A birth–death model of ageing: from individual-based dynamics to evolutive differential inclusions

Sylvie Méléard, Michael Rera, Tristan Roget

A birth-death model of ageing: from individual-based dynamics to evolutive differential inclusions

Sylvie Méléard[1], Michael Rera[2], and Tristan Roget[3]

[1]École Polytechnique
[2]Sorbonne Université
[3]Université de Montpellier

May 17, 2019

Abstract

Ageing's sensitivity to natural selection has long been discussed because of its apparent negative effect on an individual's fitness. Thanks to the recently described (Smurf) 2-phase model of ageing ([40]) we propose a fresh angle for modeling the evolution of ageing. Indeed, by coupling a dramatic loss of fertility with a high-risk of impending death - amongst other multiple so-called hallmarks of ageing - the Smurf phenotype allowed us to consider ageing as a couple of sharp transitions. The birth-death model (later called bd-model) we describe here is a simple life-history trait model where each asexual and haploid individual is described by its fertility period x_b and survival period x_d. We show that, thanks to the Lansing effect, the effect through which the "progeny of old parents do not live as long as those of young parents", x_b and x_d converge during evolution to configurations $x_b - x_d \approx 0$ in finite time.

To do so, we built an individual-based stochastic model which describes the age and trait distribution dynamics of such a finite population. Then we rigorously derive the adaptive dynamics models, which describe the trait dynamics at the evolutionary time-scale. We extend the Trait Substitution Sequence with age structure to take into account the Lansing effect. Finally, we study the limiting behaviour of this jump process when mutations are small. We show that the limiting behaviour is described by a differential inclusion whose solutions $x(t) = (x_b(t), x_d(t))$ reach the diagonal $\{x_b = x_d\}$ in finite time and then remain on it. This differential inclusion is a natural way to extend the canonical equation of adaptive dynamics in order to take into account the lack of regularity of the invasion fitness function on the diagonal $\{x_b = x_d\}$.

1 Introduction

Ageing is commonly defined as an age-dependant increase of the probability to die after the maturation phase ([19]). It affects a broad range of organisms in various ways

ranging from negligible senescence to fast post-reproductive death (reviewed in [18]). In the recent years, a new 2-phases model of ageing proposed by [40] described the ageing process not as being continuous but as made of at least 2 consecutive phases separated by a dramatic transition. This transition, dubbed "Smurf transition", was first described in drosophila ([33], [34]). In short, this transition occurs in every individuals prior to death and is marked by a series of associated phenotypes encompassing high-risk of impending death, increased intestinal permeability, loss of energy stores, reduced fertility ([34]). It was later showed to be evolutionarily conserved in *Caenorhabditis elegans* and *Danio rerio* ([9], [35]). Such broad evolutionary conservation of a marker for physiological age raises the question of an active selection of the underlying mechanisms throughout evolution. Since the beginning of ageing studies, the question of its ability to appear through evolution has been raised. In fact, since the Darwinian theory of evolution stipulates that species arise and develop thanks to the natural selection of small, inherited variations that increase an individual's ability to compete, survive, and reproduce ([10]), numerous researchers suggested that ageing - and more precisely senescence - could not be actively and directly selected thanks to evolution ([13]). One of the first to publicly address the question of the evolution of ageing was August Weismann who proposed in 1881 that the life expectancy was programmed by "the needs of the species" ([43]). Numerous theoretical works have been developed about ageing for the past 60 years in order to recenter the selection of an ageing process on the individuals more than on the population. Here we will focus our attention on the capability of a process such as ageing to be selected through evolution.

If fitness alone - as an individual's reproductive success or its average contribution to the gene pool of the next generation - were at play in the evolution process, the best adapted individuals would have infinite fertility as well as longevity. Nevertheless, this situation is never observed mainly because organisms adapted to constant variations of environmental conditions and physical limitations of resources availability. Thus, an active mechanism for the elimination of these fitness-excessive individuals would represent a selective advantage in an environment where scarcity is the rule. The Lansing effect is a good candidate for such a mechanism. It is the effect through which the "progeny of old parents do not live as long as those of young parents" first described in rotifers ([22], [23]). More recently, it has been shown that older drosophila females and in some extent males tend to produce shorter lived offspring ([31]), zebra finch males give birth to offspring with shorter telomere lengths and reduced lifespans ([29]) and finally in humans, "Older father's children have lower evolutionary fitness across four centuries and in four populations" ([2]).

In the present article, we decided to approach the problem of ageing selection and evolution by using an extremely simplified version of a living organism. It is an haploid and asexual organism carrying only two traits, x_b that defines the duration of its ability to reproduce and x_d that defines the duration of its ability to maintain its integrity - stay alive (see Figure 1). We will further discuss the properties of this simple model in the next part. Although quite simple, it allows the modeling of all types of observed ageing modes : negligible senescence, sudden post-reproductive death, or post-reproductive "menopause-like" survival as well as the smurf phase.

The main result of the present article is that a pro-senescence program can be selected through Darwinian mechanisms thanks to the Lansing effect. Indeed, our main mathe-

matical result (see Theorem 4.17) shows that evolution drives the trait (x_b, x_d) towards configurations $x_b = x_d$. It means that the individuals can enjoy all their reproductive capacity, and then are quickly removed from the population. Moreover, this result shows that after reaching the configurations $x_b = x_d$, the traits x_b and x_d continue to increase with decreasing speed, while maintaining $x_b = x_d$. This decrease in the speed of evolution is a consequence of the fitness gradients being decreasing functions of the traits (see Remark 4.4). It is related to the well-known fact that the strength of selection decreases with age, i.e that a mutation having an effect on the reproduction or mortality rates at a given age will have all the more impact as this age is small ([16], [25], [17]). Indeed, in our model a perturbation of the trait x_b (resp. x_d) is equivalent to a perturbation of the birth rates at age x_b (resp. death rates at age x_d).
We build an individual based stochastic model inspired by [39]. It describes an asexual and haploid population with a continuous age and a continuous life-history trait structure. In this model, the life-history trait of every individual is thus a pair of positive numbers $(x_b, x_d) \in \mathbb{R}_+^2$. An individual with trait (x_b, x_d) reproduces at rate one as long as it is younger than x_b and cannot die as long as it is younger than x_d (see Figure 1). This

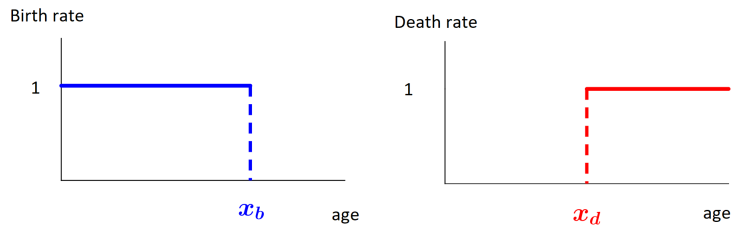

Figure 1: Life-history associated to the trait $x = (x_b, x_d)$.

model leads to three configurations: $x_d < x_b$, $x_d = x_b$ and $x_d > x_b$ (see Figure 2). From

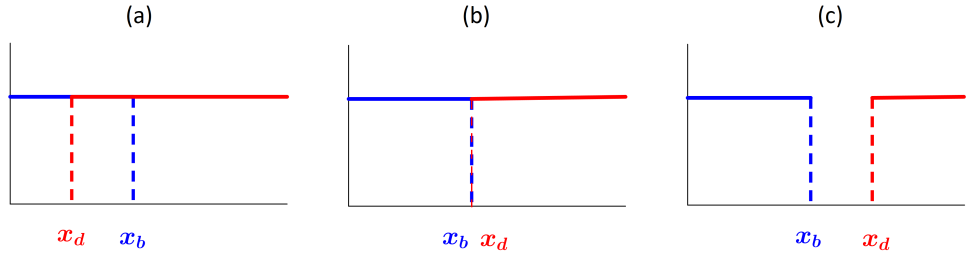

Figure 2: Three typical configurations of the model. (a) 'Too young to die' : it corresponds to configurations (x_b, x_d) which satisfies $x_d < x_b$; (b) 'Now useless': it corresponds to configurations (x_b, x_d) which satisfy $x_b = x_d$; (c) 'Menopause-like': it corresponds to configurations (x_b, x_d) which satisfies $x_d > x_b$.

one generation to the next, variation on the trait (x_b, x_d) is generated through genetic mutations. In addition, natural selection occurs through mortality due to competition for resources thanks to a logistic equation defining the maximum load of the medium. Finally, we model an epigenetic effect of senescence through the Lansing Effect. It introduces a

source of phenotypic variation at a much faster time-scale than genetic mutations. In that aim, we assume that an individual that reproduces after age x_d transmits to his descendant a shorter life expectancy (see Section 2 for details). Therefore, only individuals with trait $x_d < x_b$ are affected (see Figure 2 (c)). That creates an adaptive trade-off which impacts the phenotypic evolution of the population.
The purpose of the present article is to study the long-term evolution of the trait $x = (x_b, x_d)$ and to determine whether it concentrates on $x_b - x_d = 0$. To do so, we are inspired by the theory of adaptive dynamics ([28], [27], [11]) which studies the phenotypic eco-evolution of large populations under the assumption that genetic mutations are rare and have small effects. A central tool in that theory is the concept of *invasion fitness*. The invasion fitness is a function $1 - z(y, x)$ informally defined as the probability that an individual with trait y survives in a resident population with trait x at demographic equilibrium. In section 4, we prove that the invasion fitness satisfies the simple relation $1 - z(y, x) = (\lambda(y) - \lambda(x)) \vee 0$ where $\lambda(x)$ is the *Malthusian parameter*, describing the adaptive value associated with the trait x (see Section 3.1 (9) for the definition). This allows us to introduce the *Trait Substitution Sequence* process (TSS) which is a pure jump process describing the successive invasions of successfull mutants in monomorphic populations at the demographic equilibrium. The TSS has been heuristically introduced in [27], [11] for a trait structured population. In [5], it has been rigorously derived from an individual based model and generalised in [26], to age-structured populations. Our case differs from [26] by mainly two aspects: the additional Lansing effect and the specific form of the mutations kernel which is not absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}^2$ (as assumed in [5], [26]). In the usual case, the TSS is approximated by the solution of the *Canonical equation of adaptive dynamics* when the size of mutations is small and on a longer time-scale ([4], [7], [26]). This limit theorem requires the Lipschitz regularity of the fitness gradient. In our model this assumption is not satisfied. Nonetheless, we prove that the limiting behaviour of the TSS when mutation are small is captured by a differential inclusion, using an approach developed in [14]. A differential inclusion is an extension of ordinary differential equation to set-valued time-derivatives, which extends Cauchy-Lipschitz theory to non regular gradient cases. In our model, the gradient is smooth except on the diagonal $\{x_b = x_d\}$. We prove that the solutions are well-defined until they reach the diagonal which they do in finite time. Indeed, the drift towards the diagonal is due to on one side ($x_b < x_d$) to the fact that the individuals with larger x_b will reproduce more and thus tend to invade ; and on the other side ($x_d < x_b$) to the Lansing effect because old individuals produce short-lived offsprings. Hence, there is no advantage associated with an increasing x_b, only when associated with an increasing x_d that maximizes survival. Then the trait (x_b, x_d) evolves following the diagonal. The drift of the differential inclusion depends on the derivatives of the Malthusian parameter with respect to the trait variable. These derivatives are expressed as functions of the stable age distribution and reproductive value as in [17], [3] (see Remark 4.4).

In Section 2, we present the individual-based model. Thanks to simulations, we show what was suggested by observations: the trait distribution of the population stabilises on the diagonal $x_b - x_d = 0$.
In Section 3, we study the deterministic approximations of the stochastic dynamics under the assumption of large population and rare mutations. These approximations are nonlinear systems of partial differential equations similar to the Gurtin-McCamy Equation

([15]). We study their long-time behaviour and give some results of convergence to the stationary states.
In Section 4, we state and prove the main mathematical results of this article concerning the approximation by the TSS and the canonical inclusion of adaptive dynamics. (Theorem 4.13 and Theorem 4.17).
In Section 5, we give some comments on our results.

2 A stochastic model for the evolution of life-history traits

At time $t \geqslant 0$, the population is described by a point measure on $(\mathbb{R}_+^*)^2 \times \mathbb{R}_+$

$$Z_t^K(dx, da) = \frac{1}{K} \sum_{i=1}^{N_t^K} \delta_{(x^i(t), a^i(t))}(dx, da) \tag{1}$$

where $N_t^K = K\langle Z_t^K, 1\rangle$ is the total population size, K is the order of magnitude of a population at equilibrium, $x^i(t) = (x_b^i(t), x_d^i(t)) \in (\mathbb{R}_+^*)^2$ is the trait of the individual i and $a^i(t) \in \mathbb{R}_+$ is its age. The dynamics is defined as a piecewise deterministic Markov process which jumps as follows:

- An individual $(x, a) \in (\mathbb{R}_+^*)^2 \times \mathbb{R}_+$ reproduces at rate $\mathbb{1}_{a \leqslant x_b}$. The trait of the descendant $y = (y_b, y_d)$ is determined by the following two-steps mechanism (see Figure 3 below for an illustration):

  - Step 1: If $a \leqslant x_d$, the offspring inherits the trait $x = (x_b, x_d)$.
    Lansing Effect: if $a > x_d$, we assume that the offspring carries the trait $(x_b, 0)$.
    Let us denote by $\tilde{x}$ the trait defined as $\tilde{x} = x$ if $a \leqslant x_d$ and $\tilde{x} = (x_b, 0)$ if $a > x_d$. We observe that x_b remains unchanged and that only individuals with configurations $x_d < x_b$ are concerned by the second type $(x_b, 0)$ (see Figure 2 (c)).
  - Step 2: Genetic mutations. A mutation appears instantaneously on each trait x_b and $\tilde{x}_d$ independently with probability $p_K \in]0, 1[$. If the trait x_b mutates, the trait y_b of the descendant is $y_b = x_b + h_b$ where $h_b \in \mathbb{R}$ is chosen according to the probability measure $k(x_b, h_b)dh_b$; if the trait $\tilde{x}_d$ mutates, the trait y_d of the newborn is $y_d = \tilde{x}_d + h_d$ where $h_d \in \mathbb{R}$ is chosen according to $k(\tilde{x_d}, h_d)dh_d$, where the mutational kernel k is defined for all $u \in \mathbb{R}_+$ and $v \in \mathbb{R}$ by

$$k(u, v) = \frac{\mathbb{1}_{[0 \vee (u-1), u+1]}(u + v)e^{-\frac{v^2}{\sigma^2}}}{\int_{\mathbb{R}} \mathbb{1}_{[0 \vee (u-1), u+1]}(u + z)e^{-\frac{z^2}{\sigma^2}} dz} \tag{2}$$

    where $\sigma^2 > 0$. Note that for $u > 1$,

$$k(u, v) = k(v) = \frac{\mathbb{1}_{[-1,1]}(v)e^{-\frac{v^2}{\sigma^2}}}{\int_{\mathbb{R}} \mathbb{1}_{[-1,1]}(z)e^{-\frac{z^2}{\sigma^2}} dz}. \tag{3}$$

- An individual with trait (x, a) has a death rate $\mathbb{1}_{a>x_d} + \eta N_t^K$, with $\eta > 0$, meaning that each individual is subjected to the same competition pressure η from any individual in the population and whatever the value of its trait.
- Between jumps, individuals age at speed one: an individual with age a at time t has an age $a + s$ at time $t + s$.

Figure 3 summarizes the trait dynamics described above.

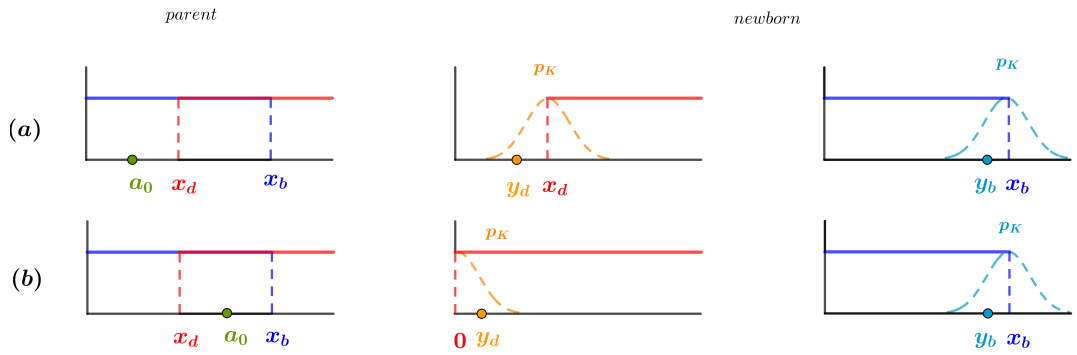

Figure 3: Picture of the reproduction and mutation mechanism. (a): the individual reproduces at age $a_0 < x_d$, there is no Lansing Effect. (b): the individual reproduces at age $a_0 > x_d$, the Lansing Effect acts.

Numerical simulation. The pictures in Figure 4 represent a simulation of the trait marginals dynamics of the process $Z_t^K(dx, da)$. We consider a monomorphic initial population with trait $x = (1.2, 2.5)$ and $N_0^K = 10000$. We consider a competition rate $\eta = 0.0005$, a probability of mutation $p_K = 0.05$ and a variance of mutations $\sigma = 0.05$. At time $t = 0$, the population is monomorphic with trait $(x_b, x_d) = (1.2, 2.5)$. We observe that before the trait x_b reaches the value of 2.5, the trait x_d remains constant. When x_b reaches $x_d \simeq 2.5$, the traits x_b and x_d continue to increase by maintaining $x_b \simeq x_d$.

3 Monomorphic and bimorphic deterministic dynamics

In this section, we study a deterministic approximation of the process Z^K when K goes to infinity. We also assume that p_K goes down to zero: almost no mutation occurs on a time interval $[0, T]$. Nonetheless, some phenotypic variation is created by Lansing Effect. Since our model is density-dependent, the deterministic approximation is a system of classical non-linear partial differential equations similar to the Gurtin MacCamy Equation ([15]). In the monomorphic case, we show that the dynamics converges to the unique non-trivial equilibrium. In the bimorphic case, we show the convergence to a monomorphic equilibrium. We will consider that a monomorphic population with trait x is composed of two subpopulations with traits (x_b, x_d) and $(x_b, 0)$.

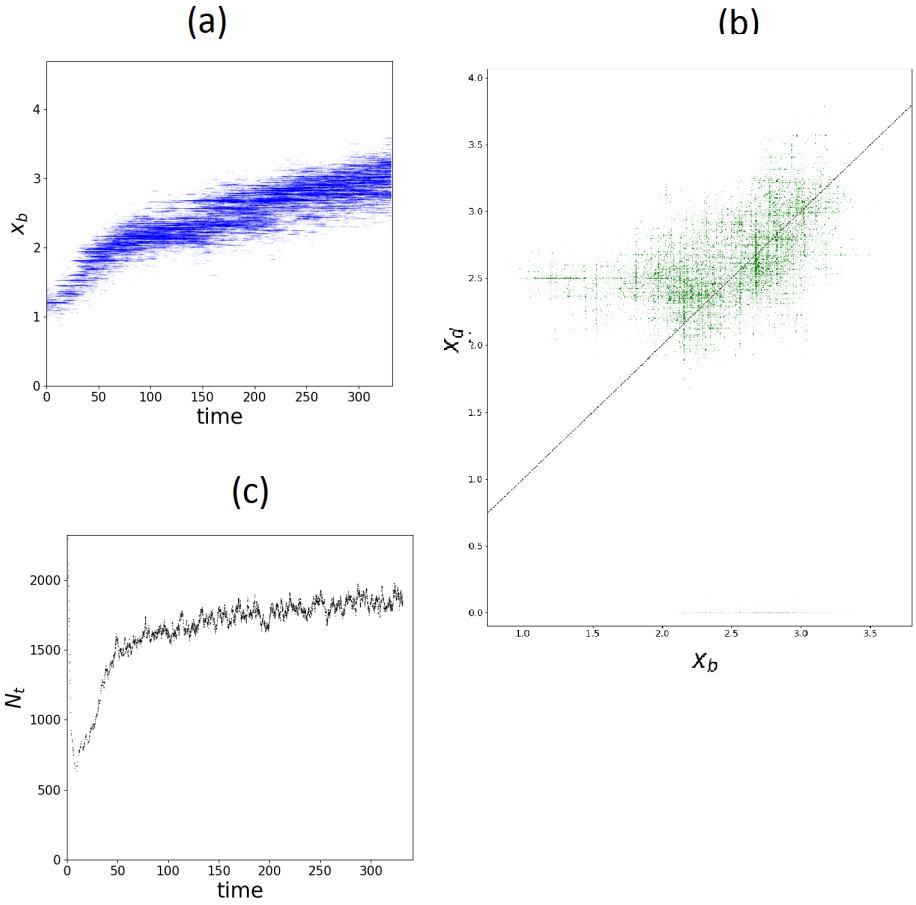

Figure 4: Simulation of the individual based model (see the script in Appendix A.4) . (a): Dynamics of the trait x_b as a function of time. (b): Dynamics of the trait x_d as a function of the trait x_b. (c): Population size as a function of time. Parameters: $N_0^K = 10000$ individuals with trait $(1.2, 2.5)$, $\eta = 0.0005$, $p = 0.05$, $\sigma = 0.1$.

Notation 3.1. *We define* $\boldsymbol{I} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$.

3.1 Monomorphic dynamics

Let $x = (x_b, x_d) \in (\mathbb{R}_+^*)^2$ be a phenotypic trait. We consider a monomorphic initial population such that the sequence $(Z_0^K)_K$ weakly converges as $K \to \infty$ to $\delta_x n_x(a)da$ which describes a monomorphic population with trait x. Then, as in [39], we can prove that the sequence of processes $(Z^K)_K$ converges in probability on any finite time interval to the weak solution $(n_x(t,.), t \geqslant 0) = ((n_x^1(t,.), n_x^2(t,.)), t \geqslant 0) \in C(\mathbb{R}_+, L^1(\mathbb{R}_+)^2)$ of the following system of partial differential equations

$$\begin{cases} \partial_t n_x(t,a) + \partial_a n_x(t,a) = -\left(\mathbf{D}_x(a) + \eta \|n_x(t,.)\|_1 \mathbf{I}\right) n_x(t,a) \\ n_x(t,0) = \int_{\mathbb{R}_+} \mathbf{B}_x(\alpha) n_x(t,\alpha) d\alpha \end{cases} \tag{4}$$

with initial condition given by $n_x(0,a) = n_x(a)$, where the densities $n_x^1(t,.)$ and $n_x^2(t,.)$ describe the population distributions with trait (x_b, x_d) and $(x_b, 0)$ respectively;

$$\|n_x(t,.)\|_1 = \sum_{i\in\{1,2\}} \int_{\mathbb{R}_+} |n_x^i(t,\alpha)| d\alpha$$

is the total population size and

$$\mathbf{B}_x(a) = \begin{pmatrix} \mathbb{1}_{a\leqslant x_b \wedge x_d} & 0 \\ \mathbb{1}_{x_d < a \leqslant x_b} & \mathbb{1}_{a\leqslant x_b} \end{pmatrix}, \quad \mathbf{D}_x(a) = \begin{pmatrix} \mathbb{1}_{a>x_d} & 0 \\ 0 & 1 \end{pmatrix} \tag{5}$$

are the birth and death interactions. We refer to [42] for the well-posedness theory of $L^1(\mathbb{R}_+)^2$ solutions of Equation (4).

Remark 3.2. *Let us comment the different terms in Equation (4). The transport term on the left-hand side of the first equation describes the aging of the individuals, the right-hand side describes the death of the individuals. The renewal condition in $a = 0$ describes the births (i.e the individuals with age 0).*

We introduce the set of viable traits

$$\mathcal{V} = \{x = (x_b, x_d) \in (\mathbb{R}_+^*)^2 : x_b \wedge x_d > 1\}. \tag{6}$$

We show that for any trait $x \in \mathcal{V}$, there exists a unique non-trivial and globally stable stationary state of (4). Indeed, the quantity $x_b \wedge x_d$ represents the mean number of descendants with trait (x_b, x_d) of an individual with trait (x_b, x_d) if there is no competition.

Proposition 3.3. *Assume that $x \in \mathcal{V}$. There exists a unique non-trivial stationary solution $\overline{n}_x \in L^1(\mathbb{R}_+)^2$ to Equation (4). Moreover, any $L^1(\mathbb{R}_+)^2$ non negative solution $n_x(t,.)$ of (4) such that $n_x^1(t,.) \neq 0$ converges to $\overline{n}_x$ in $L^1(\mathbb{R}_+)^2$ as $t \to \infty$.*

Remark 3.4. *In Proposition 3.7 below, we will give an explicit expression for the stationary state $\overline{n}_x$.*

The proof of Proposition 3.3 is based on the study of the associated linear dynamics. We introduce the linear operator

$$\begin{array}{rcl} A : D(A) \subset L^1(\mathbb{R}_+)^2 & \to & L^1(\mathbb{R}_+)^2 \\ u & \mapsto & -u' - \mathbf{D}_x(a)u \end{array} \tag{7}$$

where $D(A) = \{u \in L^1(\mathbb{R}_+)^2 : u \text{ abs. cont.}, u' \in L^1(\mathbb{R}_+)^2,\ u(0) = \int_{\mathbb{R}_+} \mathbf{B}_x(\alpha)u(\alpha)d\alpha\}$. It is well known that A is the infinitesimal generator of a strongly continuous semigroup of linear operators [42, Proposition 3.7] which describes the solutions of the linear system of McKendrick Von-Foerster Equation

$$\begin{cases} \partial_t v_x(t,a) + \partial_a v_x(t,a) = -\mathbf{D}_x(a)v_x(t,a) \\ v_x(t,0) = \int_{\mathbb{R}_+} \mathbf{B}_x(\alpha)v_x(t,\alpha)d\alpha. \end{cases} \tag{8}$$

In [8], a similar linear model is studied. The "entropy method" ([30]) allows the authors to prove the convergence of the normalised solutions to some stable distribution in some

weighted L^1-space. We need stronger convergence in order to study the long-time behaviour of the masses of the solutions of (4). Since the birth matrix $\mathbf{B}_x$ is not irreducible (it is triangular) and the parameters $\mathbf{B}_x$ and $\mathbf{D}_x$ are not smooth, we cannot apply Theorems 4.9 and 4.11 in [42]. Nonetheless, we easily extend them to our reducible and non-smooth setting.

We define the Malthusian parameter $\lambda(x)$ associated with some trait x as the unique solution of the equation

$$\int_0^{x_b \wedge x_d} e^{-\lambda(x)a} da = 1. \tag{9}$$

Proposition 3.5 below justifies this definition and shows that $\lambda(x)$ is the asymptotic growth rate of the dynamics defined by (8). Its proof is presented in the Appendix.
Let us define for all $z \in \mathbb{C}$ the 2×2 matrix

$$\mathbf{F}(z) = \int_{\mathbb{R}_+} \mathbf{B}_x(a) \exp\left(- \int_0^a (\mathbf{D}_x(\alpha) + z\mathbf{I}) d\alpha \right) da. \tag{10}$$

Note that it is well-defined since $\mathbf{B}_x$ has compact support.

Proposition 3.5. *Assume that $x \in \mathcal{V}$. Then the linear operator A admits a unique pair of simple principal eigenelements $(\lambda(x), N_x) \in \mathbb{R}_+^* \times D(A)$ where the stable age distribution N_x satisfies*

$$N_x^1(a) = e^{-(\lambda(x)a + (a - x_d) \vee 0)}, \quad N_x^2(a) = \frac{[\boldsymbol{F}(\lambda(x))]_{21}}{1 - [\boldsymbol{F}(\lambda(x))]_{22}} e^{-(1+\lambda(x))a}.$$

Moreover, for any non-negative solution $v_x(t,a)$ of (8) in $L^1(\mathbb{R}_+)^2$, there exists a positive constant $c(v_x(0,.))$ such that $e^{-\lambda(x)t} v_x(t,.) \to c(v_x(0,.)) N_x$ in $L^1(\mathbb{R}_+)^2$ as $t \to +\infty$.

Let us now give a lemma which will be used to study the long-time behaviour of the masses of the solutions of (4) and whose proof is postponed to the Appendix.

Lemma 3.6. *Let $(m_{11}, m_{12}, m_{22}) \in \mathbb{R}_+^* \times \mathbb{R}_+ \times \mathbb{R}_-^*$. Let $\mathcal{D}_{11}(t), \mathcal{D}_{12}(t), \mathcal{D}_{22}(t)$ be continuous functions from $\mathbb{R}_+$ to $\mathbb{R}$ which tend to zero as $t \to \infty$. Let us denote*

$$M = \begin{pmatrix} m_{11} & 0 \\ m_{21} & m_{22} \end{pmatrix}, \quad \mathcal{D}(t) = \begin{pmatrix} \mathcal{D}_{11}(t) & 0 \\ \mathcal{D}_{21}(t) & \mathcal{D}_{22}(t) \end{pmatrix}.$$

Then any solution $(z(t), t \geqslant 0)$ of the equation

$$\frac{dz(t)}{dt} = (M + \mathcal{D}(t)) z(t) - \eta \|z(t)\|_1 \; z(t) \tag{11}$$

started at $z(0) \in \mathbb{R}_+^* \times \mathbb{R}_+$ converges to a vector $\overline{z}$ which satisfies

$$\overline{z}_1 = \frac{m_{11}}{\eta} \frac{1}{1 + \frac{m_{21}}{m_{11} - m_{22}}}, \quad \overline{z}_2 = \frac{m_{21}}{m_{11} - m_{22}} \overline{z}_1. \tag{12}$$

We conclude this section by proving Proposition 3.3.

of Proposition 3.3. We prove the first assertion. Let $x \in \mathcal{V}$ and let $\lambda(x)$ be the principal eigenvalue of A given by Proposition 3.5. Let $\overline{n}_x$ be the (unique) principal eigenvector of A which satisfies $\eta \|\overline{n}_x\|_1 = \lambda(x)$. It is obvious that $\overline{n}_x$ is a non-trivial stationary state of (4). Reciprocally, let $\overline{n}$ be a stationary state of (4). Then we have necessarily $\lambda(x) = \eta \|\overline{n}\|_1$ and that $\overline{n}$ is an eigenvector of A associated with the eigenvalue $\lambda(x)$ that allows us to conclude. We now study the long-time behaviour of the solutions. Let us define

$$v_x(t,a) = \exp\left(\eta \int_0^t \|n_x(s,.)\|_1 ds\right) n_x(t,a).$$

It is straightforward to prove that v_x is a solution of the linear equation (8). By Proposition 3.5 we have $e^{-\lambda(x)t} v_x(t,.) \to c(v_x(0,.))N_x$ in $L^1(\mathbb{R}_+)^2$ as $t \to \infty$. We deduce that for $i \in \{1,2\}$ and denoting $\rho_x^i(t) = \|n_x^i(t,.)\|_1$,

$$\frac{n_x^i(t,.)}{\rho_x^i(t)} = \frac{e^{-\lambda(x)t} v_x^i(t,.)}{\int_{\mathbb{R}_+} e^{-\lambda(x)t} v_x^i(t,\alpha) d\alpha} \to \frac{N_x^i}{\int_{\mathbb{R}_+} N_x^i(\alpha) d\alpha} \tag{13}$$

in $L^1(\mathbb{R}_+)$ as $t \to +\infty$. We now study the behaviour of the masses $\rho_x(t)$. By taking the derivative under the integral, we obtain that for $i \in \{1,2\}$

$$\begin{aligned}
\frac{\mathrm{d}\rho_x^i(t)}{\mathrm{dt}} &= \sum_{j=1}^{2} \int_{\mathbb{R}_+} \left([\mathbf{B}_x(\alpha)]_{ij} - [\mathbf{D}_x(\alpha)]_{ij}\right) n_x^j(t,\alpha) d\alpha - \eta \rho_x^i(t) \|\rho_x(t)\|_1 \\
&= \sum_{j=1}^{2} \rho_x^j(t) \int_{\mathbb{R}_+} \left([\mathbf{B}_x(\alpha)]_{ij} - [\mathbf{D}_x(\alpha)]_{ij}\right) \frac{n_x^j(t,\alpha)}{\rho_x^j(t)} d\alpha - \eta \rho_x^i(t) \|\rho_x(t)\|_1.
\end{aligned}$$

Hence we obtain by (13) that

$$\frac{\mathrm{d}\rho_x(t)}{\mathrm{dt}} = (\mathbf{A} + \mathcal{A}(t))\rho_x(t) - \eta \|\rho_x(t)\|_1 \rho_x(t)$$

where $\mathbf{A} = (a_{ij})$ and for $(i,j) \in \{1,2\}^2$,

$$a_{ij} = \int_{\mathbb{R}_+} \left([\mathbf{B}_x(\alpha)]_{ij} - [\mathbf{D}_x(\alpha)]_{ij}\right) \frac{N_x^j(\alpha)}{\int_{\mathbb{R}_+} N_x^j} d\alpha, \tag{14}$$

and $\mathcal{A}(t)$ is a continuous function decreasing to zero as t tends to infinity. Since $a_{11} = \lambda(x) > 0$, $a_{21} \geqslant 0$, $a_{12} = 0$ and $a_{22} < 0$, Lemma 3.6 allows us to conclude that $\rho_x(t)$ converges to $\overline{\rho}_x$, which is defined as the unique solution of the equation $A\overline{\rho}_x - \eta \|\overline{\rho}_x\|_1 \overline{\rho}_x = 0$. We easily solve this system and we obtain that $\overline{\rho}_x$ satisfies

$$\overline{\rho}_x^1 = \frac{\lambda(x)}{\eta} \frac{1}{1 + \frac{a_{21}}{\lambda(x) - a_{22}}}, \quad \overline{\rho}_x^2 = \frac{a_{21}}{\lambda(x) - a_{22}} \overline{\rho}_x^1. \tag{15}$$

□

We conclude this section by writing more explicit formulas for the stationary state $\overline{n}_x$.

Proposition 3.7. *Let $x \in \mathcal{V}$ then we have*

$$\overline{n}_x^1(a) = \overline{\rho}_x^1 \frac{N_x^1(a)}{\int_{\mathbb{R}_+} N_x^1(\alpha)d\alpha}, \quad \overline{n}_x^2(a) = \overline{\rho}_x^2 \frac{N_x^2(a)}{\int_{\mathbb{R}_+} N_x^2(\alpha)d\alpha}$$

where

$$\overline{\rho}_x^1 = \frac{\lambda(x)}{\eta} \frac{1}{1 + \frac{a_{21}}{\lambda(x) - a_{22}}}, \quad \overline{\rho}_x^2 = \frac{a_{21}}{\lambda(x) - a_{22}} \overline{\rho}_x^1,$$

N_x is defined in Proposition 3.5 and a_{ij} are defined in (14).

Proof. It is a direct consequence of (13) and (15). □

Biological interpretation 3.8. *Equation (4) describes the dynamics of a large monomorphic population with trait x. Proposition 3.3 shows that the age distribution of the population stabilizes around the equilibrium $\overline{n}_x = (\overline{n}_x^1, \overline{n}_x^2)$. The equilibria $\overline{n}_x^1$ and $\overline{n}_x^2$ describe the age equilibria of the population with trait (x_b, x_d) and $(x_b, 0)$ respectively. We observe that if $x_b < x_d$, then $a_{21} = 0 = \overline{\rho}_x^2$ and Proposition 3.7 leads to the equilibrium $\overline{n}_x = (\overline{n}_x^1, 0)$.*

3.2 Bimorphic dynamics

Let $x = (x_b, x_d) \in \mathcal{V}$ and $y = (y_b, y_d) \in \mathcal{V}$. We consider a bimorphic initial population such that the sequence $(Z_0^K)_K$ weakly converges to $\delta_x n_x(a)da + \delta_y n_y(a)da$ as K tends to infinity. Using similar arguments as in [39], we can prove that the sequence of processes $(Z^K)_K$ converges in probability, on any finite time interval, to the solution $((n_x(t,.), n_y(t,.)), t \geqslant 0) \in L^1(\mathbb{R}_+)^4$ of the following system of non-linear partial differential equations

$$\begin{cases} \partial_t n_x(t,a) + \partial_a n_x(t,a) = -\left(\mathbf{D}_x(a) + \eta(\|n_x(t,.)\|_1 + \|n_y(t,.)\|_1)\mathbf{I}\right) n_x(t,a) \\ n_x(t,0) = \int_{\mathbb{R}_+} \mathbf{B}_x(\alpha) n_x(t,\alpha)d\alpha \\ \partial_t n_y(t,a) + \partial_a n_y(t,a) = -\left(\mathbf{D}_y(a) + \eta(\|n_x(t,.)\|_1 + \|n_y(t,.)\|_1)\mathbf{I}\right) n_y(t,a) \\ n_y(t,0) = \int_{\mathbb{R}_+} \mathbf{B}_y(\alpha) n_y(t,\alpha)d\alpha, \end{cases} \tag{16}$$

with initial condition given by $(n_x(0,a), n_y(0,a)) = (n_x(a), n_y(a))$. Equations (16) describe the dynamics of a dimorphic population with traits x and y interacting through competition. We prove the following proposition.

Proposition 3.9. *Let $x, y \in \mathcal{V}$ such that $x_b \wedge x_d < y_b \wedge y_d$. Then any $L^1(\mathbb{R}_+)^4$-solution $(n_x(t,.), n_y(t,.))$ of (16) satisfying $n_y^1(t,.) \neq 0$ converges to $(0, \overline{n}_y)$ in $L^1(\mathbb{R}_+)^4$ as $t \to \infty$.*

Proof. Let $x, y \in \mathcal{V}$ such that $x_b \wedge x_d < y_b \wedge y_d$. Then by (9), we have $\lambda(x) < \lambda(y)$. For $u \in \{x, y\}$ we define

$$v_u(t,a) = \exp\left(\eta \int_0^t (\|n_x(s,.)\|_1 + \|n_y(s,.)\|_1)ds\right) n_u(t,a).$$

The functions v_x and v_y are solutions of the linear systems (8). We deduce from Proposition 3.5 that $e^{-\lambda(u)t} v_u(t,.) \to c(v_u(0,.))N_u$ in $L^1(\mathbb{R}_+)^2$ as $t \to +\infty$, for a positive constant $c(v_u(0,.))$. Hence, we obtain that $e^{-\lambda(x)t} \int_{\mathbb{R}_+} v_x^1(t,\alpha)d\alpha$ converges to a positive limit. Since

$\lambda(y) > \lambda(x)$ it comes that $e^{-\lambda(y)t}\int_{\mathbb{R}_+} v_x^1(t,\alpha)d\alpha$ converges to 0 as $t \to +\infty$. We deduce that

$$\frac{\rho_y^1(t)}{\rho_x^1(t)} = \frac{e^{-\lambda(y)t}\int_{\mathbb{R}_+} v_y^1(t,\alpha)d\alpha}{e^{-\lambda(y)t}\int_{\mathbb{R}_+} v_x^1(t,\alpha)d\alpha} \to +\infty.$$

Since $\rho_y^1(t)$ is bounded we deduce that $\rho_x^1(t) \to 0$ and similarly that $\rho_x^2(t) \to 0$. Then the population with trait x becomes extinct. Using similar arguments as in the previous proof we obtain that $n_y(t,.) \to \overline{n}_y$ in $L^1(\mathbb{R}_+)^2$ which allows us to conclude. □

Biological interpretation 3.10. *Equation (16) describes a competition dynamics between two infinite monomorphic populations with trait x and y. Proposition 3.9 shows that if $x_b \wedge x_d < y_b \wedge y_d$, then the population with trait y invades and becomes fixed while the population with trait x becomes extinct. That gives us an invasion-implies-fixation criterion.*

4 Adaptive dynamics analysis

In this section, we study the model introduced in Section 2 under the different scaling of the adaptive dynamics. We generalise the Trait Substitution Sequence with age structure ([26]) to take into account the Lansing Effect. Then we study the behaviour of the TSS on a large time-scale when mutations are small. We show that the limiting behaviour of the TSS is described by a differential inclusion which generalises the canonical equation of adaptive dynamics ([11], [4], [7]) to non-regular fitness functions. We first state some properties of the demographic parameters and introduce the invasion fitness function.

4.1 Malthusian parameter and invasion fitness

Let us introduce the following sets: $U_1 = \{x \in \mathcal{V} : x_b < x_d\}$, $U_2 = \{x \in \mathcal{V} : x_d < x_b\}$ and $\mathcal{H} = \{x \in \mathcal{V} : x_b = x_d\}$.

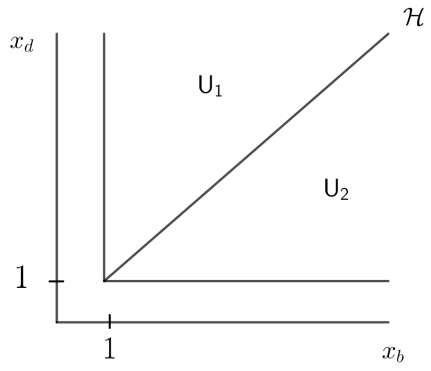

Figure 5: Picture of $\mathcal{V} = U_1 \cup \mathcal{H} \cup U_2$.

4.1.1 Malthusian parameter

We now give some properties of the Malthusian parameter $\lambda(x)$ defined in (9).

Proposition 4.1. *(i) For all $x \in \mathcal{V}$, $0 \leqslant \lambda(x) < 1$.*

(ii) The map $x \in \mathcal{V} \mapsto \lambda(x)$ is continuous. It is differentiable on $U_1 \cup U_2$ and satisfies

$$\begin{aligned} \forall x \in U_1, \quad \nabla\lambda(x) &= \left(\frac{e^{-\lambda(x)x_b}}{G(x)}, 0\right) \\ \forall x \in U_2, \quad \nabla\lambda(x) &= \left(0, \frac{e^{-\lambda(x)x_d}}{G(x)}\right) \end{aligned} \tag{17}$$

where $G(x) = \int_0^{x_b \wedge x_d} a e^{-\lambda(x)a} da$.

(iii) We have $\sup_{x \in U_1 \cup U_2} \|\nabla\lambda(x)\| < +\infty$. Moreover, for all $i \in \{1, 2\}$, the fitness gradient $\nabla\lambda$ is Lipschitz on U_i.

Remark 4.2. *Note that the Malthusian parameter $\lambda(x)$ is not differentiable on the diagonal $\mathcal{H}$, which can be easily obtained by computing left and right partial derivatives on $\mathcal{H}$.*

Notation 4.3. *For all $h \in \mathbb{R}$ we define $(h)_1 = \begin{pmatrix} h \\ 0 \end{pmatrix}$ and $(h)_2 = \begin{pmatrix} 0 \\ h \end{pmatrix}$.*

Proof. (i) Since $x_b \wedge x_d > 1$ we obtain from the definition (9) that $\lambda(x) > 0$. Assume that $\lambda(x) \geqslant 1$. Then we obtain that $1 = \int_0^{x_b \wedge x_d} e^{-\lambda(x)a} da \leqslant 1 - e^{-x_b \wedge x_d}$ which is a contradiction. (ii) We prove the continuity. Let $x \in \mathcal{V}$ and let (x_n) be a sequence of $\mathcal{V}$ such that $x^n \to x$. By (i), λ is bounded and we can extract a subsequence (still denoted x^n by simplicity) such that $\lambda(x^n) \to \lambda^*$. We deduce that $1 = \int_0^{x_b^n \wedge x_d^n} e^{-\lambda(x^n)a} da \to \int_0^{x_b \wedge x_d} e^{-\lambda^* a} da = 1$ which allows us to conclude. Differentiability properties are a direct consequence of the Implicit Function Theorem. For each $i \in \{1, 2\}$ we apply implicit function Theorem to the map

$$(\lambda, x) \in \mathbb{R} \times U_i \mapsto \int_0^{x_b \wedge x_d} e^{-\lambda a} da - 1.$$

We deduce that λ is differentiable over $U_1 \cup U_2$ and that

$$\forall i \in \{1, 2\}, \quad \forall x \in U_i, \quad \nabla\lambda(x) = \left(\frac{e^{-\lambda(x)(x_b \wedge x_d)}}{G(x)}\right)_i.$$

(iii) It is straightforward to check that for all $x \in \mathcal{V}$,

$$\frac{e^{-\lambda(x)(x_b \wedge x_d)}}{G(x)} \leqslant \frac{1}{\int_0^1 a e^{-a} da}$$

which allows us to obtain that $\sup_{x \in U_1 \cup U_2} \|\nabla\lambda(x)\| < +\infty$. Moreover, the gradient $\nabla\lambda$ is obviously differentiable on U_i. Since G is bounded below by $\int_0^1 a e^{-a} da$, we deduce that $\nabla\lambda$ has bounded derivatives on U_i and that $\nabla\lambda$ is Lipschitz on U_i. □

Remark 4.4. *Formulae (17) describe the sensitivity of the Malthusian parameter to small variations of the trait x as well as the strength of selection at ages x_b and x_d for a population with Lansing effect. The quantity $G(x)$ can be interpreted as the mean generation time associated with the trait x. Moreover (17) and Proposition 3.5 yield*

$$\forall i \in \{1,2\}, \quad \forall x \in U_i, \quad \nabla\lambda(x) = \left(\frac{N_x^1(x_b \wedge x_d)}{G(x)}\right)_i \tag{18}$$

where N_x^1 is the stable age distribution. In [3], Caswell obtains similar formulae for derivatives of the Malthusian parameter with respect to some little perturbations on the intensity of birth or death at some given age while our formulae are obtained considering a small perturbation on the duration of the reproduction phase (not on the intensity which remains constant equal to one).

The following proposition recalls a simple link between the Malthusian parameter and the stationary state of the monomorphic partial differential equation (4).

Proposition 4.5. *(i) For all $x \in \mathcal{V}$, we have $\lambda(x) = \eta \, \|\overline{n}_x\|_1$.*

(ii) The map $x \in \mathcal{V} \mapsto \overline{n}_x(0)$ is continuous and bounded.

Proof. (i) has been proved at the beginning of the proof of Proposition 3.3.
(ii) By (i) we have

$$\lambda(x) = \eta \left(\overline{n}_x^1(0) u_1(x) + \overline{n}_x^2(0) u_2(x)\right) \tag{19}$$

where

$$u_1(x) = \int_0^{+\infty} \exp\left(-\int_0^a \mathbb{1}_{\alpha > x_d} d\alpha - \lambda(x) a\right) da \ ; \ u_2(x) = \int_0^{+\infty} \exp\left(-(1+\lambda(x))a\right) da.$$

Moreover $\overline{n}_x(0)$ is a solution of

$$\overline{n}_x(0) = \mathbf{F}(\lambda(x)) \, \overline{n}_x(0) \tag{20}$$

where $\mathbf{F}$ is defined in (10). From (19) and (20) we obtain by simple computation that

$$\overline{n}_x^1(0) = \frac{\lambda(x)}{\eta} \frac{1}{u_1(x) + u_2(x) \frac{[\mathbf{F}(\lambda(x))]_{21}}{1 - [\mathbf{F}(\lambda(x))]_{22}}}$$

which is a continuous function of x. Boundedness is obvious arguing that $\overline{n}_x^1(0) \leqslant \|\overline{n}_x\|_1$. □

4.1.2 Invasion fitness

We extend the definition of the invasion fitness for age-structured populations introduced in [26, Section 3] to take into account the Lansing effect.

Definition 4.6. *For all $y \in \mathbb{R}_+^*$ and $x \in \mathcal{V}$, the invasion fitness $1 - z(y,x)$ is defined as the survival probability of a bi-type age structured branching process with birth rates and death rates defined in (5), respectively equal to*

$$\boldsymbol{B}_y(a) \quad \text{and} \quad \boldsymbol{D}_y(a) + \eta \, \|\overline{n}_x\| \boldsymbol{I}.$$

The next proposition gives a precise and precious relation between the invasion fitness and the Malthusian parameter.

Proposition 4.7. *Let $y \in \mathbb{R}_+^*$ and $x \in \mathcal{V}$, then the invasion fitness satisfies*

$$1 - z(y,x) = (\lambda(y) - \lambda(x)) \vee 0. \tag{21}$$

Proof. Let $Z_t(da) = (Z_t^1(da), Z_t^2(da))$ be an age- structured branching process with birth rates $\mathbf{B}_y(a)$ and death rates $\mathbf{D}_y(a) + \eta \|\overline{n}_x\|\mathbf{I} = \mathbf{D}_y(a) + \lambda(x)\mathbf{I}$. The process Z_t becomes extinct if and only if the process Z_t^1 becomes extinct. Indeed, if $Z_0^1 = 0$, the process Z_t^2 evolves as a sub-critical branching process. The process Z_t^1 is an age structured branching process with birth rates and death rates respectively

$$[\mathbf{B}_y(a)]_{11} = \mathbb{1}_{a \leqslant y_b \wedge y_d} \quad \text{and } [\mathbf{D}_y(a)]_{11} + \lambda(x) = \mathbb{1}_{a > y_d} + \lambda(x).$$

We deduce that $z(y,x)$ equals the smallest solution of the equation $z = F(z)$ where

$$\begin{aligned}
F(z) &= \int_{\mathbb{R}_+} e^{(z-1)\int_0^a [\mathbf{B}_y(\alpha)]_{11} d\alpha} \left([\mathbf{D}_y(a)]_{11} + \lambda(x)\right) e^{-\int_0^a \left([\mathbf{D}_y(\alpha)]_{11} + \lambda(x)\right) d\alpha} da \\
&= \int_0^{y_b \wedge y_d} \lambda(x) e^{(z-1-\lambda(x))a} da + e^{(z-1)(y_b \wedge y_d)} \int_{y_b \wedge y_d}^{y_d} \lambda(x) e^{-\lambda(x)a} da \\
&+ e^{(z-1)(y_b \wedge y_d)} (1 + \lambda(x)) \int_{y_d}^{+\infty} e^{-(a-y_d)-\lambda(x)a} da \\
&= 1 - e^{(z-1-\lambda(x))(y_b \wedge y_d)} + (z-1) \int_0^{y_b \wedge y_d} e^{(z-1-\lambda(x))a} da \\
&+ e^{(z-1)y_b \wedge y_d} \left(e^{-\lambda(x)(y_b \wedge y_d)} - e^{-\lambda(x) y_d}\right) + e^{(z-1)(y_b \wedge y_d)} + e^{(z-1)(y_b \wedge y_d)} e^{-\lambda(x) y_d} \\
&= 1 + (z-1) \int_0^{y_b \wedge y_d} e^{(z-1-\lambda(x))a} da.
\end{aligned}$$

We have obtained that the equation $z = F(z)$ is equivalent to

$$z - 1 = (z-1) \int_0^{y_b \wedge y_d} e^{(z-1-\lambda(x))a} da. \tag{22}$$

If $\lambda(y) > \lambda(x)$, Equation (22) admits two solutions $z = 1$ and $z = \lambda(x) - \lambda(y) + 1 < 1$. If $\lambda(y) < \lambda(x)$, Equation (22) admits a unique solution $z = 1$. That allows us to conclude the proof. □

Remark 4.8. *It is interesting to note that in our model, the invasion fitness (which is a concept from adaptive dynamics theory) and the Malthusian parameter are connected thanks to the simple relation (21).*

The following proposition characterises the set of traits y which can invade some given trait x.

Proposition 4.9. *For all $x \in \mathcal{V}$, $y \in \mathbb{R}_+^2$,*

$$1 - z(y,x) > 0 \Longleftrightarrow \lambda(y) > \lambda(x) \Longleftrightarrow y_b \wedge y_d > x_b \wedge x_d.$$

Proof. From the definition (9) of the Malthusian parameter, we easily deduce the following equivalences:

$$\begin{aligned}(\lambda(y)-\lambda(x))\vee 0>0 &\iff \lambda(y)-\lambda(x)>0\\ &\iff \int_0^{x_b\wedge x_d} e^{-\lambda(y)a}da<1\\ &\iff x_b\wedge x_d<y_b\wedge y_d,\end{aligned}$$

which concludes the proof. □

4.2 Trait Substitution Sequence with age structure

We first generalise the definition of the Trait Substitution Sequence (TSS) with age structure defined in [26] to take into account the Lansing Effect.

Definition 4.10. *We define the measure valued process $(T_t(dx,da), t\geqslant 0)$ by*

$$T_t(dx,da)=\delta_{X(t)}(dx)\overline{n}^1_{X(t)}(a)da+\delta_{(X_b(t),0)}(dx)\overline{n}^2_{X(t)}(a)da$$

where $(X(t), t\geqslant 0)=((X_b(t),X_d(t)), t\geqslant 0)$ is defined as the pure jump Markov process on $\mathcal{V}$ with infinitesimal generator L defined for all measurable and bounded function $\varphi:\mathcal{V}\to\mathbb{R}$ and $x\in\mathcal{V}$ by

$$L\varphi(x)=\int_{(\mathbb{R}_+)^2}(\varphi(x+h)-\varphi(x))(\lambda(x+h)-\lambda(x))\ \overline{n}^1_x(0)\ \mu(dh) \tag{23}$$

where

$$\mu(dh)=\frac{k(h_b)dh_b\otimes\delta_0(dh_d)+\delta_0(dh_b)\otimes k(h_d)dh_d}{2},$$

k being defined in (3).
The process X will be called the Trait Substitution Sequence.

Remark 4.11. *The process $(T_t, t\geqslant 0)$ describes the evolution of the phenotypic structure of the population at the mutational time-scale. At each time, and because of the Lansing effect, the population is composed of two sub-populations: the first one corresponds to viable individuals with trait $X(t)=(X_b(t),X_d(t))\in\mathcal{V}$ whose age distribution is given by $\overline{n}^1_{X(t)}(a)da$; the second one is composed of individuals generated through the Lansing effect, with trait $(X_b(t),0)$ and age distribution $\overline{n}^2_{X(t)}(a)da$.*

Remark 4.12. *Figure 6 describes the behaviour of the process $(X(t), t\geqslant 0)$.*
Any trait $x\in\mathcal{H}$ is an absorbing state for the process X. Indeed, by Proposition 4.9, for all $\varphi:\mathcal{V}\to\mathbb{R}$ measurable and bounded, for all $x\in\mathcal{H}$, $L\varphi(x)=0$. That means that when the trait (x_b,x_d) satisfies $x_b=x_d$, no mutation can invade.
By definition of the measure $\mu(dh)$ the process evolves horizontally or vertically (which means that the two traits x_b and x_d do not mutate simultaneously). By Proposition 4.9, we deduce easily that the process evolves from the left to the right on U_1 and from bottom to top on U_2.
Since the jump rates are continuous and tend to zero on $\mathcal{H}$, the process slows down as it approaches $\mathcal{H}$.

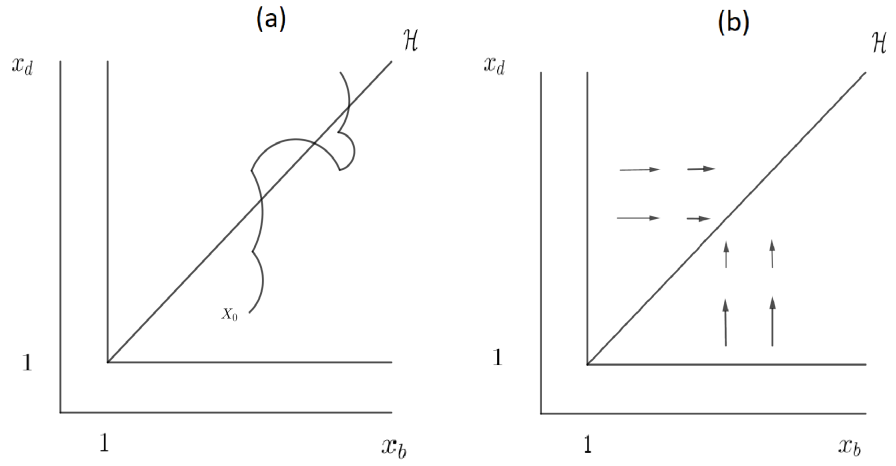

Figure 6: (a): This picture represents a trajectory of the TSS process. (b): This picture represents the drift associated with the TSS process.

We now explain the heuristics, rigorously proved in [26], which allow to obtain the TSS from the individual based model defined in Section 2.1. The main ideas have been introduced in [5], for a population without age-structure. They are based on the time-scale separation assumption on mutation probability p_K: as $K \to +\infty$

$$\forall V > 0, \quad \exp(-KV) = o(p_K), \quad p_K = o\left(\frac{1}{K\log(K)}\right), \tag{24}$$

which allows to separate the effect of the natural selection and the appearance of new mutants. Let $x \in \mathcal{V}$ and consider a sequence $(Z_0^K)_K$ converging to $\delta_x n_x^1(0,a)da$ as $K \to +\infty$.

1) Monomorphic approximation. For large K, the process Z_t^K stays close to the measure $\delta_x n_x^1(t,.) + \delta_{(x_b,0)} n_x^2(t,.)$ where $n_x(t,.) = (n_x^1(t,.), n_x^2(t,.))$ satisfies the partial differential equation (4). By Proposition 3.3, the dynamics $n_x(t,.)$ converges to $\overline{n}_x = (\overline{n}_x^1, \overline{n}_x^2)$ as t tends to infinity and hence reaches a given neighbourhood of $\overline{n}_x$ in finite time. By using large deviation results ([39]), we obtain with probability tending to one as K tends to infinity that the process Z_t^K stays in this neighbourhood of $\delta_x \overline{n}_x^1 + \delta_{(x_b,0)} \overline{n}_x^2$ during a time e^{CK} for some $C > 0$. The left-hand side in Assumption (24) ensures that the next mutation appears before the process leaves this neighbourhood.

2) Appearance of a mutant. We deduce that the monomorphic population with trait x creates a mutant with trait:

(i) $y = (x_b + h_b, x_d)$ or $y = (x_b, x_d + h_d)$ at a rate approximately equal to $2Kp_K(1 - p_K)\overline{n}_x^1(0)$;

(ii) $y = (x_b + h_b, x_d + h_d)$ at a rate approximately equal to $Kp_K^2 \overline{n}_x^1(0)$;

(iii) $y = (x_b, h_d)$ or $y = (x_b + h_b, 0)$ at at rate approximately equal to $2Kp_K(1-p_K)\overline{n}_x^2(0)$;

(iv) $y = (x_b + h_b, h_d)$ at rate approximately equal to $Kp_K^2 \overline{n}_x^2(0)$.

where the variables h_b and h_d are chosen independently with distribution k.

Since $p_K^2 = o(p_K(1-p_K))$, the cases (ii) and (iv) cannot be observed on the mutation time-scale $t/2K(1-p_K)$.

3) Effect of the natural selection. In cases (i) and (iii), the mutant population dynamics is approximated by a bi-type age structured branching process with birth rates $\mathbf{B}_y$ and death rates $\mathbf{D}_y + \eta\|\overline{n}_x\|_1\mathbf{I}$. By Proposition 4.7, the mutant population survives with probability

$$(\lambda(y) - \lambda(x)) \vee 0.$$

Let us detail the two different cases.

- Case (i). With probability 1/2, the trait of the mutant is $y = (x_b + h_b, x_d)$. From Proposition 4.7, we deduce that the mutant can survive if and only if

$$\lambda(x_b + h_b, x_d) > \lambda(x) \Longleftrightarrow h_b > 0 \text{ and } x \in U_1. \tag{25}$$

  With probability 1/2, $y = (x_b, x_d + h_d)$ and can survive if and only if

$$\lambda(x_b, x_d + h_d) > \lambda(x) \Longleftrightarrow h_d > 0 \text{ and } x \in U_2. \tag{26}$$

- Case (iii) (Lansing effect). The mutant has the trait $y = (x_b, h_d)$ or $y = (x_b + h_b, 0)$. By (3), we have $y_b \wedge y_d < 1$ which implies that $\lambda(y) < 0$ and then $(\lambda(y) - \lambda(x)) \vee 0 = 0$. In this case, the mutant population becomes extinct.

We deduce that the mutant can only survive (with positive probability) in case (i). The birth rate of such a mutant (on the time-scale $t/2p_K(1-p_K)$) is given by the intensity measure on $\mathbb{R}$

$$\overline{n}_x^1(0)\mu(dh)$$

that leads to the right hand side in (23). The probability that such a mutant survives and reaches a size of order K equals

$$(\lambda(y) - \lambda(x)) \vee 0.$$

Moreover (25) and (26) imply that

$$((\lambda(x+h) - \lambda(x)) \vee 0)\mu(dh) = (\lambda(x+h) - \lambda(x))\mathbb{1}_{\mathbb{R}_+^2}(h)\mu(dh).$$

This allows to obtain the left hand side of (23).
If the mutant population becomes extinct, the resident population remains close to its equilibrium $\overline{n}_x$.
If the mutant population survives, then it reaches a size of order K with a probability that tends to one and the population dynamics is approximated by the solution $(n_x(t,.), n_y(t,.))$ of the bimorphic system of partial differential equations (16). In this case we have necessarily $\lambda(y) > \lambda(x)$. Following Proposition 3.9, the deterministic dynamics $(n_x(t,.), n_y(t,.))$ reaches a neighbourhood of $(0, \overline{n}_y)$. By using branching processes approximations and arguments introduced in [5], we can deduce that the resident population with trait x becomes extinct. One can prove as in [5] that this competition phase has a duration of order $\log(K)$.
The right hand side of Assumption (24) ensures that the three steps of invasion are completed before the next mutation occurs.

•

The Markov property allows to reiterate the same reasoning for the next mutation occurence. In summary, the following theorem holds.

Theorem 4.13. *The following convergence holds in the sense of finite dimensional marginals:*

$$\left(Z^K_{\frac{t}{2Kp_K(1-p_K)}}, t \geqslant 0\right) \longrightarrow (T_t, t \geqslant 0), \qquad \text{as } K \to \infty,$$

where the process T is defined in Definition 4.10.

4.3 A canonical inclusion for adaptive dynamics

In this section, we assume in that mutations are small. We study the behaviour of the process X defined in (23) when the mutation scale equals $\epsilon > 0$ and the time is rescaled by $1/\epsilon^2$. To this aim, we define the rescaled trait substitution sequence process X^ϵ and study the limiting behaviour of the process X^ϵ as $\epsilon \to 0$. In the usual cases (smooth fitness functions) the canonical equation introduced by Dieckmann-Law can be derived as limit of X^ϵ as $\epsilon \to 0$ ([7]). As observed in Section 4.1, the fitness function $\lambda(x)$ does not satisfy these regularity assumptions. To overpass this difficulty, we use the approach developed in [14] based on differential inclusions. We prove in Theorem 4.17 that the set of limit points of the family X^ϵ is characterised as the set of solutions of a differential inclusion.

Definition 4.14. *The rescaled TSS process $(X^\epsilon(t), t \geqslant 0)$ is defined as a pure jump Markov process with infinitesimal generator L^ϵ defined for all measurable and bounded function $\varphi : \mathcal{V} \to \mathbb{R}$ and $x \in \mathcal{V}$ by*

$$L^\epsilon\varphi(x) = \frac{1}{\epsilon^2}\int_{(\mathbb{R}_+)^2} (\varphi(x+\epsilon h) - \varphi(x))(\lambda(x+\epsilon h) - \lambda(x))\overline{n}^1_x(0)\mu(dh) \tag{27}$$

Remark 4.15. *The process X^ϵ shows a dynamics similar to the process X. The jump rates are of order $1/\epsilon$ and the jump sizes are of order ϵ.*

We first introduce the set-valued map $F : \mathcal{V} \to \mathcal{P}(\mathbb{R}^2_+)$ defined for any $x \in \mathcal{V}$ by

$$\begin{aligned} \forall i \in \{1,2\}, \quad \forall x \in U_i, \quad & F(x) = (f(x,1))_i \\ \forall x \in \mathcal{H}, \quad & F(x) = \left\{ \frac{1}{2}\begin{pmatrix} f(x,u) \\ f(x,u) \end{pmatrix}, u \in [0,1] \right\}, \end{aligned} \tag{28}$$

where for all $(x,u) \in \mathcal{V} \times [0,1]$,

$$f(x,u) = \left(\int_0^u h^2 k(h)dh + \int_u^1 huk(h)dh\right) \frac{e^{-\lambda(x)(x_b \wedge x_d)}}{G(x)} \frac{\overline{n}^1_x(0)}{2}; \tag{29}$$

and G is defined in Proposition 4.1.

This set-valued map F somehow generalises the classical fitness gradient. It is represented by a picture in Figure 8 (b).
Let us explain the ideas leading to consider this function. Let us consider a compact

subset K of U_i. Since the Malthusian parameter λ is differentiable on U_i, the following approximation holds: for all $h \in \mathbb{R}_+$, uniformly for $x \in K$, we have

$$\lambda(x + \epsilon(h)_i) - \lambda(x) \approx \epsilon(h)_i . \nabla\lambda(x), \tag{30}$$

which leads to the definition of F on U_i. We analyse the case $x \in \mathcal{H}$ for which the approximation (30) is not true. Indeed, let $x \in \mathcal{H}$ and let $u \in [0,1]$ and let us consider a sequence $x^\epsilon = x - \epsilon(u)_i$, we have

$$\begin{aligned}\lambda(x^\epsilon + \epsilon(h)_i) - \lambda(x^\epsilon) &= \lambda(x + \epsilon(h-u)_i) - \lambda(x - \epsilon(u)_i) \\ &= \lambda(x + \epsilon(h-u)_i) - \lambda(x) + \lambda(x) - \lambda(x - \epsilon(u)_i)\end{aligned}$$

Assume $h < u$, since λ is differentiable on U_i we obtain when ϵ tends to 0 that

$$\lambda(x^\epsilon + \epsilon(h)_i) - \lambda(x^\epsilon) \approx \epsilon(h-u)_i . \nabla\lambda(x) + \epsilon(u)_i . \nabla\lambda(x) \approx \epsilon(h)_i . \nabla\lambda(x),$$

where $\nabla\lambda(x)$ is defined as the limit of $\nabla\lambda(y)$, $y \to x$, $y \in U_i$. That leads to the first integral in (29). If $h > u$, we obtain that $\lambda(x + \epsilon(h-u)_i) - \lambda(x) = 0$ and

$$\lambda(x^\epsilon + \epsilon(h)_i) - \lambda(x^\epsilon) \approx \epsilon(u)_i . \nabla\lambda(x) \leqslant \epsilon(h)_i . \nabla\lambda(x),$$

which leads to the second integral in (29). The inequality above means that when the process evolves near the diagonal $\mathcal{H}$ the adaptation slows down. Let us now define what we do mean by a solution of the differential inclusion driven by the set-valued map F.

Definition 4.16. *A solution of the differential inclusion driven by the set-valued map F is an absolutely continuous function $x : [0,T] \to \mathcal{V}$ which satisfies $x(0) = c^0$ and for almost all $t \in [0,T]$,*

$$\frac{dx(t)}{dt} \in F(x(t)). \tag{31}$$

It generalises the classical canonical equation for adaptive dynamics. For any $T > 0$ and $x^0 \in \mathcal{V}$, we denote by $\mathcal{S}_F(T, x^0)$ the set of solutions of the differential inclusion (31). The following theorem characterises the limit of the process X^ϵ as the solution of the differential inclusion (31).

Theorem 4.17. *Let $x^0 \in \mathcal{V}$. Assume that $X^\epsilon(0) \to x^0$ in probability as $\epsilon \to 0$. For all $T, \delta > 0$,*

$$\lim_{\epsilon \to 0} \mathbb{P}\left(\inf_{x \in \mathcal{S}_F(T,x^0)} \sup_{t \in [0,T]} |X^\epsilon(t) - x(t)| > \delta \right) = 0.$$

Remark 4.18. *Theorem 4.17 justifies our complete study. Let $(x(t), t \in [0,T])$ be a solution of (31). On each U_i, it satisfies*

$$\frac{dx(t)}{dt} = (f(x(t),1))_i, \quad x(t) \in U_i.$$

The map $x \in U_i \mapsto (f(x,1))_i$ is Lipschitz and bounded below by a positive constant. Hence, uniqueness holds on U_i for (31) and any solution reaches in finite time the diagonal $\mathcal{H}$. On $\mathcal{H}$, the solution satisfies

$$\frac{dx(t)}{dt} \in F(x(t)), \quad x(t) \in \mathcal{H}.$$

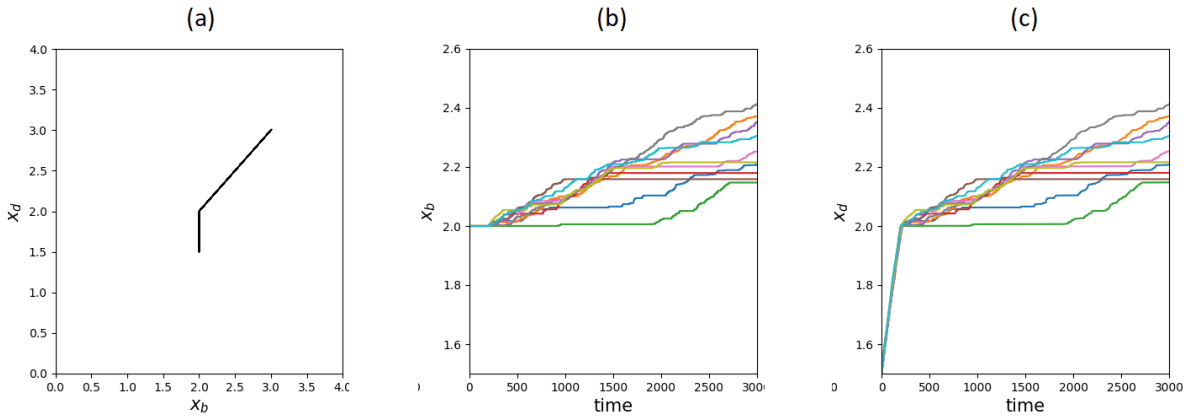

Figure 7: We represent 10 simulations of the process $(X^\epsilon(t), t \geqslant 0)$ *for* $\epsilon = 0.001$, $X^\epsilon(0) = (2, 1.5)$. *(a): We represent* $X_d^\epsilon(t)$ *as a function of* $X_b^\epsilon(t)$. *(b):* $(X_b^\epsilon(t), t \geqslant 0)$. *(c):* $(X_d^\epsilon(t), t \geqslant 0)$. *We observe on (b) and (c) that before reaching the diagonal* $\mathcal{H}$ *the limit behaviour is deterministic. On* $\mathcal{H}$, *the process evolves with speed in* $F(x)$ *for* $x \in \mathcal{H}$.

Since for all $x \in \mathcal{H}$, $F(x) \subset \mathcal{H}$, *we deduce that any solution stays in* $\mathcal{H}$.
Figure 7 illustrates Theorem 4.17. We represent some trajectories of the process $(X^\epsilon(t), t \geqslant 0)$ *started at* $X^\epsilon(0) = (2, 1.5)$ *for* $\epsilon = 0.001$.

The proof of Theorem 4.17 is based on [14, Theorem 1] recalled in the Appendix, see Theorem A.2. We start by writing the process X^ϵ as a time-changed Markov chain. We first re-write the generator L^ϵ as follows

$$L^\epsilon \varphi(x) = \int_{\mathbb{R}_+^2} (\varphi(x+h) - \varphi(x)) \frac{\overline{n}_x^1(0)}{\epsilon} k^\epsilon(x, dh)$$

where $k^\epsilon(x, dh) =$

$$\frac{\lambda(x+h) - \lambda(x)}{\epsilon} \mu^\epsilon(dh) + \left(\int_{(\mathbb{R}_+)^2} \left(1 - \frac{\lambda(x+g) - \lambda(x)}{\epsilon} \right) \mu^\epsilon(dg) \right) \delta_0(dh)$$

and μ^ϵ is the image measure of μ by the map $h \mapsto \epsilon h$. The following lemma is proved in [12, Ch. 4, S. 2, p. 163].

Lemma 4.19 ([12]). *Let* $\tau = \sup_{x \in \mathcal{V}} \overline{n}_x^1(0)$. *Let* $(Y^\epsilon(k), k \geqslant 0)$ *be a Markov chain with jump law*

$$\tilde{k}^\epsilon(x, dh) = \frac{\overline{n}_x^1(0)}{\tau} k^\epsilon(x, dh) + \left(1 - \frac{\overline{n}_x^1(0)}{\tau} \right) \delta_0(dh)$$

and Λ^ϵ *be a Poisson process with intensity* τ/ϵ. *Then the processes* X^ϵ *and* $Y^\epsilon(\Lambda^\epsilon)$ *have the same law.*

Then we are led to study the Markov chain $(Y^\epsilon(k), k \geqslant 0)$. We first define the drift of the Markov chain Y^ϵ by

$$g_\epsilon(x) = \mathbb{E}\left[Y^\epsilon(1) - Y^\epsilon(0) | Y^\epsilon(0) = x \right].$$

A simple calculation gives us

$$\forall i \in \{1,2\}, \quad \forall x \in U_i, \quad g_\epsilon(x) = \frac{\epsilon \overline{n}_x^1(0)}{2\tau} \int_{\mathbb{R}_+} \frac{\lambda(x+\epsilon(h)_i) - \lambda(x)}{\epsilon} k(h)(h)_i dh$$
$$\forall x \in \mathcal{H}, \quad g_\epsilon(x) = 0.$$

Then, we write the Markov chain Y^ϵ as a stochastic approximation algorithm

$$Y^\epsilon(k+1) = Y^\epsilon(k) + \frac{\epsilon}{\tau} U^\epsilon(k) + g^\epsilon(Y^\epsilon(k))$$

where U^ϵ is a martingale difference sequence. Assumptions of Theorem A.2 are clearly satisfied. In order to apply it, we compute the following set-valued map

$$\forall x \in \mathcal{V}, \quad H(x) = \text{conv}\left\{ \text{acc}_{\epsilon \to 0} \frac{\tau g_\epsilon(x^\epsilon)}{\epsilon} : x^\epsilon \to x \right\}$$

where $\text{conv}(A)$ denotes the smallest convex set which contains A and $\text{acc}_{\epsilon\to 0}\tau g_\epsilon(x^\epsilon)/\epsilon$ is the set of accumulation points of the sequence $\tau g_\epsilon(x^\epsilon)/\epsilon$ as ϵ tends to zero.

Lemma 4.20. *The set-valued map H satisfies*

$$\forall i \in \{1,2\}, \quad \forall x \in U_i, \quad H(x) = \{(f(x,1))_i\}$$
$$\forall x \in \mathcal{H}, \quad H(x) = \left\{ \alpha \begin{pmatrix} f(x,u) \\ 0 \end{pmatrix} + (1-\alpha) \begin{pmatrix} 0 \\ f(x,v) \end{pmatrix} : (u,v,\alpha) \in [0,1]^3 \right\}$$

where

$$f(x,u) = \frac{e^{-\lambda(x)(x_b \wedge x_d)}}{G(x)} \frac{\overline{n}_x^1(0)}{2} \left(\int_0^u h^2 k(h) dh + \int_u^1 uhk(h) dh \right).$$

Proof. Let $i \in \{1,2\}$. Let K be a compact subset of U_i. Let $\delta > 0$. We fix $\epsilon_0 > 0$ such that for all $x \in K$, $\epsilon < \epsilon_0$ and $h \in [0,1]$, we have $x + \epsilon(h)_i \in U_i$. The map λ is differentiable on U_i. Hence, for all $(x, \epsilon, h) \in K \times [0, \epsilon_0] \times [0,1]$, there exists $\theta \in [x, x + \epsilon(h)_i]$ such that

$$\lambda(x + \epsilon(h)_i) - \lambda(x) = \epsilon(h)_i . \nabla\lambda(\theta).$$

Let $x \in K$, we have

$$\left\| \frac{\tau g^\epsilon(x)}{\epsilon} - (f(x,1))_i \right\| = \frac{\overline{n}_x^1(0)}{2} \left| \int_{\mathbb{R}_+} [\nabla\lambda(\theta)]_i \, h^2 k(h) dh - [\nabla\lambda(x)]_i \int_{\mathbb{R}_+} h^2 k(h) dh \right|$$
$$\leqslant \frac{\sup_{x \in \mathcal{V}} \overline{n}_x^1(0)}{2} \int_{\mathbb{R}_+} | \, [\nabla\lambda(\theta)]_i - [\nabla\lambda(x)]_i \, | h^2 k(h) dh.$$

By Proposition 4.1, the map $\nabla\lambda$ is Lipschitz on U_i with some Lipschitz constant C. We deduce that

$$\left\| \frac{\tau g^\epsilon(x)}{\epsilon} - (f(x,1))_i \right\| \leqslant \frac{\epsilon \sup_{x \in \mathcal{V}} \overline{n}_x^1(0) C}{2} \int_{\mathbb{R}_+} h^3 k(h) dh.$$

We obtain that $\tau g^\epsilon/\epsilon$ converges uniformly on all compact subsets of U_i and we conclude that for all $x \in U_i$, $H(x) = (f(x,1))_i$.
Let $x \in \mathcal{H}$. We first show that

$$\left\{ \text{acc}_{\epsilon \to 0} \frac{\tau g_\epsilon(x^\epsilon)}{\epsilon} : x^\epsilon \to x \right\} = \bigcup_{i \in \{1,2\}} \{(f(x,u))_i : u \in [0,1]\}. \tag{32}$$

If the fitness gradient was smooth (i.e Lipschitz), the sets in (32) would be reduced to one element and we would be in the cases studied in [4], [7] or [26]. We prove the inclusion from right to left. Let $i \in \{1,2\}$, let $u \in [0,1]$, we define the sequence $x^\epsilon = x - \epsilon(u)_i$. We have

$$\begin{aligned}\frac{\tau g^\epsilon(x^\epsilon)}{\epsilon} &= \left(\int_{\mathbb{R}_+} \frac{\lambda(x-\epsilon(u)_i+\epsilon(h)_i)-\lambda(x-\epsilon(u)_i)}{\epsilon} k(h)(h)_i dh\right)\frac{\overline{n}^1_{x^\epsilon}(0)}{2}\\ &= \left(\int_0^u \frac{\lambda(x-\epsilon(u)_i+\epsilon(h)_i)-\lambda(x-\epsilon(u)_i)}{\epsilon} k(h)(h)_i dh\right.\\ &\left.+\int_u^1 \frac{\lambda(x-\epsilon(u)_i+\epsilon(h)_i)-\lambda(x-\epsilon(u)_i)}{\epsilon} k(h)(h)_i dh\right)\frac{\overline{n}^1_{x^\epsilon}(0)}{2}.\end{aligned}$$

For all $h \in [0,u]$, we have $[x-\epsilon(u)_i, x-\epsilon(u-h)_i] \subset U_i$. So we can find $\theta \in [x-\epsilon(u)_i, x-\epsilon(u-h)_i]$ such that $\lambda(x-\epsilon(u-h)_i)-\lambda(x-\epsilon(u)_i) = \epsilon(h)_i.\nabla\lambda(\theta)$. By Proposition 4.1 (ii), we deduce that

$$\int_0^u \frac{\lambda(x-\epsilon(u)_i+\epsilon(h)_i)-\lambda(x-\epsilon(u)_i)}{\epsilon}\, k(h)(h)_i dh \to \frac{e^{-\lambda(x)(x_b\wedge x_d)}}{G(x)}\int_0^u (h^2)_i\, k(h) dh$$

as ϵ tends to zero. For all $h \in [u,1]$ we have $\lambda(x+\epsilon(h-u)_i) = \lambda(x)$. We deduce similarly that

$$\int_u^1 \frac{\lambda(x-\epsilon(u)_i+\epsilon(h)_i)-\lambda(x-\epsilon(u)_i)}{\epsilon}\, k(h)(h)_i dh \to \frac{e^{-\lambda(x)(x_b\wedge x_d)}}{G(x)}\int_u^1 (h)_i u\, k(h) dh$$

as ϵ tends to zero. We conclude the proof of the first inclusion arguing that $\overline{n}^1_{x^\epsilon}(0) \to \overline{n}^1_x(0)$ as ϵ tends to zero.
We prove the inclusion from left to right. Let $x \in \mathcal{V}$. If $x \in \mathcal{H}$ then we have

$$g^\epsilon(x) = 0. \tag{33}$$

If $x \in U_i$, for some $i \in \{1,2\}$, then we have

$$\frac{\tau g^\epsilon(x)}{\epsilon} = \frac{\overline{n}^1_x(0)}{2}\int_0^1 \frac{\lambda(x+\epsilon(h)_i)-\lambda(x)}{\epsilon} k(h)(h)_i dh$$

Moreover for all ϵ, h there exists $\theta \in [x, x+\epsilon(h)_1]$ such that $\lambda(x+\epsilon(h)_i)-\lambda(x) \leqslant \frac{\partial\lambda(\theta)}{\partial x_i}\epsilon h$. We deduce that

$$\left[\frac{\tau g^\epsilon(x)}{\epsilon}\right]_i \leqslant \frac{\overline{n}^1_x(0)}{2}\int_0^1 \frac{\partial\lambda(\theta)}{\partial x_i} h^2 k(h) dh \tag{34}$$

From (33) and (34), we deduce easily the second inclusion in (32). We conclude that

$$\operatorname{conv}\left\{\operatorname{acc}_{\epsilon\to 0}\frac{\tau g_\epsilon(x^\epsilon)}{\epsilon} : x^\epsilon \to x\right\} = \operatorname{conv}\left\{\cup_{i\in\{1,2\}}\{(f(x,u))_i : u \in [0,1]\}\right\} = H(x).$$

□

In Lemma 4.22, we prove that differential inclusions associated with H and F have identical solutions. Before, we give a technical lemma.

Lemma 4.21. *Let $x \in \mathcal{H}$. Let $(u, v, \alpha) \in [0,1]^3$ and let*

$$m = \alpha \begin{pmatrix} f(x,u) \\ 0 \end{pmatrix} + (1-\alpha) \begin{pmatrix} 0 \\ f(x,v) \end{pmatrix} \in H(x).$$

Assume that $m \notin F(x)$. Then we have $\alpha f(x,u) \neq (1-\alpha) f(x,v)$.

Proof. We prove the lemma by contradiction. Assume that $\alpha f(x,u) = (1-\alpha) f(x,v)$. Then we obtain

$$m = \frac{1}{2} \begin{pmatrix} \frac{2f(x,u)f(x,v)}{f(x,u)+f(x,v)} \\ \frac{2f(x,u)f(x,v)}{f(x,u)+f(x,v)} \end{pmatrix}.$$

We assume without loss of generality that $f(x,u) \leqslant f(x,v)$. Then we obtain $0 \leqslant \frac{2f(x,u)f(x,v)}{f(x,u)+f(x,v)} \leqslant f(x,v)$. Finally we remark that the map $s \mapsto f(x,s)$ is a bijection from $[0,1]$ to $[0, f(x,1)]$. Hence there exists $w \in [0,1]$ such that $f(x,w) = \frac{2f(x,u)f(x,v)}{f(x,u)+f(x,v)}$ that allows us to obtain the contradiction. □

Lemma 4.22. *Any solution of*

$$\begin{cases} \frac{dx(t)}{dt} \in F(x(t)), & t \in [0,T] \\ x(0) = x^0. \end{cases} \tag{35}$$

is a solution of

$$\begin{cases} \frac{dx(t)}{dt} \in H(x(t)), & t \in [0,T] \\ x(0) = x^0. \end{cases} \tag{36}$$

and conversely.

Proof. For all $x \in \mathcal{V}$, we have $F(x) \subset H(x)$. We deduce that if $(x(t), t \in [0,T])$ is a solution of (35) then it is a solution of (36). Assume conversely that there exists a solution $(x(t), t \in [0,T])$ of (36) which is not a solution of (35). We deduce that there exists $t_0 \in [0,T]$ such that x is differentiable at t_0, $\mathrm{d}x(t_0)/\mathrm{d}t \in H(x(t_0))$ and $\mathrm{d}x(t_0)/\mathrm{d}t \notin F(x(t_0))$. Then we have $x(t_0) \in \mathcal{H}$ and $\mathrm{d}x(t_0)/\mathrm{d}t \notin F(x(t_0))$.

We now deduce the contradiction. By Lemma 4.21, we obtain that $\frac{\mathrm{d}x_b(t_0)}{\mathrm{d}t} \neq \frac{\mathrm{d}x_d(t_0)}{\mathrm{d}t}$. Without loss of generality, we may assume that $\frac{\mathrm{d}x_b(t_0)}{\mathrm{d}t} < \frac{\mathrm{d}x_d(t_0)}{\mathrm{d}t}$. Since $x_b(t_0) = x_d(t_0)$, there exists an interval $]t_0, t_1[$ such that for all $s \in]t_0, t_1[$, $x(s) \in U_1$. Assume that for all $s \in]t_0, t_1[$ such that x is differentiable at s, we have

$$\frac{\mathrm{d}x_d(s)}{\mathrm{d}t} = 0. \tag{37}$$

The solution x of the differential inclusion (36) is absolutely continuous. Hence for all $s \in]t_0, t_1[$, $x_d(s) = x_d(t_0) = x_b(t_0)$. Since $x(s) \in U_1$, we obtain that $x_b(s) < x_b(t_0)$. It is absurd since x_b is non-decreasing and hence it contradicts (37). So, there exists $s_0 \in]t_0, t_1[$ such that $x(s_0) \in U_1$, x is differentiable at s_0 and satisfies $\frac{\mathrm{d}x_d(s_0)}{\mathrm{d}t} > 0$. However, x is a solution of (36) that leads to the final contradiction. □

We now give the proof of Theorem 4.17. It is a direct consequence of [14, Theorem 1] recalled in Appendix A.3.

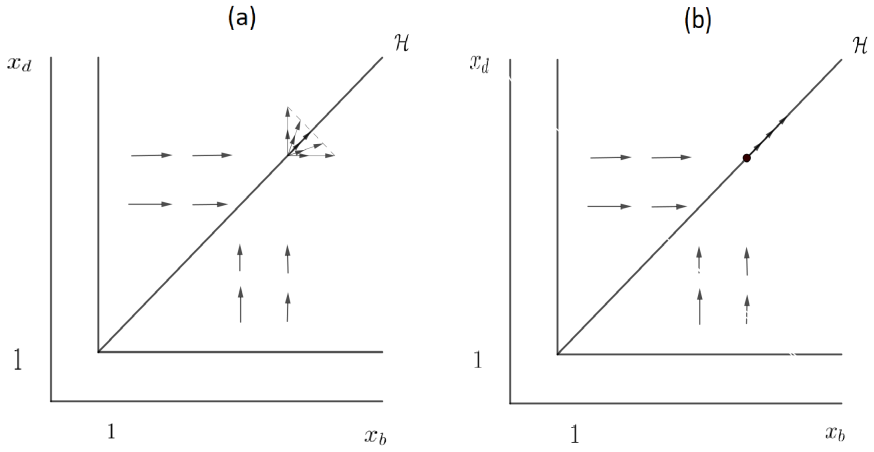

Figure 8: (a): Representation of H. (b): Representation of F.

of Theorem 4.17. Let $T > 0$ be fixed. By Lemma 4.22 and Theorem A.2, we deduce that for all $\delta > 0$ we have

$$\lim_{\epsilon \to 0} \mathbb{P}\left(\inf_{x \in \mathcal{S}_F(T,x^0)} \sup_{t \in [0,T]} |Y^\epsilon(\lfloor t\tau/\epsilon \rfloor) - x(t)| > \delta \right) = 0. \tag{38}$$

We conclude by using similar arguments as in the proof of [14, Theorem 4]. Since Λ^ϵ is a Poisson process with parameter τ/ϵ, we obtain that for all $\delta > 0$,

$$\mathbb{P}\left(\sup_{t \leqslant T} |\Lambda^\epsilon(t) - \frac{t\tau}{\epsilon}| \geqslant \frac{\tau\delta}{\epsilon} \right) \leqslant \frac{T\epsilon}{\tau\delta}. \tag{39}$$

We have $\mathbb{P}\left(\inf_{x \in \mathcal{S}_F(T,x^0)} \sup_{t \leqslant T} |X^\epsilon(t) - x(t)| > \delta\right)$

$$\begin{aligned}
&= \mathbb{P}\left(\inf_{x \in \mathcal{S}_F(T,x^0)} \sup_{t \leqslant T} |Y^\epsilon(\Lambda^\epsilon(t)) - x(t)| > \delta \right) \\
&\leqslant \mathbb{P}\left(\inf_{x \in \mathcal{S}_F(T,x^0)} \sup_{t \leqslant T} \left\{ |Y^\epsilon(\Lambda^\epsilon(t)) - x\left(\frac{\epsilon\Lambda^\epsilon(t)}{\tau}\right)| + |x\left(\frac{\epsilon\Lambda^\epsilon(t)}{\tau}\right) - x(t)| \right\} > \delta \right).
\end{aligned} \tag{40}$$

Let $x \in \mathcal{S}_F(T, x^0)$ be a solution of the differential inclusion (31). For almost all $t \in [0,T]$, we have $\mathrm{d}x(t)/\mathrm{d}t \in F(x(t))$. Since $\sup_{x \in \mathcal{V}} \sup\{F(x)\} < +\infty$, we deduce that there exists $C_T > 0$ such that for all $y \in \mathcal{S}_F(T, x^0)$, for all $s, t \in [0,T]$, $|y(t) - y(s)| \leqslant C_T|t - s|$. We deduce that (40) is less than

$$\begin{aligned}
&\mathbb{P}\left(\inf_{x \in \mathcal{S}_F(T,x^0)} \left\{ \sup_{t \leqslant T} |Y^\epsilon(\Lambda^\epsilon(t)) - x\left(\frac{\epsilon\Lambda^\epsilon(t)}{\tau}\right)| \right\} + C_T \sup_{t \leqslant T} |\frac{\epsilon\Lambda^\epsilon(t)}{\tau} - t| > \delta \right) \\
&\leqslant \mathbb{P}\left(\inf_{x \in \mathcal{S}_F(T,x^0)} \sup_{t \leqslant T} |Y^\epsilon(\Lambda^\epsilon(t)) - x\left(\frac{\epsilon\Lambda^\epsilon(t)}{\tau}\right)| > \delta \right) + \mathbb{P}\left(C_T \sup_{t \leqslant T} |\frac{\epsilon\Lambda^\epsilon(t)}{\tau} - t| > \delta \right)
\end{aligned}$$

and we conclude by using (38) and (39). □

5 Discussion

5.1 General discussion and comments

In the present article, we study the evolution of a population with a trait structure describing a simple class of life-histories. We build a stochastic individual-based model in a framework that is continuous in time, age and trait. The trait is a pair of parameters (x_b, x_d) characterising the age at end-of-reproduction x_b and the age at transition to a non-zero mortality risk x_d. The model sees two origins of phenotypic variation. First, the genetic mutations that are supposed to be rare and modify the traits symmetrically - equal probability to increase or decrease the value of the parameter. Second, we model the Lansing effect which can be considered as an epigenetic mutation affecting the progeny of an "old" individual. It is acting on a much faster time-scale - one generation - than genetic mutations and has only a negative effect on the life expectancy of the progeny. We highlight that we have chosen to model the Lansing effect by an extremely strong effect on the descendant. Indeed, it acts on each generation and degrades dramatically the life-expectancy of the progeny. Nevertheless, a more general and realistic study of epigenetic modifications on the genetic evolution has to be investigate. Some aspects of this question have been studied in [21]. The reasoning is based on the fact that epigenetic modifications are more frequent than genetic mutations ([36]). It would be interesting to extend adaptive dynamics tools in order to take into account epigenetics modifications (involving intermediary time scales).
We study the long term evolution of the trait distribution using adaptive dynamics theory. We extend the age-structured TSS to take into account the Lansing effect. The main mathematical result of the present work concerns the behaviour of the TSS when mutations are small. We show in Theorem 4.17 that the behaviour of the rescaled TSS in the limit $\epsilon \to 0$ is characterised by a differential inclusion whose solutions are not unique on the diagonal $\mathcal{H} = \{x_b = x_d\}$. This differential inclusion allows to generalize the Canonical Equation of Adaptive Dynamics [11],[7] in order to consider the non-smooth fitness gradient. The proof is based on [14]. Thanks to this approach, we show that the evolution of our model, whatever its initial configuration, leads to the apparition and maintenance of configurations (x_b, x_d) satisfying $x_b - x_d = 0$.
To our knowledge, differential inclusions have never been used before in the adaptive dynamics theory. In [4], [7], [26], the fitness gradient is assumed to be a Lipschitz function, which ensures the uniqueness of the solutions. Our approach seems useful for generalizing the canonical equation to situations where the fitness gradient is neither Lipschitz nor continuous.
The drift associated with the differential inclusion depends on the fitness gradient that satisfies (see Proposition 4.1)

$$\forall i \in \{1,2\}, \quad \forall x \in U_i, \quad \nabla\lambda(x) = \left(\frac{e^{-\lambda(x)(x_b \wedge x_d)}}{G(x)} \right)_i \tag{41}$$

where $\lambda(x)$ is the Malthusian parameter and $G(x)$ the mean generation time associated to the trait x. Hence the fitness gradient $\nabla\lambda(x)$ describes the speed of evolution of the trait x. It can be related to the seminal work of Hamilton ([17]) on the moulding of senescence. In that article, Hamilton states that senescence is unavoidable because the

strength of selection decreases with age. To show it, he defines the strength of selection at some given age a_0 as the sensitivity of the Malthusian parameter with respect to some little perturbation on the birth or death intensities at age a_0. He concludes arguing that these quantities decrease to zero as a_0 tends to infinity. Similarly, formula (41) describes the sensitivity of the Malthusian parameter with respect to some perturbation on the duration of the reproduction or the survival phase at ages x_b and x_d. Then, they can be interpreted as the strength of selection at ages x_b and x_d and describe the speed of evolution of the traits x_b and x_d in the canonical inclusion (31).

To conclude, the present article studies a case of birth and death model with a strong Lansing effect and constant competition applied to asexual and haploid individuals in order to validate mathematically the convergence of x_b and x_d observed in the numerical simulations. Further characterisation of this model is in progress, in order to better understand the influence of its different parameters on the evolution of (x_b, x_d).
Our initial motivation for developing the bd-model was an attempt to understand whether a phenomenon leading to a dramatic decrease of an individual's fitness could be selected through evolution with simple and no explicitely constraining trade-offs. Indeed, in the past years, Rera and collaborators have identified and characterised a dramatic transition preceding death in drosophila ([33], [34]) as well as other organisms ([35], [9]). We show here that, under uniform competition - i.e. environmental limitation equally affecting all genotypes - a mechanism coupling the end of reproductive capabilities and organismal homeostasis can and will be selected thanks to evolution. Thus, at least these two characteristics of senescent organisms can positively be selected through evolution. Concerning the biological interpretation of our model, our thesis is that individuals with a senescence mechanism associated to the Lansing effect tend to produce more genetic variants than those without senescence. Hence, these individuals could show a higher evolvability. This question is being investigated in a work in progress.
Here we show the positive selection of a property that limits reproduction. It is reminiscent of the article by [41] who proposed a new selective mechanism for post-reproductive life span. The latter relies on the hypothesis that the post-reproductive lifespan can be selected as an insurance against indeterminacy; a longer life expectancy reducing the risk of dying by chance before the cessation of reproductive activity. In the present article, the maintenance of individuals showing Lansing effect is the counterpart of individuals with post-reproductive survival. As discussed in [20], one of August Weismann's concepts that persisted without changes throughout his life is a conviction that "life is endowed with a fixed duration, not because it is contrary to its nature to be unlimited, but because unlimited existence of individuals would be a luxury without any corresponding advantage" ([44]). This is what we showed in the present work.

5.2 Generalizations of the model

5.2.1 Lower Lansing Effect

In this article, we model the Lansing effect by a very strong effect since we assume that the descendants of old individuals with trait (x_b, x_d) inherits of the trait $(x_b, 0)$. It could be interesting to study a more general case like $(x_b, \alpha x_d)$ for some $\alpha \in [0, 1]$.

5.2.2 Mutation kernel

It would be interesting to consider two different mutation kernels $k_b(x,h)$ and $k_d(x,h)$ for the traits x_b and x_d. This change should not modify the behaviour of the process. Indeed, on the sides off the diagonal ($x_b \neq x_d$), only the speed of evolution of the traits will be modified. Hence, the trait (x_b, x_d) will converge to the diagonal and will evolve on it at some speed depending on the variances of the kernels k_b and k_d. This case is being studied in an ongoing work mentionned in Section 5.2.1.

A Appendix

A.1 Proof of Proposition 3.5

The proof is based on classical arguments of spectral theory for strongly continuous semi-groups. Let us denote by $(P_t)_{t\geqslant 0}$ the semi-group on $L^1(\mathbb{R}_+)^2$ associated with the infinitesimal generator $(A, D(A))$. Let us denote by $\sigma(A)$ and $\sigma_e(A)$ the spectrum and the essential spectrum of the operator A respectively. Let us denote by $\alpha\left[P_t\right]$ the measure of non-compactness of P_t [42, Definition 4.14 p 165], and define $w_1(A) := \lim_{t\to\infty} t^{-1}\log(\alpha\left[P_t\right])$. We show that there exists $\omega > 0$ such that

$$\max\left(w_1(A), \sup_{z\in(\sigma(A)\backslash\sigma_e(A))\backslash\{\lambda(x)\}} \mathcal{R}e(z)\right) < \omega < \lambda(x). \tag{42}$$

By using arguments similar to [32, Section 1], we obtain that for all t large enough

$$\alpha\left[P_t\right] \leqslant e^{-t}.$$

and that

$$w_1(A) := \lim_{t\to\infty} t^{-1}\log(\alpha\left[P_t\right]) \leqslant -1.$$

By [42, Proposition 4.13 p 170], we obtain that $\sigma_e(A) \subset \{z \in \mathbb{C} : \mathcal{R}e(z) \leqslant -1\}$. Let $z \in \sigma(A)\backslash\sigma_e(A)$. Then there exists a non-zero $u \in L^1(\mathbb{R}_+, \mathbb{C})^2$ such that

$$\begin{cases} -u'(a) - \mathbf{D}_x(a)u = zu(a) \\ u(0) = \int_{\mathbb{R}_+} \mathbf{B}_x(\alpha)u(\alpha)d\alpha. \end{cases} \tag{43}$$

By solving the first equation in (43) and by injecting the result in the second equation, we obtain that $u(0)$ satisfies

$$u(0) = \mathbf{F}(z)u(0) \tag{44}$$

where

$$\mathbf{F}(z) = \int_{\mathbb{R}_+} \mathbf{B}_x(a)\exp\left(-\int_0^a (\mathbf{D}_x(\alpha) + z\mathbf{I})d\alpha\right)da. \tag{45}$$

Equation (44) admits a non-trivial solution $u(0)$ if and only if $\det\left[\mathbf{F}(z) - \mathbf{I}\right] = 0$. Since the matrix $\mathbf{F}(z)$ is triangular we have

$$\det\left[\mathbf{F}(z) - \mathbf{I}\right] = (\left[\mathbf{F}(z)\right]_{11} - 1)(\left[\mathbf{F}(z)\right]_{22} - 1) \tag{46}$$

where $[\mathbf{F}(z)]_{11} = \int_0^{x_b \wedge x_d} e^{-za} da$ and $[\mathbf{F}(z)]_{22} = \int_0^{x_b} e^{-(1+z)a} da$. We deduce that the Malthusian parameter $\lambda(x)$ is the largest real solution of $\det[\mathbf{F}(z) - \mathbf{I}] = 0$. By using similar analytical arguments as in the proof of [42, Theorem 4.10], we deduce that there exists only finitely many $z \in \mathbb{C}$ such that $\mathcal{R}e(z) > 0$ and $\det[\mathbf{F}(z) - \mathbf{I}] = 0$ which allows us to conclude for (42). We prove that $\lambda(x)$ is a simple eigenvalue of A by showing that $\lambda(x)$ is a simple zero of the equation $\det[\mathbf{F}(z) - \mathbf{I}] = 0$. Indeed we have

$$\frac{\mathrm{d}\det[\mathbf{F}(\lambda(x)) - \mathbf{I}]}{\mathrm{d}\lambda} = \frac{\mathrm{d}[\mathbf{F}(\lambda(x))]_{11}}{\mathrm{d}\lambda}([\mathbf{F}(\lambda(x))]_{22} - 1) > 0.$$

Let N_x be a principal eigenvector associated with the eigenvalue $\lambda(x)$. By (43), we have

$$N_x^1(a) = N_x^1(0)e^{-((a-x_d)\vee 0)-\lambda(x)a}, \quad N_x^2(a) = N_x^2(0)e^{-(1+\lambda(x))a}$$

and Equation (44) gives that

$$N_x^2(0) = \frac{[\mathbf{F}(\lambda(x))]_{21}}{1 - [\mathbf{F}(\lambda(x))]_{22}} N_x^1(0).$$

We conclude for the convergence by using arguments similar to proof of [42, Theorem 4.9 p187].

A.2 Proof of Lemma 3.6

Proof. Equation (11) has the form $\mathrm{d}u/\mathrm{d}t = f(t,u)$ with $f(t,u) \to g(u)$ as $t \to \infty$. So (11) is called an asymptotically autonomous differential equation ([24], [37]) with the limit equation

$$\frac{\mathrm{d}y(t)}{\mathrm{d}t} = My(t) - \eta\|y(t)\|_1 y(t). \tag{47}$$

We first show that any solution $y(t)$ of (47) started at $y(0) \in \mathbb{R}_+^* \times \mathbb{R}_+$ converges to a stationary state. In [1], the proof is given when M is irreducible. We give a slightly different proof. Let $\overline{z}$ be defined in (12). It is straightforward to prove that $\overline{z}$ is the eigenvector of M associated with the simple eigenvalue m_{11}, which satisfy the condition $\|\overline{z}\|_1 = m_{11}/\eta$. Since $z(0) \in \mathbb{R}_+^* \times \mathbb{R}_+$ there exists a positive constant $c(z(0))$ such that $e^{-m_{11}t}e^{Mt}z(0) \to c(z(0))\overline{z}$ as $t \to \infty$. We now write

$$\frac{y_1(t)}{y_2(t)} = \frac{\left[e^{-m_{11}t}e^{Mt}y(0)\right]_1}{\left[e^{-m_{11}t}e^{Mt}y(0)\right]_2} \to \frac{\overline{z}_1}{\overline{z}_2}.$$

We deduce that the ω-limit set of any solution of (47) is a subset of $\Delta = \{z \in \mathbb{R}_+^2 : z_1 = \frac{\overline{z}_1}{\overline{z}_2} z_2\}$. We conclude by proving that any solution starting from Δ converges to $\overline{z}$. Let us consider such a solution (always denoted by $y(t)$). We have

$$\begin{aligned}
\frac{\mathrm{d}y_1(t)}{\mathrm{d}t} &= y_1(t)(m_{11} - \eta\|y(t)\|_1), \\
\frac{\mathrm{d}y_2(t)}{\mathrm{d}t} &= y_2(t)\left(m_{21}\frac{y_1(t)}{y_2(t)} + m_{22} - \eta\|y(t)\|_1\right).
\end{aligned}$$

Since the ω-limit set is an invariant subset, we deduce that

$$\frac{\mathrm{d}y_2(t)}{\mathrm{dt}} = y_2(t)(m_{11} - \eta\|y(t)\|_1),$$

that $\|y(t)\|_1 \to m_{11}/\eta$ and $y(t) \to \overline{z}$ as $t \to \infty$. In order to conclude about the solutions of (11) we use [37, Theorem 1.2] arguing that $\overline{z}$ is an asymptotically stable equilibrium of (47) and that for any $y(0) \in \mathbb{R}_+^*$, the ω-limit set of any solution $y(t)$ of (11) started at $y(0)$ is not a subset of $\{0\} \times \mathbb{R}_+$. The first claim is easily proved by showing that the Jacobian matrix has negative eigenvalues. For the second claim, let us assume it is not satisfied. Then $y_1(t) \to 0$ as $t \to \infty$. Let us show the contradiction. Let ϵ be sufficiently small and t_0 such that for any $t \geqslant t_0$

$$\begin{aligned} 0 < m_{11} - \epsilon \leqslant m_{11} + \mathcal{D}_{11}(t) \leqslant m_{11} + \epsilon \\ m_{22} - \epsilon \leqslant m_{22} + \mathcal{D}_{22}(t) \leqslant m_{22} + \epsilon < 0 \\ 0 < m_{12} - \epsilon \leqslant m_{12} + \mathcal{D}_{12}(t) \leqslant m_{12} + \epsilon. \end{aligned}$$

We introduce

$$\begin{aligned} P^\epsilon(y_1, y_2) &= y_1(m_{11} - \epsilon - \eta(y_1 + y_2)) \\ Q^\epsilon(y_1, y_2) &= y_2(m_{22} + \epsilon - \eta(y_1 + y_2)) + (m_{12} + \epsilon)y_1 \end{aligned}$$

and

$$\begin{aligned} \mathcal{A}^\epsilon &= \left\{y \in \mathbb{R}_+^2 : y_1 + y_2 \leqslant m_{11} - \epsilon\right\} \\ \mathcal{B}^\epsilon &= \left\{y \in \mathbb{R}_+^2 : y_2 \geqslant \frac{1}{2\eta}\left(m_{22} + \epsilon - \eta y_1 + \sqrt{(m_{22} + \epsilon - \eta y_1)^2 + 4\eta(m_{12} + \epsilon)y_1}\right)\right\}. \end{aligned}$$

We deduce that there exists t_1 such that for any $t \geqslant t_1$, $\frac{\mathrm{d}y_2(t)}{\mathrm{dt}} < 0$ on $\mathcal{B}^\epsilon$ and $y_1(t) < \epsilon$. We deduce that there exists t_2 such that for all $t \geqslant t_2$, $y_1(t) + y_2(t) \leqslant \frac{m_{11}-\epsilon}{\eta}$. So for all $t \geqslant t_2$, $\frac{\mathrm{d}y_1(t)}{\mathrm{dt}} \geqslant 0$ which is absurd. □

A.3 Differential inclusions

In this appendix, we recall the results of [14] which concern the approximation of Markov chains by differential inclusions.
Let $\epsilon > 0$ be a scale parameter. Let $(Y^\epsilon(k), k \in \mathbb{N})$ be a Markov chain with values in $\mathbb{R}^d$. The drift of the Markov chain Y^ϵ is defined by

$$g^\epsilon(x) = \mathbb{E}\left[Y^\epsilon(k+1) - Y^\epsilon(k) | Y^\epsilon(k) = x\right].$$

Let $(\gamma^\epsilon)_{\epsilon>0}$ be such that $\lim_{\epsilon\to 0} \gamma^\epsilon = 0$ and let us denote

$$f^\epsilon(x) = \frac{g^\epsilon(x)}{\gamma^\epsilon}.$$

One can write the evolution of the Markov chain as a stochastic approximation algorithm with constant step size γ^ϵ

$$Y^\epsilon(k+1) = Y^\epsilon(k) + \gamma^\epsilon\left(f^\epsilon(Y^\epsilon(k)) + U^\epsilon(k+1)\right)$$

where U^ϵ is a martingale difference sequence with respect to the filtration associated with the process Y^ϵ.
Let us define

$$F(x) = \text{conv}\left(\left\{\text{acc}_{\epsilon\to 0} f^\epsilon(x^\epsilon) \text{ for all } x^\epsilon \text{ such that } \lim_{\epsilon\to 0} x^\epsilon = x\right\}\right)$$

where $\text{conv}(A)$ denotes the convex hull of the set A and $\text{acc}_{\epsilon\to 0} f^\epsilon(x^\epsilon)$ denotes the set of accumulation points of the sequence $f^\epsilon(x^\epsilon)$ as $\epsilon \to 0$. Let us denote by $\mathcal{S}_F(T, x^0)$ the set of solutions $(x(t), t \in [0,T])$ of the differential inclusion

$$\begin{cases} \frac{\mathrm{d}x(t)}{dt} \in F(x(t)), \quad t \in [0,T] \\ x(0) = x^0. \end{cases} \tag{48}$$

Let us recall the definition of a solution of (48).

Definition A.1. *A map $x : [0,T] \mapsto \mathbb{R}^d$ is a solution of (48) if there exists a map $\varphi : [0,T] \mapsto \mathbb{R}^d$ such that:*

(i) For all $t \in [0,T]$, $x(t) = x^0 + \int_0^t \varphi(s)ds$,

(ii) For almost every $t \in [0,T]$, $\varphi(t) \in F(x(t))$.

In particular (i) is equivalent to saying that x is absolutely continuous. (i) and (ii) imply that x is differentiable at almost every $t \in [0,T]$ with $\mathrm{d}x(t)/\mathrm{d}t \in F(x(t))$.

We define the continuous process $\overline{Y}^\epsilon(t)$ as the piecewise interpolation of Y^ϵ whose time has been accelerated by $1/\gamma^\epsilon$: for all $k \in \mathbb{N}$, $\overline{Y}^\epsilon(k\gamma^\epsilon) = Y^\epsilon(k)$ and $\overline{Y}^\epsilon$ is linear on $[k\gamma^\epsilon, (k+1)\gamma^\epsilon]$. We have the following theorem proved in [14, Theorem 1].

Theorem A.2. *Assume that:*

- *There exists a constant $c > 0$ such that for all $y \in \mathbb{R}^d$, $\quad \|f^\epsilon(y)\| \leqslant c(1 + \|y\|)$.*
- *U^ϵ is a martingale difference sequence which is uniformly integrable.*

If $Y^\epsilon(0)$ tends to x^0 in probability as ϵ tends to zero, then

$$\inf_{y\in\mathcal{S}_F(T,x^0)} \sup_{t\in[0,T]} \|\overline{Y}^\epsilon(t) - x(t)\| \longrightarrow 0$$

in probability as ϵ tends to zero.

A.4 Numerical simulation

We give the Python script for the numerical simulations of the individual based model described in Section 2. The algorithm is based on a classical acceptation/reject method ([6], [38])

```
#parameters

#intensity of competition
competition = 0.0005

#probability of mutation
p_mut = 0.05

#variance of mutations
var_mut = 0.05

#number of jumps
number_of_jumps = 1000000

#initial population size
population_size = 10000


#traits in the initial population

trait = numpy.ones((population_size, 2))
for k in range(len(trait[:, 0])):
    trait[k, :] = [1.2, 1.6]


#initial matrix population: [x_b,x_d, living/dead, birth date, death date, id, id parent,
     parent senescent]


population_0 = numpy.zeros((len(trait[:, 0]), 8), order='C', dtype=numpy.float32)
population_0[:, 0:2] = trait
population_0[:, 2] = 1
population_0[:, 3: 5] = 0
for l in range(len(population_0[:, 0])):
    population_0[l, 5] = l+1


#birth rates

def b(x, a):
    if a <= x[0]:
        u = 1
    else:
        u = 0.0
    return u

#death rates

def d(x, a):
    if a > x[1]:
        u = 1
    else:
        u = 0.0
    return u


# maximal jump intensity by individual  (for the acceptation/reject method)

intmax = 2.0 + competition



#Lansing effect

def lansing_effect(x, u):
    if x[1] > 0:
        if u < x[1]:
            r = x
        else:
```

```
            r = [x[0], 0]
    else:
        r = x
    return r


#mutation kernel

def dm(x):
    g = random.gauss(0, var_mut)
    while x + g < 0:
        g = random.gauss(0, var_mut)
    return x + g


#functions for acceptation/reject method

def acceptation_rejet_clone(t, a, n):
    return b(t, a)*(1-(p_mut*p_mut+2*p_mut*(1-p_mut)))/(intmax*n)

def acceptation_rejet_mutant_1(t, a, n):
    return acceptation_rejet_clone(t, a, n)+(b(t, a)*2*p_mut*(1-p_mut))/(intmax*n)

def acceptation_rejet_mutant_2(t, a, n):
    return acceptation_rejet_mutant_1(t, a, n) + (b(t, a)*p_mut*p_mut)/(intmax*n)

def acceptation_rejet_mort(t, a, n):
    return acceptation_rejet_mutant_2(t, a, n) + (d(t, a) + (n-1)*intensite_competition)
        /(intmax*n)



#transition of the process: it is based on a classical acceptation/reject method

def transition(p, time):

#initialisation for acceptation/reject method

#living individuals
        viv = p[(p[:, 2] == 1), :]

#population size
    n = len(viv[:, 0])

#jump time
    jump_time = random.expovariate(1)/(intmax*n*n)

#uniform law on (0,1)
    u = random.uniform(0, 1)

#random sampling of one individual
    ind = random.randint(0, n-1)


    w = time - viv[ind, 3] + jump_time #age of this individual

#acceptation/reject method
    while u > acceptation_rejet_mort(viv[ind, :2], w, n):
        jump_time += (random.expovariate(1)/(intmax*n*n))
        u = random.uniform(0, 1)
        ind = random.randint(0, n-1)
        w = time - viv[ind, 3] + jump_time

#when accepted
    s = viv[ind, :]
    a1 = acceptation_rejet_clone(s[:2], w, n)
    a2 = acceptation_rejet_mutant_1(s[:2], w, n)
    a3 = acceptation_rejet_mutant_2(s[:2], w, n)

    a = 0
```

```
    if w > s[1]:
        a = 1

#Lansing effect
    s[:2] = lansing_effect(s[:2], w)

#if clonal birth
    if u <= a1:
        c1 = [1, time + jump_time, 0.0, p[-1, 5]+1, s[5], a]
        p = numpy.vstack((p, numpy.append(numpy.array(s[:2]), numpy.array(c1))))

#if birth with mutation (1 trait)
    elif a1 < u <= a2:
        x1 = int(numpy.random.randint(1, 3, 1))
        z1 = [[dm(s[0]), s[1]], [s[0], dm(s[1])]]
        c2 = [1, time + jump_time, 0.0, p[-1, 5]+1, s[5], a]
        p = numpy.vstack((p, numpy.append(numpy.array(z1[x1-1]), numpy.array(c2))))

#if birth with mutation (2 traits)
    elif a2 < u <= a3:
        c3 = [1, time + jump_time, 0.0, p[-1, 5]+1, s[5], a]
        p = numpy.vstack((p, numpy.append(numpy.array([dm(s[0]), dm(s[1])]), numpy.array(
            c3))))

#if death
    else:
        p[(p[:, 5] == s[5]), 2] = 0
        p[(p[:, 5] == s[5]), 4] = time + jump_time

    time += jump_time

    return p, time


#for simulating one trajectory of the process of number_of_jumps jumps

def trajectoire(p):
    population_size = sum(p[:, 2])                                          time = 0
    for cs in range(0, number_of_jumps):
        if population_size >= 1:
            x = transition(p, time)
            p = x[0]
            time = x[1]
            population_size = sum(p[:, 2])
            print cs
        else:
            break
    p[(p[:, 4] == 0), 4] = time
    return p, time
```

acknowledgements. We acknowledge partial support by the Chaire Modélisation Mathématique et Biodiversité of Veolia Environment - École Polytechnique - Museum National d'Histoire Naturelle - FX. We acknowledge partial support by CNRS and the ATIP/Avenir-Aviesan kick-starting group leaders program. This work was also supported by a public grant as part of the investissement d'avenir project, reference ANR-11-LABX-0056-LMH, LabEx LMH. Finally, we acknowledge the reviewers for their interesting and constructive suggestions.

References

[1] A Ackleh and Shuhua Hu. Comparison between stochastic and deterministic selection-mutation models. *Mathematical Biosciences and Engineering*, 4(2):133, 2007.

[2] Ruben C Arslan, Kai P Willführ, Emma M Frans, Karin JH Verweij, Paul-Christian Bürkner, Mikko Myrskylä, Eckart Voland, Catarina Almqvist, Brendan P Zietsch, and Lars Penke. Older fathers' children have lower evolutionary fitness across four centuries and in four populations. *Proc. R. Soc. B*, 284(1862):20171562, 2017.

[3] Hal Caswell. Reproductive value, the stable stage distribution, and the sensitivity of the population growth rate to changes in vital rates. *Demographic Research*, 23:531–548, 2010.

[4] N Champagnat, R Ferrière, and G Ben Arous. The canonical equation of adaptive dynamics: a mathematical view. *Selection*, 2(1-2):73–83, 2002.

[5] Nicolas Champagnat. A microscopic interpretation for adaptive dynamics trait substitution sequence models. *Stochastic processes and their applications*, 116(8):1127–1160, 2006.

[6] Nicolas Champagnat, Régis Ferrière, and Sylvie Méléard. Unifying evolutionary dynamics: from individual stochastic processes to macroscopic models. *Theoretical population biology*, 69(3):297–321, 2006.

[7] Nicolas Champagnat and Sylvie Méléard. Polymorphic evolution sequence and evolutionary branching. *Probability Theory and Related Fields*, 151(1-2):45–94, 2011.

[8] Frédérique Clément, Frédérique Robin, and Romain Yvinec. Analysis and calibration of a linear model for structured cell populations with unidirectional motion: Application to the morphogenesis of ovarian follicles. *arXiv preprint arXiv:1712.05372*, 2017.

[9] Emilie Dambroise, Lea Monnier, Lu Ruisheng, Hugo Aguilaniu, J-S Joly, Herve Tricoire, and Michael Rera. Two phases of aging separated by the smurf transition as a public path to death. *Scientific reports*, 6:23523, 2016.

[10] Charles Darwin. *The origin of species by means of natural selection, or, the preservation of favoured races in the struggle for life*. London, John Murray, 1871.

[11] Ulf Dieckmann and Richard Law. The dynamical theory of coevolution: a derivation from stochastic ecological processes. *Journal of mathematical biology*, 34(5-6):579–612, 1996.

[12] Stewart N Ethier and Thomas G Kurtz. *Markov processes: characterization and convergence*, volume 282. John Wiley & Sons, 2009.

[13] Daniel Fabian and Thomas Flatt. The evolution of aging. *Nature Education Knowledge*, 3(9), 2011.

[14] Nicolas Gast and Bruno Gaujal. Markov chains with discontinuous drifts have differential inclusion limits. *Performance Evaluation*, 69(12):623–642, 2012.

[15] Morton E Gurtin and Richard C MacCamy. Non-linear age-dependent population dynamics. *Archive for Rational Mechanics and Analysis*, 54(3):281–300, 1974.

[16] John Burdon Sanderson Haldane. *New paths in genetics*. George allen & Unwin, 1942.

[17] William D Hamilton. The moulding of senescence by natural selection. *Journal of theoretical biology*, 12(1):12–45, 1966.

[18] Owen R Jones, Alexander Scheuerlein, Roberto Salguero-Gómez, Carlo Giovanni Camarda, Ralf Schaible, Brenda B Casper, Johan P Dahlgren, Johan Ehrlén, María B García, Eric S Menges, et al. Diversity of ageing across the tree of life. *Nature*, 505(7482):169, 2014.

[19] Thomas BL Kirkwood and Steven N Austad. Why do we age? *Nature*, 408(6809):233, 2000.

[20] Thomas BL Kirkwood and Thomas Cremer. Cytogerontology since 1881: a reappraisal of august weismann and a review of modern progress. *Human genetics*, 60(2):101–121, 1982.

[21] Filippos D Klironomos, Johannes Berg, and Sinéad Collins. How epigenetic mutations can affect genetic evolution: model and mechanism. *Bioessays*, 35(6):571–578, 2013.

[22] Albert I Lansing. A transmissible, cumulative, and reversible factor in aging. *Journal of Gerontology*, 2(3):228–239, 1947.

[23] Albert I Lansing. A nongenic factor in the longevity of rotifers. *Annals of the New York Academy of Sciences*, 57(1):455–464, 1954.

[24] L. Markus. Asymptotically autonomous differential systems. In *Contributions to the theory of nonlinear oscillations, vol. 3*, Annals of Mathematics Studies, no. 36, pages 17–29. Princeton University Press, Princeton, N. J., 1956.

[25] Peter Brian Medawar. *An unsolved problem of biology*. College, 1952.

[26] Sylvie Méléard and Viet Chi Tran. Trait substitution sequence process and canonical equation for age-structured populations. *Journal of mathematical biology*, 58(6):881–921, 2009.

[27] J.A.J. Metz, S.A.H. Geritz, G. Meszéna, F.A.J. Jacobs, and J.S. van Heerwaarden. Adaptive dynamics: a geometrical study of the consequences of nearly faithful reproduction. *Stochastic and Spatial Structures of Dynamical Systems. Elsevier Science, Burlington, MA*, pages 183–231, 1996.

[28] Johan AJ Metz, Roger M Nisbet, and Stefan AH Geritz. How should we define 'fitness' for general ecological scenarios? *Trends in Ecology & Evolution*, 7(6):198–202, 1992.

[29] José C Noguera, Neil B Metcalfe, and Pat Monaghan. Experimental demonstration that offspring fathered by old males have shorter telomeres and reduced lifespans. *Proc. R. Soc. B*, 285(1874):20180268, 2018.

[30] Benoît Perthame. *Transport equations in biology*. Springer Science & Business Media, 2006.

[31] Nicholas K Priest, Benjamin Mackowiak, and Daniel EL Promislow. The role of parental age effects on the evolution of aging. *Evolution*, 56(5):927–935, 2002.

[32] Jan Prüß. Equilibrium solutions of age-specific population dynamics of several species. *Journal of Mathematical Biology*, 11(1):65–84, 1981.

[33] Michael Rera, Sepehr Bahadorani, Jaehyoung Cho, Christopher L Koehler, Matthew Ulgherait, Jae H Hur, William S Ansari, Thomas Lo Jr, D Leanne Jones, and David W Walker. Modulation of longevity and tissue homeostasis by the drosophila pgc-1 homolog. *Cell metabolism*, 14(5):623–634, 2011.

[34] Michael Rera, Rebecca I Clark, and David W Walker. Intestinal barrier dysfunction links metabolic and inflammatory markers of aging to death in drosophila. *Proceedings of the National Academy of Sciences*, 109(52):21528–21533, 2012.

[35] Michael Rera, Céline Vallot, and Christel Lefrançois. The smurf transition: new insights on ageing from end-of-life studies in animal models. *Current opinion in oncology*, 30(1):38–44, 2018.

[36] Robert J Schmitz, Matthew D Schultz, Mathew G Lewsey, Ronan C O'Malley, Mark A Urich, Ondrej Libiger, Nicholas J Schork, and Joseph R Ecker. Transgenerational epigenetic instability is a source of novel methylation variants. *Science*, 334(6054):369–373, 2011.

[37] Horst R Thieme. Asymptotically autonomous differential equations in the plane. *The Rocky Mountain Journal of Mathematics*, pages 351–380, 1994.

[38] Viet Chi Tran. *Modèles particulaires stochastiques pour des problèmes d'évolution adaptative et pour l'approximation de solutions statistiques*. PhD thesis, Université de Nanterre-Paris X, 2006.

[39] Viet Chi Tran. Large population limit and time behaviour of a stochastic particle model describing an age-structured population. *ESAIM: Probability and Statistics*, 12:345–386, 2008.

[40] Hervé Tricoire and Michael Rera. A new, discontinuous 2 phases of aging model: lessons from drosophila melanogaster. *PloS one*, 10(11):e0141920, 2015.

[41] Thomas Tully and Amaury Lambert. The evolution of postreproductive life span as an insurance against indeterminacy. *Evolution*, 65(10):3013–3020, 2011.

[42] Glenn F Webb. *Theory of nonlinear age-dependent population dynamics*. CRC Press, 1985.

[43] A Weismann. The origin of the markings of caterpillars. *Studies in the theory of descent/Tr. and ed. R. Meldola. London: Sampson Low. Marston. Searle Rivington*, pages 161–389, 1881.

[44] August Weismann, Edward Poulton, Sir Bagnall, S. Schönland, A. E Shipley, and Sir Arthur Everett. *Essays upon heredity and kindred biological problems*, volume v.1. Oxford,Clarendon Press, 1891.