, all B-cell traits can be divided in exactly $N+1$ distinct affinity classes, named 0 to N.

Definition 1 Let $\overline{\mathbf{x}}$ be the antigen target trait. Given a B-cell trait $\mathbf{x}$, we denote by $a_{\overline{\mathbf{x}}}(\mathbf{x})$ the affinity class it belongs to with respect to $\overline{\mathbf{x}}$, $a_{\overline{\mathbf{x}}}(\mathbf{x}) \in \{0,\dots,N\}$. The maximal affinity corresponds to the first class, 0, and the minimal one to N.

Definition 2 Let $\mathbf{x}$ be a B-cell trait belonging to the affinity class $a_{\overline{\mathbf{x}}}(\mathbf{x})$ with respect to $\overline{\mathbf{x}}$. We say that its affinity with $\overline{\mathbf{x}}$ is given by:

$$\mathrm{aff}(\mathbf{x},\overline{\mathbf{x}}) = N - a_{\overline{\mathbf{x}}}(\mathbf{x})$$

Of course, this is not the only possible choice of affinity. Typically affinity is represented as a Gaussian function [40,22], having as argument the distance between the B-cell trait and the antigen in the shape space of possible traits. In our model this distance corresponds to the index of the affinity class the B-cell belongs to (0 being the minimal distance, N the maximal one). Nevertheless the choice of the affinity function does not affect our model.

During the GC reaction B-cells are submitted to random mutations. This implies switches from one affinity class to another with a given probability. Setting these probability means defining a mutational rule on the state space $\{0,\dots,N\}$ of affinity classes indices (the formal mathematical definition will be given in Section 3.2).

The main model we study in this paper is represented schematically in Figure 1. It is defined as follows:

Definition 3 The process starts with $z_0 \geq 1$ B-cells entering the GC, belonging to some affinity classes in $\{0,\dots,N\}$. In case they are all identical, we denote by a_0 the affinity class they belong to, with respect to the antigen target cell $\overline{\mathbf{x}}$. At each time step, each GC B-cell can eventually undertake three distinct processes: division, mutation and selection. First of all, each GC B-cell can die with a given rate $r_d \in [0,1]$. If not, each B-cell can divide with rate $r_{div} \in [0,1]$: each daughter cell may have a mutated trait, according to the mutational rule allowed. Hence it eventually belongs to a different affinity class than its mother cell. Clearly, it also happens that a B-cell stays in the GC without dying nor dividing. Finally, with rate $r_s \in [0,1]$ each B-cell can

be submitted to selection, which is made according to its affinity with $\overline{\mathbf{x}}$. A threshold $\overline{a}_s$ is fixed: if the B-cell belongs to an affinity class with index greater than $\overline{a}_s$, the B-cell dies. Otherwise, the B-cell exits the GC pool and reaches the selected pool. Therefore, for any GC B-cell and at any generation, we have:

$$\begin{cases} \text{Probability of cellular apoptosis:} & \mathbb{P}(\text{death}) & = r_d \\ \text{Probability of cellular division:} & \mathbb{P}(\text{division}) & = r_{div} \\ \text{Probability of selection challenge:} & \mathbb{P}(\text{selection}) & = r_s \end{cases}$$

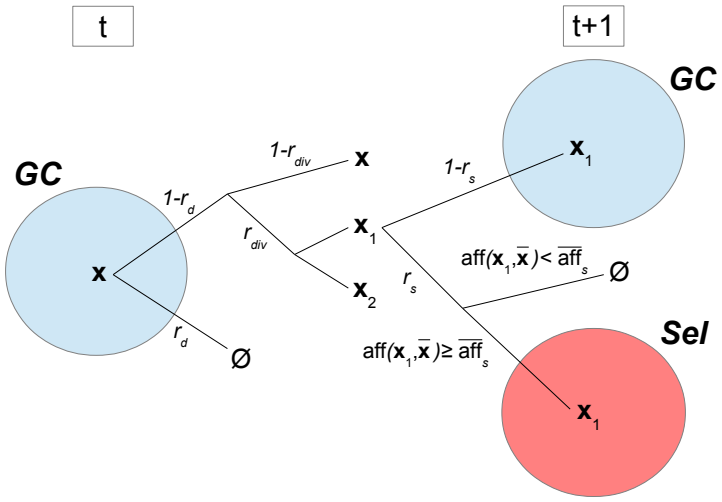


Figure 1: Schematic representation of model described by Definition 3. Here we denote by $\overline{\text{aff}}_s := N - \overline{a}_s$, the fitness of each B-cell in the affinity class whose index is $\overline{a}_s$ (see Definitions 1 and 2).

Once the GC reaction is fully established ($\sim$ day 7 after immunization), it is polarized into two compartments, named Dark Zone (DZ) and Light Zone (LZ) respectively. The DZ is characterized by densly packed dividing B-cells, while the LZ is less densely populated and contains FDCs and Tfh cells. The LZ is the preferential zone for selection [10]. The transition of B-cells from the DZ to the LZ seems to be determined by a timed cellular program: over a 6 hours period about 50% of DZ B-cells transit to the LZ, where they compete for positive selection signaling [6,36].

Through the entire paper one should keep in mind the following main modeling assumptions:

Modeling assumption 1 In our simplified mathematical model we do not take into account any spatial factor and in a single time step a GC B-cell can eventually undergo both division (with mutation) and selection. Hence the time unit has to be chosen big enough to take into account both mechanisms.

Modeling assumption 2 In this paper we are considering discrete-time models. The symbol t always denote a discrete time step, hence it is an integral value. We will refer to t as time, generation, or even maturation cycle to further stress the fact that in a single time interval $[t, t+1]$ each B-cell within the GC population is allowed to perform a complete cycle of division, mutation and selection.

Modeling assumption 3 Throughout the entire paper, when we talk about death rate (respectively division rate or selection rate) we are referring to the probability that each cell has of dying (respectively dividing or being submitted to selection) in a single time step.

3 Results

In this Section we formalize mathematically the model introduced above. This enables the estimation of various qualitative and quantitative measures of the GC evolution and of the selected pool as well. In Section 3.1 we show that a simple GW process describes the evolution of the size of the GC and determine a condition for its extinction. In order to do this we do not need to know the mutational model. Nevertheless, if we want to understand deeply the whole reaction we need to consider a $(N+3)$-type GW process, which we introduce in Section 3.2. Therefore we determine explicitly other quantities, such as the average affinity in the GC and the selected pool, or the evolution of the size of the latter. We conclude this section by numerical simulations (Section 3.4).

3.1 Evolution of the GC size

The aim of this section is to estimate the evolution of the GC size and its extinction probability. In order to do so we define a simple GW process, with respect to parameters r_d, r_{div} and r_s. Indeed, each B-cell submitted to selection exits the GC pool, independently from its affinity with $\overline{\mathbf{x}}$. Hence we apply some classical results about generating functions and GW processes ([14], Chapter I), which we recall in Appendix A. Proposition 1 gives explicitly the expected size of the GC at time t and conditions for the extinction of the GC.

Definition 4 Let $Z_t^{(z_0)}$, $t \geq 0$ be the random variable (rv) describing the GC-population size at time t, starting from $z_0 \geq 1$ initial B-cells. $(Z_t^{(z_0)})_{t\in\mathbb{N}}$ is a Markov Chain (MC) - since each cell behaves independently from the others and from previous generations - on $\{0, 1, 2, \dots\}$.

If $z_0 = 1$ and there is no confusion, we denote $Z_t := Z_t^{(1)}$. By Definition 4, Z_1 corresponds to the number of cells in the GC at the first generation, starting from a single seed cell. Thanks to Definition 3 one can claim that $Z_1 \in \{0,1,2\}$, with the following probabilities:

$$\begin{cases} p_0 := \mathbb{P}(Z_1 = 0) = r_d + (1-r_d)r_s(1-r_{div}+r_{div}r_s) \\ p_1 := \mathbb{P}(Z_1 = 1) = (1-r_d)(1-r_s)(1-r_{div}+2r_{div}r_s) \\ p_2 := \mathbb{P}(Z_1 = 2) = r_{div}(1-r_d)(1-r_s)^2 \end{cases} \tag{1}$$

As far as next generations are concerned, conditioning to $Z_t = k$, *i.e.* at generation t there are k B-cells in the GC, Z_{t+1} is distributed as the sum of k independent copies of Z_1: $\mathbb{P}(Z_{t+1} = k' \,|\, Z_t = k) = \mathbb{P}\left(\sum_{i=1}^{k} Z_1 = k'\right)$.

Equalities in (1) are derived by identifying the events leading to 0, 1 or 2 offspring in the GC coming from a single clone. Since these events are independent and disjoint, the result follows. For instance there will be 0 new individuals in the GC if either the mother cell dies, or it does not die, does not divide and is submitted to selection, or it does not die, it does divide and both daughter cells are submitted to selection :

$$\begin{aligned} \mathbb{P}(Z_1 = 0) &= \mathbb{P}\Big(\text{death} \cup (\text{death}^C \cap \text{division}^C \cap \text{selection}) \cup (\text{death}^C \cap \text{division} \cap \text{selection} \cap \text{selection})\Big) \\ &= \mathbb{P}(\text{death}) + \mathbb{P}(\text{death}^C)\mathbb{P}(\text{division}^C)\mathbb{P}(\text{selection}) + \mathbb{P}(\text{death}^C)\mathbb{P}(\text{division})\mathbb{P}(\text{selection})^2 \\ &= r_d + (1-r_d)(1-r_{div})r_s + (1-r_d)r_{div}r_s^2 \end{aligned}$$

We have denoted by A^C the complement of A. Expressions for p_1 and p_2 are obtained proceeding as before.

Definition 5 Let X be an integer valued rv, $p_k := \mathbb{P}(X = k)$ for all $k \geq 0$. Its probability generating function (pgf) is given by:

$$F_X(s) = \sum_{k=0}^{+\infty} p_k s^k$$

The pgf for Z_1:

$$\begin{aligned} F(s) &= p_0 + p_1 s + p_2 s^2 \\ &= r_d + (1-r_d)r_s(1-r_{div}+r_{div}r_s) \\ &+ (1-r_d)(1-r_s)(1-r_{div}+2r_{div}r_s)s + r_{div}(1-r_d)(1-r_s)^2 s^2 \end{aligned} \tag{2}$$

By using classical results on Galton-Watson processes (see Appendix A), one can prove:

Proposition 1

(i) The expected size of the GC at time t and starting from z_0 initial B-cells is given by:

$$\mathbb{E}(Z_t^{(z_0)}) = z_0\left((1-r_d)(1+r_{div})(1-r_s)\right)^t \tag{3}$$

(ii) Denoted by η_{z_0} the extinction probability of the GC population starting from z_0 initial B-cells, one has:

- *if* $\mathbb{E}(Z_1^{(1)}) \leq 1 \Leftrightarrow r_s \geq 1 - \dfrac{1}{(1-r_d)(1+r_{div})}$, *then* $\eta_{z_0} = 1$*: the process is subcritical*
- *otherwise* $\eta_{z_0} = \eta^{z_0} < 1$, η *being the smallest fixed point of* (2)*: the process is supercritical*

In particular, the initial number of seed cells z_0 does not affect the criticality of the process. Nevertheless, in the supercritical case, increasing the number of seed B-cells at the beginning of the process makes the probability of extinction decrease. More precisely, in the case $\eta < 1$, then $\eta_{z_0} \to 0$ if $z_0 \to \infty$, but we recall that GCs seem to be typically seeded by few B-cells, varying from ten to hundreds [34].

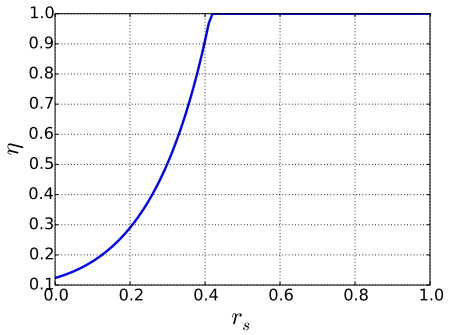


Figure 2: Numerical estimation of the extinction probability η of the GC with respect to r_s for $r_d = 0.1$ and $r_{div} = 0.9$.

This section shows that a classical use of a simple GW process enables to understand quantitatively the GC growth. Moreover, Proposition 1 (ii) gives a condition on the main parameters for the extinction of the GC: if the selection pressure is too high, with probability 1 the GC size goes to 0, independently from the initial number of seed cells. Intuitively, a too high selection pressure prevents those B-cells with bad affinity to improve their fitness undergoing further rounds of mutation and division. Most B-cells will be rapidly submitted to selection, hence either exit the GC as output cells or die by apoptosis if they fail to receive positive selection signals [20]. In Figure 2 we plot the extinction probability of a GC initiated from a single seed cell as a function of r_s (r_d and r_{div} are fixed), in order to stress the presence of a threshold effect of the selection probability over the extinction probability. The extinction probability of the GC process can give us some further insights on factors which are potentially involved in determining the success or failure of a GC reaction. This simplified mathematical model suggests that if the

selection pressure is too high compared to the division rate (*c.f.* due to Tfh signals in the LZ), the GC will collapse with probability 1, preventing the generation of high affinity antibodies against the presented antigen, hence an efficient immune response.

3.2 Evolution of the size and fitness of GC and selected pools

The GW process defined in the previous Section only describes the size of the GC. Indeed, we are not able to say anything about the average fitness of GC clones, or the expected number of selected B-cells, or their average affinity. Hence, we need to consider a more complex model and take into account the threshold for positive selection $\overline{a}_s$, and the transition probability matrix characterizing the mutational rule. Indeed, the mutational process is described as a Random Walk (RW) on the state space $\{0,\dots,N\}$ of affinity classes indices. The mutational rule reflects the edge set associated to the state-space $\{0,\dots,N\}$: this is given by a transition probability matrix.

Definition 6 Let $(\mathbf{X}_t)_{t\geq 0}$ be a RW on the state-space of B-cell traits describing a pure mutational process of a B-cell during the GC reaction. We denote by $\mathcal{Q}_N = (q_{ij})_{0\leq i,j\leq N}$ the transition probability matrix over $\{0,\dots,N\}$ which gives the probability of passing from an affinity class to another during the given mutational model. For all $0 \leq i,j \leq N$:

$$q_{ij} = \mathbb{P}(a_{\overline{\mathbf{x}}}(\mathbf{X}_{t+1}) = j \,|\, a_{\overline{\mathbf{x}}}(\mathbf{X}_t) = i)$$

From a biological point of view, these probabilities could be obtained *e.g.* by identifying which key mutations are the most relevant in determining changes in the fitness of a clone to a specific antigen and at which frequency they are produced.

We introduce a multi-type GW Process (see for instance [3], chapter V).

Definition 7 Let $\mathbf{Z}_t^{(\mathbf{i})} = (Z_{t,0}^{(\mathbf{i})},\dots,Z_{t,N+2}^{(\mathbf{i})})$, $t \geq 0$ be a MC where for all $0 \leq j \leq N$, $Z_{t,j}^{(\mathbf{i})}$ describes the number of GC B-cells belonging to the j^{th}-affinity class with respect to $\overline{\mathbf{x}}$, $Z_{t,N+1}^{(\mathbf{i})}$ the number of selected B-cells and $Z_{t,N+2}^{(\mathbf{i})}$ the number of dead B-cells at generation t, when the process is initiated in state $\mathbf{i} = (i_0,\dots,i_N,0,0)$.

Let $m_{ij} := \mathbb{E}[Z_{1,j}^{(i)}]$ the expected number of offspring of type j of a cell of type i in one generation. We collect all m_{ij} in a matrix, $\mathcal{M} = (m_{ij})_{0\leq i,j\leq N+2}$. We have:

$$\mathbb{E}[\mathbf{Z}_t^{(\mathbf{i})}] = \mathbf{i}\mathcal{M}^t \tag{4}$$

Supposing matrix $\mathcal{Q}_N$ given (Definition 6), describing the probability to switch from one affinity class to another thanks to a single mutation event, one can explicitly derive the elements of $\mathcal{M}$.

Proposition 2 *$\mathcal{M}$ is a $(N+3)\times(N+3)$ matrix defined as a block matrix:*

$$\mathcal{M} = \begin{pmatrix} \mathcal{M}_1 & \mathcal{M}_2 \\ \mathbf{0}_{2\times(N+1)} & \mathcal{I}_2 \end{pmatrix}$$

Where:

- $\mathbf{0}_{2\times(N+1)}$ *is a* $2\times(N+1)$ *matrix with all entries 0;*
- $\mathcal{I}_n$ *is the identity matrix of size* n*;*
- $\mathcal{M}_1 = 2(1-r_d)r_{div}(1-r_s)\mathcal{Q}_N + (1-r_d)(1-r_{div})(1-r_s)\mathcal{I}_{N+1}$
- $\mathcal{M}_2 = (m_{2,ij})$ *is a* $(N+1)\times 2$ *matrix where for all* $i \in \{0,\dots,N\}$*:*
  - *if* $i \leq \overline{a}_s$*:*

$$m_{2,i1} = (1-r_d)(1-r_{div})r_s + 2(1-r_d)r_{div}r_s \sum_{j=0}^{\overline{a}_s} q_{ij},$$

$$m_{2,i2} = r_d + 2(1-r_d)r_{div}r_s \sum_{j=\overline{a}_s+1}^{N} q_{ij}$$

  - *if* $i > \overline{a}_s$*:*

$$m_{2,i1} = 2(1-r_d)r_{div}r_s \sum_{j=0}^{\overline{a}_s} q_{ij},$$

$$m_{2,i2} = r_d + (1-r_d)(1-r_{div})r_s + 2(1-r_d)r_{div}r_s \sum_{j=\overline{a}_s+1}^{N} q_{ij}$$

The proof of Proposition 2 is available in Appendix B. It is based on the computation of the probability generating function of $\mathbf{Z}_1$.

Remark 1 Independently from the given mutational model, the expected number of selected or dead B-cells that each GC B-cell can produce in a single time step is given by $\alpha := r_d + (1-r_d)(1+r_{div})r_s$. All rows of $\mathcal{M}_2$ sum to α independently from the probability that each clone submitted to selection has of being positive selected, which we recall is 1 if it belongs to the i^{th} affinity class, $i \leq \overline{a}_s$, zero otherwise.

Of course in the multi-type context we recover again results from Section 3.1, such as the extinction probability of the GC (detailed in Appendix C).

In order to determine the expected number of selected cells at a given time t, we need to introduce another multi-type GW process.

Definition 8 Let $\widetilde{\mathbf{Z}}_t^{(\mathbf{i})} = (\widetilde{Z}_{t,0}^{(\mathbf{i})},\dots,\widetilde{Z}_{t,N+2}^{(\mathbf{i})})$, $t \geq 0$ be a MC where for all $0 \leq j \leq N$, $\widetilde{Z}_{t,j}^{(\mathbf{i})}$ describes the number of GC B-cells belonging to the j^{th}-affinity class with respect to $\overline{\mathbf{x}}$, $\widetilde{Z}_{t,N+1}^{(\mathbf{i})}$ the number of selected B-cells and $\widetilde{Z}_{t,N+2}^{(\mathbf{i})}$ the number of dead B-cells at generation t, when the process is initiated in state $\mathbf{i} = (i_0,\dots,i_N,0,0)$ and before the selection mechanism is performed for the t^{th}-generation.

Proceeding as we did for $\mathbf{Z}_t^{(\mathbf{i})}$, we can determine a matrix $\widetilde{\mathcal{M}}$ whose elements are $\widetilde{m}_{ij} := \mathbb{E}[\widetilde{Z}_{1,j}^{(i)}]$ for all $i, j \in \{0,\dots,N+2\}$.

Proposition 3 *$\widetilde{\mathcal{M}}$ is a $(N+3)\times(N+3)$ matrix, which only depends on matrix $\mathcal{Q}_N$, r_d and r_{div} and can be defined as a block matrix as follows:*

$$\widetilde{\mathcal{M}} = \begin{pmatrix} \widetilde{\mathcal{M}}_1 & \widetilde{\mathcal{M}}_2 \\ \mathbf{0}_{2\times(N+1)} & \mathcal{I}_2 \end{pmatrix}$$

Where:

- $\widetilde{\mathcal{M}}_1 = 2(1-r_d)r_{div}\mathcal{Q}_N + (1-r_d)(1-r_{div})\mathcal{I}_{N+1}$
- $\widetilde{\mathcal{M}}_2 = (\mathbf{0}_{N+1}, r_d\cdot\mathbf{1}_{N+1})$, *where $\mathbf{0}_{N+1}$ (resp. $\mathbf{1}_{N+1}$) is a $(N+1)$-column vector whose elements are all 0 (resp. 1).*

One could prove that:

$$\mathbb{E}\left[\widetilde{\mathbf{Z}}_t^{(\mathbf{i})}\right] = \mathbf{i}\mathcal{M}^{t-1}\widetilde{\mathcal{M}} \tag{5}$$

Proposition 4 *Let $\mathbf{i}$ be the initial state, $|\mathbf{i}|$ its 1-norm ($|\mathbf{i}| := \sum_{j=0}^{N+2}\mathbf{i}_j$).*

- *The expected size of the GC at time t:*

$$\sum_{k=0}^{N}(\mathbf{i}\mathcal{M}^t)_k \left(= |\mathbf{i}|\left((1-r_d)(1+r_{div})(1-r_s)\right)^t\right) \tag{6}$$

- *The average affinity in the GC at time t:*

$$\frac{\sum_{k=0}^{N}(N-k)(\mathbf{i}\mathcal{M}^t)_k}{\sum_{k=0}^{N}(\mathbf{i}\mathcal{M}^t)_k} \tag{7}$$

- *Let S_t, $t\geq 1$ denotes the random variable describing the number of selected B-cells at time t. By hypothesis $S_0 = 0$. $(S_t)_{t\in\mathbb{N}}$ is a MC on $\{0,1,2,\dots\}$. The expected number of selected B-cells at time t, $t\geq 1$:*

$$\mathbb{E}(S_t) = r_s\sum_{k=0}^{\overline{a}_s}\left(\mathbf{i}\mathcal{M}^{t-1}\widetilde{\mathcal{M}}\right)_k \tag{8}$$

- *The expected number of selected B-cells produced until time t:*

$$\mathbb{E}\left[\sum_{n=0}^{t}S_n\right] = \mathbb{E}\left[\left(\mathbf{Z}_t^{(\mathbf{i})}\right)_{N+1}\right] = \left(\mathbf{i}\mathcal{M}^t\right)_{N+1} \tag{9}$$

– *The average affinity of selected B-cells at time t, $t \geq 1$:*

$$\frac{\sum_{k=0}^{\overline{a}_s}(N-k)\left(\mathbf{i}\mathcal{M}^{t-1}\widetilde{\mathcal{M}}\right)_k}{\sum_{k=0}^{\overline{a}_s}\left(\mathbf{i}\mathcal{M}^{t-1}\widetilde{\mathcal{M}}\right)_k} \tag{10}$$

– *The average affinity of selected B-cells until time t:*

$$\frac{r_s\sum_{n=1}^{t}\sum_{k=0}^{\overline{a}_s}(N-k)\left(\mathbf{i}\mathcal{M}^{n-1}\widetilde{\mathcal{M}}\right)_k}{\left(\mathbf{i}\mathcal{M}^{t}\right)_{N+1}} \tag{11}$$

Proof Equations (6) and (9) are a direct application of what stated in Equation (17). Indeed, Equation (17) states that $\mathbf{i}\mathcal{M}^t$ contains the expectation of the number of all types cells at generation t when the process is started in $\mathbf{i}$. Hence the expectation of the size of the GC at the t^{th} generation is given by $\sum_{k=0}^{N}(\mathbf{i}\mathcal{M}^t)_k$, since the GC at generation t contains all alive non-selected B-cells, irrespectively from their affinity. Similarly, the expected number of selected B-cells untill time t (9) corresponds to the expectation of the $(N+1)^{\text{th}}$-type cell, $\left(\mathbf{i}\mathcal{M}^t\right)_{N+1}$.

The proof of Equation (8) is based on Equation (5), which allows to estimate the number of GC B-cells at generation t which are susceptible of being challenged by selection. One can remark that the expected number of selected B-cells at time t is obtained from the expected number of B-cells in GC at time t (before the selection mechanism is performed) having fitness good enough to be positive selected. This is given by $\sum_{k=0}^{\overline{a}_s}\left(\mathbf{i}\mathcal{M}^{t-1}\widetilde{\mathcal{M}}\right)_k$, thanks to (5). The result follows by multiplying this expectation by the probability that each of these B-cells is submitted to mutation, *i.e.* r_s. Finally, results about the average affinity in both the GC and the selected pool (Equations (7), (10) and (11)) are obtained from the previous ones (*c.f.* (6), (8) and (9)) by multiplying the number of individuals belonging to the same class by their fitness (Definition 2), and dividing by the total number of individuals in the considered pool. The definition of affinity as a function of the affinity classes, determines Equations (7), (10) and (11). Indeed, the affinity of the k^{th}-affinity class is given by $N-k$. □

Remark 2 The expected size of the GC at time t can be obtained applying a simple GW process (Section 3.1) and is given by (3). It is possible to prove the equality in brackets in Equation (6) starting from the $(N+3)$-type GW process. The interested reader can address to Appendix D for the detailed proof.

3.3 Optimal value of r_s maximizing the expected number of selected B-cells at time t

What is the behavior of the expected number of selected B-cells as a function of the model parameters? In particular, is there an optimal value of the selection rate which maximizes this number? In this section we show that, indeed, the answer is positive.

To do so we detail hereafter the computation of $\mathbb{E}(S_t)$ (Equation (8)), given by Proposition 4.

Let us suppose, for the sake of simplicity, that $\mathcal{Q}_N$ is diagonalizable:

$$\mathcal{Q}_N = R\Lambda_N L \ , \tag{12}$$

where $\Lambda_N = \mathrm{diag}(\lambda_0, \ldots, \lambda_N)$, and $R = (r_{ij})$ (resp. $L = (l_{ij})$) is the transition matrix whose rows (resp. lines) contain the right (resp. left) eigenvectors of $\mathcal{Q}_N$, corresponding to $\lambda_0, \ldots, \lambda_N$.

Proposition 5 *Let us suppose that at $t = 0$ there is a single B-cell entering the GC belonging to the i^{th}-affinity class with respect to the target cell. Moreover, let us suppose that $\mathcal{Q}_N = R\Lambda_N L$. For all $t \in \mathbb{N}$, the expected number of selected B-cells at time t, is:*

$$\mathbb{E}(S_t) = r_s(1-r_s)^{t-1}(1-r_d)^t \sum_{\ell=0}^{N} (2\lambda_\ell r_{div} + 1 - r_{div})^t \sum_{k=0}^{\overline{a}_s} r_{i\ell} l_{\ell k} \ ,$$

The proof of Proposition 5 is detailed in Appendix E.

As an immediate consequence of Proposition 5, we can claim:

Proposition 6 *For all $t^* \in \mathbb{N}$ fixed, the value $r_s^* := r_s(t^*)$ which maximizes the expected number of selected B-cells at the $t^{*\,th}$ maturation cycle is:*

$$r_s^* = \frac{1}{t^*}$$

Proof Since $(1-r_d)^t \sum_{\ell=0}^{N} (2\lambda_\ell r_{div} + 1 - r_{div})^t \sum_{k=0}^{\overline{a}_s} r_{i\ell} l_{\ell k}$ is a non negative quantity independent from r_s, the value of r_s which maximizes $\mathbb{E}(S_{t^*})$ is the one that maximizes $r_s(1-r_s)^{t^*-1}$. The result trivially follows. □

This result suggests that the selection rate in GCs is tightly related to the timing of the peak of a GC response, *i.e.* the timing corresponding to the maximal production of output cells (this timing can be determined *e.g.* by observing the concentration in blood of produced specific B-cells after infection or vaccination). In particular, following this model, GCs which peak early (*e.g.* for whom the maximal output cell production is reached in a few days) are

possibly characterized by a higher selection pressure than GCs peaking later. The peak of a typical GC reaction, measured as the average GC volume, has been estimated to be close to day 12 post immunization or a few days before [42], which is consistent with the observation of plasma cell response peak after immunization, *e.g.* [24]. Moreover, an high selection rate could also prevent a correct and efficient establishment of an immune response (*c.f.* results about extinction probability - Proposition 1). In addition, from a biological viewpoint, a too demanding selection pressure could avoid the generation of advantageous mutations, hence their fixation.

Remark 3 Under certain hypotheses about the mutational model and the GC evolution, one could justify the claim of Proposition 6 by heuristic arguments, without considering the $(N+3)$-type GW process. This leads to approximately estimate the expected number of selected B-cells at time t (Appendix F). Figure 4 (a) shows the peak of positive selected B-cells at generation t for a certain set of parameters.

3.4 Numerical simulations

We evaluate numerically results of Proposition 4. The $(N+3)$-type GW process allows a deeper understanding of the dynamics of both populations: inside the GC and in the selected pool. Through numerical simulations we emphasize the dependence of the quantities defined in Proposition 4 on parameters involved in the model.

In previous works [5,4] we have modeled B-cells and antigens as N-length binary strings, hence their traits correspond to elements of $\{0,1\}^N$. In this context we have characterized affinity using the Hamming distance between B-cell and antigen representing strings. The idea of using a N-dimensional shape space to represent antibodies traits and their affinity with respect to a specific antigen has already been employed (*e.g.* [28,22,17]), and N typically varies from 2 to 4. In the interests of simplification, we chose to set $N=2$. Moreover, from a biological viewpoint, this choice means that we classify the amino-acids composing B-cell receptors strings into 2 classes, which could represent amino-acids negatively and positively charged respectively. Charged and polar amino-acids are the most responsible in creating bonds which determine the antigen-antibody interaction [26].

While performing numerical simulations (Sections 3.4 and 4.2) we refer to the following transition probability matrix on $\{0,\dots,N\}$:

Definition 9 For all $i, j \in \{0,\dots,N\}$:

$$q_{ij} = \mathbb{P}(a_{\overline{\mathbf{x}}}(\mathbf{X}_{t+1}) = j \,|\, a_{\overline{\mathbf{x}}}(\mathbf{X}_t) = i) = \begin{cases} i/N & \text{if} \quad j = i-1 \\ (N-i)/N & \text{if} \quad j = i+1 \\ 0 & \text{if} \quad |j-i| \neq 1 \end{cases}$$

$\mathcal{Q}_N := (q_{ij})_{0\leq i,j\leq N}$ is a tridiagonal matrix where the main diagonal consists of zeros.

If we model B-cell traits as vertices of the state-space $\{0,1\}^N$, this corresponds to a model of simple point mutations (see [5] for more details and variants of this basic mutational model on binary strings).

Example 1 One can give explicitly the form of matrix $\mathcal{M}_2$ (Proposition 2) corresponding to the mutational model defined in Definition 9:

$$\mathcal{M}_2 = \begin{array}{c} 0 \\ \vdots \\ \overline{a}_s - 1 \\ \overline{a}_s \\ \overline{a}_s + 1 \\ \overline{a}_s + 2 \\ \vdots \\ N \end{array} \begin{pmatrix} \alpha & r_d \\ \vdots & \vdots \\ \alpha & r_d \\ \alpha - \beta + \beta\frac{\overline{a}_s}{N} & r_d + \beta\frac{N-\overline{a}_s}{N} \\ \beta\frac{\overline{a}_s+1}{N} & r_d + \alpha - \beta + \beta\frac{N-(\overline{a}_s+1)}{N} \\ 0 & r_d + \alpha \\ \vdots & \vdots \\ 0 & r_d + \alpha \end{pmatrix},$$

where:

- $\alpha := (1-r_d)(1+r_{div})r_s$
- $\beta := 2(1-r_d)r_{div}r_s$

Indeed, due to the particular form of matrix $\mathcal{Q}_N$ one has straightforward:

- $$\sum_{j=0}^{\overline{a}_s} q_{ij} = \begin{cases} 1 & \text{if } i < \overline{a}_s \\ \overline{a}_s/N & \text{if } i = \overline{a}_s \\ (\overline{a}_s+1)/N & \text{if } i = \overline{a}_s + 1 \\ 0 & \text{if } i > \overline{a}_s + 1 \end{cases}$$
- $$\sum_{j=\overline{a}_s+1}^{N} q_{ij} = \begin{cases} 0 & \text{if } i < \overline{a}_s \\ (N-\overline{a}_s)/N & \text{if } i = \overline{a}_s \\ (N-(\overline{a}_s+1))/N & \text{if } i = \overline{a}_s + 1 \\ 1 & \text{if } i > \overline{a}_s + 1 \end{cases}$$

Remark 4 Note that all mathematical results obtained in previous sections are independent from the mutation model defined in Definition 9.

We suppose that at the beginning of the process there is a single B-cell entering the GC belonging to the affinity class a_0. Of course, the model we set allows to simulate any possible initial condition. Indeed, by fixing the initial vector $\mathbf{i}$, we can decide to start the reaction with more B-cells, in different affinity classes. When it is not stated otherwise, the employed parameter set for simulations is given in Table 1.

Table 1: Parameter choice for simulations in Sections 3.4 (unless stated otherwise).

N	r_s	r_d	r_{div}	a_0	$\overline{a}_s$
10	0.1	0.1	0.9	3	3

We perform numerical simulations to better appreciate how the dynamics of the GC and positive selected clones populations are related and evolve depending on model parameters. In the case of a subcritical GC, by model definition selected clones stabilize at a given level once the GC becomes extinct. Hence we conveniently chose a parameter set (Table 1) which implies a supercritical GC (Proposition 1): with great probability the simulated GC goes through explosion, and so the selected population does.

3.4.1 Evolution of the GC population

The evolution of the size of the GC can be studied by using the simple GW process defined in Section 3.1. Equation (3), in the case of a single initial B-cell, evidences that the expected number of B-cells within the GC for this model only depends on r_d, r_{div} and r_s and it is not driven by the initial affinity, nor by the threshold chosen for positive selection $\overline{a}_s$, nor by the mutational rule.

Equation (3) evidences that, independently from the transition probability matrix defining the mutational mechanism, the GC size at time t increases with r_{div} and decreases for increasing r_s and r_d. Moreover, the impact of these last two parameters is the same for the growth of the GC. One could expect this behavior since the effect of both the death and the selection on a B-cell is the exit from the GC.

In order to study the evolution of the average affinity within the GC, we need to refer to the $(N+3)$-type GW process defined in Section 3.2.

Proposition 7 *Let us suppose that $\mathcal{Q}_N = R\Lambda_N L$. The average affinity within the GC at time t, starting from a single B-cell belonging to the i^{th}-affinity class with respect to $\overline{\mathbf{x}}$ is given by:*

$$N - \frac{\sum_{\ell=0}^{N}(2\lambda_\ell r_{div} + 1 - r_{div})^t \sum_{k=0}^{N} k \cdot r_{i\ell} l_{\ell k}}{(1+r_{div})^t} \ ,$$

Proof It follows directly from Equations (7) and by considering the eigendecomposition of matrix $\mathcal{Q}$. One has to consider the expression of the t^{th} power of matrix $\mathcal{M}$ (which can be obtained recursively, see Appendix E): one can

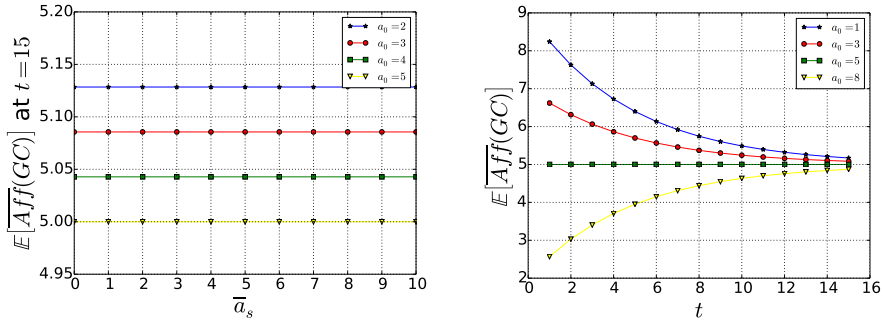


Figure 3: (a) Dependence of the expected average affinity in the GC on $\overline{a}_s$ at time $t=15$, for different values of a_0. The average affinity in the GC is constant with respect to $\overline{a}_s$. (b) The evolution during time of the expected average affinity in the GC for different values of a_0. The average affinity converges through $N/2$, due to the stationary distribution of Q_N, the binomial probability distribution.

prove that the first $N+1$ components of the i^{th}-row of matrix $\mathcal{M}^t$ are the elements of the i^{th}-row of matrix RD^tL, where $D=2(1-r_d)r_{div}(1-r_s)\Lambda_N+(1-r_d)(1-r_{div})(1-r_s)\mathcal{I}_{N+1}$ is a diagonal matrix. □

It is obvious from Proposition 7 that this quantity only depends on the initial affinity with the target trait, the transition probability matrix $\mathcal{Q}_N$ and the division rate r_{div}. The average affinity within the GC does not depend on $\overline{a}_s$ (as one can clearly see in Figure 3 (a)), nor by r_s or r_d. One can intuitively understand this behavior: independently from their fitness, all B-cells submitted to mutation exit the GC. Moreover, r_s and r_d impact the GC size, but not its average affinity, as selection and death affect all individuals of the GC independently from their fitness.

It can be interesting to observe the evolution of the expected average affinity within the GC during time. Numerical simulations of our model show that the expected average affinity in the GC converges through $N/2$, independently from the affinity of the first naive B-cell (Figure 3 (b)). This depends on the mutational model we choose for these simulations. Indeed, providing that the GC is in a situation of explosion, for t big enough the distribution of GC clones within the affinity classes is governed by the stationary distribution of matrix $\mathcal{Q}_N$. Since for $\mathcal{Q}_N$ given by Definition 9 one can prove that the stationary distribution over $\{0,\dots,N\}$ is the binomial probability distribution [5], the average affinity within the GC will quickly stabilizes at a value of $N/2$. Note that for these simulations we chose $N=10$, hence affinities are in the range $[0,10]$.

3.4.2 Evolution of the selected pool

The evolution of the number of selected B-cells during time necessarily depends on the evolution of the GC. In particular, let us suppose we are in the supercritical case, *i.e.* the extinction probability of the GC is strictly smaller than 1. Than, with positive probability, the GC explodes and so does the selected pool. On the other hand, if the GC extinguishes, the number of selected B-cells will stabilize at a constant value, as once a B-cell is selected it can only stay unchanged in the selected pool.

As demonstrated in Section 3.3, there exists an optimal value of the parameter r_s which maximizes the expected number of selected B-cells at time t. Figure 4 (a) evidences this fact. Moreover, as expected, simulations show that the expected size of selected B-cells at a given time t increases with the threshold $\overline{a}_s$ chosen for positive selection (Figure 4 (b)). This is a consequence of Proposition 5: $\overline{a}_s$ determines the number of elements of the sum $\sum_{k=0}^{\overline{a}_s} r_{i\ell} l_{\ell k}$.

Figure 4 (c) underlines the correspondence between theoretical results given by Proposition 4 and numerical values obtained by simulating the evolutionary process described by Definition 3. In particular Figure 4 (c) shows the expected (resp. average) number of selected B-cells produced until time $t = 15$ depending on the threshold chosen for positive selection, $\overline{a}_s$.

Remark 5 We recall that values expressed on y-axes of all graphs in Figure 4 (and later in Figures 8 to 10) describe the expected number of some groups of B-cells (*e.g.* GC B-cells, output B-cells) generated at a given time step or after a given number of maturation cycles. Henceforth this is an adimensional number. It is of course envisageable to translate these values into concentrations of some specific B-cell phenotypes into *e.g.* blood or tissue samples in order to interpret theoretical results and compare them to biological data.

4 Extensions of the model

Proceeding as in Section 3.2, we can define and study many different models of affinity-dependent selection. Here we propose a model in which we perform only positive selection and a model reflecting a Darwinian evolutionary system, in which the selection is only negative. For the latter, we will take into account only $N + 2$ types instead of $N + 3$: we do not have to consider a selected pool. Indeed the selected population remains in the GC. Here below we give the definitions of both models. In Section 4.1 we formalize these problems mathematically, then in Section 4.2 we show some numerical results.

4.1 Definitions and results

Let us consider the process described in Definition 3. We change only the selection mechanism.

Definition 10 (Positive selection) If a B-cell submitted to selection belongs to an affinity class with index greater than $\overline{a}_s$, nothing happens. Otherwise, the B-cell exits the GC pool and reaches the selected pool.

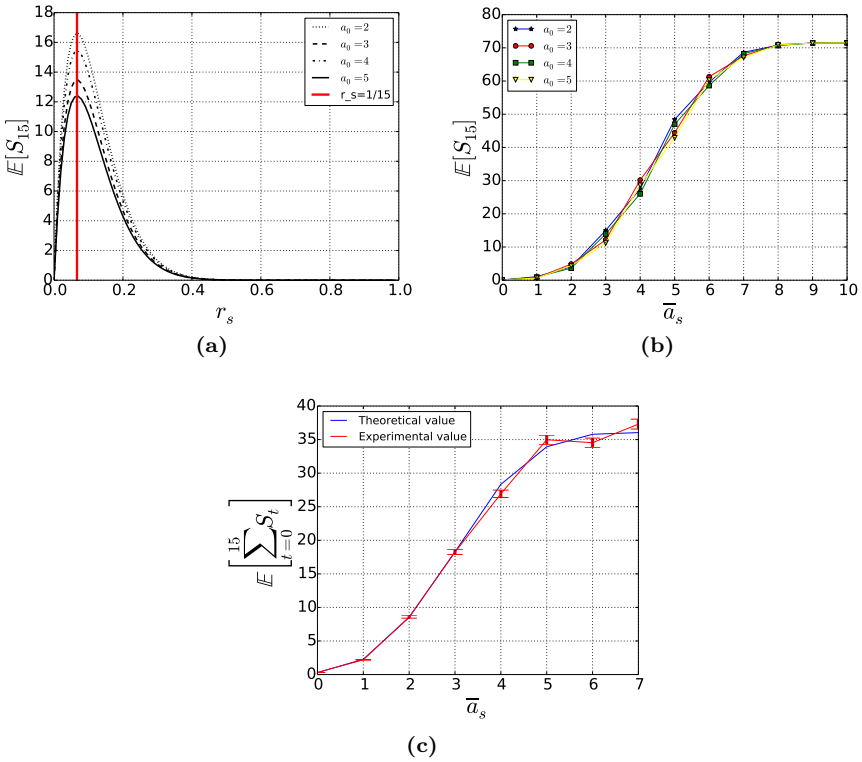
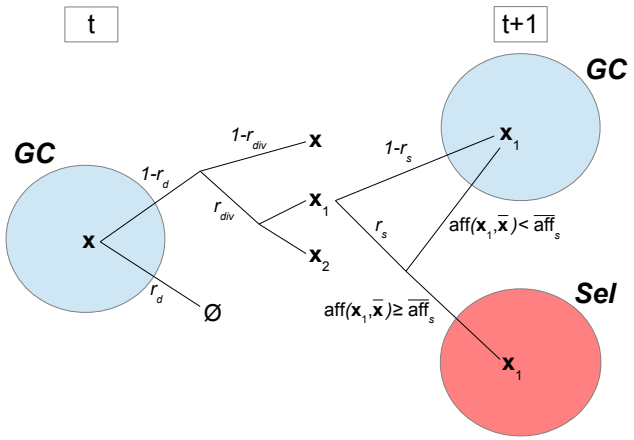
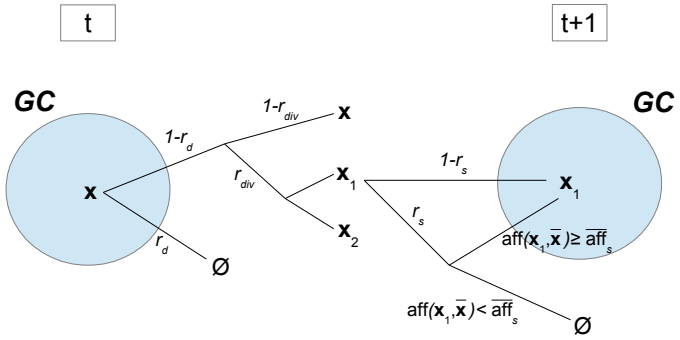


Figure 4: (a-b) Expected number of selected B-cells for the time step $t = 15$ for different values of a_0, depending on r_s and $\overline{a}_s$ respectively. There exists an optimal value of r_s maximizing the expected number of selected B-cells for a given generation. This value is independent from a_0 and is equal to $1/t$ as demonstrated in Proposition 6: the red vertical line in (a) corresponds to this value. (c) Comparison between the expected number of selected B-cells until time t given by evaluation of the theoretical formula (Equation (9)), and the empirical value obtained as the mean over 4000 simulations. Vertical bars denotes the corresponding estimated standard deviations. Here $N = 7$ and $r_s = 0.3$.



(a) Positive selection



(b) Negative selection

Figure 5: Schematic representations of models described (a) by Definitions 10 and (b) by Definitions 11 of exclusively positive (resp. exclusively negative) selection.

Definition 11 (Negative selection) If a B-cell submitted to selection belongs to an affinity class with index greater than $\overline{a}_s$, it dies. Otherwise, nothing happens.

In Figure 5 we represent schematically both processes of positive selection and of negative selection. It is clear from Figure 5 (b) that in the case of Definition 11 we do not need to consider the selected pool anymore.

Positive selection

Definition 12 Let $\mathbf{Z}_t^{+\,(\mathbf{i})} = (Z_{t,0}^{+\,(\mathbf{i})}, \ldots, Z_{t,N+2}^{+\,(\mathbf{i})})$, $t \geq 0$ be a MC where for all $0 \leq j \leq N$, $Z_{t,j}^{+\,(\mathbf{i})}$ describes the number of GC B-cells belonging to the j^{th}-affinity class with respect to $\overline{\mathbf{x}}$, $Z_{t,N+1}^{+\,(\mathbf{i})}$ the number of selected B-cells and $Z_{t,N+2}^{+\,(\mathbf{i})}$ the number of dead B-cells at generation t, when the process is initiated in state $\mathbf{i} = (i_0, \ldots, i_N, 0, 0)$, and following the evolutionary model described by Definition 10.

Let us denote by $\mathcal{M}^+ = (m_{ij}^+)_{0 \leq i,j \leq N+2}$ the matrix containing the expected number of type-j offspring of a type-i cell corresponding to the model defined by Definition 10. We can explicitly write the value of all m_{ij}^+ depending on r_d, r_{div}, r_s, and the elements of matrix $\mathcal{Q}_N$.

Proposition 8 *$\mathcal{M}^+$ is a $(N+3)^2$ matrix, which we can define as a block matrix in the following way:*

$$\mathcal{M}^+ = \begin{pmatrix} \mathcal{M}_1^+ & \mathcal{M}_2^+ \\ \mathbf{0}_{2\times(N+1)} & \mathcal{I}_2 \end{pmatrix}$$

Where:

- $\mathcal{M}_1^+ = (m_{1,ij}^+)$ *is a* $(N+1)^2$ *matrix. For all* $i \in \{0, \ldots, N\}$*:*
  - $\forall j \leq \overline{a}_s$*:* $m_{1,ij}^+ = 2(1-r_d)r_{div}(1-r_s)q_{ij} + (1-r_d)(1-r_{div})(1-r_s)\delta_{ij}$
  - $\forall j > \overline{a}_s$*:* $m_{1,ij}^+ = 2(1-r_d)r_{div}q_{ij} + (1-r_d)(1-r_{div})\delta_{ij}$

  where δ_{ij} *is the Kronecker delta.*
- $\mathcal{M}_2^+ = (m_{2,ij}^+)$ *is a* $(N+1)\times 2$ *matrix where for all* $i \in \{0, \ldots, N\}$, $m_{2,i1}^+ = m_{2,i1}$*, and* $m_{2,i2}^+ = r_d$*. We recall that* $m_{2,i1}$ *is the* i^{th}*-component of the first column of matrix* $\mathcal{M}_2$*, given in Proposition 2.*

Negative selection

Definition 13 Let $\mathbf{Z}_t^{-\,(\mathbf{i})} = (Z_{t,0}^{-\,(\mathbf{i})}, \ldots, Z_{t,N+1}^{-\,(\mathbf{i})})$, $t \geq 0$ be a MC where for all $0 \leq j \leq N$, $Z_{t,j}^{-\,(\mathbf{i})}$ describes the number of GC B-cells belonging to the j^{th}-affinity class with respect to $\overline{\mathbf{x}}$ and $Z_{t,N+1}^{-\,(\mathbf{i})}$ the number of dead B-cells at generation t, when the process is initiated in state $\mathbf{i} = (i_0, \ldots, i_N, 0)$, and following the evolutionary model described by Definition 11.

Let us denote by $\mathcal{M}^- = (m_{ij}^-)_{0 \leq i,j \leq N+1}$ the matrix containing the expected number of type-j offspring of a type-i cell corresponding to the model defined by Definition 13.

Proposition 9 *$\mathcal{M}^-$ is a $(N+2)^2$ matrix, which we can define as a block matrix in the following way:*

$$\mathcal{M}^- = \begin{pmatrix} \mathcal{M}_1^- & \mathbf{m}_2^- \\ \mathbf{0}'_{N+1} & 1 \end{pmatrix}$$

Where:

- *$\mathcal{M}_1^- = (m_{1,ij}^-)$ is a $(N+1)^2$ matrix. For all $i \in \{0,\dots,N\}$:*
  - *$\forall j \leq \overline{a}_s$: $m_{1,ij}^- = 2(1-r_d)r_{div}q_{ij} + (1-r_d)(1-r_{div})\delta_{ij}$*
  - *$\forall j > \overline{a}_s$: $m_{1,ij}^- = 2(1-r_d)r_{div}(1-r_s)q_{ij} + (1-r_d)(1-r_{div})(1-r_s)\delta_{ij}$*
- *$\mathbf{m}_2^-$ is a $(N+1)$ column vector s.t. for all $i \in \{0,\dots,N\}$ $m_i^+ = m_{2,i2}$, $m_{2,i2}$ being the i^{th}-component of the second column of matrix $\mathcal{M}_2$, given in Proposition 2.*
- *$\mathbf{0}'_{N+1}$ is a $(N+1)$ row vector composing of zeros.*

We do not prove Propositions 8 and 9, since the proofs are the same as for Proposition 2 (Appendix B).

Results stated in Proposition 4 hold true for these new models, by simply replacing matrix $\mathcal{M}$ with $\mathcal{M}^+$ (resp. $\mathcal{M}^-$). Of course, in the case of negative selection, as we do not consider the selected pool, we only refer to (6) and (7) quatifying the growth and average affinity of the GC. Matrix $\widetilde{\mathcal{M}}$ is the same for both models as only selection principles change.

Because of peculiar structures of matrices $\mathcal{M}^+$ and $\mathcal{M}^-$, we are not able to compute explicitly their spectra. Henceforth we can not give an explicit formula for the extinction probability or evaluate the optimal values of the selection rate r_s as we did in Sections 3.2 and 3.3.

Nevertheless, by using standard arguments for positive matrices, the greatest eigenvalue of both matrices $\mathcal{M}_1^+$ and $\mathcal{M}_1^-$ can be bounded, and hence give sufficient conditions for extinction. Indeed, form classical results about multi-type GW processes, the value of the greatest eigenvalue allows to discriminate between subcritical case (*i.e.* extinction probability equal to 1) and supercritical case (*i.e.* extinction probability strictly smaller than 1) [3].

Proposition 10 *Let $\mathbf{q}^+$ (resp. $\mathbf{q}^-$) be the extinction probability of the GC for the model corresponding to matrix $\mathcal{M}_1^+$ (resp. $\mathcal{M}_1^-$).*

- *If $r_{div} \leq \dfrac{r_d}{1-r_d}$, then $\mathbf{q}^+ = \mathbf{q}^- = \mathbf{1}$.*
- *If $r_s < 1 - \dfrac{1}{(1-r_d)(1+r_{div})}$, then $\mathbf{q}^+ < \mathbf{1}$ and $\mathbf{q}^- < \mathbf{1}$.*

Proof Since both matrices $\mathcal{M}_1^+$ and $\mathcal{M}_1^-$ are strictly positive matrices, the Perron Frobenius Theorem insures that the spectral radius is also the greatest eigenvalue. Then the following classical result holds [23]:

Theorem 1 *Let $A=(a_{ij})$ be a square nonnegative matrix with spectral radius $\rho(A)$ and let $r_i(A)$ denote the sum of the elements along the i^{th}-row of A. Then:*

$$\min_i r_i(A) \leq \rho(A) \leq \max_i r_i(A)$$

Simple calculations provide:

$$\min_i r_i(\mathcal{M}_1^+) = (1-r_d)(1+r_{div}) - r_s(1-r_d)\left(2r_{div}\min_i \sum_{j=0}^{\overline{a}_s} q_{ij} + 1 - r_{div}\right)$$

$$\max_i r_i(\mathcal{M}_1^+) = (1-r_d)(1+r_{div}) - 2r_s r_{div}(1-r_d)\max_i \sum_{j=0}^{\overline{a}_s} q_{ij}$$

$$\min_i r_i(\mathcal{M}_1^-) = (1-r_d)(1+r_{div}) - r_s(1-r_d)\left(2r_{div}\min_i \sum_{j=\overline{a}_s+1}^{N} q_{ij} + 1 - r_{div}\right)$$

$$\max_i r_i(\mathcal{M}_1^-) = (1-r_d)(1+r_{div}) - 2r_s r_{div}(1-r_d)\max_i \sum_{j=\overline{a}_s+1}^{N} q_{ij}$$

The result follows by observing that for all $i \in \{0,\ldots,N\}$, $0 \leq \sum_{j=0}^{\overline{a}_s} q_{ij}$, $\sum_{j=\overline{a}_s+1}^{N} q_{ij} \leq 1$, and applying Theorem 3. □

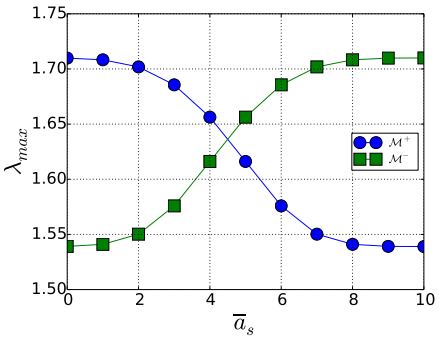


Figure 6: Dependence of greatest eigenvalues of matrices $\mathcal{M}^+$ (blue circles) and $\mathcal{M}^-$ (green squares) respectively on $\overline{a}_s$ for $N=10$, $r_{div}=0.9$, $r_d=r_s=0.1$. Hence $(1-r_d)(1+r_{div})(1-r_s)=1.539$ and $(1-r_d)(1+r_{div})=1.71$.

Remark 6 One can intuitively obtain the second claim of Proposition 10, as this condition over the parameters implies that the probability of extinction of the GC for the model underlined by matrix $\mathcal{M}_1$ of positive and negative selection is strictly smaller than 1 (Proposition 1). Indeed keeping the same

parameters for all models, the size of the GC for the model of positive and negative selection is smaller than the size of GCs corresponding to both models of only positive and only negative selection. Consequently if the GC corresponding to $\mathcal{M}$ has a positive probability of explosion, it will be necessarily the same for $\mathcal{M}^+$ and $\mathcal{M}^-$.

Remark 7 The values of both $\rho(\mathcal{M}_1^+)$ and $\rho(\mathcal{M}_1^-)$ depend on $\overline{a}_s$, varying from a minimum of $(1-r_d)(1+r_{div})(1-r_s)$ and a maximum of $(1-r_d)(1+r_{div})$. Figure 6 evidences the dependence on $\overline{a}_s$ of the spectral radius of $\mathcal{M}_1^+$ and $\mathcal{M}_1^-$, using matrix $\mathcal{Q}_N$ given by Definition 9 as transition probability matrix.

Remark 7 and Figure 6 evidences that, conversely to the previous case of positive and negative selection, in both cases of exclusively positive (resp. exclusively negative) selection the parameter $\overline{a}_s$ plays an important role in the GC dynamics, affecting its extinction probability. In particular, keeping unchanged all other parameters, if $\overline{a}_s \to N$ (resp. $\overline{a}_s \to 0$), then $\rho(\mathcal{M}_1^+)$ (resp. $\rho(\mathcal{M}_1^-)) \to (1-r_d)(1+r_{div})(1-r_s)$, which implies $\mathbf{q}^+$ (resp. $\mathbf{q}^-) \to \mathbf{1}$. From a biological viewpoint we expect that the GC dynamics should be influenced by the threshold required for selection. B-cell affinity determines the ability of a B-cell to internalize antigen, and present it to Tfh cells to receive appropriate rescue signals. Experimental evidence indicates that B-cell affinity is extremely important to determine differential decision in GCs, *i.e.* if a B-cell submitted to selection is committed to become either a plasma cell or a memory B-cell, recycle back to the dark zone to perform further rounds of somatic hypermutations, or die [16].

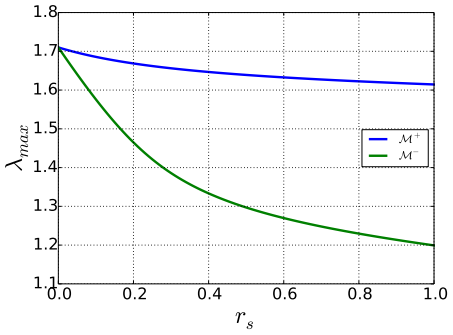


Figure 7: Dependence of greatest eigenvalues of matrices $\mathcal{M}^+$ (blue) and $\mathcal{M}^-$ (green) respectively on r_s for $N = 10$, $r_{div} = 0.9$, $r_d = 0.1$, $\overline{a}_s = 3$.

In Figure 7 we plot the dependence of greatest eigenvalues of both matrices $\mathcal{M}^+$ and $\mathcal{M}^-$ with respect to r_s. We fix $r_d = 0.1$ and $r_{div} = 0.9$ as for Figure 2. One can note that with this parameter set and if the threshold for positive selection $\overline{a}_s$ is chosen not "too small" nor "too large" with respect to N, then

the greatest eigenvalue for both matrices is always greater than 1 independently from r_s, *i.e.* the extinction probability is always strictly smaller than 1. From a biological viewpoint we expect that a phisiological threshold for positive selection should not be too strict nor to weak. Indeed, a too demanding threshold for positive selection is not optimal since B-cells should have gained an extremely high affinity in order to be positive selected, which would at least require too much time, avoiding a prompt immune response against the invading pathogen. On the other hand, a too weak threshold results in an unchallenging affinity maturation process: almost any B-cell would be positive selected, irrespective from its affinity level with respect to the presented antigen. This could also entail the generation of auto-reactive clones.

4.2 Numerical simulations

The evolution of GCs corresponding to matrices $\mathcal{M}^+$ and $\mathcal{M}^-$ respectively are complementary. Moreover, in both cases, keeping all parameters fixed one expects a faster expansion if compared to the model of positive and negative selection, since the selection acts only positively (resp. negatively) on good (resp. bad) clones. In particular, the model of negative selection corresponds to the case of 100% of recycling, meaning that all positively selected B-cells stay in the GC for further rounds of mutation, division and selection.

Figure 8 shows the dependence on $\overline{a}_s$ of the GC size and fitness, comparing $\mathcal{M}^+$ (left column) and $\mathcal{M}^-$ (right column). Indeed, for these models the GC dynamics depends on the selection threshold, conversely to the previous case of positive and negative selection, and not only on the selection rate. The effects of $\overline{a}_s$ on the GC are perfectly symmetric: it is interesting to observe that when both selection mechanisms are coupled, then $\overline{a}_s$ does not affect the GC dynamics anymore, as shown for instance in Figure 3 (a). Moreover, Figures 8 (c,d) evidence the existence of a value of $\overline{a}_s$ that minimizes (resp. maximizes) the expected average affinity in the GC for $\mathcal{M}^+$ (resp. $\mathcal{M}^-$). In both cases this value is approximately $N/2$. This certainly depends on the transition probability matrix chosen for the mutational model, which converges to a binomial probability distribution over $\{0,\dots,N\}$.

The evolution of the selected pool for the model of positive selection have some important differences if compared to the model described in Section 3. For instance, it is not easy to identify an optimal value of r_s which maximizes the expected number of selected B-cells at time t. Indeed it depends both on a_0 and $\overline{a}_s$: if $a_0 \leq \overline{a}_s$ we find curves similar to those plotted in Figure 4 (a), otherwise Figure 9 (a) shows a substantial different behavior. Indeed, if $a_0 > \overline{a}_s$, choosing a big value for r_s does not negatively affect the number of selected B-cells at time t. In this case, for the first time steps no (or a very few) B-cells will be positively selected, since they still need to improve their affinity to the

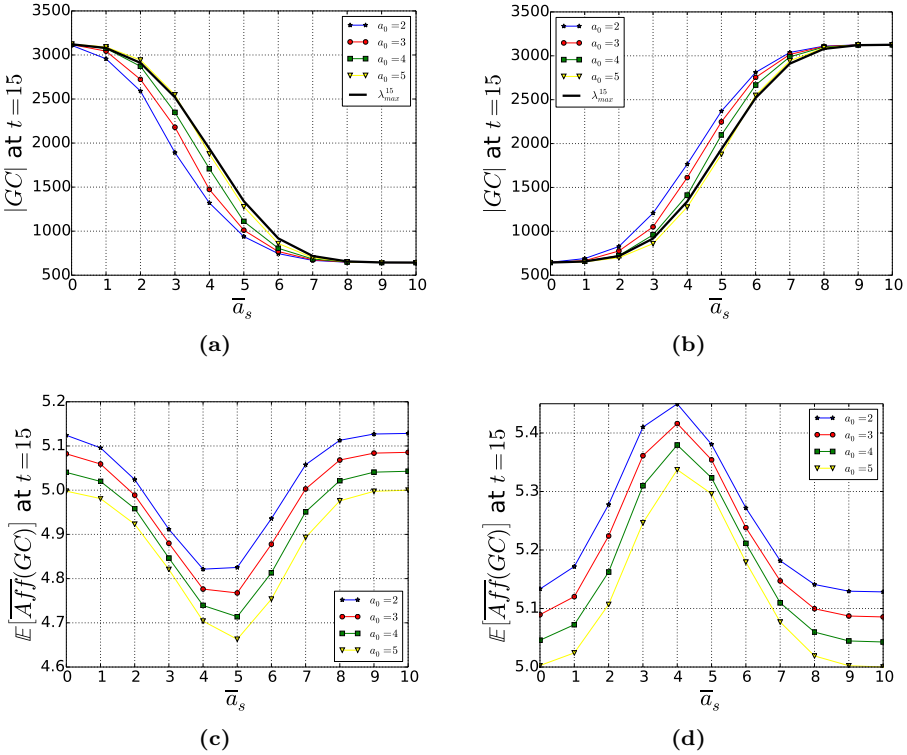


Figure 8: (a,b) Dependence of the expected size of the GC after 15 time steps on $\overline{a}_s$ for different values of a_0. The thick black line corresponds in both figures to the value of the greatest eigenvalue of matrices $\mathcal{M}_1^+$ and $\mathcal{M}_1^-$ respectively, raised to the power of $t = 15$ (see Figure 6). Note that thanks to Proposition 1 we know that for this parameter choice the expected size of the GC for the model of positive and negative selection corresponds to $((1-r_d)(1+r_{div})(1-r_s))^{15}$, which is equivalently λ_{max}^{15} for $\overline{a}_s = 10$ in Figure 8 (a) or λ_{max}^{15} for $\overline{a}_s = 0$ in Figure 8 (b). (c,d) Dependence of the expected average affinity in the GC after $t = 15$ time steps on $\overline{a}_s$ for different values of a_0. The left column of Figure 8 refers to the model of positive selection, while the right column to the model of negative selection.

target. Therefore, they stay in the GC and continue to proliferate for next generations. This fact is further underlined in Figure 9 (b), where we estimate numerically the optimal r_s^* which maximizes the expected number of selected B-cells at time t. Simulations show that for $a_0 \leq \overline{a}_s$ the value of r_s^* for the model of positive selection is really close to the one obtained by Proposition 6. On the other hand if we start from an initial affinity class $a_0 > \overline{a}_s$ the result we obtain is substantially different from the previous one, especially for small

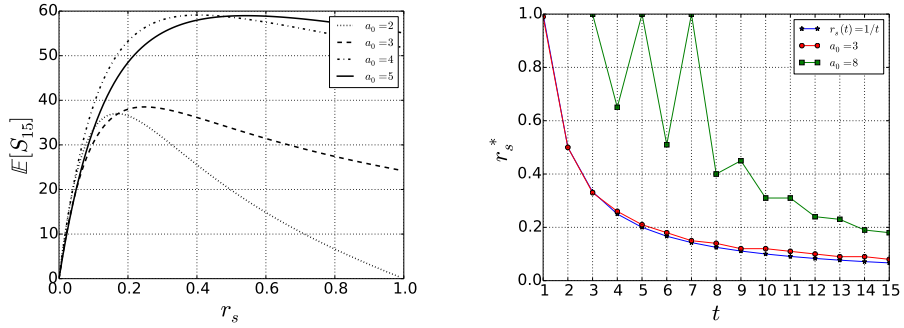
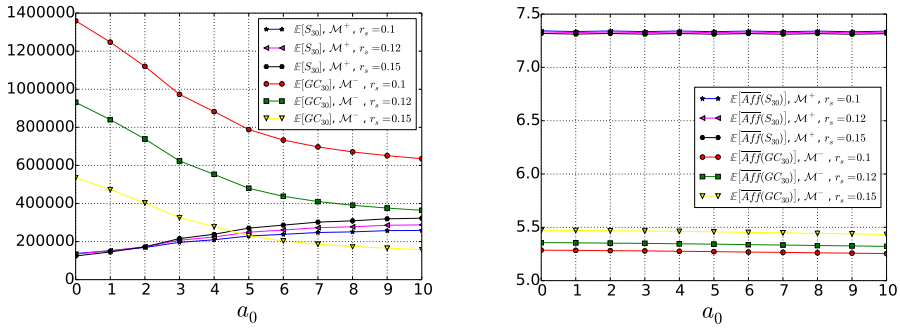


Figure 9: Model of positive selection. (a) Expected number of selected B-cells for the time step $t = 15$ for different values of a_0, depending on r_s. (b) Estimation of the optimal r_s^* maximizing the expected number of selected B-cells for a given generation, comparing the model of positive selection for different values of a_0 and the model described in Section 3 (we plot the exact value, $r_s(t) = 1/t$, as obtained by Proposition 6). In (b), for simulations corresponding to the model of positive selection we set $\overline{a}_s = 5$.

t. Moreover we observe important oscillations, which are probably due to the mutational model, and to the fact that the total GC size is still small for small t, since the process starts from a single B-cell. Nevertheless, it seems that for t big enough also in this case the value of r_s^* tends to approach $1/t$.

Since in the case of negative selection there is no selected pool, one can suppose that at a given time t the process stops and all clones in the GC pool exit the GC as selected clones. Hence it can be interesting to compare the selected pool of the model of positive selection and the GC pool of the model of negative selection at time t. Clearly to make these two compartments comparable, the main parameters of both systems have to be opportunely chosen. In Figure 10 we compare the size and average fitness of the selected pool for $\mathcal{M}^+$ and the GC for $\mathcal{M}^-$ at time $t = 30$. We test different values of the parameter r_s. In particular, we observe that increasing r_s the GC size for the model of negative selection decreases and its average fitness increases. For the parameter choices we made for these simulations, Figure 10 (a) shows that the size of the GC for $\mathcal{M}^-$ is comparable to the size of the selected pool for $\mathcal{M}^+$ at time $t = 30$ if, keeping all other parameters fixed, $r_s = 0.15$ for $\mathcal{M}^-$. Nevertheless, this does not implies a comparable value for the average affinity: the clones of the selected pool for $\mathcal{M}^+$ have a significantly greater average affinity than those of the GC for $\mathcal{M}^-$. In order to increase the average fitness in the GC for the model of negative selection one has to consider greater values for the parameter r_s, but this affects the probability of extinction of the process.



(a) **(b)**

Figure 10: (a) Expected number of B-cells which have been selected until time $t = 30$ for $\mathcal{M}^+$ compared to the expected size of the GC for $\mathcal{M}^-$ for different values of r_s. (b) Expected corresponding average affinity for the selected pool (case of positive selection) and the GC (case of negative selection). For some choice of the parameter r_s, the size of the selected pool for $\mathcal{M}^+$, and the GC for $\mathcal{M}^-$, are comparable. Nevertheless, the corresponding average affinities are significantly different.

We can expect this discrepancy between the average affinity for the selected pool for $\mathcal{M}^+$ and the one of the GC for $\mathcal{M}^-$. Indeed, in the first case we are looking to all those B-cells which have been positive selected, hence belong at most to the $\overline{a}_s^{\text{th}}$-affinity class. On the contrary in the case of $\mathcal{M}^-$, we consider the average affinity of all B-cells which are still alive in the GC at a given time step. Among these clones, if $r_s < 1$, with positive probability there are also individuals with affinity smaller than the one required for escaping negative selection. They remain in the GC because they have not been submitted to selection. These B-cells make the average affinity decrease. Of course r_s is not the only parameter affecting the quantities plotted in Figure 10. In particular, one can observe that choosing a greater value for $\overline{a}_s$ also have a significant effect over the growth of both pools, as discussed in Remark 7.

5 Conclusions and perspectives

In this paper we formalize and analyze a mathematical model describing an evolutionary process with affinity-dependent selection. We use a multi-type GW process, obtaining a discrete-time probabilistic model, which includes division, mutation, death and selection. This is employed in the context of antibody affinity maturation in GCs. We believe that a probabilistic approach is well suited to the study of Darwinian-like processes such as the one taking place in GCs during an immune response. Indeed, this kind of approaches allows to better take into account local inhomogeneities related to the discrete nature

of cells and stochastic fluctuations intrinsic to these processes, conversely to more popular deterministic continuum approaches. There, cell concentrations are described by a set of coupled ODEs changing deterministically and continuously during time, which has many computational advantages and has often been employed to model biological systems (*e.g.* [21,15,25] for applications to the GC reaction). In the main model developed here, we choose a selection mechanism which acts both positively and negatively on individuals submitted to selection. This choice is motivated by the fact that there are biological evidence supporting both kind of selection mechanisms: positive affinity-based selection by antigen binding as well as selection-dependent apoptosis [38,16]. The simplified mathematical framework proposed here allow to investigate how different kind of B-cell population evolves during the immune response both in the initial explosion phase and in the later relaxation phase of typical GCs. Indeed, mathematical analysis of the model leads to build matrix $\mathcal{M}$, which contains the expectations of each type (Proposition 2) and enables to describe the average behavior of all components of the process. Moreover, thanks to the spectral decomposition of $\mathcal{M}$ we were able to obtain explicitly some formulas giving the expected dynamics of all types. In addition, we exhibited an optimal value of the selection rate maximizing the expected number of selected clones for the t^{th}-generation (Proposition 6).

This is one possible choice of the selection mechanism. From a mathematical point of view, matrix $\mathcal{M}$ is particularly easy to manipulate, as we can obtain explicitly its spectrum. On the other hand, the positive and negative selection model leads, for example, to a selection threshold that does not have any impact on the evolution of the GC size. From a biological point of view this seems counterintuitive, since we could expect that the GC dynamics is sensible to the minimal fitness required for positive selection. Moreover, this process does not take into account any recycling mechanism, which has been confirmed by experiments [39] and which improves GCs' efficiency. In addition, we considered that only the selection mechanism is affinity dependent, while in the GC reaction other mechanisms, such as the death and proliferation rate, may depend on fitness [13,1]. Of course it is possible to define models with affinity-dependent division and death mechanisms with our formalism. This would clearly lead to a more complicated model, which can be at least studied numerically.

Mathematical tools used in Section 3 can be applied to define and study other selection mechanisms. For instance in Section 4 we propose two variants of the model analyzed in Section 3, in which selection acts only positively, resp. only negatively. This Section shows how our mathematical environment can be modified to describe different selection mechanisms, which can be studied at least numerically. Moreover, it gives a deeper insight of the previous model of positive and negative selection, by highlighting the effects of each selection mechanism individually, when they are not coupled.

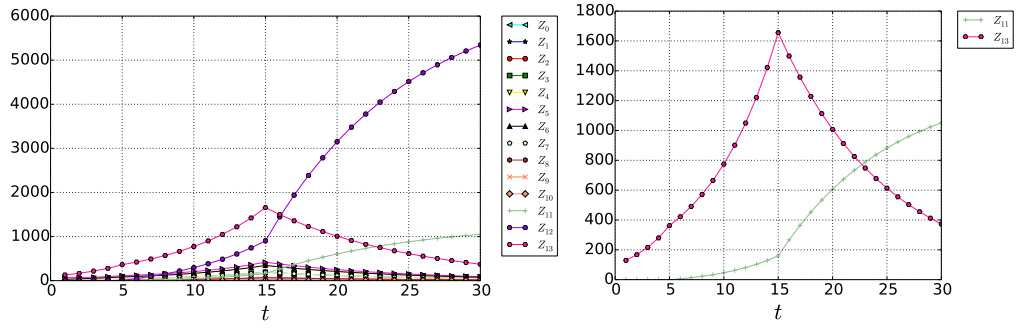
From a biological viewpoint there exist many possibilities to improve the models proposed in this paper. First of all it is extremely important to fix the system parameters, which have to be consistent with the real biological process. The choice of N defines the number of affinity level with respect to a given antigen. This value can be interpreted in different ways. On the one hand it can correspond to the number of key mutations observed during the process of Antigen Affinity Maturation, hence be even smaller than 10. On the other hand, each mutational event implies a change in the B-cell affinity, slight or not if it is a key mutation. In this case the affinity can be modeled as a continuous function, hence N corresponds to a possible discretization [41,43]. To this choice corresponds an appropriate choice of the transition probability matrix defining the mutational model over the affinity classes, $\mathcal{Q}_N$. In most numerical simulations we set $N = 10$, which is a sensible value since experimentalists observe that high-affinity B-cells differ in their BCR coding gene by about 9 mutations from germline genes [15,45]. Nevertheless all mathematical results are independent from this choice and hold true for all $N \geq 1$. The selection, division and death rates have also an important impact in the GC and selected pool dynamics: in the simulations we set them in order to be in a case of explosion of the GC hence appreciate the effects of all parameters over the main quantities, but they are not biologically justified. For instance, the typical proliferation rate of a B-cell has been estimated between 2 and 4 per day and in the literature we found B-cell death rates of the order of 0.5-0.8 per day [22,45,18]. Hence, if we suppose that a single time step corresponds *e.g.* to 6 hours, a consistent proliferation rate would be $r_{div} \simeq 0.75$, while the death rate r_d should be around 0.175. Since over a 6 hours period about 50% of B-cells transit from the DZ to the LZ, where they compete for positive selection signaling [6,36], we should choose $r_s \leq 0.5$. It could be further characterized taking into account its tightly relation with the time of GC peak, as highlighted in Section 3.3.

In Section 3.3 we have explicitly determined the optimal value of the selection rate maximizing the production of output cells at time t for the main model of positive and negative selection. It is equal to $1/t$ independently from all other parameters. Moreover, numerical estimations for the model of positive selection (Section 4.2) suggest that also in this case there exists an optimal value of $r_s(t)$, which tends to $1/t$ at least for t big enough. One has to interpret this result as the ideal optimal strength of the selection pressure to obtain a peak of the GC production of output cells at a given time step. For example, let us suppose again that a time step corresponds to 6 hours. The peak of the GC reaction has been measured to be close to day 12 [42], *i.e.* after ~ 48 maturation cycles in our model: for the kind of models we built and analyzed in this paper, a constant selection pressure r_s of $1/48 \simeq 0.02$ assures that the production of plasma and memory B-cells at the GC peak is maximized. Note that with the parameter choice $r_d = 0.175$, $r_{div} = 0.75$ and $r_s = 0.02$, the extinction probability of the GC is $\simeq 0.3^{z_0}$, z_0 being the number of initial seed cells. Since the extinction probability is strictly smaller than 1, such a GC

will explode with high probability and will be able to assure an intense and efficient immune response.

The particular form $1/t$ of the optimal selection rate for the t^{th} generation obtained in Section 3.3 certainly derives from the simplified structure of the model of positive and negative selection, even if this trend is further confirmed in the model of exclusively positive selection. Nevertheless it should be interesting to test the existence of an inverse relation between the selection rate and the timing of GC peak. Selection pressure can be quantified *e.g.* through comparative analysis between groups of sequences derived from different germline V(D)J segments, as proposed by the statistical framework for BAyesian estimation of Antigen-driven SELectIoN (BASELINe) [44]. BASELINe takes into account both mutation targeting bias and substitution bias and identifies point mutations grouped by location. Moreover it adresses the question of positive versus negative selection: positive selection is identified by an increased frequency of replacements, while a decreased frequency indicates negative selection. According to [44], the selection strength seems to vary and also switch from positive to negative in a different way depending on the location, *i.e.* if we are looking to complementary determining regions (CDRs), which are more significant for functional selection, or to framework regions. This gives stronger motivation to analyse both kind of selection mechanisms, acting both separately and simultaneously, and observe their effects over affinity maturation, as we have done in this paper using our simplified mathematical framework.

In our models the selection pressure is constant. Since the optimal selection rate above depends on time, this suggests to go further in this direction. Moreover, a time-dependent selection pressure would allow to take into account, for instance, the early GC phase in which simple clonal expansion of B-cells with no selection occurs [10]. The hypothesis of a selection pressure changing over time can be easily integrated in our model. Indeed let us suppose that a selection rate $r_{s,1}$ until time t_1 and $r_{s,2}$ for all $t > t_1$ are fixed. Starting from the initial condition $\mathbf{i}$ the expectations of each type at time t are given by $(\mathbf{i}\mathcal{M}^t_{r_{s,1}})$ if $t \leq t_1$ and $(\mathbf{i}\mathcal{M}^{t_1}_{r_{s,1}}\mathcal{M}^{t-t_1}_{r_{s,2}})$ if $t > t_1$, where $\mathcal{M}_{r_{s,i}}$ is the matrix containing the expectations of each type for an evolutionary process with constant selection rate $r_{s,i}$, $i = 1, 2$. In Figure 11 we plot the expected evolution during time of all types considering an increasing selection rate. We evaluate the expectations of all types following a process with positive and negative selection. We set $r_s = 0$ until $t = 5$, $r_s = 0.1$ from $t = 6$ to $t = 15$ and $r_s = 0.3$ for $t > 15$. Numerical simulations show that a time dependent selection rate allows initial explosion of the GC, and then progressive extinction, while when parameters are fixed, a GW process gives only rise either to explosion or to extinction, as shown above. The regulation and termination of the GC reaction has not yet been fully understood. In the literature, an increasing differentiation rate of GC B-cells is thought to be a good explanation [25], here we show that other



(a) **(b)**

Figure 11: (a) Evolution during time of the expected value of all types for the model of positive and negative selection, with r_s varying during time and $N = 10$. In particular we set $r_s = 0$ until $t = 5$, $r_s = 0.1$ from $t = 6$ to $t = 15$ and $r_s = 0.3$ from $t = 16$ to $t = 30$. Z_{13} denotes the total size of the GC (*i.e.* $\sum_{k=0}^{N} Z_k$), and we recall that Z_{11} corresponds to selected B-cells and Z_{12} to dead B-cells. We set $r_{div} = 0.3$, $r_d = 0.005$ and $z_0 = 100$ initial naive B-cells. All initial B-cells belong to $a_0 = 5$, and the selection threshold is $\overline{a}_s = 3$. (b) Evolution during time of the expected total size of the GC and the selected pool respectively, for the same set of parameters as in Figure 11 (a).

reasons could be of importance as well. Similarly, we can let other parameters vary for fixed time intervals, as well as decide to alternatively switch on and off the mutational mechanism, as already proposed in [29]. This can be obtained by alternatively use the identity matrix in place of $\mathcal{Q}_N$.

Applications of the models presented here to real biological problems and data should be further investigated. We propose here some contexts for which we believe that our kind of modeling approach could be employed to address biologically relevant questions.
Even if it is still extremely hard to have precise experimental information about the evolution of Antibody Affinity Maturation inside GCs, new refined techniques start to be available to measure clonal diversity in GCs. As an example, in [34] the authors combine multiphoton microscopy and sequencing to understand how different clonal diversification patterns can lead to efficient affinity maturation. The models we propose could be used to infer which are reasonable mutational transitional probability matrices and selection mechanisms/pressure to obtain such different scenario and infer if the tendency of GC to go or not through homogenizing selection is solely due to the hazard or if this is dependent on the kind of antigenic challenge and/or some specific characteristics of the host. If this is the case, these results could be particularly relevant *e.g.* in the context of vaccination design, where we are interested in find new way to improve the quality of the immune response after vaccination

challenge.
Another potential interesting application field is the study of some diseases entailing a dysfunction of the immune system, such as in particular Chronic Lymphocytic Leukemia (CLL), derived from antigen-experienced B-cells that differ in the level of mutations in their receptors [8]. This is the commonest form of leukemia in the Western world [12]. In CLL, leukemia B-cells can mature partially but not completely, are unable to opportunely undergo mutations in GCs, and survive longer than normal cells, crowding out healthy B-cells. Prognosis varies depending on the ability of host B-cells to mutate their antibody gene variable region. Even if major progresses have been made in the identification of molecular and cellular markers predicting the expansion of this disease in patients, the pathology remains incurable [11,12]. Our modeling approach could be employed to understand how an "healthy" mutational matrix is modified in patients affected by CLL, and if other mechanisms could contribute to get the prognosis worse. This could eventually provide suggestions about the causes that lead to CLL, and motivation for further research on possible treatments.

6 Acknowledgements

This work was supported by the Labex inflamex, ANR project 10-LABX-0017.

References

1. Anderson, S.M., Khalil, A., Uduman, M., Hershberg, U., Louzoun, Y., Haberman, A.M., Kleinstein, S.H., Shlomchik, M.J.: Taking advantage: high-affinity b cells in the germinal center have lower death rates, but similar rates of division, compared to low-affinity cells. The Journal of Immunology **183**(11), 7314–7325 (2009)
2. Ansari, H.R., Raghava, G.P.: Identification of conformational b-cell epitopes in an antigen from its primary sequence. Immunome research **6**(1), 1 (2010)
3. Athreya, K.B., Ney, P.E.: Branching processes, vol. 196. Springer Science & Business Media (2012)
4. Balelli, I., Milisic, V., Wainrib, G.: Branching random walks on binary strings for evolutionary processes. arXiv preprint arXiv:1607.00927 (2016)
5. Balelli, I., Milišić, V., Wainrib, G.: Random walks on binary strings applied to the somatic hypermutation of b-cells. Mathematical biosciences **300**, 168–186 (2018)
6. Bannard, O., Horton, R.M., Allen, C.D., An, J., Nagasawa, T., Cyster, J.G.: Germinal center centroblasts transition to a centrocyte phenotype according to a timed program and depend on the dark zone for effective selection. Immunity **39**(5), 912–924 (2013)
7. Castro, L.N.D., Zuben, F.J.V.: Learning and optimization using the clonal selection principle. Evolutionary Computation, IEEE Transactions on **6**(3), 239–251 (2002)
8. Chiorazzi, N., Rai, K.R., Ferrarini, M.: Chronic lymphocytic leukemia. New England Journal of Medicine **352**(8), 804–815 (2005)
9. Currin, A., Swainston, N., Day, P.J., Kell, D.B.: Synthetic biology for the directed evolution of protein biocatalysts: navigating sequence space intelligently. Chemical Society Reviews **44**(5), 1172–1239 (2015)
10. De Silva, N.S., Klein, U.: Dynamics of b cells in germinal centres. Nature Reviews Immunology **15**(3), 137–148 (2015)
11. Dighiero, G., Hamblin, T.: Chronic lymphocytic leukaemia. The Lancet **371**(9617), 1017–1029 (2008)

12. Eichhorst, B., Robak, T., Montserrat, E., Ghia, P., Hillmen, P., Hallek, M., Buske, C.: Chronic lymphocytic leukaemia: Esmo clinical practice guidelines for diagnosis, treatment and follow-up. Annals of Oncology **26**(suppl 5), v78–v84 (2015)
13. Gitlin, A.D., Shulman, Z., Nussenzweig, M.C.: Clonal selection in the germinal centre by regulated proliferation and hypermutation. Nature (2014)
14. Harris, T.E.: The theory of branching processes. Springer-Verlag (1963)
15. Iber, D., Maini, P.K.: A mathematical model for germinal centre kinetics and affinity maturation. Journal of theoretical biology **219**(2), 153–175 (2002)
16. Inoue, T., Moran, I., Shinnakasu, R., Phan, T.G., Kurosaki, T.: Generation of memory b cells and their reactivation. Immunological reviews **283**(1), 138–149 (2018)
17. Kauffman, S.A., Weinberger, E.D.: The nk model of rugged fitness landscapes and its application to maturation of the immune response. Journal of theoretical biology **141**(2), 211–245 (1989)
18. Keşmir, C., De Boer, R.J.: A mathematical model on germinal center kinetics and termination. The Journal of Immunology **163**(5), 2463–2469 (1999)
19. Kringelum, J.V., Lundegaard, C., Lund, O., Nielsen, M.: Reliable b cell epitope predictions: impacts of method development and improved benchmarking. PLoS Comput Biol **8**(12), e1002,829 (2012)
20. MacLennan, I.C., de Vinuesa, C.G., Casamayor-Palleja, M.: B-cell memory and the persistence of antibody responses. Philosophical Transactions of the Royal Society of London B: Biological Sciences **355**(1395), 345–350 (2000)
21. Meyer-Hermann, M., Mohr, E., Pelletier, N., Zhang, Y., Victora, G.D., Toellner, K.M.: A theory of germinal center b cell selection, division, and exit. Cell reports **2**(1), 162–174 (2012)
22. Meyer-Hermann, M.E., Maini, P.K., Iber, D.: An analysis of b cell selection mechanisms in germinal centers. Mathematical Medicine and Biology **23**(3), 255–277 (2006)
23. Minc, H.: Nonnegative matrices, 1988 (1988)
24. Mitchell, R., Kelly, D.F., Pollard, A.J., Trück, J.: Polysaccharide-specific b cell responses to vaccination in humans. Human vaccines & immunotherapeutics **10**(6), 1661–1668 (2014)
25. Moreira, J.S., Faro, J.: Modelling two possible mechanisms for the regulation of the germinal center dynamics. The Journal of Immunology **177**(6), 3705–3710 (2006)
26. Murphy, K.M., Travers, P., Walport, M., et al.: Janeway's immunobiology, vol. 7. Garland Science New York, NY, USA (2012)
27. Pang, W., Wang, K., Wang, Y., Ou, G., Li, H., Huang, L.: Clonal selection algorithm for solving permutation optimisation problems: A case study of travelling salesman problem. In: International Conference on Logistics Engineering, Management and Computer Science (LEMCS 2015). Atlantis Press (2015)
28. Perelson, A.S., Oster, G.F.: Theoretical studies of clonal selection: minimal antibody repertoire size and reliability of self-non-self discrimination. Journal of theoretical biology **81**(4), 645–670 (1979)
29. Perelson, A.S., Weisbuch, G.: Immunology for physicists. Reviews of modern physics **69**(4), 1219–1267 (1997)
30. Phan, T.G., Paus, D., Chan, T.D., Turner, M.L., Nutt, S.L., Basten, A., Brink, R.: High affinity germinal center b cells are actively selected into the plasma cell compartment. The Journal of experimental medicine **203**(11), 2419–2424 (2006)
31. Shannon, M., Mehr, R.: Reconciling repertoire shift with affinity maturation: the role of deleterious mutations. The Journal of Immunology **162**(7), 3950–3956 (1999)
32. Shen, W.J., Wong, H.S., Xiao, Q.W., Guo, X., Smale, S.: Towards a mathematical foundation of immunology and amino acid chains. arXiv preprint arXiv:1205.6031 (2012)
33. Shlomchik, M., Watts, P., Weigert, M., Litwin, S.: Clone: a monte-carlo computer simulation of b cell clonal expansion, somatic mutation, and antigen-driven selection. In: Somatic Diversification of Immune Responses, pp. 173–197. Springer (1998)
34. Tas, J.M., Mesin, L., Pasqual, G., Targ, S., Jacobsen, J.T., Mano, Y.M., Chen, C.S., Weill, J.C., Reynaud, C.A., Browne, E.P., et al.: Visualizing antibody affinity maturation in germinal centers. Science **351**(6277), 1048–1054 (2016)
35. Timmis, J., Hone, A., Stibor, T., Clark, E.: Theoretical advances in artificial immune systems. Theoretical Computer Science **403**(1), 11–32 (2008)

36. Victora, G.D.: Snapshot: the germinal center reaction. Cell **159**(3), 700–700 (2014)
37. Victora, G.D., Mesin, L.: Clonal and cellular dynamics in germinal centers. Current opinion in immunology **28**, 90–96 (2014)
38. Victora, G.D., Nussenzweig, M.C.: Germinal centers. Annual review of immunology **30**, 429–457 (2012)
39. Victora, G.D., Schwickert, T.A., Fooksman, D.R., Kamphorst, A.O., Meyer-Hermann, M., Dustin, M.L., Nussenzweig, M.C.: Germinal center dynamics revealed by multiphoton microscopy with a photoactivatable fluorescent reporter. Cell **143**(4), 592–605 (2010)
40. Wang, P., Shih, C.m., Qi, H., Lan, Y.h.: A stochastic model of the germinal center integrating local antigen competition, individualistic t–b interactions, and b cell receptor signaling. The Journal of Immunology p. 1600411 (2016)
41. Weiser, A.A., Wittenbrink, N., Zhang, L., Schmelzer, A.I., Valai, A., Or-Guil, M.: Affinity maturation of b cells involves not only a few but a whole spectrum of relevant mutations. International immunology **23**(5), 345–356 (2011)
42. Wollenberg, I., Agua-Doce, A., Hernández, A., Almeida, C., Oliveira, V.G., Faro, J., Graca, L.: Regulation of the germinal center reaction by foxp3+ follicular regulatory t cells. The Journal of Immunology **187**(9), 4553–4560 (2011)
43. Xu, H., Schmidt, A.G., O'Donnell, T., Therkelsen, M.D., Kepler, T.B., Moody, M.A., Haynes, B.F., Liao, H.X., Harrison, S.C., Shaw, D.E.: Key mutations stabilize antigen-binding conformation during affinity maturation of a broadly neutralizing influenza antibody lineage. Proteins: Structure, Function, and Bioinformatics **83**(4), 771–780 (2015)
44. Yaari, G., Uduman, M., Kleinstein, S.H.: Quantifying selection in high-throughput immunoglobulin sequencing data sets. Nucleic acids research **40**(17), e134–e134 (2012)
45. Zhang, J., Shakhnovich, E.I.: Optimality of mutation and selection in germinal centers. PLoS Comput Biol **6**(6), e1000,800 (2010)

Appendix

A Few reminders of classical results on GW processes

We recall here some classical results about GW processes we employed to derive Proposition 1 (Section 3.1). For further details the reader can refer to [14].

Definition 14 Let X be an integer valued rv, $p_k := \mathbb{P}(X = k)$ for all $k \geq 0$. Its probability generating function (pgf) is given by:

$$F_X(s) = \sum_{k=0}^{+\infty} p_k s^k$$

F_X is a convex monotonically increasing function over $[0,1]$, and $F_X(1) = 1$. If $p_0 \neq 0$ and $p_0 + p_1 < 1$ then F is a strictly increasing function.

Definition 15 Given F, the pgf of a rv X, the iterates of F are given by:

$$\begin{aligned} F_0(s) &= s \\ F_1(s) &= F(s) \\ F_t(s) &= F(F_{t-1}(s)) \text{ for } t \geq 2 \end{aligned}$$

Proposition 11
(i) If $\mathbb{E}(X)$ exists (respectively $\mathbb{V}(X)$), then $\mathbb{E}(X) = F'_X(1)$ (respectively $\mathbb{V}(X) = F''_X(1) - (\mathbb{E}(X))^2 + \mathbb{E}(X)$).
(ii) If X and Y are two integer valued independent rvs, then $X+Y$ is still an integer valued rv and its pgf is given by $F_{X+Y} = F_X F_Y$.

Definition 16 We denote by η the extinction probability of the process $(Z_t)_{t\in\mathbb{N}}$:

$$\eta := \lim_{t\to\infty} F_t(0)$$

Theorem 2

(i) The pgf of $Z_t^{(z_0)}$, $t \in \mathbb{N}$, which represents the population size of the t^{th}-generation starting from $z_0 \geq 1$ seed cells, is $F_t^{(z_0)} = (F_t)^{z_0}$, F_t being the t^{th}-iterate of F (Equation (2)).

(ii) The expected size of the GC at time t and starting from z_0 B-cells is given by:

$$\mathbb{E}(Z_t^{(z_0)}) = z_0\,(\mathbb{E}(Z_t)) = z_0\,(\mathbb{E}(Z_1))^t \ , \tag{13}$$

(iii) η is the smallest fixed point of the generating function F, i.e. η is the smallest s s.t. $F(s) = s$.

(iv) If $\mathbb{E}(Z_1) =: m$ is finite, then:

- *if $m \leq 1$ then F has only 1 as fixed point and consequently $\eta = 1$;*
- *if $m > 1$ then F as exactly a fixed point on $[0,1[$ and then $\eta < 1$.*

(v) Denoted by η_{z_0} the probability of extinction of $(Z_t^{(z_0)})$, one has:

$$\eta_{z_0} = \eta^{z_0}$$

where η is given by (iii).

Proposition 1 of Section 3.1 follows by applying Theorem 2 and Equation (1).

B Proof of Proposition 2

For all $j \in \{0,\dots,N+2\}$ the generating function of Z_j gives the number of offspring of each type that a type j particle can produce. It is defined as follows:

$$f^{(j)}(s_0,\dots,s_{N+2}) = \sum_{k_0,\dots,k_{N+2}\geq 0} p^{(j)}(k_0,\dots,k_{N+2}) s_0^{k_0}\cdots s_{N+2}^{k_{N+2}}, \tag{14}$$

$$0 \leq s_\alpha \leq 1 \text{ for all } \alpha \in \{0,\dots,N+2\}$$

where $p^{(j)}(k_0,\dots,k_{N+2})$ is the probability that a type j cell produces k_0 cells of type 0, k_1 of type 1, …, k_{N+2} of type $N+2$ for the next generation.
We denote:

- $\mathbf{p}(\mathbf{k}) = (p^{(0)}(\mathbf{k}),\dots,p^{(N+2)}(\mathbf{k}))$, for $\mathbf{k} = (k_0,\dots,k_{N+2}) \in \mathbb{Z}_+^{N+3}$
- $\mathbf{f}(\mathbf{s}) = (f^{(1)}(\mathbf{s}),\dots,f^{(N+1)}(\mathbf{s}))$, for $\mathbf{s} = (s_0,\dots,s_{N+2}) \in \mathcal{C}^{N+3} := [0,1]^{N+3}$

Then the probability generating function of $\mathbf{Z}_1$ is given by:

$$\mathbf{f}(\mathbf{s}) = \sum_{\mathbf{k}\in\mathbb{Z}_+^{N+3}} \mathbf{p}(\mathbf{k})\mathbf{s}^{\mathbf{k}},\ \mathbf{s} \in \mathcal{C}^{N+3} \tag{15}$$

Again, the generating function of $\mathbf{Z}_t$, $\mathbf{f}_t(\mathbf{s})$, is obtained as the t^{th}-iterate of $\mathbf{f}$, and it holds true that:

$$\mathbf{f}_{t+r}(\mathbf{s}) = \mathbf{f}_t[\mathbf{f}_r(\mathbf{s})],\ \mathbf{s} \in \mathcal{C}^{N+3}.$$

Let $m_{ij} := \mathbb{E}[Z_{1,j}^{(i)}]$ the expected number of offspring of type j of a cell of type i in one generation. We collect all m_{ij} in a matrix, $\mathcal{M} = (m_{ij})_{0\leq i,j\leq N+2}$. We have [3]:

$$m_{ij} = \frac{\partial f^{(i)}}{\partial s_j}(\mathbf{1})$$

and:

$$\mathbb{E}[Z_{t,j}^{(i)}] = \frac{\partial f_t^{(i)}}{\partial s_j}(\mathbf{1}) \tag{16}$$

Finally:

$$\mathbb{E}[\mathbf{Z}_t^{(\mathbf{i})}] = \mathbf{i}\mathcal{M}^t \tag{17}$$

One can explicitly derive the elements of matrix $\mathcal{M}$ for the process described in Definition 13.

Proposition *$\mathcal{M}$ is a $(N+3)\times(N+3)$ matrix defined as a block matrix:*

$$\mathcal{M} = \begin{pmatrix} \mathcal{M}_1 & \mathcal{M}_2 \\ \mathbf{0}_{2\times(N+1)} & \mathcal{I}_2 \end{pmatrix}$$

Where:

- $\mathbf{0}_{2\times(N+1)}$ *is a $2\times(N+1)$ matrix with all entries 0;*
- $\mathcal{I}_n$ *is the identity matrix of size n;*
- $\mathcal{M}_1 = 2(1-r_d)r_{div}(1-r_s)\mathcal{Q}_N + (1-r_d)(1-r_{div})(1-r_s)\mathcal{I}_{N+1}$
- $\mathcal{M}_2 = (m_{2,ij})$ *is a $(N+1)\times 2$ matrix where for all $i \in \{0,\dots,N\}$:*
  - *if $i \le \overline{a}_s$:*

$$m_{2,i1} = (1-r_d)(1-r_{div})r_s + 2(1-r_d)r_{div}r_s \sum_{j=0}^{\overline{a}_s} q_{ij},$$

$$m_{2,i2} = r_d + 2(1-r_d)r_{div}r_s \sum_{j=\overline{a}_s+1}^{N} q_{ij}$$

  - *if $i > \overline{a}_s$:*

$$m_{2,i1} = 2(1-r_d)r_{div}r_s \sum_{j=0}^{\overline{a}_s} q_{ij},$$

$$m_{2,i2} = r_d + (1-r_d)(1-r_{div})r_s + 2(1-r_d)r_{div}r_s \sum_{j=\overline{a}_s+1}^{N} q_{ij}$$

Proof One has to compute all $f^{(i)}(\mathbf{s})$ for $i = 0,\dots,N+2$, which depend on r_d, r_{div}, r_s, $\overline{a}_s$ and the elements of $\mathcal{Q}_N$. First, the elements of the $(N+2)^{\text{th}}$ and $(N+3)^{\text{th}}$-lines are obviously determined: all selected (resp. dead) cells remain selected (resp. dead) for next generations, as they can not give rise to any other cell type offspring (we do not take into account here any type of recycling mechanism). Let $i \in \{0,\dots,N\}$ be a fixed index: we evaluate m_{ij} for all $j \in \{0,\dots,N+2\}$. The first step is to determine the value of $p^{(i)}(\mathbf{k})$ for $\mathbf{k} = (k_0,\dots,k_{N+2}) \in \mathbb{Z}_+^{N+3}$. There exists only a few cases in which $p^{(i)}(\mathbf{k}) \neq 0$, which can be explicitly evaluated:

- $p^{(i)}(0,\dots,0,1) = \begin{cases} r_d & \text{if } i \le \overline{a}_s \\ r_d + (1-r_d)(1-r_{div})r_s & \text{otherwise} \end{cases}$
- $p^{(i)}(0,\dots,0,1,0) = \begin{cases} (1-r_d)(1-r_{div})r_s & \text{if } i \le \overline{a}_s \\ 0 & \text{otherwise} \end{cases}$
- $p^{(i)}(0,\dots,0,\underset{i}{1},0,\dots,0,0) = (1-r_d)(1-r_{div})(1-r_s)$
- $p^{(i)}(0,\dots,0,2) = (1-r_d)r_{div}r_s^2 \sum_{j_1=\overline{a}_s+1}^{N} q_{ij_1} \sum_{j_2=\overline{a}_s+1}^{N} q_{ij_2}$
- $p^{(i)}(0,\dots,0,2,0) = (1-r_d)r_{div}r_s^2 \sum_{j_1=0}^{\overline{a}_s} q_{ij_1} \sum_{j_2=0}^{\overline{a}_s} q_{ij_2}$

- $p^{(i)}(0,\dots,0,1,1) = 2(1-r_d)r_{div}r_s^2 \sum_{j_1=0}^{\overline{a}_s} q_{ij_1} \sum_{j_2=\overline{a}_s+1}^{N} q_{ij_2}$
- For all $j_1 < j_2 \in \{0,\dots,N\}$:
  - $p^{(i)}(0,\dots,0,\underset{j_1}{2},0,\dots,0,0) = (1-r_d)r_{div}(1-r_s)^2 q_{ij_1}^2$
  - $p^{(i)}(0,\dots,0,\underset{j_1}{1},0,\dots,0,\underset{j_2}{1},0,\dots,0,0) = 2(1-r_d)r_{div}(1-r_s)^2 q_{ij_1}q_{ij_2}$
  - $p^{(i)}(0,\dots,0,\underset{j_1}{1},0,\dots,0,1) = 2(1-r_d)r_{div}r_s(1-r_s)q_{ij_1}\sum_{j_2=\overline{a}_s+1}^{N} q_{ij_2}$
  - $p^{(i)}(0,\dots,0,\underset{j_1}{1},0,\dots,0,1,0) = 2(1-r_d)r_{div}r_s(1-r_s)q_{ij_1}\sum_{j_2=0}^{\overline{a}_s} q_{ij_2}$
- $p^{(i)}(\mathbf{k}) = 0$ otherwise

We can therefore evaluate $f^{(i)}(\mathbf{s})$, with $\mathbf{s} = (s_0,\dots,s_{N+2}) \in \mathcal{C}^{N+3}$.

For all $i \leq \overline{a}_s$:

$$
\begin{aligned}
f^{(i)}(\mathbf{s}) &= r_d s_{N+2} + (1-r_d)(1-r_{div})r_s s_{N+1} + (1-r_d)(1-r_{div})(1-r_s)s_i \\
&+ (1-r_d)r_{div}r_s^2 \left(\sum_{j_1=\overline{a}_s+1}^{N} q_{ij_1} \sum_{j_2=\overline{a}_s+1}^{N} q_{ij_2} s_{N+2}^2 \right. \\
&\left. + \sum_{j_1=0}^{\overline{a}_s} q_{ij_1} \sum_{j_2=0}^{\overline{a}_s} q_{ij_2} s_{N+1}^2 + 2\sum_{j_1=0}^{\overline{a}_s} q_{ij_1} \sum_{j_2=\overline{a}_s+1}^{N} q_{ij_2} s_{N+1}s_{N+2} \right) \\
&+ (1-r_d)r_{div}(1-r_s)^2 \left(\sum_{j_1=0}^{N} q_{ij_1}^2 s_{j_1}^2 + 2\sum_{j_1=0}^{N} q_{ij_1} \sum_{j_2<j_1=0}^{N} q_{ij_2} s_{j_1} s_{j_2} \right) \\
&+ 2(1-r_d)r_{div}r_s(1-r_s) \sum_{j_1=0}^{N} q_{ij_1} \left(\sum_{j_2=\overline{a}_s+1}^{N} q_{ij_2} s_{N+2} + \sum_{j_2=0}^{\overline{a}_s} q_{ij_2} s_{N+1} \right) s_{j_1}
\end{aligned}
\tag{18}
$$

If $i > \overline{a}_s$ then $f^{(i)}(\mathbf{s})$ is the same except for the first line, which becomes:

$$(r_d + (1-r_d)(1-r_{div})r_s)s_{N+2} + (1-r_d)(1-r_{div})(1-r_s)s_i$$

The values of each m_{ij} are now obtained by evaluating all partial derivatives of $f^{(i)}(\mathbf{s})$ in $\mathbf{1}$, keeping in mind that for all $i \in \{0,\dots,N\}$, $\sum_{j=0}^{N} q_{ij} = 1$. □

C Deriving the extinction probability of the GC from the multi-type GW process (Section 3.2)

Let us recall some results about the extinction probability for multi-type GW processes [3].

Definition 17 Let $q^{(i)}$ be the probability of eventual extinction of the process, when it starts from a single type i cell. As above bold symbols denote vectors *i.e.* $\mathbf{q} := (q^{(0)},\dots,q^{(N+2)}) \geq 0$.

Definition 18 We say that $(\mathbf{Z}_t)$ is singular if each particle has exactly one offspring, which implies that the branching process becomes a simple MC.

Definition 19 Matrix $\mathcal{M}$ is said to be strictly positive if it has non-negative entries and there exists a t s.t. $\left(\mathcal{M}^t\right)_{ij} > 0$ for all i, j. $(\mathbf{Z}_t)$ is called positive regular iff $\mathcal{M}$ is strictly positive.

Notation 1 Let $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$. We say that $\mathbf{u} \leq \mathbf{v}$ if $u_i \leq v_i$ for all $i \in \{1, \ldots, n\}$. Moreover, we say that $\mathbf{u} < \mathbf{v}$ if $\mathbf{u} \leq \mathbf{v}$ and $\mathbf{u} \neq \mathbf{v}$.

Theorem 3 *Let $(\mathbf{Z}_t)$ be non singular and strictly positive. Let ρ be the maximal eigenvalue of $\mathcal{M}$. The following three results hold:*

1. *If $\rho < 1$ (subcritical case) or $\rho = 1$ (critical case) then $\mathbf{q} = \mathbf{1}$. Otherwise, if $\rho > 1$ (supercritical case), then $\mathbf{q} < \mathbf{1}$.*
2. *$\lim_{t \to \infty} \mathbf{f}_t(\mathbf{s}) = \mathbf{q}$, for all $\mathbf{s} \in \mathcal{C}^{N+3}$.*
3. *$\mathbf{q}$ is the only solution of $\mathbf{f}(\mathbf{s}) = \mathbf{s}$ in $\mathcal{C}^{N+3}$.*

The spectrum of matrix $\mathcal{M}$ defined in Definition 2 (and recalled in Appendix B) is obtained as follows:

Proposition 12 *Let $\mathcal{M}$ be defined as a block matrix as in Proposition 2. Let $\lambda_{\mathcal{M},i}$ be its i^{th}-eigenvalue. The spectrum of $\mathcal{M}$ is given by:*

- *For all $i \in \{0, \ldots, N\}$, $\lambda_{\mathcal{M},i} = (1-r_d)(1-r_s)(1+r_{div}(2\lambda_i - 1))$, where λ_i is the i^{th}-eigenvalue of matrix $\mathcal{Q}_N$.*
- *whereas $\lambda_{\mathcal{M},N+1} = 1$ with multiplicity 2.*

Proof As $\mathcal{M}$ is a block matrix with the lower left block composed of zeros, then $Spec(\mathcal{M}) = Spec(\mathcal{M}_1) \cup Spec(\mathcal{I}_2)$. The result follows. □

Therefore we obtain the same condition as in Proposition 1 for the extinction probability in the GC:

Proposition 13 *Let $\mathbf{q}$ be the extinction probability for the process $(\mathbf{Z}_t)$ defined in Definition 13 and restricted to the first $N+1$ components (*i.e. *we refer only to matrix $\mathcal{M}_1$, which defines the expectations of GC B-cells). Therefore:*

- *if $r_s \geq 1 - \dfrac{1}{(1-r_d)(1+r_{div})}$, then $\mathbf{q} = \mathbf{1}$*
- *otherwise $\mathbf{q} < \mathbf{1}$ is the smallest fixed point of $\mathbf{f}(\mathbf{s})$ in $\mathcal{C}^{N+3}$.*

Proof $\mathcal{Q}_N$ is a stochastic matrix, therefore its largest eigenvalue is 1. The corresponding eigenvalue of matrix $\mathcal{M}_1$ is: $\lambda_{\mathcal{M}_1,1} = (1-r_d)(1-r_s)(1+r_{div})$. The proposition is proved by observing that $\lambda_{\mathcal{M}_1,1} \leq 1 \Leftrightarrow r_s \geq 1 - \dfrac{1}{(1-r_d)(1+r_{div})}$ and applying Theorem 3 (note that $\mathcal{M}_1$ is positive regular: this is not the case for matrix $\mathcal{M}$). □

D Expected size of the GC derived from the multi-type GW process (Section 3.2)

Proposition *Let $\mathbf{i}$ be the initial state, $z_0 := |\mathbf{i}|$ its 1-norm ($|\mathbf{i}| := \sum_{j=0}^{N+2} \mathbf{i}_j$). The expected size of the GC at time t:*

$$\sum_{k=0}^{N} (\mathbf{i}\mathcal{M}^t)_k = |\mathbf{i}| \left((1-r_d)(1+r_{div})(1-r_s)\right)^t$$

Proof For the sake of simplicity, let us suppose that the process starts from a single B-cell belonging to the affinity class $a_0 = i$ with respect to the target trait. We do not need to specify the transition probability matrix used to define the mutational model allowed.

We recall the expression of $\mathcal{M}^t$ obtained by iteration:

$$\mathcal{M}^t = \begin{pmatrix} \mathcal{M}_1^t & \sum_{k=0}^{t-1} \mathcal{M}_1^k \mathcal{M}_2 \\ \mathbf{0}_{2\times(N+1)} & \mathcal{I}_2 \end{pmatrix}$$

Therefore we can claim that $(\mathbf{i}\mathcal{M}^t)_k$ corresponds to the k^{th}-component of the i^{th}-row of matrix $\mathcal{M}_1^t = (2(1-r_d)r_{div}(1-r_s)\mathcal{Q}_N + (1-r_d)(1-r_{div})(1-r_s)\mathcal{I}_{N+1})^t$, where $\mathcal{Q}_N$ is a stochastic matrix. Matrices $\mathcal{A} := 2(1-r_d)r_{div}(1-r_s)\mathcal{Q}_N$ and $\mathcal{B} := (1-r_d)(1-r_{div})(1-r_s)\mathcal{I}_{N+1}$ clearly commute, therefore:

$$(\mathcal{A}+\mathcal{B})^t = \sum_{j=0}^{t} C_t^j \mathcal{A}^{t-j}\mathcal{B}^j \tag{19}$$

For all j, $0 \le j \le t$:

$$\begin{aligned}\mathcal{A}^{t-j}\mathcal{B}^j &= 2^{t-j}(1-r_d)^{t-j} r_{div}^{t-j}(1-r_s)^{t-j}(1-r_d)^j(1-r_{div})^j(1-r_s)^j \mathcal{Q}_N^{t-j} \\ &= (1-r_d)^t(1-r_s)^t(2r_{div})^{t-j}(1-r_{div})^j \mathcal{Q}_N^{t-j}\end{aligned}$$

Hence:

$$(\mathcal{A}+\mathcal{B})^t = (1-r_d)^t(1-r_s)^t \sum_{j=0}^{t} C_t^j (2r_{div})^{t-j}(1-r_{div})^j \mathcal{Q}_N^{t-j}$$

And consequently:

$$\begin{aligned}\sum_{k=0}^{N}(\mathbf{i}\mathcal{M}^t)_k &= \sum_{k=0}^{N}\left(\mathbf{i}(\mathcal{A}+\mathcal{B})^t\right)_k \\ &= (1-r_d)^t(1-r_s)^t \sum_{j=0}^{t} C_t^j (2r_{div})^{t-j}(1-r_{div})^j \sum_{k=0}^{N}\left(\mathbf{i}\mathcal{Q}_N^{t-j}\right)_k\end{aligned}$$

Since $\mathcal{Q}_N$ is a stochastic matrix, for all n, $\mathcal{Q}_N^n$ is still a stochastic matrix, *i.e.* the entries of each row of $\mathcal{Q}_N^n$ sum to 1. Therefore:

$$\begin{aligned}\sum_{k=0}^{N}(\mathbf{i}\mathcal{M}^t)_k &= (1-r_d)^t(1-r_s)^t \sum_{j=0}^{t} C_t^j (2r_{div})^{t-j}(1-r_{div})^j \\ &= (1-r_d)^t(1-r_s)^t(2r_{div}+1-r_{div})^t = (1-r_d)^t(1-r_s)^t(1+r_{div})^t \ ,\end{aligned}$$

as stated by Equation (3) for $z_0 = 1$. This result can be easily generalized to the case of $z_0 \ge 1$ initial B-cells.

E Proof of Proposition 5

Proposition *Let us suppose that at time $t = 0$ there is a single B-cell entering the GC belonging to the i^{th}-affinity class with respect to the target cell. Moreover, let us suppose that $\mathcal{Q}_N = R\Lambda_N L$. For all $t \ge 1$, the expected number of selected B-cells at time t, is:*

$$\mathbb{E}(S_t) = r_s(1-r_s)^{t-1}(1-r_d)^t \sum_{\ell=0}^{N}(2\lambda_\ell r_{div} + 1 - r_{div})^t \sum_{k=0}^{\overline{a}_s} r_{i\ell} l_{\ell k} \ ,$$

Proof Let us suppose, for the sake of simplicity, that $\mathcal{Q}_N$ is diagonalizable:

$$\mathcal{Q}_N = R\Lambda_N L \ , \tag{20}$$

We can prove by iteration that:

$$\mathcal{M}^t = \begin{pmatrix} \mathcal{M}_1^t & \sum_{k=0}^{t-1} \mathcal{M}_1^k \mathcal{M}_2 \\ \mathbf{0}_{2\times(N+1)} & \mathcal{I}_2 \end{pmatrix} \tag{21}$$

It follows from (20) and (21) that for all $t \geq 1$, $\mathcal{M}^t$ can be written as:

$$\mathcal{M}^t = \begin{pmatrix} RD^tL & \left(R\sum_{k=0}^{t-1} D^kL\right)\mathcal{M}_2 \\ \mathbf{0}_{2\times(N+1)} & \mathcal{I}_2 \end{pmatrix} \ , \tag{22}$$

where $D = 2(1-r_d)r_{div}(1-r_s)\Lambda_N + (1-r_d)(1-r_{div})(1-r_s)\mathcal{I}_{N+1}$ is a diagonal matrix. We obtain its expression thanks to Proposition 2.

Moreover, by Proposition 3 and Equation (20) we have:

$$\widetilde{\mathcal{M}} = \begin{pmatrix} R\widetilde{D}L & \widetilde{\mathcal{M}}_2 \\ \mathbf{0}_{2\times(N+1)} & \mathcal{I}_2 \end{pmatrix} \ , \tag{23}$$

where $\widetilde{D} = 2(1-r_d)r_{div}\Lambda_N + (1-r_d)(1-r_{div})\mathcal{I}_{N+1}$ is a diagonal matrix.

Proposition 4 claims:

$$\mathbb{E}(S_t) = r_s \sum_{k=0}^{\overline{a}_s} \left(\mathbf{i}\mathcal{M}^{t-1}\widetilde{\mathcal{M}}\right)_k$$

From Equations (22) and (23):

$$\mathcal{M}^{t-1}\widetilde{\mathcal{M}} = \begin{pmatrix} RD^{t-1}\widetilde{D}L & RD^{t-1}L\widetilde{\mathcal{M}}_2 + \left(R\sum_{k=0}^{t-2} D^kL\right)\mathcal{M}_2 \\ \mathbf{0}_{2\times(N+1)} & \mathcal{I}_2 \end{pmatrix}$$

Since, by hypothesis, $\mathbf{i} = (0,\ldots,0,1,0,\ldots,0,0)$, with the only 1 being at position i, $0 \leq i \leq N$, then $\left(\mathbf{i}\mathcal{M}^{t-1}\widetilde{\mathcal{M}}\right)$ denotes the i^{th}-row of matrix $\mathcal{M}^{t-1}\widetilde{\mathcal{M}}$. Therefore, we are interested in the sum between 0 and $\overline{a}_s$ of the elements of the i^{th}-row of matrix $\mathcal{M}^{t-1}\widetilde{\mathcal{M}}$, *i.e.* of the i^{th}-row of matrix $RD^{t-1}\widetilde{D}L$, since clearly $\overline{a}_s \leq N$. $D^{t-1}\widetilde{D}$ is a diagonal matrix whose ℓ^{th}-diagonal element is given by:

$$\begin{aligned} \left(D^{t-1}\widetilde{D}\right)_\ell &= (2(1-r_d)r_{div}(1-r_s)\lambda_\ell + (1-r_d)(1-r_{div})(1-r_s))^{t-1} \\ &\quad \cdot(2(1-r_d)r_{div}\lambda_\ell + (1-r_d)(1-r_{div})) \\ &= (1-r_s)^{t-1}(1-r_d)^t\left(2\lambda_\ell r_{div} + 1 - r_{div}\right)^t \end{aligned}$$

The result follows observing that: $\left(RD^{t-1}\widetilde{D}L\right)_{ik} = \sum_{\ell=0}^{N}\left(D^{t-1}\widetilde{D}\right)_\ell r_{i\ell}l_{\ell k}$. □

F Heuristic proof of Proposition 6

Proposition *For all $t \in \mathbb{N}$ the value $r_s(t)$ which maximizes the expected number of selected B-cells at the t^{th} maturation cycle is:*

$$r_s(t) = \frac{1}{t}$$

Hypothesis 1 $\mathcal{Q}_N$ converges through its stationary distribution, denoted by $\mathbf{m} = (m_i)$, $i \in \{0, \dots, N\}$.

Hypothesis 2 Z_t explodes, where $(Z_t)_{t\in\mathbb{N}}$ is given by Definition 4.

Let $\widetilde{Z}_t$, $t \geq 0$ be the random variable describing the GC-population size at time t before the selection mechanism is performed for this generation. For the sake of simplicity, let us suppose $\widetilde{Z}_0 = 1$. $(\widetilde{Z}_t)_{t\in\mathbb{N}}$ is a MC on $\{0,1,2,\dots\}$. Denoted by $\tilde{p}_k := \mathbb{P}(\widetilde{Z}_1 = k)$, $k \in \{0,1,2\}$:

$$\begin{cases} \tilde{p}_0 = r_d \\ \tilde{p}_1 = (1-r_d)(1-r_{div}) \\ \tilde{p}_2 = (1-r_d)r_{div} \end{cases} \tag{24}$$

It follows: $\tilde{m} := \mathbb{E}(\widetilde{Z}_1) = (1-r_d)(1-r_{div}) + 2(1-r_d)r_{div} = (1-r_d)(1+r_{div})$.

Conditioning to $Z_t = k$, $\widetilde{Z}_{t+1}$ is distributed as the sum of k independent copies of $\widetilde{Z}_1$, which gives:

$$\mathbb{E}(\widetilde{Z}_t) = \mathbb{E}(Z_{t-1})\mathbb{E}(\widetilde{Z}_1) = \mathbb{E}(Z_1)^{t-1}\mathbb{E}(\widetilde{Z}_1) = (1-r_d)^t(1+r_{div})^t(1-r_s)^{t-1} \tag{25}$$

Thanks to Hypotheses 1 and 2, if t is big enough, there is approximately a proportion of m_i elements in the i^{th}-affinity class with respect to $\overline{\mathbf{x}}$. Therefore, on average at time t there are approximately $\sum_{i=0}^{\overline{a}_s} m_i \mathbb{E}(\widetilde{Z}_t)$ B-cells in the GC belonging to an affinity class with index at most equal to $\overline{a}_s$ with respect to $\overline{\mathbf{x}}$, before the selection mechanism is performed for this generation. Each one of these cells can be submitted to selection with probability r_s, and in this case it will be positively selected. Hence:

$$\mathbb{E}(S_t) \simeq r_s \sum_{i=0}^{\overline{a}_s} m_i \mathbb{E}(\widetilde{Z}_t) = (1-r_d)^t(1+r_{div})^t(1-r_s)^{t-1} r_s \sum_{i=0}^{\overline{a}_s} m_i \ , \tag{26}$$

which is maximized at time $t \geq 1$ for $r_s(t) = 1/t$.

Remark 8 One observes that the approximation in (26) gives the same value for the optimal $r_s(t)$ as in Proposition 6. Nevertheless, it does not allow to describe exactly the behavior of $\mathbb{E}(S_t)$, since it is obtained by approximating the distribution of B-cells in the GC with their stationary distribution.