The evolution of variance in sequential defences

Lingzi Wang*, Stephen J Cornell & *Michael P. Speed*

Department of Evolution, Ecology and Behaviour, Institute of Integrative Biology,

Faculty of Health & Life Sciences, University of Liverpool, Liverpool, L69 7ZB, UK

Abstract

The defences used by organisms against predators display a great degree of variability. Defence phenotypes can differ substantially among individuals of the same species, and a single individual can itself deploy a variety of defences. Here, we use a mathematical model that includes mutation and selection to understand the evolutionary origin of this variability in a population of a species that deploys defences sequentially ("first" and "second" defences). Typically, the first defence evolves to have lower variance, i.e. appears more closely accumulated around the ideal phenotype, than the second defence (even when the breaching the first defence incurs more fitness loss than breaching the second defence with the other parameters the same for both defences). However, if the first defence is much less effective in repelling predators, or is much less tolerant of deviation from the ideal phenotype, then the first defence can evolve to have higher variance than the second. Other factors like mutation strength and the losses in the fitness when each defence fails also influence the defence variance. Larger mutation rate incurs larger equilibrium variances, and when the comparative importance in fitness of one defence increases, then the ratio between the variances of this defence and the other defence decreases. Sequentially acting defences are found in many organisms, so we encourage empirical research to test our theoretical predictions.

keywords: Phenotypic variation, multiple defences, predation, selection, herbivory

*correponding author, email address: lingzi.wang@liverpool.ac.uk

1 Introduction

Protective defences against organisms' enemies, such as predators, parasites and pathogens, are ubiquitous (Cott 1944, Caro 2004, Ruxton *et al.* 2004, Schoonhoven *et al.* 2005, Walters 2011, Magoli 2016) and the study of adaptations for defence is consequently a major theme in adaptive evolutionary biology.

Evolutionary studies of defences often focus on one or more perspectives, including: the evolutionary history of defence mechanisms (Futuyma & Agrawal 2009, Agrawal *et al.* 2009, Magoli 2016), their roles in macroevolutionary patterns (Arbuckle & Speed 2015; Harriss & Arbuckle 2016; Blanchard & Moreau 2017), the variety of forms of defences used in taxonomic groups (e.g. Edmunds 1974, Caro 2004), the influence of life-history variation on defence (Norris & Evans, 2000, Zuk & Stoehr 2002, Higginson & Ruxton 2009, Higginson *et al.* 2011, 2012), coevolution (Gilman *et al.* 2012, Britton *et al.* 2007) and strategies for optimal investment in defences (Svennungsen *et al.* 2011, Broom *et al.* 2010).

Despite the extensive research in the biology of defence, an area that has received relatively little attention is the nature of defensive variation between individuals and between species. Thus, many studies which seek to understand the function and mode of action of defensive phenotypes focus (rightly) on species typical defences, rather than variation within species. The notable exception to this is seen when frequency dependent evolution causes stable polymorphisms in defences, for example those that give the greatest net benefit when rare, such as parasitic Batesian mimicry (Ruxton *et al.* 2004). Some classes of defence are however, very variable within populations. Chemical defences of plants (More *et al.* 2014) and animals (Speed, 2012) are, for example, notoriously variable, both in terms of the concentrations of compounds that can repel and deter predatory enemies, and even in the mixtures of compounds that are present in different individuals (Speed, 2012). Arguably, less is known about variation in other

forms of defence in animals such as camouflage or warning signals, because of an emphasis on species-typical traits. However, the recent onset of methods for measuring colour patterns is enabling some evaluation of levels of variation in animal colouration, but overall conclusions cannot be made at present. Similarly, chemical ecologists have for a long period been able to evaluate (and demonstrate) variation in secondary metabolites in plants (Goodger, Capon & Woodrow, 2002). In addition variation in physical defences (density of protective trichome hairs, thickness of cuticles and waxes etc), can be measured, and reveal the level of variation there is within populations (Mauricio, 1998).

One reason for the interest in the variation of defences is, as described above, that they can be very variable indeed. There is an apparent paradox here; traits that are viewed as vital to survival of individuals are none the less highly variable, suggesting that some individuals are poorly protected in populations. Several explanations have been proposed including frequency dependence (rare toxins work best) because of predator-counter adaptation and coevolution (Speed *et al.*, 2006). A second compelling explanation is that the effectiveness of some forms of defence saturate at levels that are phenotypically cheap to achieve by organisms, hence a lot of observed variation is above a threshold of effectiveness-saturation, of little effect on survival and with little variation in costs between individuals (Speed *et al.* 2012). It might be for example that some defensive chemicals are cheaply synthesised and stored, and the observed levels of variation in concentration imply nothing about variation in survival from attack.

Here we propose an additional and potentially predictive explanation for different levels of variation in different kinds of defence.

We reason that many defences often work in what Frank (1993) calls "sequential layers". Defences are in effect ordered as a set of barriers surrounding the organism: each one must be crossed in turn by an enemy before it can reach the valuable core tissues of the victim. As Endler (1991) and others point out (Broom *et al*, 2010), those components of sequential defence suites which are met first will be challenged by enemies

more frequently than those that are met only later in a sequence. A general conclusion is then that those defences met or deployed early in an encounter with an enemy will have a larger contribution to the protection of a victim than those met later. Suppose we have two defences that act in sequence, and the probability that an enemy successfully crosses each is 0.5. For each time the first defence is challenged, the second defence is challenged only half as often, and its contribution to survival is half that of the first.

Put simply then, selection is likely to be weaker on later acting, than on earlier acting defences. We may then predict that the mutation-selection equilibrium for a defensive trait is different depending at what stage in encounters with an enemy it is deployed. For example, an organism whose only defence is chemical in nature relies very strongly on that defence and selection to keep it at an optimal value will be very strong. Should however the organism evolve an effective physical defence that acts before the chemical defence, then the chemical defence is used less often and makes a smaller contribution to survival from an encounter. The "corrupting" effects of mutation will make more headway against the unifying force of selection toward the optimal value of the trait.

Though it is easy to argue this verbally, here we seek a quantitative analysis to evaluate the effects of order of deployment on mutation. We present a model that is simple in structure (with only two stages) and investigates the dynamical evolution of paired, sequential defences, seeking out the conditions in which there will be inequalities in variation between them arising from mutation-selection balance. A key point is that while we do confirm that the later acting defence may often evolve to be more variable, we can identify conditions in which the later-acting defences are the least variable.

Several other theoretical papers look at sequential defences, and though none focus on the question we ask here about variability, we will briefly comment on their relevance to our model here.

Broom *et al*'s (2010) sequential defence model gave different benefit and cost values to both defences, and found the optimal strategies (none/preattack defence/post attack

defence/both defences) in regards to these different benefit and cost values. In the model due to the order in defence, the relation of benefits and costs of the first defence can influence the condition when the second defence is used or not; but the relation of benefits and costs of the second defence can not influence the condition when the first defence is used or not. So the first defence might be relatively more influential in the optimal decision making.

Speed *et al* (2012), Gilman *et al* (2012) and Sasaki (2000) gave coevolutionary models to explore the investments in different defences. In Speed *et al* (2012), victims could invest in one or more defences, and coevolution could be the reason for more than one defence, since when there is not coevolution, plants evolve to invest in only one toxin trait. Gilman *et al* (2012)'s paper showed that increasing the number of defence traits, and the correlation between traits could help the victims to win the evolutionary contest, so different defence traits functioning interactively might be the reason why more than one defence is profitable. Sasaki (2000) found that when the effects of defence genes acts multiplicatively, different resistant defences exist in either coevolutionary cycle or static equilibria depending on the cost of resistance and virulence values. These models gave reasons for the existence of more than one defences, although these models did not show the defence variance evolution of each defences as we did.

2 The Model

We consider a prey species that mounts two sequential defences against predation. We assume that each individual has a phenotype x describing its first-level defense and a phenotype y characterising its second-level defense. These phenotypes determine the success of each defense repelling predation, so not breached by enemies, and we denote by $p_1(x)$ the probability that the first defense holds and by $p_2(y)$ the probability that the second defense holds. We assume that there are ideal values a and b for these

phenotypes, so that $p_1(x)$ is maximal at $x = a$, and $p_2(y)$ is maximal when $y = b$, and that the defense will be less likely to hold when the phenotypic values are further away from these ideal values. Specifically, we assume functional forms

$$p_1(x) = e^{-\epsilon_1 - \frac{(x-a)^2}{\alpha}}, \tag{1}$$

$$p_2(y) = e^{-\epsilon_2 - \frac{(y-b)^2}{\beta}}, \tag{2}$$

so that the first (respectively, second) defense will hold with probability $e^{-\epsilon_1}$ (respectively, $e^{-\epsilon_2}$) when the corresponding phenotype is at its ideal value $x = a$ (respectively, $y = b$), and that the tolerance of phenotypic deviations from the ideal will be wide when α (respectively, β) is large.

Since these defenses are met sequentially, there are three mutually exclusive scenarios: (1) defense 1 holds, which occurs with probability $p_1(x)$; (2) defense 1 fails, but defense 2 holds, which occurs with probability $(1 - p_1(x))p_2(y)$; (3) defenses 1 and 2 both fail, which occurs with probability $(1-p_1(x))(1-p_2(y))$. We assume that the prey's fitnesses under these three scenarios are f_1, f_2 and f_3 respectively, and note that these represent increasingly adverse outcomes for the prey so that $f_1 \geq f_2 \geq f_3 \geq 0$. The average fitness of an individual with phenotype (x, y) is then given by

$$\Phi(x, y) = f_1 p_1(x) + f_2(1 - p_1(x))p_2(y) + f_3(1 - p_1(x))(1 - p_2(y)) \tag{3}$$

We now consider how the population distribution of the phenotypes evolves in time. We assume non-overlapping generations, and let $N_t(x_t, y_t)$ represent the density of individuals with vector of phenotypes (x_t, y_t) at generation t. We first consider the case where the phenotype is completely heritable with no mutation. In that case, the abundance of

individuals with phenotypes (x_t, y_t) simply changes by $\Phi(x_t, y_t)$ at each generation:

$$N_{t+1}(x_{t+1}, y_{t+1}) = N_t(x_t, y_t)\Phi(x_t, y_t). \tag{4}$$

Secondly, we consider the case where phenotype mutates between generations. If $M(x_t, y_t, x_{t+1}, y_{t+1})$ is the mutation kernel, i.e. the probability density that a parent with phenotype (x_t, y_t) has offspring of phenotype (x_{t+1}, y_{t+1}), then

$$N_{t+1}(x_{t+1}, y_{t+1}) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} N_t(x_t, y_t)\Phi(x_t, y_t)M(x_t, y_t, x_{t+1}, y_{t+1})dx_t\, dy_t. \tag{5}$$

We assume that the phenotypes mutate independently with a Gaussian mutation kernel of the form

$$M(x_t, y_t; x_{t+1}, y_{t+1}) = \frac{1}{\pi\mu}e^{-\frac{(x_t-x_{t+1})^2-(y_t-y_{t+1})^2}{\mu}}, \tag{6}$$

so that $\mu/2$ is the variance in the mutation per generation in either phenotype.

We assume that, at the first generation, the phenotypes are independently normally distributed with variances $v_1/2$ and $w_1/2$ and means $\bar{x}_1$ and $\bar{y}_1$ for the first and second defenses respectively, and n_1 the total population number.

$$N_1(x_1, y_1) = n_1\frac{1}{\pi\sqrt{v_1 w_1}}e^{-(\frac{(x_1-\bar{x}_1)^2}{v_1}+\frac{(y_1-\bar{y}_1)^2}{w_1})}. \tag{7}$$

The evolution of the distribution of phenotypes in the population is therefore obtained by starting with the initial distribution given in eqn. (7) and iterating eqns (4) or (5), substituting for Φ from eqn. (3), p_1 and p_2 from eqns. (1) and (2), and M from eqn. (6). We characterise the population distribution by the means $\overline{x}_t$ and $\overline{y}_t$ and variances

$\frac{v_t}{2}$ and $\frac{w_t}{2}$ of the phenotypic values, defined as

$$\begin{aligned}
\overline{x}_t &= E_t(x_t) \\
\overline{y}_t &= E_t(y_t) \\
\frac{v_t}{2} &= E_t((x - \overline{x}_t)^2) \\
\frac{w_t}{2} &= E_t((y - \overline{y}_t)^2),
\end{aligned}$$

where $E_t(\cdot)$ represents the expectation value at time t, defined as

$$E_t(f(x_t, y_t)) = \frac{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x_t, y_t) N_t(x_t, y_t) dx_t dy_t}{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} N_t(x_t, y_t) dx_t dy_t}$$

for any function $f(x, y)$. Note that the total number of population at t,

$$n_t = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} N_t(x_t, y_t) dx_t dy_t.$$

Also note that the scale of the variance in defence phenotypes is set by the parameters μ, α, and β. That is, if we increase these three parameters by a common factor, the equilibrium value of v_t and w_t will change by the same factor (Appendix B).

To summarise, the parameters, variables and functions are shown in the tables 1,2,3.

Parameters	
μ	mutation strength
a, b	ideal phenotypes
ϵ_1, ϵ_2	effectiveness
α, β	Tolerance of phenotypic deviations from the ideal
f_1, f_2, f_3	conditional fitness

Table 1: Parameter table

Variables	
x_t, y_t	defence phenotype values at time t
$\bar{x}_t, \bar{y}_t$	means of defence phenotype values at time t
$\frac{v_t}{2}, \frac{w_t}{2}$	variances of defence phenotype values at time t
n_t	total population numbers at time t

Table 2: Variable table

Functions	
$p_1(x), p_2(y)$	probability each defence holds
$\Phi(x, y)$	average fitness
$M(x_t, y_t; x_{t+1}, y_{t+1})$	Gaussian Mutation Kernal from (x_t, y_t) to (x_{t+1}, y_{t+1})
$N_t(x_t, y_t)$	population density function about (x_t, y_t)

Table 3: Function table

3 Methods

We use both numerical and analytical approaches to explore how different factors affect the variances of both defences. The section 3.1 is the numerical approach and the section 3.2 is the analytical approach.

3.1 Numerical integration

We are not able to find exact closed-form analytical expressions for the mean or variance of the phenotypes at generation t, so we approximate the continuous distribution of phenotypic values by a discrete set and iterate (4) or (5) numerically. A each t we replace

(x_t, y_t) by the grid of pairs of values $\{(x_{ti}, y_{tj}); i \in \{1, 2, \ldots, n\}, j \in \{1, 2, \ldots, n\}\}$, where

$$x_{ti} = x_{t1} + (i-1)\Delta x, \qquad i \in \{1, 2, \ldots, n\}$$
$$y_{tj} = y_{t1} + (j-1)\Delta y, \qquad j \in \{1, 2, \ldots, n\}.$$

In all cases we start with a population with means $(\bar{x}_1, \bar{y}_1) = (1, 1)$ and variances $(v_1/2, w_1/2) = (2, 2)$, and total population number $n_1 = 10000$. The fitnesses are set to $(f_1, f_2, f_3) = (2, 1, 0.2)$, the ideal phenotypes to $(a, b) = (0, 0)$, the selection forces are $(\alpha, \beta) = (5, 5)$, and the grid is defined by $\Delta x = \Delta y = 0.2$, $n = 101$. All integrals are approximated as follows:

$$\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} F(x_t, y_t) dx_t \, dy_t \approx \sum_{i=1}^{n} \sum_{j=1}^{n} F(x_{ti}, y_{tj}) \Delta x \Delta y,$$

for any function F. The grid of values extends over a range of $(n-1)\Delta x = (n-1)\Delta y = 20$ units, which (for the variance values under consideration) is sufficient for this finite sum to approximate the infinite range of integration. To avoid numerical overflow, after each iteration we replace N_t by

$$N_t(x_{ti}, y_{tj}) \to N'_t(x_{ti}, y_{tj}) = n_1 \frac{N_t(x_{ti}, y_{tj})}{\sum_{i=1}^{n} \sum_{j=1}^{n} N_t(x_{ti}, y_{tj}) \Delta x \Delta y},$$

so that the population is always normalised to contain $n_1 = 10000$ individuals; this does not affect our results, as we are only interested in the relative abundance of different phenotypes.

For the case where there is no mutation, we use $x_{i1} = y_{i1} = -9$ (so that x extends over the range $\overline{x}_1 \pm 10$ and y over the range $\overline{y}_1 \pm 10$), and iterate eqn. (4) at this grid of values for 1000 generations.

For the case where there is mutation, we allow the range of integration to vary as the mean phenotypic values change to ensure that the range of phenotypic values in the

population does not stray too close to (or beyond) the edges of the range of integration. To do this, at each generation we calculate $\overline{x}_t$ and $\overline{y}_t$, and then set

$$x_{t+1,1} = \overline{x}_t - \frac{(n-1)}{2}\Delta x$$
$$y_{t+1,1} = \overline{y}_t - \frac{(n-1)}{2}\Delta y,$$

so that the ranges of x and y grid values at the next generation are centred on $\overline{x}_t$ and $\overline{y}_t$ respectively. Since the grid of (x, y) values changes between generations, we need to determine the density of phenotypes evaluated on the current generation's grid from the density evaluated on the previous generation's grid. We do this by assuming that the density is constant within a range $(\pm\frac{\Delta x}{2}, \pm\frac{\Delta y}{2})$ from the points on the previous generation's grid. We set $\mu = 0.02$ and iterate eqn. (5) for 1000 generations.

3.2 Normal approximation

An alternative method for calculating approximately the evolution in time of the population is to use a moment closure assumption, which is a well established approximation method for stochastic systems that cannot be solved exactly. Moment closure assumes that the distribution of a random variable is well approximated by a particular parametric form (Whittle 1957), and then derives (approximate) equations for the parameters of the distribution. Here, we perform a normal moment closure by assuming that the traits are normally distributed at a generation t, and then calculate the mean and covariance matrix at generation $t + 1$ in terms of the mean and covariance at time t. While the trait will not in general be normally distributed (except at the first generation, where this is assumed), it is reasonable to assume that the iteration equations for the mean and covariance provide a good enough approximation to the true time evolution of the system for the purposes of understanding the general behaviour of the model. This is an uncontrolled approximation, by contrast with the direct numerical solution described

in the previous section (which will describe the dynamics exactly in the limit where the integrals are approximated by sums over a very large and very fine grid), but has the advantage of being much quicker to evaluate and therefore permits a much wider exploration of parameter space. We have tested our approximation scheme against simulation results, and find that it reproduces the patterns in the result well for a wide range of parameters.

We begin by assuming that the traits are normally distributed at time t, and write the distribution of traits as

$$N_t(x_t, y_t) = n_t \frac{\sqrt{|W_t|}}{\pi} \exp\Big(- (z_t - \bar{z}_t)^T W_t (z_t - \bar{z}_t)\Big), \tag{8}$$

where $z_t = \begin{pmatrix} x_t \\ y_t \end{pmatrix}$ is the vector of defence phenotypes, $\bar{z}_t = \begin{pmatrix} \bar{x}_t \\ \bar{y}_t \end{pmatrix}$ is the mean vector of defence phenotypes, and $W_t = (2\Sigma_t)^{-1}$, where

$$\Sigma_t = E_t \begin{pmatrix} x_t^2 & x_t y_t \\ x_t y_t & y_t^2 \end{pmatrix} - E_t \begin{pmatrix} x_t \\ y \end{pmatrix} E_t \begin{pmatrix} x_t & y_t \end{pmatrix}$$

is the covariance matrix for the trait.

We can find the population distribution at the next generation by applying the iteration equation (5), where Φ is defined by equations (1–3) and M from eqn. (6). After performing the integrals over (x_t, y_t) (the details are shown in Appendix A), this leads to

$$N_{t+1}(x_{t+1}, y_{t+1}) = n_t \sqrt{|W_t|} \Theta_t \sum_{j=1}^{4} \theta_{t,j} \frac{1}{2\pi\sqrt{|\Sigma_{t+1,j}|}} \exp\Big(- \frac{1}{2}(z_{t+1} - z'_{t+1,j})^T \Sigma_{t+1,j}^{-1} (z_{t+1} - z'_{t+1,j})\Big), \tag{9}$$

where

$$
\begin{aligned}
z'_{t+1,j} &= (W_t + F_j)^{-1}(W_t \bar{z}_t + F_j \hat{a}) \\
\Sigma_{t+1,j} &= \frac{1}{2}\Big(U^{-1} + (\Sigma_t^{-1} + F_j)^{-1}\Big) \\
\gamma_1 &= f_3 \\
\gamma_2 &= (f_1 - f_3)e^{-\epsilon_1} \\
\gamma_3 &= (f_2 - f_3)e^{-\epsilon_2} \\
\gamma_4 &= (f_3 - f_2)e^{-(\epsilon_1+\epsilon_2)} \\
F_1 &= \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} \\
F_2 &= \begin{pmatrix} \frac{1}{\alpha} & 0 \\ 0 & 0 \end{pmatrix} \\
F_3 &= \begin{pmatrix} 0 & 0 \\ 0 & \frac{1}{\beta} \end{pmatrix} \\
F_4 &= F_2 + F_3 \\
U &= \begin{pmatrix} \frac{1}{\mu} & 0 \\ 0 & \frac{1}{\mu} \end{pmatrix} \\
\theta_{t,j} &= \frac{\gamma_j s_{t,j}}{\Theta_t \sqrt{|W_t + F_j|}} \\
\Theta_t &= \sum_{j=1}^{4} \frac{\gamma_j s_{t,j}}{\sqrt{|W_t + F_j|}} \\
s_{t,j} &= \exp\Big(\frac{1}{\mu}(W_t \bar{z}_t + F_j \hat{a})^T \Big((W_t + F_j)^{-1} + \mu I\Big) K_{t,j}^{-1}(W_t \bar{z}_t + F_j \hat{a}) - \bar{z}_t{}^T W_t \bar{z}_t - \hat{a}^T F_j \hat{a}\Big) \\
\hat{a} &= \begin{pmatrix} a \\ b \end{pmatrix}
\end{aligned}
$$

This shows that N_{t+1} is the sum of four normal distributions with different means and co-variance matrices, so cannot be expressed as a single normal distribution. We can, however, use this expression to compute the mean and covariance of the traits at generation $t+1$ (also see equations (A.9), (A.11) in Appendix A):

$$\begin{aligned}
\overline{z}_{t+1} &= E_{t+1}(z_{t+1}) \\
&= \sum_{j=1}^{4} \theta_{t,j}(W_t + F_j)^{-1}(W_t \bar{z}_t + F_j \hat{a}) \qquad (10) \\
\Sigma_{t+1} &= E_{t+1}\begin{pmatrix} x_{t+1}^2 & x_{t+1}y_{t+1} \\ x_{t+1}y_{t+1} & y_{t+1}^2 \end{pmatrix} - E_t\begin{pmatrix} x_t \\ y \end{pmatrix} E_t\begin{pmatrix} x_t & y_t \end{pmatrix} \\
&= \frac{1}{2}U^{-1} + \sum_{j=1}^{4} \theta_{t,j}(\Sigma_t^{-1} + 2F_j)^{-1} + \sum_{j=1}^{4} \theta_{t,j} \cdot z'_{t+1,j} z'^{T}_{t+1,j} - \bar{z_{t+1}} \bar{z_{t+1}}^{T}. \qquad (11)
\end{aligned}$$

From the expression of F_j $(j = 1, 2, 3, 4)$, we can tell from eqn. (10) that the mean will evolve over time towards the ideal phenotype $\hat{a}$. Also $z'_{t+1,j} = (W_t + F_j)^{-1}(W_t \bar{z}_t + F_j \hat{a})$ $(j = 1, 2, 3, 4)$ approaches to the ideal phenotype $\hat{a}$, therefore the term $\sum_{j=1}^{4} \theta_{t,j} \cdot z'_{t+1,j} z'^{T}_{t+1,j} - \bar{z_{t+1}} \bar{z_{t+1}}^{T}$ in (11) approaches to zero, and the covariance matrix approaches to the following equation (also see in equation (A.12) in Appendix A),

$$\Sigma_{t+1} = \frac{1}{2}U^{-1} + \sum_{j=1}^{4} \theta_{t,j}(\Sigma_t^{-1} + 2F_j)^{-1}. \qquad (12)$$

These equations can be iterated rapidly over time to give an approximation to the time evolution and equilibrium values of $\overline{z}_t$ and Σ_t (We can use (11) to do the iteration for the evolution of $\overline{z}_t$ and Σ_t, and both (11) and (12) can be used to generate equilibrium value of Σ_t).

In the limit $\mu \to 0$, Equation (12) approaches the limit

$$\Sigma_{t+1} = \sum_{j=1}^{4} \theta_{t,j} (\Sigma_t^{-1} + 2F_j)^{-1} \tag{13}$$

from which it can be shown that variance of both the first and second defence evolve towards zero as time t grows (details in Appendix C). If $U \neq 0$, however, it can be shown from (12) that the covariance matrix evolves to a non-zero equilibrium. This shows that mutation is necessary in order for the traits to be variable.

4 Results

Because of selection, the means of both defences evolve towards the ideal phenotype (this can be shown analytically for the normal approximation— see the iteration equation for the mean (10) — and also see numerical results in Figure 2 (a), (b)). We are interested in the evolution of the distribution of phenotypes within the population, but in particular in the variance of the phenotypic values. There are five factors that will influence these variances.

1. Mutation

In this model, mutation must be present for phenotypic variance to be maintained — when there is no mutation, the variances of both traits evolve to be zero. This is is visible in the numerical results Figure 1 (a), (b), and can be shown analytically for the Normal approximation (see Appendix C).

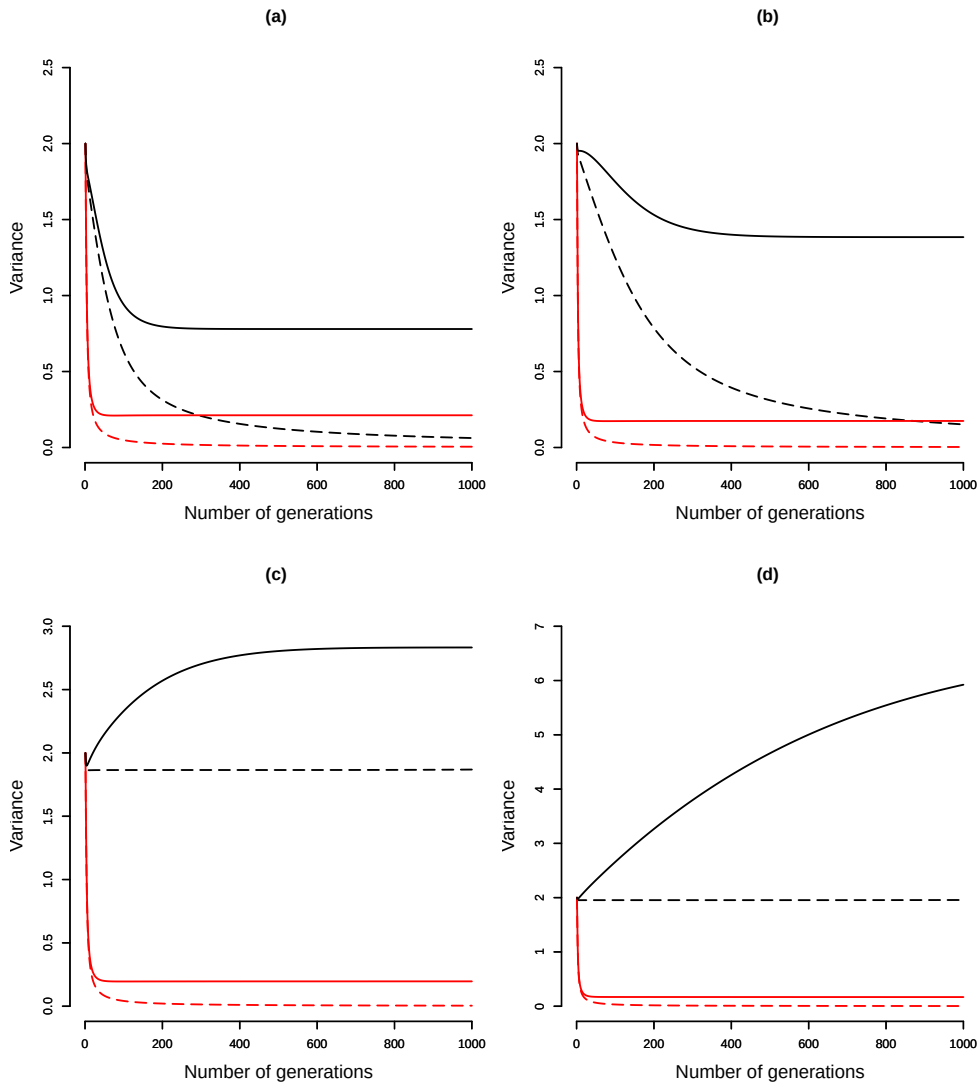


Figure 1: The evolution of phenotypic variances, obtained using numerical integration. Dashed lines indicate the case of no mutation, solid lines the case when $\mu = 0.02$. Red lines indicate the variance in the first defense, black lines the variance in the second defense. (a) $\epsilon_1 = 0.1, \epsilon_2 = 0$, (b) $\epsilon_1 = 0.1, \epsilon_2 = 0.9$, (c) $\epsilon_1 = 0, \epsilon_2 = 0$, (d) $\epsilon_1 = 0, \epsilon_2 = 0.9$. Other parameters: $\alpha = 5$, $\beta = 5$, $(f_1, f_2, f_3) = (2, 1, 0.2)$.

A special case is that when mutation is zero and the first defence is perfectly effective ($\epsilon_1 = 0$), i.e. when the defence will succeed with probability 1 if the trait is at its ideal value $x = a$. As the first defence evolves close to the ideal phenotype, and first defence

variance evolves to be zero, the first defence protects all the victims from the enemies so that the second defence is hardly ever tested and evolves very slowly (Figures 1 (c)(d), 2 (c)(d), 3 (b)).

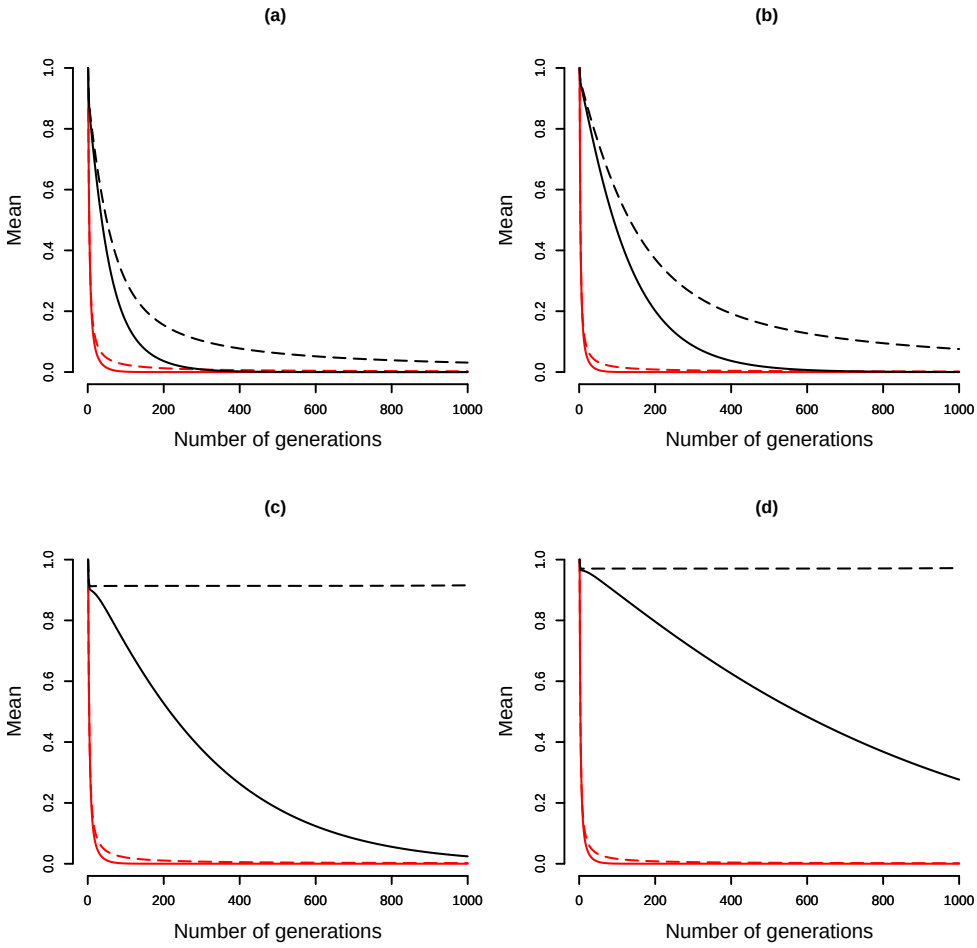


Figure 2: The evolution of phenotypic means, obtained from numerical integration. Dashed lines indicate the case of no mutation, solid lines the case when $\mu = 0.02$. Red lines indicate the mean in the first defence, black lines the mean in the second defense. (a) $\epsilon_1 = 0.1, \epsilon_2 = 0$, (b) $\epsilon_1 = 0.1, \epsilon_2 = 0.9$, (c) $\epsilon_1 = 0, \epsilon_2 = 0$, (d) $\epsilon_1 = 0, \epsilon_2 = 0.9$. Other parameters: $\alpha = 5$, $\beta = 5$, $(f_1, f_2, f_3) = (2, 1, 0.2)$.

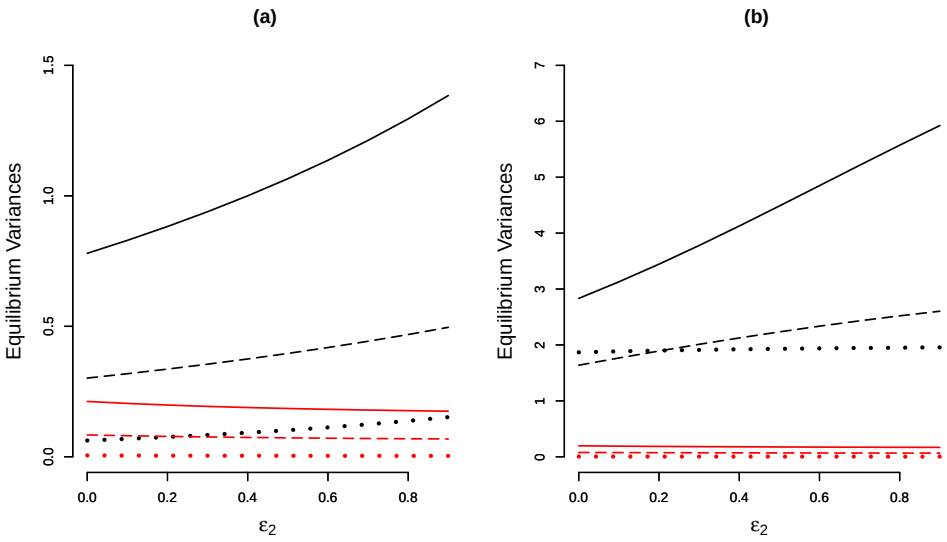


Figure 3: The equilibrium values of the variances, obtained from numerical integration. Red lines: first defence; black lines: second defence. Dotted lines: $\mu = 0$; dashed lines: $\mu = 0.01$; solid lines: $\mu = 0.02$. (a) $\epsilon_1 = 0.1$, (b) $\epsilon_1 = 0$. Other parameters: $\alpha = 5$, $\beta = 5$, $(f_1, f_2, f_3) = 2, 1, 0.2$.

When there is mutation, the variance of both traits evolve to have positive values (Figures 1, 3). This is proved for the moment closure approximation in Appendix A. Stronger mutation lead to higher variances (Figure 3).

2. Order of defence in the sequence

When the first and second defence have the same effectiveness ($\epsilon_1 = \epsilon_2$) and the tolerance range is the same for both defences ($\alpha = \beta$), then the first defence variance is always lower than the second defence variance (i.e., the first defence clusters more closely than the second around its ideal phenotype), no matter what the conditional fitness values are. This is shown in figure 6, where var1/var2<1 (in which var1 stands for the first defence variance, and var2 stands for the second defence variance) along the line $\epsilon_1 = \epsilon_2$; even in figure 6 (a), where f_1 is only a little higher than f_2 so that it makes little difference whether the first defence holds or does not, the first defence variance is still a little smaller than the second defence when $\epsilon_1 = \epsilon_2$.

3. Effectiveness of defences

(1) If the first defence is less effective than the second defence ($\epsilon_1 > \epsilon_2$), then the first defence variance can be larger than the second defence variance (in Figure 6, var1/var2> 1 when $\epsilon_1 > \epsilon_2$). The threshold value for the ineffectiveness ϵ_1 of the first defence, above which the first defence has higher variance than the second, depends also on the conditional fitness values (see the contour lines above the red contour line given different fitness values in Figure 6 (a-c) described also in "Conditional fitness" below).

(2) When the effectiveness of the first defence increases, the first defence variance decreases and the second defence variances increases. When the effectiveness of the second defence increases, the opposite occurs (see Figure 4).

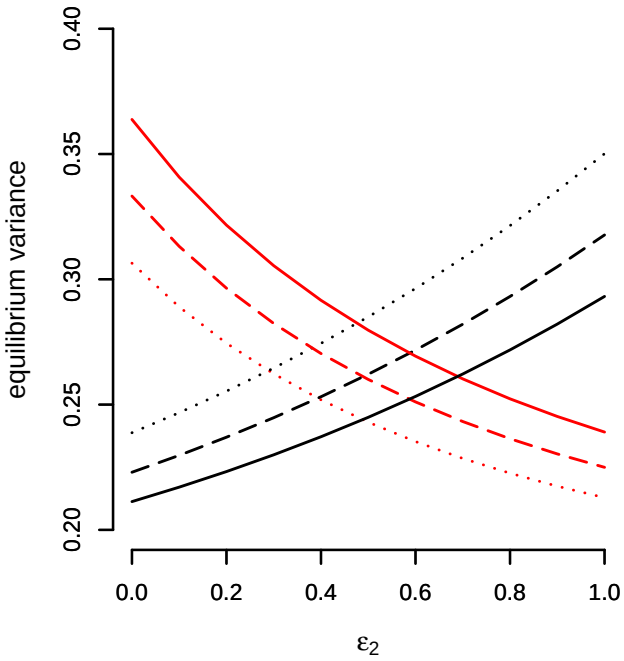


Figure 4: The effect of effectiveness on equilibrium variance. Red lines: first defence; black lines: second defence. Dotted lines: $\epsilon_1 = 1.1$; dashed lines: $\epsilon_1 = 1.3$; solid lines: $\epsilon_1 = 1.5$. Other parameters: $\alpha = 5$, $\beta = 5$, $(f_1, f_2, f_3) = (2, 1, 0.2)$, $\mu = 0.02$.

4. Tolerance of phenotypic deviations from the ideal

We refer to the quantities α and β as the "tolerance of phenotypic deviations from the ideal" on the two defensive traits, because they quantify how sensitive the fitness is to deviations from the ideal trait value. However, the variances of the traits do not depend on these quantities in a straightforward way. When the tolerance of deviation from the ideal on a trait is wide, the variance in that trait has a positive relationship with the tolerance as would usually be expected in a mutation-selection balance (Figure 5 (a), (e), large values of α or β). However, when the tolerance is narrowed beyond a threshold value, the variance in that trait starts to increase. This is because mutation limits how small the variance in a trait can become, so that as α (for example) decreases more individuals have a maladapted first defence, which as a result is increasingly likely to fail. Since this defence is very likely to fail anyway, its importance in determining the animal's relative fitness actually decreases, and the variance of that trait increases, as α decreases further. (Figure 5 (a), (e), small values of α or β). Increasing the ineffectiveness of a defence (ϵ_1 or ϵ_2) makes this effect stronger, so that the positive relationship starts at a smaller value of the tolerance. Since narrowing one defence's tolerance makes it more likely to fail, and therefore makes the other defence more important, the variance in the other defence consequentially decreases (Figure 5 (b), (d)). Because the first defence variance and second defence variance can either increase or decrease as the tolerance values change, the ratio of these variances can either increase or decrease (Figure 5 (c), (f)).

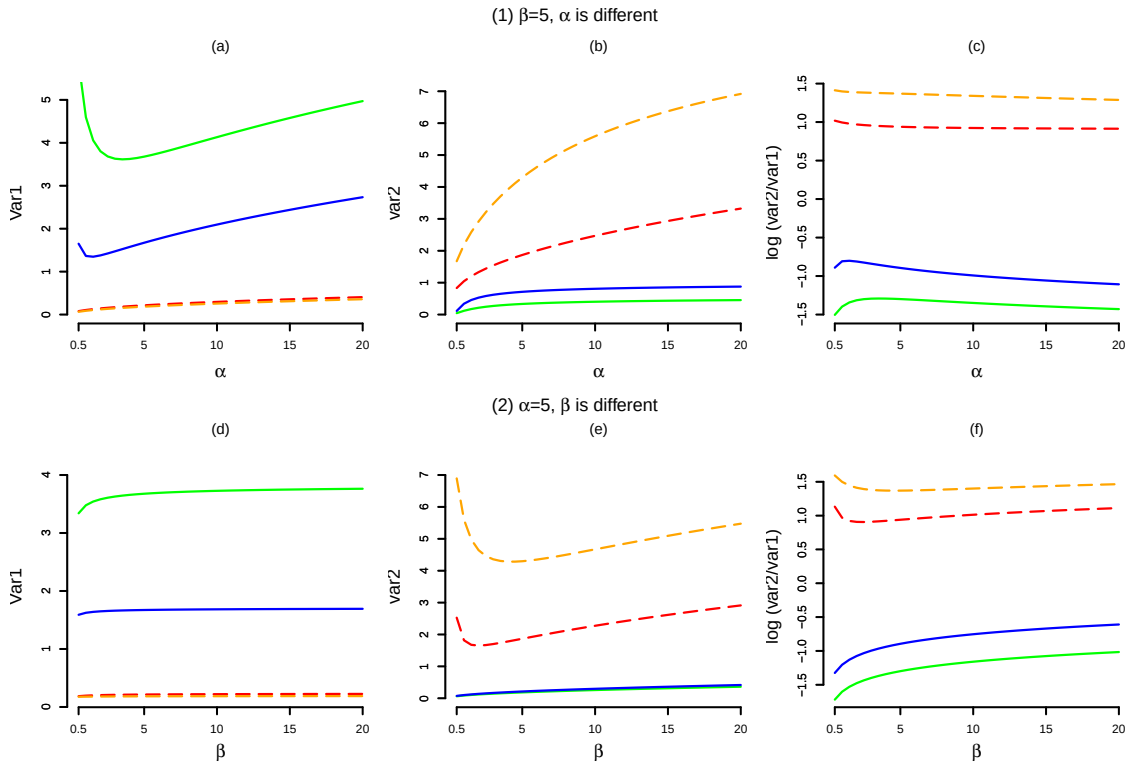
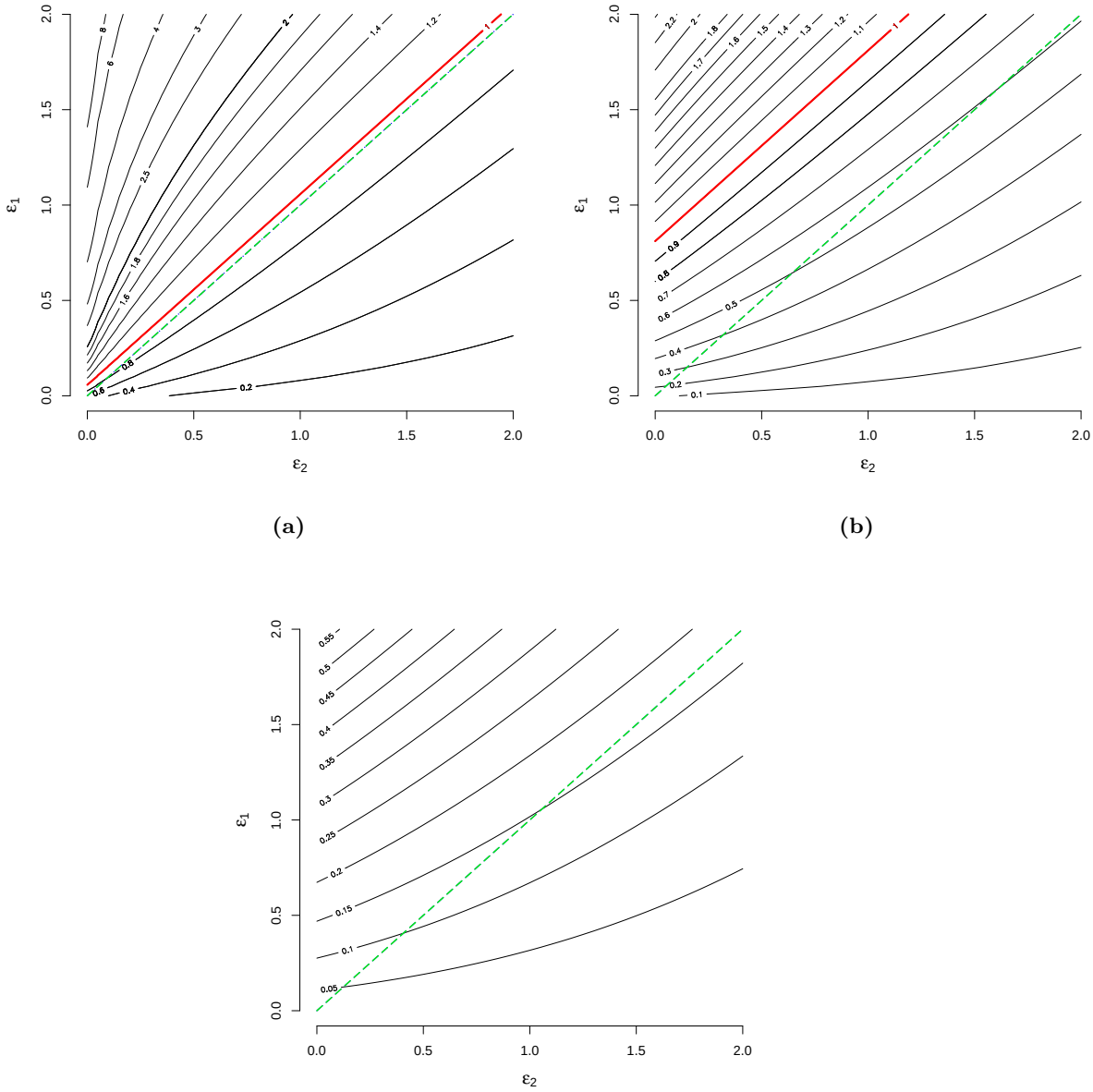


Figure 5: The effect of tolerance of phenotypic deviations from the ideal on equilibrium variances. Red line: $\epsilon_1 = 0$, $\epsilon_2 = 0$; yellow line: $\epsilon_1 = 0$, $\epsilon_2 = 0.9$; green line: $\epsilon_1 = 5$, $\epsilon_2 = 0$; black line: $\epsilon_1 = 5$, $\epsilon_2 = 0.9$; In (a,b,c) $\beta = 5$, and in (d,e,f) $\alpha = 5$. Other parameters: $(f_1, f_2, f_3) = (2, 1, 0.2)$, $\mu = 0.02$. These results were obtained by the the Normal Approximation (12), which is much faster than the numerical iteration.

5. Conditional fitness

We change the relative values of $\frac{f_1}{f_2}$ (the ratio between the conditional fitness f_1 while the first defence holds and the conditional fitness f_2 while the first defence fails) and $\frac{f_2}{f_3}$ (the ratio between the conditional fitness f_2 while the second defence holds and the conditional fitness f_3 while the second defence fails) to see the relative importance of the first and second defences. When the relative fitness value $\frac{f_1}{f_2}$ increases and $\frac{f_2}{f_3}$ decreases, meaning that the first defence becomes more important, then var1/var2 decreases (given the same ϵ_1 and ϵ_2). This can be seen in Figure 6(a), where $\frac{f_1}{f_2}$ is lowest and $\frac{f_2}{f_3}$ is highest, the value of var1/var2 (keeping the same values of (ϵ_1, ϵ_2))is highest; and in Figure 6(c)

where $\frac{f_1}{f_2}$ is highest and $\frac{f_2}{f_3}$ is lowest, the value of var1/var2 is lowest. As seen in Figure 6, the second defence must be much more effective than the first defence ($\epsilon_2 \ll \epsilon_1$, the upper-left side of red solid lines) for the first defence variance to be larger than the second defence variance. Note that here only three typical cases of fitness values are showed (in Figure 6(a)(b)(c) respectively) because for the other values (e.g. $f_2 = 0$ which may correspond to that a victim animal is killed when the second defence is breached, or the other values of $f_2 > 0$ which may correspond to that a victim plant is still alive when the second defence is breached), the figures are similar and the relation showed above keep the same.



(a) **(b)** **(c)**

Figure 6: Contours of var1/var2 in the (ϵ_1, ϵ_2) plane, for different conditional fitness values. Red solid line: the contour line var1/var2=1 (above which var1/var2> 1, below which var1/var2 < 1) (red line is not visible in (c) as it occurs only when $\epsilon_1 > 2$, which is out off the range of ϵ_1-axis); green dashed line: the line $\epsilon_1 = \epsilon_2$. (a) $(f_1, f_2, f_3) = (2, 1.9, 0.2)$, (b) $(f_1, f_2, f_3) = (2, 1, 0.2)$, (c) $(f_1, f_2, f_3) = (2, 0.3, 0.2)$. Other parameters: $\alpha = 5$, $\beta = 5$, $\mu = 0.02$. These results were obtained by the the Normal Approximation (12), which is much faster than the numerical iteration.

5 Discussion

In this paper we aimed to predict and explain patterns in the variation of anti-predator defences, when those defences are deployed in a predictable sequence. It is well known that defences can be variable in a population, but there is relatively little systematic evaluation of patterns of variation, even though diversifying evolutionary mechanisms are easily identified (Speed *et al.*, 2012; Moore *et al.*, 2014; Barnett, 2014; Speed, 2015; Jeckel *et al.*, 2015). It is our contention that the sequential nature of defence may often cause predictable patterns of diversity, allowing testable hypotheses about defence variation. Hence, we built and interrogated a model representing both the selection and mutation mechanisms on the evolution of population distribution of two sequential defences. By using both analytical and numerical methods, we get the evolution processes and equilibrium evolution values of the variances in both defences. We first briefly account for the major determinants of defence variation in our model and subsequently relate its general findings to a wider set of defences and to other theoretical treatments of defence evolution.

Factors predicted to be influential in defence variance

(1) Mutation. Mutation is the reason of why there are defence variances in our model. In the absence of mutation, both defence variances will evolve toward zero, whereas if there is mutation, equilibria of mutation and selection that give variances > 0. Unsurprisingly, the stronger the mutation is, the larger the two defence variances evolve to be.

(2) Order of defence in the sequence. If the first defence and the second defence are as effective as each other, and the tolerance range is the same for both defences, then the first defence distribution evolves to have smaller variance than the second defence distribution (see Results, Figure 6). That means the first defence is more closely

gathered around the ideal phenotype and therefore has more influence in protecting the victims from being attacked. Hence, the model demonstrates our verbal argument in the introduction: that earlier acting defences can often evolve to lower levels of variation than later acting defences.

(3) Effectiveness of defences. Whether the defences are effective enough in the environment (in the sense of successfully repelling an enemy) is also important to the evolution of population variance. If the first defence in a sequence is not as effective in repelling predators as the second defence, then the force of selection can be felt most strongly on the second defence, with the consequence that it has a lower equilibrium variance than the first defence. This is counter to the intuition in our Introduction, that defences deployed earlier are less variable than those deployed later in sequence, and shows the value of a formal model.

(4) Tolerance of phenotypic deviations from the ideal. We consider that effectiveness of a defence in repelling enemies becomes weaker as the phenotype diverges from the ideal value for the relevant trait. A key measure in this model is therefore how much defensive effectiveness is lost for an incremental deviation from the ideal phenotypic value: in effect the tolerance of the phenotype in relation to its defensive function (α, β). If tolerance of phenotypic deviation is narrow, then even when the phenotype is similar to the ideal phenotype, the defence is likely to fail and be breached. On the contrary, if the tolerance is very wide and permissive, even the phenotype is quite dissimilar to the ideal phenotype, the defence is likely to hold. Both the first and second defence variances will evolve to be high when the tolerance is very narrow or wide. When the tolerance is very wide, the phenotypes quite different from the ideal phenotypes are effective to protect the victims, then the population variance could evolve to be very large. When the tolerance is very narrow, even the phenotypes are quite similar as the ideal phenotypes are useless in protecting the victims, then it will not be profitable for the phenotypes to evolve to be similar to the ideal phenotypes, so the population variance will also be

very wide. An interesting result pertains now if the first defence is subject to narrow tolerance and the second defence to wider tolerance. Here the first defence can be of little use, and contributes little to prey survival, hence mutation accumulates and the phenotype becomes variable. Variation in the second defence however is fundamental to prey survival, hence the model predicts a lower equilibrium value for mutation (Figure 5). This gives us an additional scenario in which the first defence may evolve to a higher level of variation than the second.

(5) Conditional Fitness. In our model, the relative importance of whether the first defence holds to whether the second defence holds are described by the relative conditional fitness values. When the relative conditional fitness value $\frac{f_1}{f_2}$ increases (which means that the importance of holding the first defence increases) and the relative conditional fitness value $\frac{f_2}{f_3}$ decreases (which means that the importance of holding the second defence increases), then the ratio between the first defence variance and the second defence variance decreases. The contrary is true when the conditions are reversed.

Application to biological and other contexts

Sequentially-layered defences are very common in biological and other contexts. Many plants and animals present their enemies with layered defences. John Endler for example (1991) argued that an attack by a predator on its (animal) prey is typically composed of a sequence of six stages: (i) encounter (spatial proximity), (ii) detection, (iii) identification, (iv) approach, (v) subjugation and ultimately (vi) consumption. At each stage in this sequence the prey organism can put up one or more lines of defence with the aim of preventing, interrupting and stopping the attack. An animal prey may for example hide (to prevent encounter, i, and detection, ii), use masquerade and cryptic colouration (to prevent detection and identification, ii, iii), perhaps form aggressive defensive groups (to prevent approach, iv). They may alternatively have a startle display or use vigilance and rapid escape behaviours (to prevent approach, iv). They may vio-

lently retaliate (to prevent subjugation, v.) perhaps using stings, spines or bites and/or deploy irritating or toxic chemicals (to prevent subjugation and consumption, v, vi). At each stage in the sequence Endler identified, one or more defences could be deployed by a prey animal, and they could often operate sequentially, some defences typically used only if earlier-acting defences have failed to stop the predation event. Here we have simplified to two layers, but the model could be extended to larger set of defences. An important point is, however, that we expect sequentially acting defences to be very common in organismal defence, hence our model has generality.

One very general result is greater variation in later-acting defences. There is some evidence supportive of a key feature of the model, that later-acting defences are used less often than earlier defences, and thus contribute less to fitness. A meta-analysis of studies of plant-herbivore interactions shows that variation in earlier-acting defences (physical, morphological, physiology, chemical defences) in plants is better in predicting herbivores' damage than later defences (toxic secondary metabolites; Carmona *et al*, 2011, Cardenas *et al.*, 2014). A number of authors have remarked on the high levels of variability in defensive toxins (see review in Speed *et al.* 2012), but we do not know of studies that measure the variability of sets of defences and relate these to their use in a sequence. We suggest that this is an interesting area for valuable empirical research.

We note that anti-pathogen systems (skin, immune responses) are also usually layered in their organisation, hence the model could be elaborated to consider these kinds of defensive systems. There are also interesting parallels between the organisation of biological and human military defences. Both concern protection of valuable yet vulnerable targets, seeking optimal deployment of costly defensive "assets". A relevant military tactic is "layered defence" in which sets of defensive resources, such as inter-ballistic missiles, are deployed in sequence; when a first line of defence fails against an incoming threat a second line of defence activates to minimise further risk, and after that perhaps a third or fourth defence, and so on. In the military theory literature, layered defence has

been described and modelled by Wilkening (2000). We suggest that it might be an interesting question to determine whether, in military contexts and perhaps cyber-security, later-acting defences are more variable in their form and effectiveness than earlier-acting defences.

Developments of the model

We would draw the reader's attention to some key assumptions in the model. First, we assume that there is an optimal value for each defence, and any deviation from that is punished by reduced efficiency in repelling enemies. This assumption does simplify implementation, giving us a clear set of results, but it does bring some limitations. On the one hand, this assumption may fit morphological defences well - for example defensive spines may need to be the right size to repel certain enemies. It does not represent some kinds of chemical defence as well however. Here concentrations that are too low may lead to reduced efficiency, but higher and higher values probably become more effective at anti-predator defence, albeit in a saturating manner, not less. In this case the model would have to be modified to incorporate this asymmetry in defensive benefit. An interesting question is whether the distribution of naturally occurring defensive toxins is asymmetric in this manner. Secondly, we assumed that once each defence is breached, it cannot be healed (e.g. the spines of golden barrel cactus *Echinocactus grusonii* once moved from areole cannot grow back), in comparison to that for some organisms, defences can replenish when damaged (e.g. the claw of Florida stone crab, *Menippe mercenaria* can grow back when broken with the diaphragm at the claw joint intact). Also for the case such as different parts of plants maybe attacked independently, defences can be breached several times in these different parts. Since our model is based on an average fitness, it can describe qualitatively the dependence of the relative variances on fitness costs and tolerances in the above cases, but it would be good to be further modified in the case of multiple successive breaches of each defence (where the organism heals be-

tween each attack) and may also for the case when different parts of plants are breached independently. Finally, we have deliberately excluded costs of defences in the model, in part to keep the structure simple and predictions tractable, but there is profitable scope for including costs in a more complex development. We note that studies of prey defence can not always identify measurable costs to defence in any case (Zvereva & Kozlov 2016).

6 Conclusion

We aimed to explore the patterns of the defence variations when defences are deployed in sequence. We built a model with two sequential defences, and use both selection and mutation as the evolutionary mechanisms on the evolution of population distribution of the two defences. Through both analytical and numerical methods, we found that typically the earlier defence has lower variance than the later defence, which means that the earlier defence phenotypes are more closely accumulated around the ideal phenotype than the later defence phenotypes. This matches with intuition and some research that the earlier defences have higher probability in use and therefore probably have higher anti-predator effect. Besides, our formal model also gives a broader explanation that when the first defence is less effective in repelling the predators, or the first defence is less tolerant of phenotypic deviations from the ideal, then the first defence could evolve to have higher variance than the second defence. Sequential defences are widely seen in different defence systems, therefore our model might be predictive in a wide range of areas. Since the empirical research of sequential defence variances is rare, related research could be valuable.

7 References

Agrawal, A. A., Salminen, J. P., & Fishbein, M. (2009). Phylogenetic trends in phenolic metabolism of milkweeds (Asclepias): evidence for escalation. Evolution, 63(3), 663-673.

Arbuckle, K., & Speed, M. P. (2015). Antipredator defenses predict diversification rates. Proceedings of the National Academy of Sciences, 112(44), 13597-13602.

Barnett, C. A., Bateson, M., & Rowe, C. (2014). Better the devil you know: avian predators find variation in prey toxicity aversive. Biology letters, 10(11), 20140533.

Blanchard, B. D., & Moreau, C. S. (2017). Defensive traits exhibit an evolutionary trade-off and drive diversification in ants. Evolution, 71(2), 315-328.

Britton, N. F., Planqué, R., & Franks, N. R. (2007). Evolution of defence portfolios in exploiter-victim systems. Bulletin of mathematical biology, 69(3), 957-988.

Broom, M., Higginson, A. D., & Ruxton, G. D. (2010). Optimal investment across different aspects of anti-predator defences. Journal of theoretical biology, 263(4), 579-586.

Cardenas, R. E., Valencia, R., Kraft, N. J., Argoti, A., & Dangles, O. (2014). Plant traits predict inter and intraspecific variation in susceptibility to herbivory in a hyper-diverse Neotropical rain forest tree community. Journal of Ecology, 102(4), 939-952.

Carmona, D., Lajeunesse, M. J., & Johnson, M. T. (2011). Plant traits that predict resistance to herbivores. Functional Ecology, 25(2), 358-367.

Caro, T. (2005). Antipredator defenses in birds and mammals. University of Chicago Press.

Cott, H. B. (1940). Adaptive coloration in animals. Methuen; London.

Edmunds, M. (1974). Defence in animals. Longman.

Endler, J. A. (1991). Interactions between predators and prey. Behavioural ecology: an evolutionary approach 3:169-196.

Frank, S. A. (1993). Evolution of host-parasite diversity. Evolution, 47(6), 1721-

1732.

Futuyma, D. J., & Agrawal, A. A. (2009). Macroevolution and the biological diversity of plants and herbivores. Proceedings of the National Academy of Sciences, 106(43), 18054-18061.

Gilman, R. T., Nuismer, S. L., & Jhwueng, D. C. (2012). Coevolution in multidimensional trait space favours escape from parasites and pathogens. Nature, 483(7389), 328.

Goodger, J. Q. D., Capon, R. J., & Woodrow, I. E. (2002). Cyanogenic polymorphism in Eucalyptus polyanthemos Schauer subsp. vestita L. Johnson and K. Hill (Myrtaceae). Biochemical Systematics and Ecology, 30(7), 617-630.

Harris, R. J., & Arbuckle, K. (2016). Tempo and mode of the evolution of venom and poison in tetrapods. Toxins, 8(7), 193.

Higginson, A. D., & Ruxton, G. D. (2009). Dynamic models allowing for flexibility in complex life histories accurately predict timing of metamorphosis and antipredator strategies of prey. Functional ecology, 23(6), 1103-1113.

Higginson, A. D., Delf, J., Ruxton, G. D., & Speed, M. P. (2011). Growth and reproductive costs of larval defence in the aposematic lepidopteran Pieris brassicae. Journal of Animal Ecology, 80(2), 384-392.

Higginson, A. D., McNamara, J. M., & Houston, A. I. (2012). The starvation-predation trade-off predicts trends in body size, muscularity, and adiposity between and within taxa. The American Naturalist, 179(3), 338-350.

Jeckel, A. M., Saporito, R. A., & Grant, T. (2015). The relationship between poison frog chemical defenses and age, body size, and sex. Frontiers in zoology, 12(1), 27.

Malagoli, D. (Ed.). (2016). The evolution of the immune system: conservation and diversification. Academic Press.

Mauricio, R. (1998). Costs of resistance to natural enemies in field populations of the annual plant Arabidopsis thaliana. The American Naturalist, 151(1), 20-28.

Moore, B. D., Andrew, R. L., Külheim, C., & Foley, W. J. (2014). Explaining intraspecific diversity in plant secondary metabolites in an ecological context. New Phytologist, 201(3), 733-750.

Norris, K., & Evans, M. R. (2000). Ecological immunology: life history trade-offs and immune defense in birds. Behavioral Ecology, 11(1), 19-26.

Ruxton, G. D., Sherratt, T. N., & Speed, M. P. (2004). Avoiding attack: the evolutionary ecology of crypsis, warning signals and mimicry. Oxford University Press.

Sasaki, A. (2000). Host-parasite coevolution in a multilocus gene-for-gene system. Proceedings of the Royal Society of London B: Biological Sciences, 267(1458), 2183-2188.

Schoonhoven, L. M., Van Loon, J. J., & Dicke, M. (2005). Insect-plant biology. Oxford University Press on Demand.

Speed, M. P., Fenton, A., Jones, M. G., Ruxton, G. D., & Brockhurst, M. A. (2015). Coevolution can explain defensive secondary metabolite diversity in plants. New Phytologist, 208(4), 1251-1263.

Speed, M. P., Ruxton, G. D., & Broom, M. (2006). Automimicry and the evolution of discrete prey defences. Biological Journal of the Linnean Society, 87(3), 393-402.

Speed, M. P., Ruxton, G. D., Mappes, J., & Sherratt, T. N. (2012). Why are defensive toxins so variable? An evolutionary perspective. Biological Reviews, 87(4), 874-884.

Svennungsen, T. O., Holen, Ø. H., & Leimar, O. (2011). Inducible defenses: continuous reaction norms or threshold traits?. The American Naturalist, 178(3), 397-410.

Walters, D. (2011). Plant defense: warding off attack by pathogens, herbivores and parasitic plants. John Wiley & Sons.

Whittle, P. (1957). On the Use of the Normal Approximation in the Treatment of Stochastic Processes. Journal of the Royal Statistical Society. 19 (2): 268-281

Wilkening, D. A. (2000). A simple model for calculating ballistic missile defense effectiveness. Science & Global Security, 8(2), 183-215.

Zuk, M., & Stoehr, A. M. (2002). Immune defense and host life history. the american naturalist, 160(S4), S9-S22.

Zvereva, E. L., & Kozlov, M. V. (2016). The costs and effectiveness of chemical defenses in herbivorous insects: a meta-analysis. Ecological Monographs, 86(1), 107-124.

8 Appendix

8.1 Appendix A

To get the population distribution at the next generation $N_{t+1}(x_{t+1}, y_{t+1})$ (equation (9)) we first write the fitness function in matrix form. From (1) and (2), the fitness function (3) can be written as

$$\Phi(x_t, y_t) = f_3 + (f_1 - f_3)e^{-\epsilon_1 - \frac{(x_t-a)^2}{\alpha}} + (f_2 - f_3)e^{-\epsilon_2 - \frac{(y_t-b)^2}{\beta}} + (f_3 - f_2)e^{-\epsilon_1 - \frac{(x_t-a)^2}{\alpha}} e^{-\epsilon_2 - \frac{(y_t-b)^2}{\beta}}$$

$$= \sum_{j=1}^{4} \gamma_j \exp\Big(- (z_t - \hat{a})^T F_j (z_t - \hat{a})\Big) \tag{A.1}$$

where,

$$\gamma_1 = f_3, \quad \gamma_2 = (f_1 - f_3)e^{-\epsilon_1}, \quad \gamma_3 = (f_2 - f_3)e^{-\epsilon_2}, \quad \gamma_4 = (f_3 - f_2)e^{-(\epsilon_1+\epsilon_2)},$$

$$F_1 = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}, \quad F_2 = \begin{pmatrix} \frac{1}{\alpha} & 0 \\ 0 & 0 \end{pmatrix}, \quad F_3 = \begin{pmatrix} 0 & 0 \\ 0 & \frac{1}{\beta} \end{pmatrix}, \quad F_4 = \begin{pmatrix} \frac{1}{\alpha} & 0 \\ 0 & \frac{1}{\beta} \end{pmatrix}$$

$$z_t - \hat{a} = \begin{pmatrix} x_t - a \\ y_t - b \end{pmatrix}, \quad \hat{a} = \begin{pmatrix} a \\ b \end{pmatrix}$$

The mutation function (6) can be written as

$$M(x_t, y_t; x_{t+1}, y_{t+1}) = \frac{1}{\pi\mu} \exp\Big(- (z_t - z_{t+1})^T U (z_t - z_{t+1})\Big) \tag{A.2}$$

where $z_t - z_{t+1} = \begin{pmatrix} x_t - x_{t+1} \\ y_t - y_{t+1} \end{pmatrix}$, $U = \begin{pmatrix} \frac{1}{\mu} & 0 \\ 0 & \frac{1}{\mu} \end{pmatrix}$.

Then from (8), (A.1) and (A.2), the population distribution density iteration function (5) is as follows.

$$N_{t+1}(x_{t+1}, y_{t+1}) = \frac{n_t\sqrt{|W_t|}}{\pi^2\mu} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \sum_{j=1}^{4} \gamma_j \exp\Big(-(z_t - \bar{z}_t)^T W_t (z_t - \bar{z}_t) - (z_t - \hat{a})^T F_j (z_t - \hat{a})$$

$$-(z_t - z_{t+1})^T U (z_t - z_{t+1})\Big) dx_t dy_t \tag{A.3}$$

Since z_t is a normal distributed vector, we collect terms for z_t, let $K_{t,j} = W_t + F_j + U$, and complete the square for z_t:

$$N_{t+1}(x_{t+1}, y_{t+1}) = \frac{n_t\sqrt{|W_t|}}{\pi^2\mu} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \sum_{j=1}^{4} \gamma_j \exp\Big(-z_t^T K_{t,j} z_t + 2(W_t\bar{z}_t + F_j\hat{a} + U z_{t+1})^T z_t$$

$$-(\bar{z}_t{}^T W_t \bar{z}_t + \hat{a}^T F_j \hat{a} + z_{t+1}^T U z_{t+1})\Big) dx_t dy_t$$

$$= \frac{n_t\sqrt{|W_t|}}{\pi^2\mu} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \sum_{j=1}^{4} \gamma_j \exp\Big(-(z_t - K_{t,j}^{-1}(W_t\bar{z}_t + F_j\hat{a} + U z_{t+1}))^T K_{t,j} (z_t - K_{t,j}^{-1}(W_t\bar{z}_t + F_j\hat{a} + U z_{t+1}))$$

$$+(W_t\bar{z}_t + F_j\hat{a} + U z_{t+1})^T K_{t,j}^{-1} (W_t\bar{z}_t + F_j\hat{a} + U z_{t+1}) - (\bar{z}_t{}^T W_t \bar{z}_t + \hat{a}^T F_j \hat{a} + z_{t+1}^T U z_{t+1})\Big) dx_t dy_t$$

From the fact that the integral of normal density function equals 1, the above is equivalent to

$$N_{t+1}(x_{t+1}, y_{t+1}) = \frac{n_t\sqrt{|W_t|}}{\pi^2\mu}\sum_{j=1}^{4}\gamma_j\frac{\pi}{\sqrt{|K_{t,j}|}}\exp\Big((W_t\bar{z}_t+F_j\hat{a}+Uz_{t+1})^T K_{t,j}^{-1}(W_t\bar{z}_t+F_j\hat{a}+Uz_{t+1})$$

$$-(\bar{z}_t{}^T W_t\bar{z}_t + \hat{a}^T F_j\hat{a} + z_{t+1}^T U z_{t+1})\Big) \tag{A.4}$$

Since z_{t+1} is the next generation defence phenotype vector, if we collect terms for z_{t+1}, given that $U = \begin{pmatrix} \frac{1}{\mu} & 0 \\ 0 & \frac{1}{\mu} \end{pmatrix}$, we have

$$N_{t+1}(x_{t+1}, y_{t+1}) = \frac{n_t}{\pi\mu}\sqrt{|W_t|}\sum_{j=1}^{4}\frac{\gamma_j}{\sqrt{|K_{t,j}|}}\exp\Big(-\frac{1}{\mu}z_{t+1}^T(I-(\mu K_{t,j})^{-1})z_{t+1}$$

$$+2\frac{1}{\mu}(W_t\bar{z}_t + F_j\hat{a})^T K_{t,j}^{-1}z_{t+1} + (W_t\bar{z}_t + F_j\hat{a})^T K_{t,j}^{-1}(W_t\bar{z}_t + F_j\hat{a}) - \bar{z}_t{}^T W_t\bar{z}_t - \hat{a}^T F_j\hat{a}\Big)$$

To complete a square, the above is equivalent to

$$N_{t+1}(x_{t+1}, y_{t+1}) = \frac{n_t}{\pi\mu}\sqrt{|W_t|}\sum_{j=1}^{4}\frac{\gamma_j}{\sqrt{|K_{t,j}|}}\exp\Big(-\frac{1}{\mu}(z_{t+1}-z'_{t+1,j})^T(I-(\mu K_{t,j})^{-1})(z_{t+1}-z'_{t+1,j})$$

$$+\frac{1}{\mu}z'^T_{t+1,j}(I-(\mu K_{t,j})^{-1})z'_{t+1,j} + (W_t\bar{z}_t + F_j\hat{a})^T K_{t,j}^{-1}(W_t\bar{z}_t + F_j\hat{a}) - \bar{z}_t{}^T W_t\bar{z}_t - \hat{a}^T F_j\hat{a}\Big)$$

where $z'_{t+1,j} = (I-(\mu K_{t,j})^{-1})^{-1}K_{t,j}^{-1}(W_t\bar{z}_t + F_j\hat{a}) = (K_{t,j} - \frac{1}{\mu}I)^{-1}(W_t\bar{z}_t + F_j\hat{a}) = (W_t + F_j)^{-1}(W_t\bar{z}_t + F_j\hat{a})$. If we put the expression of $z'_{t+1,j}$ into the second term in the exponential bracket, and combine the second term with the third term, we have

$$N_{t+1}(x_{t+1}, y_{t+1}) = \frac{n_t}{\pi\mu}\sqrt{|W_t|}\sum_{j=1}^{4}\frac{\gamma_j}{\sqrt{|K_{t,j}|}}\exp\Big(-\frac{1}{\mu}(z_{t+1}-z'_{t+1,j})^T(I-(\mu K_{t,j})^{-1})(z_{t+1}-z'_{t+1,j})$$

$$+\frac{1}{\mu}(W_t\bar{z}_t + F_j\hat{a})^T\big((K_{t,j}-\frac{1}{\mu}I)^{-1}K_{t,j}^{-1} + \mu K_{t,j}^{-1}\big)(W_t\bar{z}_t + F_j\hat{a}) - \bar{z}_t{}^T W_t\bar{z}_t - \hat{a}^T F_j\hat{a}\Big)$$

Since $K_{t,j} = W_t + F_j + U$, we have

$$N_{t+1}(x_{t+1}, y_{t+1}) = \frac{n_t}{\pi\mu}\sqrt{|W_t|}\sum_{j=1}^{4}\frac{\gamma_j}{\sqrt{|K_{t,j}|}}\exp\Big(-\frac{1}{\mu}(z_{t+1}-z'_{t+1,j})^T(I-(\mu K_{t,j})^{-1})(z_{t+1}-z'_{t+1,j})$$

$$+\frac{1}{\mu}(W_t\bar{z}_t + F_j\hat{a})^T\big((W_t+F_j)^{-1}K_{t,j}^{-1} + \mu K_{t,j}^{-1}\big)(W_t\bar{z}_t + F_j\hat{a}) - \bar{z}_t{}^T W_t\bar{z}_t - \hat{a}^T F_j\hat{a}\Big)$$

If we let $s_{t,j} = \exp\Big(\frac{1}{\mu}(W_t\bar{z}_t + F_j\hat{a})^T\Big((W_t+F_j)^{-1} + \mu I\Big)K_{t,j}^{-1}(W_t\bar{z}_t + F_j\hat{a}) - \bar{z}_t{}^T W_t\bar{z}_t - \hat{a}^T F_j\hat{a}\Big)$, then

$$N_{t+1}(x_{t+1}, y_{t+1}) = \frac{n_t}{\pi\mu}\sqrt{|W_t|}\sum_{j=1}^{4}\frac{\gamma_j}{\sqrt{|K_{t,j}|}}s_{t,j}\exp\Big(-\frac{1}{\mu}(z_{t+1}-z'_{t+1,j})^T(I-(\mu K_{t,j})^{-1})(z_{t+1}-z'_{t+1,j})\Big)$$

Then we can write it into the normal distribution form:

$$N_{t+1}(x_{t+1}, y_{t+1}) = n_t\sqrt{|W_t|}\sum_{j=1}^{4}\frac{\gamma_j s_{t,j}}{\sqrt{|W_t+F_j|}}$$

$$\frac{\sqrt{|(I-(\mu K_{t,j})^{-1})|}}{\pi\mu}\exp\Big(-\frac{1}{\mu}(z_{t+1}-z'_{t+1,j})^T(I-(\mu K_{t,j})^{-1})(z_{t+1}-z'_{t+1,j})\Big)$$

This is a combination of four normal distribution functions, with mean $z'_{t+1,j} =$

$(W_t + F_j)^{-1}(W_t\bar{z}_t + F_j\hat{a})$, and variance

$$\Sigma_{t+1,j} = \frac{\mu}{2}(I - (\mu K_{t,j})^{-1})^{-1} = \frac{\mu}{2}(K_{t,j} - \frac{1}{\mu}I)^{-1}K_{t,j} = \frac{\mu}{2}(W_t + F_j)^{-1}(W_t + F_j + U)$$

$$= \frac{\mu}{2}I + \frac{1}{2}(W_t + F_j)^{-1} = \frac{1}{2}\Big(U^{-1} + (W_t + F_j)^{-1}\Big) = \frac{1}{2}\Big(U^{-1} + (\frac{1}{2}\Sigma_t^{-1} + F_j)^{-1}\Big)$$

$$= \frac{1}{2}U^{-1} + (\Sigma_t^{-1} + 2F_j)^{-1} \qquad j = 1, 2, 3, 4 \tag{A.5}$$

So the next generation population function can be written as

$$N_{t+1}(x_{t+1}, y_{t+1}) = n_t\sqrt{|W_t|}\sum_{j=1}^{4}\frac{\gamma_j s_{t,j}}{\sqrt{|W_t + F_j|}}\frac{1}{2\pi\sqrt{|\Sigma_{t+1,j}|}}\exp\Big(-\frac{1}{2}(z_{t+1}-z'_{t+1,j})^T\Sigma_{t+1,j}^{-1}(z_{t+1}-z'_{t+1,j})\Big)$$

Let $\Theta_t = \sum_{j=1}^{4}\frac{\gamma_j s_{t,j}}{\sqrt{|W_t+F_j|}}$, and $\theta_{t,j} = \frac{\gamma_j s_{t,j}}{\Theta_t\sqrt{|W_t+F_j|}}$ $(j = 1, 2, 3, 4)$ then $\sum_{j=1}^{4}\theta_{t,j} = 1$. Then the above is equivalent to

$$N_{t+1}(x_{t+1}, y_{t+1}) = n_t\sqrt{|W_t|}\Theta_t\sum_{j=1}^{4}\theta_{t,j}\frac{1}{2\pi\sqrt{|\Sigma_{t+1,j}|}}\exp\Big(-\frac{1}{2}(z_{t+1}-z'_{t+1,j})^T\Sigma_{t+1,j}^{-1}(z_{t+1}-z'_{t+1,j})\Big)$$

Therefore, the **probability density function** in the next generation can be written as

$$f(x_{t+1}, y_{t+1}) = N_{t+1}(x_{t+1}, y_{t+1})/\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}N_{t+1}(x_{t+1}, y_{t+1})dz_{t+1}$$

$$= \sum_{j=1}^{4}\theta_{t,j}\frac{1}{2\pi\sqrt{|\Sigma_{t+1,j}|}}\exp\Big(-\frac{1}{2}(z_{t+1} - z'_{t+1,j})^T\Sigma_{t+1,j}^{-1}(z_{t+1} - z'_{t+1,j})\Big) \tag{A.6}$$

$$= \sum_{j=1}^{4}\theta_{t,j}f_j(x_{t+1}, y_{t+1})$$

Therefore the population distribution probability density function in the $(t+1)$-th generation is written as a combination of four normal probability density functions

$$z'_{t+1,j} = (W_t + F_j)^{-1}(W_t\bar{z}_t + F_j\hat{a}) \tag{A.7}$$

and covariance matrix

$$\Sigma_{t+1,j} = \frac{1}{2}U^{-1} + (\Sigma_t^{-1} + 2F_j)^{-1} \qquad j = 1,2,3,4 \tag{A.8}$$

Note that the integral of $f(x_{t+1}, y_{t+1})$ in respect of (x_{t+1}, y_{t+1}) equals 1, which is the property of the probability density function. We use $E_{N_j}(\cdot)$ to denote the expectation of each of the four corresponding normal population distribution function.

Therefore the **mean in the next generation** is

$$\begin{aligned}
\bar{z_{t+1}} = E(z_{t+1}) &= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} z_{t+1} f(x_{t+1}, y_{t+1}) dz_{t+1} \\
&= \sum_{j=1}^{4} \theta_{t,j} \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} z_{t+1} f_i(x_{t+1}, y_{t+1}) dz_{t+1} \\
&= \sum_{j=1}^{4} \theta_{t,j} E_{N_j}(z_{t+1}) \\
&= \sum_{j=1}^{4} \theta_{t,j} z'_{t+1,j} \\
&= \sum_{j=1}^{4} \theta_{t,j} (W_t + F_j)^{-1}(W_t\bar{z}_t + F_j\hat{a})
\end{aligned} \tag{A.9}$$

This is the iteration equations for the mean between generations.

The above also equals

$$\bar{z}_{t+1} = \sum_{j=1}^{4} \theta_{t,j}(W_t + F_j)^{-1} W_t \bar{z}_t + \sum_{j=1}^{4} \theta_{t,j}(W_t + F_j)^{-1} F_j \hat{a} \tag{A.10}$$

The above is a first-order difference equation. Since $\sum_{j=1}^{4} \theta_{t,j}(W_t + F_j)^{-1} W_t + \sum_{j=1}^{4} \theta_{t,j}(W_t + F_j)^{-1} F_j = I$, the mean $\bar{z}_{t+1}$ will gradually approach to the equilibrium –the ideal phenotype $\hat{a}$ as t increases. This equilibrium value can be got by letting both $\bar{z}_{t+1}$ and z_t in the above equation equal z_T and solve the equation, we will have that $z_T = \hat{a}$.

For variance, we have the **covariance matrix in the next generation**

$$\Sigma_{t+1} = E\begin{pmatrix} x_{t+1}^2 & x_{t+1}y_{t+1} \\ x_{t+1}y_{t+1} & y_{t+1}^2 \end{pmatrix} - E\begin{pmatrix} x_{t+1} \\ y_{t+1} \end{pmatrix} E\begin{pmatrix} x_{t+1} & y_{t+1} \end{pmatrix}$$

$$= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \begin{pmatrix} x_{t+1}^2 & x_{t+1}y_{t+1} \\ x_{t+1}y_{t+1} & y_{t+1}^2 \end{pmatrix} \cdot \sum_{j=1}^{4} \theta_{t,j} f_i(x_{t+1}, y_{t+1}) dz_{t+1} - \bar{z_{t+1}} \bar{z_{t+1}}^T$$

$$= \sum_{j=1}^{4} \theta_{t,j} \cdot \begin{pmatrix} E_{N_j}(x_{t+1}^2) & E_{N_j}(x_{t+1}y_{t+1}) \\ E_{N_j}(x_{t+1}y_{t+1}) & E_{N_j}(y_{t+1}^2) \end{pmatrix} - \bar{z_{t+1}} \bar{z_{t+1}}^T$$

$$= \sum_{j=1}^{4} \theta_{t,j} \cdot \begin{pmatrix} Var_{N_j}(x_{t+1}) + E_{N_j}(x_{t+1})^2 & Cov_{N_j}(x_{t+1}, y_{t+1}) + E_{N_j}(x_{t+1})E_{N_j}(y_{t+1}) \\ Cov_{N_j}(x_{t+1}, y_{t+1}) + E_{N_j}(x_{t+1})E_{N_j}(y_{t+1}) & Var_{N_j}(y_{t+1}) + E_{N_j}(y_{t+1})^2 \end{pmatrix}$$

$$- \bar{z_{t+1}} \bar{z_{t+1}}^T$$

$$= \sum_{j=1}^{4} \theta_{t,j} \cdot \Sigma_{t+1,j} + \sum_{j=1}^{4} \theta_{t,j} \cdot z'_{t+1,j} z'^T_{t+1,j} - \bar{z_{t+1}} \bar{z_{t+1}}^T$$

$$= \sum_{j=1}^{4} \theta_{t,j} \Big(\frac{1}{2} U^{-1} + (\Sigma_t^{-1} + 2F_j)^{-1} \Big) + \sum_{j=1}^{4} \theta_{t,j} \cdot z'_{t+1,j} z'^T_{t+1,j} - \bar{z_{t+1}} \bar{z_{t+1}}^T$$

$$= \frac{1}{2}U^{-1} + \sum_{j=1}^{4} \theta_{t,j}(\Sigma_t^{-1} + 2F_j)^{-1} + \sum_{j=1}^{4} \theta_{t,j} \cdot z'_{t+1,j} z'^{T}_{t+1,j} - \bar{z_{t+1}} \bar{z_{t+1}}^{T} \tag{A.11}$$

where $z'_{t+1,j}$ is the j-th mean shown in (A.7).

Now we can use the variance iteration equations from (A.11) to get the evolution of variance across generations. As $\bar{z}_t$ approaches to the ideal phenotype $\hat{a}$ as $t \to +\infty$, so does $\bar{z_{t+1}}$ and $z'_{t+1,j} = (W_t + F_j)^{-1}(W_t \bar{z}_t + F_j \hat{a})$ $(j = 1, 2, 3, 4)$ also approaches to the ideal phenotype $\hat{a}$. Therefore the term $\sum_{j=1}^{4} \theta_{t,j} \cdot z'_{t+1,j} z'^{T}_{t+1,j} - \bar{z_{t+1}} \bar{z_{t+1}}^{T}$ in (A.11) approaches to zero, so covariance matrix (A.11) approaches to the following equation as time grows.

$$\Sigma_{t+1} = \frac{1}{2}U^{-1} + \sum_{j=1}^{4} \theta_{t,j}(\Sigma_t^{-1} + 2F_j)^{-1} \tag{A.12}$$

which is larger than $\frac{1}{2}U^{-1}$, since $\frac{1}{2}\sum_{j=1}^{4} \theta_{t,j}(\Sigma_t^{-1} + F_j)^{-1}$ is larger than zero, so as t goes to infinity, the equilibrium value of Σ_t will be larger than $\frac{1}{2}U^{-1}$. Therefore, if there is mutation ($U > 0$), variances of both defences will be positive (larger than $\frac{\mu}{2}$).

8.2 Appendix B

The scale of the variance in defence phenotypes is set by the parameters μ, α, and β. In this appendix, we show that if we increase these three parameters by a common factor, the equilibrium value of v_t and w_t will change by the same factor.

Let $P_t(x, y)$ denote the probability distribution of traits at time t, where

$$P_t(x, y) = \frac{N_t(x, y)}{\int\int N_t(x, y) dx\, dy}.$$

The integration limits are from $-\infty$ to ∞, and are suppressed throughout this section

for brevity. From eqn. (5), the dynamics of P_t is determined by

$$\begin{aligned} P_{t+1}(x,y) &= \frac{N_{t+1}(x,y)}{\int\int N_{t+1}(x,y)dx\,dy} \\ &= \frac{\int\int P_t(x',y')\Phi(x',y')M(x',y',x,y)dx'\,dy'}{\int\int P_t(x',y')\Phi(x',y')dx'\,dy'}, \end{aligned}$$

where we have used the fact that $\int\int M(x',y',x,y)dx\,dy = 1$. Over time, P_t will approach an equilibrium $P_*(x,t) = \lim_{t\to\infty} P_t(x,t)$, where

$$P_*(x,y) = \frac{\int\int P_*(x',y')\Phi(x',y')M(x',y',x,y)dx'\,dy'}{\int\int P_*(x',y')\Phi(x',y')dx'\,dy'}. \tag{B.1}$$

From Eqns. (1–3) and (6), Φ and M can be written in the form

$$\Phi(x,y) = \tilde{\Phi}\left(\frac{x-a}{\alpha^{1/2}}, \frac{y-b}{\beta^{1/2}}\right) \tag{B.2}$$

$$M(x',y',x,y) = \frac{1}{\mu}\tilde{M}\left(\frac{x-x'}{\mu^{1/2}}, \frac{y-y'}{\mu^{1/2}}\right), \tag{B.3}$$

where

$$\tilde{\Phi}(u,v) = f_1 e^{-\epsilon_1-u^2} + \left(1-e^{-\epsilon_1-u^2}\right)\left((f_2-f_3)\,e^{-\epsilon_2-v^2}+f_3\right) \tag{B.4}$$

$$\tilde{M}(u,v) = \frac{1}{\pi}e^{-u^2-v^2}. \tag{B.5}$$

We now substitute eqns. (B.2) and (B.3) into Eqn. (B.1), make the change of

variables

$$\xi = \frac{x-a}{\mu^{1/2}}$$
$$\xi' = \frac{x'-a}{\mu^{1/2}}$$
$$\eta = \frac{y-b}{\mu^{1/2}}$$
$$\eta' = \frac{y'-b}{\mu^{1/2}},$$

and further define (without loss of generality)

$$P_*(x,y) = \frac{1}{\mu}\tilde{P}_*(\xi,\eta), \tag{B.6}$$

to give

$$\tilde{P}_*(\xi,\eta) = \frac{\int\int \tilde{P}_*(\xi',\eta')\tilde{\Phi}\left(\left(\frac{\mu}{\alpha}\right)^{1/2}\xi', \left(\frac{\mu}{\beta}\right)^{1/2}\eta'\right)\tilde{M}(\xi-\xi',\eta-\eta')d\xi'\,d\eta'}{\int\int \tilde{P}_*(\xi',\eta')\tilde{\Phi}\left(\left(\frac{\mu}{\alpha}\right)^{1/2}\xi', \left(\frac{\mu}{\beta}\right)^{1/2}\eta'\right)d\xi'\,d\eta'}. \tag{B.7}$$

Note that, from eqns. (B.4,B.5), neither $\tilde{\Phi}$ nor $\tilde{M}$ have any explicit dependence on μ, α, or β so these parameters only enter into eqn. (B.7) through the ratios $\frac{\mu}{\alpha}$ and $\frac{\mu}{\beta}$ in the arguments to $\tilde{\Phi}$. This means that the solution $\tilde{P}_*(\xi,\eta)$ to eqn. (B.7) can be written in the form

$$\tilde{P}_*(\xi,\eta) = p\left(\xi,\eta,\frac{\mu}{\alpha},\frac{\mu}{\beta}\right),$$

where p does not depend explicity on μ, α, or β except through its third and fourth arguments. This means that P_* takes the form

$$P_*(x,y) = \frac{1}{\mu}p\left(\frac{x-a}{\mu^{1/2}},\frac{y-b}{\mu^{1/2}},\frac{\mu}{\alpha},\frac{\mu}{\beta}\right).$$

In other words, when traits are measured as a difference from their optimum in units of

$\mu^{1/2}$, their distribution depends only on the ratios of α and β to μ.

We note that P_* is a probability density so we have $\int\int P_*(x,y)\,dx\,dy = 1$. Also, we can show that $\int\int (x-a)P_*(x,y)dx\,dy = \int\int \xi\tilde{P}_*(\xi,\eta)\,d\xi\,d\eta = 0$. This follows because $\tilde{M}$and $\tilde{\Phi}$ are even functions of their arguments, so from eqn (B.7) if $P_*(\xi,\eta) = \hat{P}(\xi,\eta)$ is a solution then so is $P_*(\xi,\eta) = \hat{P}(-\xi,\eta)$. Since this solution is unique we must have $P_*(\xi,\eta) = P_*(-\xi,\eta)$, which implies $\int\int \xi\tilde{P}_*(\xi,\eta)\,d\xi\,d\eta = 0$. Therefore, the mean of the first defensive trait at equilibrium is

$$\begin{aligned} x_* &= \int\int xP_*(x,y)dx\,dy \\ &= a\int\int P_*(x,y)dx\,dy + \int\int (x-a)P_*(x,y)dx\,dy \\ &= a. \end{aligned}$$

The equilibrium variance of the first defensive trait is then

$$\begin{aligned} v_* &= \int\int (x-x_*)^2 P_*(x,y)dx\,dy \\ &= \frac{1}{\mu}\int\int (x-a)^2\, p\left(\frac{x-a}{\mu^{1/2}}, \frac{y-b}{\mu^{1/2}}, \frac{\mu}{\alpha}, \frac{\mu}{\beta}\right) dx\,dy \\ &= \mu\int\int \xi^2 p\left(\xi,\eta,\frac{\mu}{\alpha},\frac{\mu}{\beta}\right) d\xi\,d\eta. \end{aligned}$$

Therefore, $\frac{v_*}{\mu}$ depends on μ, α, or β through the ratios $\frac{\mu}{\alpha}$ and $\frac{\mu}{\beta}$ only. This means that, if μ, α, and β are increased by a common factor λ, which means that $\frac{\mu}{\alpha}$ and $\frac{\mu}{\beta}$ are unchanged, then v_* increases by the same factor λ. A similar argument can be made for the variance of the second trait.

8.3 Appendix C When there is no mutation force, the variances of both the first and second defences approaches to zero

When there is no mutation, the population distribution function in the $(t+1)$-th generation is the density iteration equations (4):

$$N_{t+1}(x_{t+1}, y_{t+1}) = N_t(x_t, y_t)\Phi(x_t, y_t)$$

From (8) and (A.1),

$$N_{t+1}(x_{t+1}, y_{t+1}) = \frac{n_t\sqrt{|W_t|}}{\pi}\sum_{j=1}^{4}\gamma_j \exp\Big(- (z_t - \bar{z}_t)^T W_t(z_t - \bar{z}_t) - (z_t - \hat{a})^T F_j(z_t - \hat{a})\Big)$$

Collecting the terms for z_t,

$$N_{t+1}(x_{t+1}, y_{t+1}) = \frac{n_t\sqrt{|W_t|}}{\pi}\sum_{j=1}^{4}\gamma_j \exp\Big(-z_t^T(W_t+F_j)z_t+2(\bar{z}_t{}^T W_t+\hat{a}^T F_j)z_t-(\bar{z}_t{}^T W_t\bar{z}_t+\hat{a}^T F_j\hat{a})\Big)$$

If we complete a square for z_t, the above is equivalent to

$$N_{t+1}(x_{t+1}, y_{t+1}) = \frac{n_t\sqrt{|W_t|}}{\pi}\sum_{j=1}^{4}\gamma_j \exp\Big(- (z_t - z'_{t+1,j})^T(W_t + F_j)(z_t - z'_{t+1,j})$$

$$+(W_t\bar{z}_t + F_j\hat{a})^T(W_t + F_j)^{-1}(W_t\bar{z}_t + F_j\hat{a}) - (\bar{z}_t{}^T W_t\bar{z}_t + \hat{a}^T F_j\hat{a})\Big)$$

where

$$z'_{t+1,j} = (W_t + F_j)^{-1}(W_t\bar{z}_t + F_j\hat{a})$$

If we let $s_{t,j} = \exp\Big((W_t\bar{z}_t + F_j\hat{a})^T(W_t + F_j)^{-1}(W_t\bar{z}_t + F_j\hat{a}) - (\bar{z}_t{}^T W_t\bar{z}_t + \hat{a}^T F_j\hat{a})\Big)$, then the above is equivalent to

$$= \frac{n_t}{\pi}\sqrt{|W_t|}\sum_{j=1}^{4}\gamma_j s_{t,j}\exp\Big(-(z_t - z'_{t+1,j})^T(W_t + F_j)(z_t - z'_{t+1,j})\Big)$$

If we make it in a normal distribution form,

$$= n_t\sqrt{|W_t|}\sum_{j=1}^{4}\frac{\gamma_j s_{t,j}}{\sqrt{|W_t + F_j|}}\frac{\sqrt{|W_t + F_j|}}{\pi}\exp\Big(-(z_{t+1} - z'_{t+1,j})^T(W_t + F_j)(z_{t+1} - z'_{t+1,j})\Big)$$

which is equivalent to

$$= n_t\sqrt{|W_t|}\Theta_t\sum_{j=1}^{4}\theta_{t,j}\frac{\sqrt{|W_t + F_j|}}{\pi}\exp\Big(-(z_{t+1} - z'_{t+1,j})^T(W_t + F_j)(z_{t+1} - z'_{t+1,j})\Big)$$

$$t = 1, 2, 3, ... \tag{C.1}$$

where $\Theta_t = \sum_{j=1}^{4}\frac{\gamma_j s_{t,j}}{\sqrt{|W_t+F_j|}}$, and $\theta_{t,j} = \frac{\gamma_j s_{t,j}}{\Theta_t\sqrt{|W_t+F_j|}}$ $(j = 1, 2, 3, 4)$. Note that $\sum_{j=1}^{4}\theta_{t,j} = 1$

From the above population distribution function, the **probability density function** in the next generation can be written as

$$f(x_{t+1}, y_{t+1}) = N(x_{t+1}, y_{t+1})/\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}N(x_{t+1}, y_{t+1})$$

$$= \sum_{j=1}^{4}\theta_{t,j}\frac{\sqrt{|W_t + F_j|}}{\pi}\exp\Big(-(z_{t+1} - z'_{t+1,j})^T(W_t + F_j)(z_{t+1} - z'_{t+1,j})\Big)dz_{t+1} \tag{C.2}$$

$$= \sum_{j=1}^{4}\theta_{t,j}f_j(x_{t+1}, y_{t+1})$$

where

$$\theta_{t,j} = \frac{\gamma_j s_{t,j}}{\Theta_t \sqrt{|W_t + F_j|}}$$

$$\Theta_t = \sum_{j=1}^{4} \frac{\gamma_j s_{t,j}}{\sqrt{|W_t + F_j|}}$$

$$s_{t,j} = \exp\Big((W_t \bar{z}_t + F_j \hat{a})^T (W_t + F_j)^{-1} (W_t \bar{z}_t + F_j \hat{a}) - (\bar{z}_t{}^T W_t \bar{z}_t + \hat{a}^T F_j \hat{a})\Big)$$

Therefore the population distribution probability density function in the $(t+1)$-th generation is written as a combination of four normal probability density functions (each has mean $z'_{t+1,j} = (W_t + F_j)^{-1}(W_t \bar{z}_t + F_j \hat{a})$, covariance matirx $\Sigma_{t+1,j} = \frac{1}{2}(W_t + F_j)^{-1} = \frac{1}{2}($ $frac12\Sigma_t^{-1} + F_j)^{-1} = (\Sigma_t^{-1} + 2F_j)^{-1}$, $j = 1, 2, 3, 4)$. Note that the integral of $f(x_{t+1}, y_{t+1})$ in respect of (x_{t+1}, y_{t+1}) equals 1, which is the property of the probability density function. Let $E_{N_j}(\cdot)$ denote the expectation of each of the four corresponding normal population distribution function. Then the **mean in the next generation** is

$$\bar{z_{t+1}} = E(z_{t+1}) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} z_{t+1} f(x_{t+1}, y_{t+1}) dz_{t+1} = \sum_{j=1}^{4} \theta_{t,j} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} z_{t+1} f_i(x_{t+1}, y_{t+1}) dz_{t+1}$$

$$= \sum_{j=1}^{4} \theta_{t,j} E_{N_j}(z_{t+1}) = \sum_{j=1}^{4} \theta_{t,j} z'_{t+1,j} = \sum_{j=1}^{4} \theta_{t,j} (W_t + F_j)^{-1} (W_t \bar{z}_t + F_j \hat{a}) \tag{C.3}$$

The above also equals to

$$\bar{z_{t+1}} = \sum_{j=1}^{4} \theta_{t,j} (W_t + F_j)^{-1} W_t \bar{z}_t + \sum_{j=1}^{4} \theta_{t,j} (W_t + F_j)^{-1} F_j \hat{a}$$

As mentioned in Appendix A, since $\sum_{j=1}^{4} \theta_{t,j} (W_t + F_j)^{-1} W_t + \sum_{j=1}^{4} \theta_{t,j} (W_t + F_j)^{-1} F_j = I$, the mean z_{t+1} will gradually approach to the ideal phenotype as t increases.

Now for variance, we have the **covariance matrix in the next generation** is

$$\Sigma_{t+1} = E\begin{pmatrix} x_{t+1}^2 & x_{t+1}y_{t+1} \\ x_{t+1}y_{t+1} & y_{t+1}^2 \end{pmatrix} - E\begin{pmatrix} x_{t+1} \\ y_{t+1} \end{pmatrix} E\begin{pmatrix} x_{t+1} & y_{t+1} \end{pmatrix}$$

$$= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \begin{pmatrix} x_{t+1}^2 & x_{t+1}y_{t+1} \\ x_{t+1}y_{t+1} & y_{t+1}^2 \end{pmatrix} \cdot \sum_{j=1}^{4} \theta_{t,j} f_i(x_{t+1}, y_{t+1}) dz_{t+1} - \bar{z_{t+1}}\bar{z_{t+1}}^T$$

$$= \sum_{j=1}^{4} \theta_{t,j} \cdot \begin{pmatrix} E_{N_j}(x_{t+1}^2) & E_{N_j}(x_{t+1}y_{t+1}) \\ E_{N_j}(x_{t+1}y_{t+1}) & E_{N_j}(y_{t+1}^2) \end{pmatrix} - \bar{z_{t+1}}\bar{z_{t+1}}^T$$

$$= \sum_{j=1}^{4} \theta_{t,j} \cdot \begin{pmatrix} Var_{N_j}(x_{t+1}) + E_{N_j}(x_{t+1})^2 & Cov_{N_j}(x_{t+1}, y_{t+1}) + E_{N_j}(x_{t+1})E_{N_j}(y_{t+1}) \\ Cov_{N_j}(x_{t+1}, y_{t+1}) + E_{N_j}(x_{t+1})E_{N_j}(y_{t+1}) & Var_{N_j}(y_{t+1}) + E_{N_j}(y_{t+1})^2 \end{pmatrix}$$

$$-\bar{z_{t+1}}\bar{z_{t+1}}^T$$

$$= \sum_{j=1}^{4} \theta_{t,j} \cdot \Sigma_{t+1,j} + \sum_{j=1}^{4} \theta_{t,j} \cdot z'_{t+1,j} z'^T_{t+1,j} - \bar{z_{t+1}}\bar{z_{t+1}}^T$$

where $\Sigma_{t+1,j}$ is the jth covariance matrix from the jth integral, $z'_{t+1,j}$ is jth mean from the jth integral. So the above is equivalent to

$$= \sum_{j=1}^{4} \theta_{t,j} \cdot (\Sigma_t^{-1} + 2F_j)^{-1} + \sum_{j=1}^{4} \theta_{t,j} \cdot z'_{t+1,j} z'^T_{t+1,j} - \bar{z_{t+1}}\bar{z_{t+1}}^T \tag{C.4}$$

Now we can use the variance iteration equations from (C.4) to get the evolution of variance across generations. As $\bar{z}_t$ approaches to the ideal phenotype $\hat{a}$ as $t \to +\infty$, so does $\bar{z_{t+1}}$ and $z'_{t+1,j} = (W_t + F_j)^{-1}(W_t\bar{z}_t + F_j\hat{a})$ $(j = 1, 2, 3, 4)$ also approaches to the ideal phenotype $\hat{a}$. Therefore the term $\sum_{j=1}^{4} \theta_{t,j} \cdot z'_{t+1,j} z'^T_{t+1,j} - \bar{z_{t+1}}\bar{z_{t+1}}^T$ in (A.11) approaches to zero, so covariance matrix (A.11) approaches to the following equation as

time grows.

$$\Sigma_{t+1} = \sum_{j=1}^{4} \theta_{t,j} \cdot (\Sigma_t^{-1} + 2F_j)^{-1} \tag{C.5}$$

If there is an equilibrium, as time t grows as large as T (a very large number), the above equation (C.5) approaches to

$$\Sigma_T = \sum_{j=1}^{4} \theta_{T,j} \cdot (\Sigma_T^{-1} + 2F_j)^{-1} \tag{C.6}$$

If we write $\Sigma_T^{-1} = \begin{pmatrix} a & c \\ c & b \end{pmatrix}$, then the equation (C.6) can be written as

$$\begin{aligned}
\begin{pmatrix} a & c \\ c & b \end{pmatrix}^{-1} &= \theta_{T,1} \begin{pmatrix} a & c \\ c & b \end{pmatrix}^{-1} + \theta_{T,2} \begin{pmatrix} a + \frac{2}{\alpha} & c \\ c & b \end{pmatrix}^{-1} \\
&+ \theta_{T,3} \begin{pmatrix} a & c \\ c & b + \frac{2}{\beta} \end{pmatrix}^{-1} + \theta_{T,4} \begin{pmatrix} a + \frac{2}{\alpha} & c \\ c & b + \frac{2}{\beta} \end{pmatrix}^{-1}
\end{aligned}$$

which is equivalent to

$$\begin{aligned}
(2 - \theta_{T,1}) \frac{1}{ab - c^2} \begin{pmatrix} b & -c \\ -c & a \end{pmatrix} &= \theta_{T,2} \frac{1}{(a + \frac{2}{\alpha})b - c^2} \begin{pmatrix} b & -c \\ -c & a + \frac{2}{\alpha} \end{pmatrix} \\
&+ \theta_{T,3} \frac{1}{a(b + \frac{2}{\beta}) - c^2} \begin{pmatrix} b + \frac{2}{\beta} & -c \\ -c & a \end{pmatrix} \\
&+ \theta_{T,4} \frac{1}{(a + \frac{2}{\alpha})(b + \frac{2}{\beta}) - c^2} \begin{pmatrix} b + \frac{2}{\beta} & -c \\ -c & a + \frac{2}{\alpha} \end{pmatrix}
\end{aligned} \tag{C.7}$$

Since $\alpha > 0$ and $\beta > 0$ and the determinant of the covariance matrix $ab - c^2 > 0$, so

$$(a + \frac{2}{\alpha})b - c^2 > ab - c^2 > 0, \qquad a(b+\frac{2}{\beta})-c^2 > ab - c^2 > 0, \qquad (a+\frac{2}{\alpha})(b+\frac{2}{\beta})-c^2 > ab - c^2 > 0$$

Comparing the coefficient of term $-c$ in the matrix for both side of the equation, the coefficient in the left-hand side is larger than the coefficient in the right-hand side, so we have $c = 0$, and the equation (C.7) is equivalent to

$$\begin{aligned}(\theta_{T,2} + \theta_{T,3} + \theta_{T,4})\frac{1}{ab}\begin{pmatrix} b & 0 \\ 0 & a \end{pmatrix} &= \theta_{T,2}\frac{1}{(a+\frac{2}{\alpha})b}\begin{pmatrix} b & 0 \\ 0 & a+\frac{2}{\alpha} \end{pmatrix} \\ &+ \theta_{T,3}\frac{1}{a(b+\frac{2}{\beta})}\begin{pmatrix} b+\frac{2}{\beta} & 0 \\ 0 & a \end{pmatrix} \\ &+ \theta_{T,4}\frac{1}{(a+\frac{2}{\alpha})(b+\frac{2}{\beta})}\begin{pmatrix} b+\frac{2}{\beta} & 0 \\ 0 & a+\frac{2}{\alpha} \end{pmatrix} \end{aligned} \tag{C.8}$$

which is equivalent to

$$\begin{aligned}(\theta_{T,2} + \theta_{T,3} + \theta_{T,4})\begin{pmatrix} \frac{1}{a} & 0 \\ 0 & \frac{1}{b} \end{pmatrix} &= \theta_{T,2}\begin{pmatrix} \frac{1}{a+\frac{2}{\alpha}} & 0 \\ 0 & \frac{1}{b} \end{pmatrix} + \theta_{T,3}\begin{pmatrix} \frac{1}{a} & 0 \\ 0 & \frac{1}{b+\frac{2}{\beta}} \end{pmatrix} \\ &+ \theta_{T,4}\begin{pmatrix} \frac{1}{a+\frac{2}{\alpha}} & 0 \\ 0 & \frac{1}{b+\frac{2}{\beta}} \end{pmatrix} \end{aligned} \tag{C.9}$$

Since $\alpha > 0$ and $\beta > 0$, the above equation cannot hold unless a and b approaches to $+\infty$ as t grows, so the matrix Σ_t grows to zero matrix as t grows. Therefore the variances of both the first and second defences approaches to zero.