

## The canonical equation of adaptive dynamics for Mendelian diploids and haplo-diploids

Johan A. J. Metz<sup>1,3,4</sup> and Carolien G. F. de Kovel<sup>2</sup>

<sup>1</sup> (= Hans)

Mathematical Institute & Institute of Biology, Leiden University, NL-2333CA  
Leiden, Netherlands

email: [j.a.j.metz@biology.leidenuniv.nl](mailto:j.a.j.metz@biology.leidenuniv.nl)

<sup>2</sup> Universitair Medisch Centrum Utrecht, Sectie Complex Genetics, Department of  
Medical Genetics, Utrecht, Netherlands

email: [C.deKovel@umcutrecht.nl](mailto:C.deKovel@umcutrecht.nl)

<sup>3</sup> Evolution and Ecology Program, International Institute for Applied Systems  
Analysis, A-2361 Laxenburg, Austria

<sup>4</sup> Naturalis Biodiversity Center, NL-2333CR Leiden, Netherlands

**Abstract:** One of the powerful tools of adaptive dynamics is its so-called Canonical Equation, a differential equation describing how the prevailing trait vector changes over evolutionary time. The derivation of the CE is based on two simplifying assumptions, separation of population dynamical and mutational timescales, and small mutational steps. (It may appear that these two conditions rarely go together. However, for small step sizes the timescale separation need not be very strict.) Dieckmann and Law have derived the CE in 1996, with Champagnat in 2003 adding mathematical rigour. Both papers consider only well-mixed clonal populations with the simplest possible life-histories. Durinx, Metz and Meszéna have in 2008 heuristically extended the CE's reach to locally well-mixed populations with general life-histories. We, again heuristically, extend it further to Mendelian diploids and haplo-diploids. Away from strict timescale separation the CE does an even better approximation job in the Mendelian than in the clonal case due to gene substitutions occurring effectively in parallel, which obviates slowing down by clonal interference.

**Keywords:** meso-evolution, adaptive dynamics, canonical equation, haplo-diploids, invasion probability, effective reproductive variance

### 1. Introduction

For context it is useful to distinguish between micro-, meso-, and macro-evolution. The term micro-evolution customarily refers to changes in gene frequencies on a population dynamical timescale. The evolution of quantitative traits through the repeated substitution of new mutations, including the splitting of the evolutionary path into separate evolutionary lines, we will refer to as meso-evolution. The term macro-evolution then becomes restricted to larger scale changes like anatomical innovations, where one cannot even speak in terms of a fixed set of traits.

Adaptive dynamics (AD) was devised as a mathematical framework for dealing with meso-evolution in 'realistic' ecological settings. It differs from more classical approaches to modelling evolutionary change, which generally assume constant fitnesses, by its focus on the population dynamical basis for those fitnesses, and hence on their inevitable change over evolutionary time. Of course, the greater

realism at the ecological end is bought by making different simplifying assumptions, this time genetically unrealistic ones. The main simplification is (i) *separation of the population dynamical and mutational timescales*. In order to concentrate on ecological aspects, unencumbered by the complexities of the genetic architecture and genotype to phenotype map, most AD research moreover assumes (ii) *clonal inheritance*.

One of the powerful tools of AD is its so-called Canonical Equation (CE), a differential equation describing how the prevailing trait vector changes over evolutionary time. The derivation of the CE is based on a subsequent further simplifying assumption: (iii) *small mutational steps*. Since the speed at which a mutation substitutes is proportional to its effect, conditions (i) and (iii) will only rarely be met together. In Section 5 we give arguments why for small step sizes the timescale separation need not be very strict.

The CE has been derived first in [1] for well-mixed clonal populations with the simplest possible life-histories. This was underpinned on a level of rigour sufficient to satisfy hard probabilists in [2] and [3]. The initial results have been extended in [4] to, possibly spatially distributed, locally well mixed populations with general life-histories, with [5] giving a probabilistically rigorous underpinning for the case of simple age-dependence. In this paper we describe the extension to Mendelian diploids and haplo-diploids, once again on a physicist level of rigour. As it turns out, the CE for the simplest and most complex life-histories differ only in a scalar factor summarising how the intricacies of the life-history feed through to the invasion probabilities of advantageous mutations. Mendelian diploidy brings in an additional factor 2, due to the doubling of the number of mutant alleles per individual over a substitution.

[6] also derives a CE for Mendelian diploids, in full rigour. However, this in essence is a CE for a single locus trait, selected by a two-tiered ecology, within and between diploid bodies, the latter with the simplest possible life-history. ([6] also lists the early applications of AD to Mendelian models.) We consider general life-histories and phenotypic traits underlain by many loci.

In Section 2 we give a general heuristic derivation of the CE for clonal and for haploid and diploid Mendelian populations. In Section 3 we work out the details and in Section 4 we consider haplo-diploids. The technicalities can be found in a suite of appendices. In the final Section 5 we discuss the strengths and weaknesses of the CE as tool for evolutionary understanding.

## 2. Deriving the CE for the textbook genetic scenarios

Mathematically, the CE is derived by taking two subsequent limits: (i) letting the system size  $K$  (and hence the average population size  $\bar{n}$ ) go to infinity (to make the community dynamics deterministic) and the mutation probability per birth event  $\epsilon$  go to zero in such a manner that (a)  $\epsilon K \ln(K) \rightarrow 0$  (to make the time for a substitution shorter than the time between mutations) and (b)  $\epsilon K \exp(\alpha K) \rightarrow \infty$  for sufficiently small positive  $\alpha$  (to keep the population from going extinct on the timescale of the trait changes), followed by (ii) letting the mutational step sizes go to zero, all the while keeping the trait changes in view by rescaling time. Biologically, the CE is best seen as an approximation. From that perspective it is expedient to express the result in the original timescale so that the basic biological parameters are kept in view.

Under the assumptions that  $K$  is large,  $\epsilon K \ln(K)$  small and  $\epsilon K \exp(\alpha K)$  large for a sufficiently small positive  $\alpha$ , that mutations are unbiased, and that the

environment as perceived by the individuals does not fluctuate (note that this implies a non-fluctuating resident population), the rate of change of a trait vector  $X$  can in the clonal case be expressed approximatively as

$$\begin{aligned} \frac{dX}{dt} &= \left\{ \begin{array}{l} \text{[rate at which mutants are produced]} \times \\ \text{[AVERAGE OVER THE MUTATION DISTRIBUTION OF} \\ \text{(effect of mutation} \times \text{linear approximation for its invasion probability)]} \end{array} \right\} \quad (2.1) \\ &= \left[ \frac{\varepsilon \bar{n}}{T_s} \right] \times \left[ \frac{T_r}{\sigma_e^2} \mathbf{C} \partial_1^T s(X|X) \right], \end{aligned}$$

with  $s(Y|X)$  the invasion fitness of  $Y$  mutants in the environment generated by  $X$  residents (see [7,8]),  $\partial_1 s$  the derivative of  $s$  for its first argument, a row vector, and  $\partial_1^T s$  the corresponding column vector,  $T_s$  the mean survival time of the residents,  $T_r$  their average age at reproduction,  $\sigma_e^2$  a measure for the variability of their lifetime offspring production (detailed in Section 3), and  $\mathbf{C}$  the covariance matrix of the mutational steps. (In the AD literature the quantity  $\partial_1^T s(X|X)$  is known as the selection gradient.)

The first term in square brackets after the second equal sign in (2.1) comes from multiplying the probability of a mutation per birth with the population birth rate,  $b = \bar{n}/T_s$ , a formula derived from the consistency relation  $\bar{n} = bT_s$  (the average number of particles in a "reservoir" equals the average entrance rate times the average residence time).

The second term comes from combining three approximation formulas to calculate the probability  $p$  that a mutant invades into the resident population, all coming from a branching process approximation for the invasion phase of the mutant dynamics, followed by averaging the product of the resulting expression and the mutational effect  $Z := Y - X$  over the distribution  $g$  of  $Z$ .

$$(i) \quad p = \left( \frac{2 \ln(R_0)}{\sigma_e^2} \right)_+ + O(\ln(R_0)^2), \quad (2.2)$$

with  $(x)_+ := x$  if  $x \geq 0$  and  $(x)_+ := 0$  if  $x \leq 0$ , and  $R_0$  the average lifetime offspring number of the mutant. (Note that the general  $R_0$  concept allows for multiple birth states as for example in spatially distributed populations [9].) (2.2) is derived through a perturbation expansion from an equation for the invasion probabilities of a branching process ([10,11,12,13,14,15,16,17,18], Appendix B).

$$(ii) \quad \ln(R_0(Y|X)) = \partial_1 R_0(X|X)Z + O(\|Z\|^2). \quad (2.3)$$

(2.3) follows from  $R_0(X|X) = 1$ .

$$(iii) \quad s = \frac{\ln(R_0)}{T_r} + O(\ln(R_0)^2). \quad (2.4)$$

(2.4) is derived through a perturbation expansion from the characteristic equation for the Malthusian parameter of a branching process ([19,20,4], Appendix A). (Note that  $\dim(R_0) = 1$  and  $\dim(s) = 1/\text{time}$ .) Together (i) to (iii) result in

$$p = \left( \frac{2T_r}{\sigma_e^2} \partial_1 s(X|X) Z \right)_+ + O(\|Z\|^2). \quad (2.5)$$

When multiplying (2.5) with the mutational effect, it pays first to take transposes to make use of

$$\int Z (Z^T \partial_1^T s(X|X))_+ g(Z) dZ = \frac{1}{2} \mathbf{C} \partial_1^T s(X|X). \quad (2.6)$$

(2.6) follows from the facts that for unbiased mutational effects  $\int ZZ^T g(Z) dZ = \mathbf{C}$  and that the  $(\ )_+$  makes that we effectively integrate only over the half space where  $\partial_1 s(X|X) Z > 0$ .

In the derivation of (2.6) we have used the very strong interpretation of "unbiased" that  $Z$  not only has mean zero but also is distributed symmetrically around that mean. Relaxing these assumptions gives an expression with considerably less appeal [1,2,3]. Since relaxing them would make the following arguments less easy to follow while their essence stays the same, we have chosen to stick to the time-honoured simplification.

Moreover, in writing down (2.1) we have tacitly assumed that an invading mutant that makes it through the stochastic boundary layer, where its population dynamics can be approximated with a branching process, also makes it to a full substitution. This 'invasion implies substitution' rule presumably holds good for small mutational steps, away from population dynamical bifurcation points and evolutionarily singular strategies (characterized by  $\partial_1 s(X|X) = 0$ ). A hard proof is available for the case where the community dynamics allows a finite dimensional representation [21,22,23]. However, it looks as if with the right mathematical expertise the rule should be extendable to the required generality by combining the approaches in [24] and [25,26].

The two terms in square brackets in (2.1) connect to the simplifying assumptions in the following manner. The first term is contingent on the timescale separation assumption. Only when substitutions do not interfere, the speed of trait movement becomes proportional to the number of invasion attempts per time unit. The derivation of the second term is contingent on the smallness of the mutational steps, and also on the timescale separation to produce the constant environment underlying the branching process calculations (which assume fixed probability distributions for the life-time offspring numbers). More in particular, timescale separation makes that in between the negligibly short substitution events the environment is stationary and resident populations are genetically homogeneous. (More about the latter, seemingly unrealistic, consequence in Section 5.)

We now consider the Mendelian case. For chromosomal sex determination we focus on the autosomes, deferring the allosomal contributions to Section 4.

Thanks to the genetic homogeneity of the residents, the argument in (2.1) extends seamlessly to haploids. The only difference with the clonal case is that thinking genetics points one to the concept of genotype to phenotype map and

mutations that occur on multiple loci. However, when substitutions occur singly the latter multiplicity becomes phenotypically inconsequential.

Getting on with diploids, we first take a closer look at genotype to phenotype maps. The prevalent view in Evo-Devo nowadays is that the trait changes that AD attempts to model are mostly caused not by changes in the coding regions of genes but at their regulatory regions (e.g. [27,28]). The latter determine the activity of the genes in different parts of the body, at different times during development and under different micro-environmental conditions. This scenario allows us to look at the genotype as a sequence of vectors  $(U_{a_1}, U_{a_2})$  of expression levels, with a  $(a_i)$  a placeholder for the name of (an allele on) a generic locus. The genotype to phenotype map  $\Phi$  transforms this sequence into phenotypic traits. It is from this perspective that one should judge the assumption of smallness of mutational steps: the influence of any specific regulatory site among its many colleagues tends to be relatively minor. And it is this perspective that allows us to assume that (iv) *genotype to phenotype maps are smooth*.

**Lemma** ([29], Andrea Pugliese personal communication): When there are no parental effects, smooth genotype to phenotype maps are locally additive, i.e., if at a the expression vector  $U_a$  mutates to  $U_A$ , then

$$\begin{aligned} \Phi(\dots; U_A, U_A; \dots) - \Phi(\dots; U_a, U_a; \dots) = \\ 2(\Phi(\dots; U_a, U_A; \dots) - \Phi(\dots; U_a, U_a; \dots)) + O(\|U_A - U_a\|^2). \end{aligned} \quad (2.7)$$

**Proof:** Without parental effects

$$\left. \frac{\partial \Phi(\dots; U_a, U_A; \dots)}{\partial U_a} \right|_{U_A=U_a} = \left. \frac{\partial \Phi(\dots; U_a, U_A; \dots)}{\partial U_A} \right|_{U_A=U_a} =: \Phi'_a(\dots; U_a, U_a; \dots). \quad (2.8)$$

Hence

$$\begin{aligned} X_{aA} &= \Phi(\dots; U_a, U_a; \dots) + \Phi'_a(\dots; U_a, U_a; \dots)(U_A - U_a) + O(\|U_A - U_a\|^2), \\ X_{AA} &= \Phi(\dots; U_a, U_a; \dots) + 2\Phi'_a(\dots; U_a, U_a; \dots)(U_A - U_a) + O(\|U_A - U_a\|^2). \end{aligned} \quad (2.9)$$

In diploids an invading mutant allele  $A$  practically always shares a body with a resident allele  $a$  and this  $aA$  reproduces through backcrossing with a resident  $aa$ . Hence, the allele population initially grows like clonally reproducing  $aA$ -s (producing  $aa$ -s on the side), and its invasion fitness corresponds to the asymptotic average per capita growth of such a clonal population in an environmental background provided by, also seemingly clonally reproducing (for identical), resident  $aa$ -s. The invasion implies substitution theorem also applies unchanged (thanks to the local additivity). However, after substitution the population consists of mutant homozygotes, making the resulting step in phenotype space twice as large as in the clonal and haploid cases. We thus conclude that the CE for Mendelian diploids reads

$$\frac{dX}{dt} = 2\bar{n} \frac{T_r}{T_s} \frac{1}{\sigma_c^2} \varepsilon C \partial_1^T s(X|X). \quad (2.10)$$

On the right hand side we have put first the ecologically determined number of resident haplotypes (sets of chromosomes as present in gametes), followed by the life-

history statistics controlling the initial demographic stochasticity of allelic substitutions, followed by an expression quantifying the per birth mutational variability generated by the genetic architecture and genotype to phenotype map, to conclude with the selection gradient, summarising the ecology's current tendency for filtering novel genetic variation. In a paper in preparation [30] the first author and Vincent Jansen argue that  $\bar{n}$  times the second group of quantities precisely equals the effective population size from population genetics.

### 3. Filling in the details

It may seem that with (2.10) we are done. However, the devil is in the details, to wit the calculation of  $s$ ,  $T_s$ ,  $T_r$ ,  $\sigma_c^2$  and  $\mathbf{C}$ . For background material on the covered ecological models see [20,31] for the resident and [32,33] for the invader dynamics.

In principle also  $\bar{n}$  presents a problem, although only when the resident population fluctuates. This is even so in the clonal case as  $\bar{n}$  is not just a time average but a peculiarly weighted one (e.g. [34,35]). To keep things simple we have confined the argument to non-fluctuating residents.

We first consider  $R_0$ , although this quantity does not appear in our (2.1)-derived list. The reason that it does not do so is that we wanted to write the CE in the form customary in the literature. However, for most structured populations  $R_0$  is far easier to calculate than  $s$ . Therefore we might just as well in (2.1) and (2.10) drop  $T_r$  and replace  $s$  with  $R_0$ .

$R_0$  equals the dominant eigenvalue of the next generation matrix  $\mathbf{L}(Y|X)$  (or operator in case of infinitely many birth states; see e.g. [9]). For the clonal case  $\mathbf{L}$  is constructed by calculating from a model for the behaviour of individuals how many offspring in different birth states they produce on average dependent on their own birth state.

Given the next generation matrix, we can introduce two further quantities for later use: the stationary birth state distribution, i.e., normalised positive right eigenvector of  $\mathbf{L}(X|X)$  going with  $R_0=1$ ,  $U = (u_1 \ \dots \ u_n)^T$ ,  $\mathbf{1}^T U = 1$  with  $\mathbf{1}^T := (1 \ \dots \ 1)$ , and the corresponding reproductive values, i.e., co-normalised left eigenvector,  $V = (v_1 \ \dots \ v_n)$ ,  $VU = 1$ .  $v_i$  equals the expected contribution of an  $X$  newborn with birth-state  $i$  to future birth rates.

How to define  $R_0$  for evolving sexual diploids? A first answer is that in hermaphrodites one can just add the numbers of kids that individuals father and mother over their life (i.e., produce through the micro- and macro-gametic routes) and divide by 2. The factor  $\frac{1}{2}$  comes from the fact that in the Mendelian process each allele is only transferred with that probability. When hermaphrodites are born stochastically equal (i.e., their birth states have the same probability distribution),

$$R_0 = \frac{1}{2}(f + m), \quad (3.1)$$

with  $m$  and  $f$  the average number of offspring fathered or mothered by a randomly chosen individual. For later use we moreover note that for the resident

$$m = f = 1, \quad (3.2)$$

since resident densities are supposedly constant and every individual has one father

and one mother.

(3.1) and (3.2) also hold good for dioecious organisms, but with a twist. For later use we rewrite (3.1) and (3.2) by letting  $p_f$  and  $p_m$  denote the fractions of newly produced females and males and  $f_+$  and  $m_+$  the average lifetime numbers of offspring begotten by a female or male. Then  $f = p_f f_+$  and  $m = p_m m_+$ , so that

$$R_0 = \frac{1}{2}(p_f f_+ + p_m m_+), \quad (3.3)$$

and for the residents

$$f_+ = 1/p_f \quad \text{and} \quad m_+ = 1/p_m. \quad (3.4)$$

The above results are not completely trivial, as, in contrast to hermaphrodites, individuals of dioecious species get born in different flavours. To account for this,  $\mathbf{L}(Y|X)$  should be properly extended. If no other birth state distinctions are needed

$$\mathbf{L} = \frac{1}{2} \begin{pmatrix} \ell_{ff} & \ell_{fm} \\ \ell_{mf} & \ell_{mm} \end{pmatrix}, \quad (3.5)$$

with  $\ell_{ff}$ ,  $\ell_{fm}$  the average lifetime numbers of daughters of a female, male, and  $\ell_{mf}$ ,  $\ell_{mm}$  the corresponding average lifetime numbers of sons, all for mutant heterozygotes in the environmental and genetic background provided by the resident. The simplest case is when there is no connection between the traits and sex determination so that  $\ell_{ff} = p_f f_+$ ,  $\ell_{mf} = p_m f_+$ ,  $\ell_{fm} = p_f m_+$ ,  $\ell_{mm} = p_m m_+$ . Then  $\mathbf{L}$  has rank one and  $R_0 = \frac{1}{2}(p_f f_+ + p_m m_+)$ . Another way of getting the latter result is by observing that in this case everybody is born stochastically equal, having each the same probability of being born female. We then can do as if sex is determined after birth and calculate  $R_0$  by averaging over the possible courses of a life. When there is a connection, we end up with the same formula by defining  $p_m$  and  $p_f$  as the asymptotic probabilities of being born male or female, i.e., by choosing for  $p_m$  and  $p_f$  the components of the right eigenvector  $U$  of  $\mathbf{L}$ , and defining  $m_+$  and  $f_+$  again as the average numbers of offspring fathered or mothered over a lifetime given the parental sex,  $f_+ = \ell_{ff} + \ell_{mf}$ ,  $m_+ = \ell_{fm} + \ell_{mm}$ . By using  $R_0 = \mathbf{1}^T \mathbf{L} U$ , we again get  $R_0 = \frac{1}{2}(p_f f_+ + p_m m_+)$ . However, only the similarity of the expression is pleasing as this time it is not explicit, since to calculate  $p_m$  and  $p_f$  one first needs to calculate  $R_0$ .

Appendix C indicates how the preceding considerations extend in the presence of additional birth state distinctions.

Next we consider  $T_s$ , the quantity that had to be combined with  $\bar{n}$  to get the population birth rate.

$$T_s = \int_0^{\infty} a \Gamma(a) U da, \quad (3.6)$$

with  $\Gamma = (\gamma_1 \ \cdots \ \gamma_m)$ ,  $\gamma_i$  the probability density of time till death of a resident born in state  $i$ . In the dioecious case with everybody born stochastically equal

$$T_s = p_f T_{s,f} + p_m T_{s,m}. \quad (3.7)$$

The mean age at reproduction  $T_r$  needs more thought as also the offspring may differ in their birth states. The perturbation expansion for  $s$  in Appendix A tells that we should weight those offspring with their reproductive values.

$$T_r = \int_0^{\infty} a V\Lambda(a) U da, \quad (3.8)$$

with  $\Lambda(a)$  composed of  $\frac{1}{2}$  times the average per capita parenting rates of age  $a$  residents split according to the birth states of offspring and parent. This formula generalises the usual definition of the age at reproduction for everybody born equal. (Note that  $\int_0^{\infty} V\Lambda(a) U da = V\mathbf{L}(X|X)U = 1$  so that  $V\Lambda(a)U$  is a probability density.)

When everybody is born stochastically equal  $T_r = \int_0^{\infty} a \lambda(a) da = \frac{1}{2}(T_{r,f} + T_{r,m})$  with  $\lambda = \lambda_f + \lambda_m$ ,  $\lambda_f$  and  $\lambda_m$  half the average per capita mothering and fathering rates of the residents. For dioecious organisms  $\lambda_f = p_f \lambda_{f+}$ ,  $\lambda_m = p_m \lambda_{m+}$  with  $\lambda_{f+}$ ,  $\lambda_{m+}$  half the average female, male parenting rates. Yet,  $T_{r,f+} = T_{r,f}$ ,  $T_{r,m+} = T_{r,m}$ , thanks to the normalisation of  $\lambda_{f+}$  and  $\lambda_{m+}$  before calculating mean parenting ages.

$\sigma_e^2$ , appearing as coefficient in the approximation formula for the invasion probability  $p$ , is the most complicated beast. We first give its formula for the clonal case

$$\sigma_e^2 := \sum_j \text{var} \left( \sum_i v_i \underline{k}_i \mid j \right) u_j. \quad (3.9)$$

Here the  $\underline{k}_i$  are the lifetime numbers of  $i$ -offspring of residents and  $\text{var}(\underline{h} \mid j)$  means the variance of the random quantity  $\underline{h}$  for an individual born in state  $j$ . (3.9) emerges from the perturbation expansion (Appendix B).

To calculate  $\sigma_e^2$  for the Mendelian diploid case we apply (3.9) to the alleles. We only consider the case with everybody born stochastically equal.

$$\sigma_e^2 = \frac{1}{4} [\sigma_f^2 + 2(c+1) + \sigma_m^2], \quad (3.10)$$

with  $\sigma_f^2$  and  $\sigma_m^2$  the variances of the lifetime offspring numbers  $\underline{k}_f$  and  $\underline{k}_m$  produced through the macro- and micro-gametic routes, and  $c$  their covariance (Appendix D). The 1 inside the inner brackets comes from the Mendelian sampling of alleles.

For dioecious organisms  $c = -1$  (since  $\mathbf{E}k_f k_m = 0$ , and  $\mathbf{E}k_f = \mathbf{E}k_m = 1$ ), and  $\sigma_f^2 = (\theta_{f+}^2 + 1 - p_f) / p_f$ ,  $\sigma_m^2 = (\theta_{m+}^2 + 1 - p_m) / p_m$ ,  $\theta_{f+} = \sigma_{f+} / f_+$ ,  $\theta_{m+} = \sigma_{m+} / m_+$ . The latter result is obtained from the following Lemma together with the observation that  $\mathbf{E}_+ k_f = f_+$  and  $\mathbf{E}_+ k_m = m_+$ , with  $\mathbf{E}_+$  the expectation conditional on the parental sex.

**Lemma:** Let the discrete non-negative random variable  $\underline{k}$  be constructed from a random variable  $\underline{k}_+$  by setting  $\underline{k} = \underline{k}_+$  with probability  $p$  and  $\underline{k} = 0$  with probability  $1 - p$ , then  $\mathbf{E}\underline{k} = p\mathbf{E}\underline{k}_+$  and  $\text{var}(\underline{k}) = \mathbf{E}\underline{k}^2 - (\mathbf{E}\underline{k})^2 = p[\text{var}(\underline{k}_+) + (\mathbf{E}\underline{k}_+)^2] - (p\mathbf{E}\underline{k}_+)^2$ .

When moreover  $\mathbf{E}\underline{k} = 1$  so that  $\mathbf{E}\underline{k}_+ = p^{-1}$ ,



$$\text{var}(\underline{k}) = \mathbf{E}\underline{k}^2 - 1 = p \left[ \text{var}(\underline{k}_+) + p^{-2} \right] - 1 = \frac{\text{var}(\underline{k}_+) / (\mathbf{E}\underline{k}_+)^2 + 1 - p}{p}. \quad (3.11)$$

The mutational covariance matrix  $\mathbf{C}$  is primarily a phenomenological quantity, although it in principle can be expressed in terms of per locus statistics and the genotype to phenotype map,

$$\begin{aligned} \mathbf{C} &= \sum_a p_a \int \Phi(\dots; U_a, U_a + Z_a; \dots) \Phi^T(\dots; U_a, U_a + Z_a; \dots) g_a(Z_a) dZ_a \\ &\approx \sum_a p_a \Phi'_a(\dots; U_a, U_a; \dots) \left[ \int Z_a Z_a^T g_a(Z_a) dZ_a \right] \Phi_a'^T(\dots; U_a, U_a; \dots), \end{aligned} \quad (3.12)$$

with  $p_a$  the relative frequency with which a mutation occurs at locus  $a$  and  $g_a$  the probability density of mutational steps in the allelic trait vector of that locus.

As a final point we need to say something about the traits and the concept of phenotype. In general the phenotypes of AD should be seen as reaction norms, i.e., maps from micro-environmental conditions to characteristics of individuals (another term is conditional strategies). Only in the simplest cases a reaction norm is degenerate, taking only a single value. The dioecious case is similar in that the development of the sexes need not be, and in fact rarely are, equal. Hence, trait vectors will generally consist of two components, traits of the male,  $X_m$ , and of the female,  $X_f$ . In general the traits of the two sexes evolve dependently as they are coupled by their mutational covariances. In the extreme case that the mutational covariances between male and female traits are all zero the female and male co-evolve as if they are separate species. At the opposite end of the spectrum  $X_m = G_m(X_f)$ , or  $X_f = G_f(X_m)$ , and the mutational variation in the male and female traits is fully correlated. The upshot is that, except in the fully correlated case, we cannot work with a monolithic trait vector influencing both macro- and micro-gametic reproduction, as in hermaphrodites. Instead we should take account of the fact that when the sexes are separate they can evolve in their own ways.

#### 4. Haplo-diploids: a not uncommon, but often-neglected reproductive mode

In addition to the haploid and diploid ones there exist all sorts of other lifestyles. One common type is where haploid and diploid phases alternate (as in, e.g., ferns, mosses and a great variety of algae). It is then necessary, as in dioecious diploids, to introduce trait vectors for each separate phase. We can then consider a diploid plus its haploid offspring as a single generalised individual ([36] gives a relaxed introduction to this useful concept) and apply the theory of the previous two sections, where for the diploid phase traits we again have to put in a factor 2.

Still another lifestyle is the so-called haplo-diploid one, where one sex is diploid and the other haploid (the supplementary material to [37] lists the many known haplo-diploid taxa). Although this case can also be treated through the mental construction of appropriate generalised individuals, we follow the strategy of Section 3, and treat the two sexes as different birth states, as this is simpler and allows us to illustrate further tricks of the trade.

Although the opposite also occurs, we shall for definiteness take the hymenopteran situation as reference and assume that females are diploid. In that case only the female reproductive output needs to be discounted with  $\frac{1}{2}$ . We allow any

physiological structure, and only assume that within each sex everybody is born equal. The next-generation matrix then becomes

$$\mathbf{L}(Y|X) = \begin{pmatrix} \frac{1}{2}f_f(Y|X) & m(Y|X) \\ \frac{1}{2}f_m(Y|X) & 0 \end{pmatrix} \quad (4.1)$$

with  $f_f$  and  $f_m$  the average lifetime numbers of female and male offspring of a female, and  $m$  the average lifetime number of (female) offspring of a male.  $f_f$  and  $f_m$  depend only on the female traits,  $m$  only on the male traits.

For the resident  $f_f = 1$  (since the density of females is constant) and hence  $f_m = r$ , the sex-ratio at birth (the number of males born relative to females), and  $m = r^{-1}$  (since all females have a father and one female is born for  $r$  males). Hence the resident's next generation matrix is

$$\mathbf{L}(X|X) = \begin{pmatrix} \frac{1}{2} & r^{-1} \\ \frac{1}{2}r & 0 \end{pmatrix}, \quad (4.2)$$

with co-normalised eigenvectors

$$U = \frac{1}{2+r} \begin{pmatrix} 2 \\ r \end{pmatrix}, \quad V = \frac{2+r}{3r} (r \ 1). \quad (4.3)$$

As is commonly the case in models with more than one birth state, the resulting expression for  $R_0$  does not excel in transparency. However, we do not need  $R_0$  but its derivative, for which there exists a simple formula given in the following Lemma, derived by differentiating through the characteristic equation and solving for  $\partial R_0 / \partial Y$  (see [38,39] and in particular [40,8]).

**Lemma:** Let  $P(Y|X)(z) := -\det(z\mathbf{I} - \mathbf{L}(Y|X))$  and  $Q(Y|X) := P(Y|X)(1)$ , then

$$\partial_1 R_0(X|X) = \left[ -\frac{dP(X|X)}{dz}(1) \right]^{-1} [\partial_1 Q(X|X)]. \quad (4.4)$$

So far the additional ideas. The rest is work, following the route outlined in Sections 2 and 3.

We kick off with the selection gradient

$$\begin{aligned} \partial_1^T R_0(X|X) &= \frac{1}{3} (\partial_1^T f_f(X|X) + r^{-1} \partial_1^T f_m(X|X) + r \partial_1^T m(X|X)) \\ &= \frac{1}{3} (\partial_1^T \ln(f_f)(X|X) + \partial_1^T \ln(f_m)(X|X) + \partial_1^T \ln(m)(X|X)). \end{aligned} \quad (4.5)$$

Since we start from  $R_0$  instead of  $s$  there is no need for calculating  $T_r$ , and since  $T_s$  has nothing to do with the reproductive system, (3.7) holds good as for any other sex differentiated population. Remains

$$\sigma_e^2 = \frac{2+r}{18r} [2r^2 \sigma_{f_{lm}}^2 + r \sigma_{f_{lf}}^2 + \sigma_{m_{lf}}^2 + (2c_{lf} + 1 + r)], \quad (4.6)$$

where the  $f$ ,  $m$  in the indices indicate that the quantity is calculated conditional on the parent being female or male. There are nicer expressions than (4.6). If you wish you can check it by slogging through Appendix E. The term  $1+r$  inside the inner brackets derives from the Mendelian sampling in the females. Furthermore, for the male traits the factor 2 in front of  $\bar{n}$  in (2.10) should be removed. Lastly, a factor accounting for the reduced number of haplotypes per individual has to be inserted - for sex-independent per haplotype mutation frequencies  $(1 + \frac{1}{2}r)/(1+r)$ .

Our arguments for diploids only considered autosomes. To get the full CE in the case of chromosomal sex determination it suffices to add the allosomal contributions to the autosomal one. The contribution of the chromosome occurring only in the heterogametic sex follows the rules for clonal reproduction, that of the chromosome of the homogametic sex those for haplo-diploids.

## 5. Discussion

Away from evolutionarily singular strategies (where  $\partial_1 s(X|X) = 0$ ), the timescale separation assumption combined with the assumption of small mutational steps guarantees practically permanent genetic homogeneity of the resident population. More in particular, a population with sufficiently restricted variability will become homogeneous on the timescale of the substitutions, thanks to the close to additive genetics, and will effectively stay so till the next substitution. The exceptions to this homogeneity occur when an evolutionary trajectory comes close to an evolutionarily singular strategy. The reason is that, where everywhere else we have an approximately linear selection regime, near the singular strategies the quadratic terms in the local expansion of the fitness landscape start to dominate, creating epistasis (non-additivity) at the level of the fitnesses.

Genetic homogeneity of the resident population lies at the base of the approximations made to reach the CE and of its easy extension to a Mendelian world. At first sight this seems a strong argument against the CE's applicability, as genetic variability appears to be rampant. There are two arguments for yet thinking that the CE may often be a fair description of reality. The first one is that we need mutation limitation only for genes affecting the trait. The thrust of our theoretical deductions is not affected by selectively protected variability that is developmentally and selectively unconnected to the focal traits or variability on neutral loci subject to mutation, drift and draft (hitchhiking with loci under selection). The only effect that selectively kept variability at unconnected loci may have is that it makes the lifetime offspring numbers of the substituting alleles more variable, enlarging  $\sigma_c^2$ . However, since we treated the components of  $\sigma_c^2$  as empirical quantities, any genetic causes of this variability are automatically included.

The second argument bears on variability due to a lack of strict timescale separation. As long as that variability stays sufficiently limited it should in Mendelian populations have little effect on the quality of the CE approximation. The reason is that the effect of such variability on the invasion fitness is of higher order in the mutational step size than the terms retained in the derivation of the CE. (The argument for the latter statement may be found in [41]. This paper admittedly only considers unstructured populations. However, the nature of the argument suggests that with the right mathematical expertise it should be extendable to structured ones.) In clonal populations a lack of mutation limitation yet presents an obstacle to the quantitative applicability of the CE due to clonal interference: a substituting clone may be ousted

en route by a better mutant so that only a fraction of the mutants that invade contribute to longer run evolutionary change. We may thus expect 'reality' to be slower than the CE. Mendelian populations do not suffer this slowing down as mutants on different loci effectively substitute in parallel without interfering thanks to recombination (and the approximate additivity coming from the smallness of mutational effects).

The picture of multiple substituting mutants with small additive effect may seem close to the one considered by Lande [42,43]. However, there is a difference as in our picture only a small and variable number of loci is simultaneously substituting, with new mutants continually coming and old ones going. In contrast, the quantitative genetics picture underlying Lande's work has its basis in Fisher's picture of evolutionary change coming from changes in the allele frequencies on a large number of loci with all alleles present from the start [44]. (An argument against that picture is that the genetic differences between, say, a choanoflagellate and a human are so many that the attendant polymorphism would contain more genotypes than the number of atoms in the earth, invalidating the classical population genetical calculations.) Of course, Lande did not subscribe to this simplified picture (e.g. [45,46]), and neither did Fisher, at least not wholeheartedly [47, chapter IV]. However, the mathematical details of the connection between the approximations introduced in Lande's various papers remain to be worked out.

One of the consequences of having only a few loci substituting at any one time is that the within population variance of the trait fluctuates stochastically, and appears to be highly variable (Jörgen Ripa, pers. com.). This notwithstanding a negative feedback: when the standing variability is high, mutant alleles will find it more difficult to pass through the stochastic boundary layer due to the larger fluctuations in their offspring numbers. Unfortunately, this picture does not easily lead to explicit expressions, as the fluctuations in variability occur on a timescale similar to that of the substitutions so that we cannot use our earlier branching process calculation. However, we surmise that in reality the allelic reproductive variability coming from co-substituting alleles can usually be neglected relative to the variability in offspring numbers caused by alleles not affecting the traits and by micro-environmental variability like some individuals running into a predator and others not.

Although the previous paragraphs may give the impression that we are quite optimistic about the applicability of the CE, there is one problem that in applications usually is swept under the carpet: there is no need for  $\mathbf{C}$  to stay constant. It may perhaps seem, when the mutations with small and hence additive effect occur all on different loci, that also their cumulative effect will be additive. However, this is not the case. This can be seen by looking at the developmental process as a sequence of maps each of which transforms the genetic information plus environmental inputs further towards the final phenotype. Even if linearity were to prevail early on, the accumulated change will be appreciable and when the output from those early stages is fed into a nonlinear map to get to the next stage, approximate linearity is lost and the derivatives of  $\Phi$  that appear in (3.12) change over evolutionary time.

In principle it is possible to write down a CE for the extended trait vector  $(X, \mathbf{C})$  to obtain the change in  $\mathbf{C}$  as correlated response to selection on  $X$ . However, going through the calculations in the manner of (3.12) shows that the expression for the covariances between the mutational steps in  $X$  and  $\mathbf{C}$  generically involves higher moments. We should thus see the CE as the first step in a truncated moment expansion.

All this does not mean that using the CE with fixed  $\mathbf{C}$  is never more than a

heuristic tool, with no strong connection to reality. There are scenarios where one may with impunity assume  $\mathbf{C}$  to be constant, in particular when investigating the behaviour of evolutionary trajectories close to evolutionarily singular strategies. In a linearised stability analysis as in [48], or in the analysis of scenarios for initial divergence close to an adaptive branching point (as defined in [49,50,51]) as in [4,52,53,54], only situations are considered where the change in  $X$  and hence the associated change in  $\mathbf{C}$  is small. (Such arguments involve two trait scales, a small one for linearising the CE and a smaller one for the mutational steps.)

Acknowledgements: We thank Jacques van Alphen for inviting JAJM to a workshop on insect parasitoids, thus sparking of the research reported in Section 4. CGFdK was supported by the Netherlands Organization of Scientific Research (NWO), grant 813.04.001. JAJM's research benefited from the support of the Chair Modélisation Mathématique et Biodiversité VEOLIA-École Polytechnique-MNHN-F.X.

### Appendix A. Approximating a mutant's invasion fitness

For the sake of the exposition we first consider the case of a single birth state. Then  $s$  satisfies the characteristic equation (Lotka's equation),

$$1 = \int_0^{\infty} e^{-sa} \lambda(a) da \quad \text{or} \quad 1 = \sum_{a=1}^{\infty} e^{-sa} \lambda(a), \quad (\text{A.1})$$

depending on whether we are dealing with continuous or discrete time community dynamics. Here  $\lambda(a)$  is the average birth rate or ratio at age  $a$ , where the average is taken over whatever stochastic trajectories individuals may follow during their life.

We concentrate on the continuous time case. Rewriting (A.1) by introducing the probability density  $\ell$  of the age at reproduction,  $\underline{a}$ , and its cumulant generating function  $\phi$ ,

$$\ell(a) := R_0^{-1} \lambda(a), \quad \phi(s) := \ln \left( \int_0^{\infty} e^{sa} \ell(a) da \right) \quad (\text{A.2})$$

gives

$$0 = \ln(R_0) + \phi(-s). \quad (\text{A.3})$$

Expanding  $\phi$  as

$$\phi(-s) = \sum_{i=1}^n \frac{\kappa_i (-s)^i}{i!} + O(s^{n+1}) \quad (\text{A.4})$$

with  $\kappa_i$  the  $i^{\text{th}}$  cumulant,  $\kappa_1 = \mathbf{E}\underline{a} =: T_r$ ,  $\kappa_2 = \text{Var}(\underline{a}) =: \sigma^2$ , and solving subsequently for the first and second order terms (on the assumption that  $|\ln(R_0)|$  is small) gives

$$s = \frac{\ln(R_0)}{T_r} \left( 1 + \frac{1}{2} \left( \frac{\sigma}{T_r} \right)^2 \ln(R_0) + O(\ln(R_0)^2) \right). \quad (\text{A.5})$$

(A.5) often performs far better than the estimate of the error term suggests. The reason is that a similar result appears for birth kernels with potentially large  $R_0$  but narrow spread. When the  $i^{\text{th}}$  central moment  $\mu_i$  of  $\underline{a}$  scales like  $\sigma^n$ , (A.1) can be written as

$$1 = R_0 e^{-sT_r} \left( 1 + \sum_{i=2}^n \frac{(-r)^i \mu_i}{i!} + O(\sigma^{n+1}) \right). \quad (\text{A.6})$$

Taking logarithms and solving subsequently for first and second order terms gives

$$s = \frac{\ln(R_0)}{T_r} \left( 1 + \frac{1}{2} (\sigma/T_r)^2 \ln(R_0) + O(\sigma^3) \right). \quad (\text{A.7})$$

The two approximations agree up to their second order terms (but not in higher order ones).

For more than one birth state we introduce  $\Lambda(a) = (\lambda_{ij}(a))$  with  $\lambda_{ij}(a)$  the average rate at which an individual born in state  $j$  gives birth to offspring in state  $i$ . Then  $s$  can be calculated from the characteristic equation

$$\left[ \text{dominant eigenvalue of } \tilde{\Lambda}(s) \right] = 1 \quad (\text{A.8})$$

with

$$\tilde{\lambda}_{ij}(s) = \int_0^\infty e^{-sa} \lambda_{ij}(a) da, \quad (\text{A.9})$$

or, equivalently, as the rightmost solution of

$$\det(\mathbf{I} - \tilde{\Lambda}(s)) = 0. \quad (\text{A.10})$$

Let  $\lambda_d$  stand for "dominant eigenvalue of" and  $Z := Y - X$  be the mutational step. Then (A.8) can be rewritten as

$$\psi(s; Z) := \ln \left( \lambda_d \left( \int_0^\infty e^{-sa} \Lambda(a; Y | X) da \right) \right) = 0. \quad (\text{A.11})$$

Expanding  $\psi$  as a function of its first argument gives

$$\psi(s; Z) = \psi(0; Z) + \partial_1 \psi(0; Z) s + O(s^2). \quad (\text{A.12})$$

On the assumption that  $s = O(\|Z\|)$  for small  $\|Z\|$  we can rewrite (A.12) as

$$0 = \psi(s; Z) = \psi(0; Z) + \partial_1 \psi(0; Z) s + O(s^2) = \ln(R_0) + \partial_1 \psi(0; 0) s + O(\|Z\|^2). \quad (\text{A.13})$$

Hence

$$s = \frac{-\ln(R_0)}{\partial_1 \psi(0; 0)} + O(\|Z\|^2). \quad (\text{A.14})$$

Remains to calculate

$$\partial_1 \psi(0; 0) = \frac{1}{\lambda_d \left( \int_0^\infty \Lambda(a; X | X) da \right)} \frac{\partial}{\partial s} \lambda_d \left( \int_0^\infty e^{-sa} \Lambda(a; X | X) da \right) \Bigg|_{s=0}. \quad (\text{A.15})$$

The denominator in the first factor equals  $R_0(X | X) = 1$ . The second factor equals (see e.g. [55])

$$-V \left( \left. \frac{\partial}{\partial s} \int_0^\infty e^{-sa} \mathbf{\Lambda}(a; X | X) da \right|_{s=0} \right) U = - \int_0^\infty a V \mathbf{\Lambda}(a; X | X) U da = -T_r. \quad (\text{A.16})$$

Hence

$$s = \frac{\ln(R_0)}{T_r} + O(\|Z\|^2). \quad (\text{A.17})$$

In Appendix B of [4] it is moreover proved that when  $\ln(R_0) = O(\|Z\|^2)$ , as is the case around evolutionarily singular strategies, it is still possible to use (A.17) with  $O(\|Z\|^2)$  replaced by  $O(\|Z\|^3)$ .

### Appendix B. Approximating a mutant's invasion probability

(3.9) has a long history, starting with [10,11] for single birth states, with refinements in [12,13,14]. [15] starts the treatment for multiple birth states, with refinements in [16,17,18]. All later papers are rather complicated as they aim at the strongest possible results. However, when the offspring numbers have third moments a standard perturbation expansion does the job.

As a warming up we consider the case of a single birth state. The following calculation can be found in any textbook devoting space to branching processes. Denote the lifetime number of offspring of a representative individual as  $\underline{k}$ , where the underlining signifies that  $\underline{k}$  is a random variable, and its so-called generating function as

$$g(w) := \mathbf{E}w^{\underline{k}} = \sum_{k=0}^{\infty} q_k w^k, \quad (\text{B.1})$$

with  $q_k$  the probability that an individual begets  $k$  kids. By differentiating one finds for  $n = 0, 1, \dots$

$$\frac{d^n g}{dz^n}(w) = \mathbf{E} \underline{k}^{[n]} w^{(k-n)} \quad \text{and hence} \quad \frac{d^n g}{dz^n}(1) = \mathbf{E} \underline{k}^{[n]}, \quad (\text{B.2})$$

with  $k^{[n]} := k(k-1)\dots(k-n+1)$  and  $k^{[0]} := 1$ .  $\mathbf{E} \underline{k}^{[n]}$  is known as the  $n^{\text{th}}$  factorial moment. Let  $p$  denote the probability of invasion. When an individual is known to beget  $k$  kids, the chance that its line goes extinct is  $(1-p)^k$ . Hence

$$1-p = \sum_0^\infty q_k (1-p)^k = g(1-p). \quad (\text{B.3})$$

We know that  $p = 0$  for  $R_0 \leq 1$ . For small  $R_0 - 1$  and hence small  $p$

$$1-p = g(1) - g'(1)p + \frac{1}{2} g''(1)p^2 + O(p^3) = 1 - R_0 p + \frac{1}{2} (\sigma^2 + R_0^2 - R_0) p^2 + O(p^3), \quad (\text{B.4})$$

$\sigma^2 := \text{Var}(\underline{k})$ . Substituting the Ansatz  $p = c(R_0 - 1) + O((R_0 - 1)^2)$  and solving for  $c$  gives

$$p = \frac{2(R_0 - 1)_+}{\sigma^2} + O((R_0 - 1)^2) = \frac{2 \ln(R_0)}{\sigma^2} + O(\ln(R_0)^2). \quad (\text{B.5})$$

The calculation for more than one birth state starts from the vector of generating functions  $G(W, Z) = (\dots, g_j(W, Z), \dots)^T$  with

$$g_j(W, Z) = \mathbf{E}(\dots w_{i-1}^{k_{i-1}} w_i^{k_i} w_{i+1}^{k_{i+1}} \dots | j). \quad (\text{B.6})$$

$j$  the birth state of the parent. By a similar argument as before (see e.g. [33])

$$G'(\mathbf{1}, Z)^T = \mathbf{L}_m := \mathbf{L}(Y | X) \quad (\text{B.7})$$

with  $\mathbf{1} := (1, \dots, 1)^T$  and a prime denoting the derivative for  $Z$  ( $g'_j$  is thus a row vector with components  $\partial g_j / \partial z_i$  and  $G'$  a matrix), and

$$\begin{aligned} g''_j(\mathbf{1}, Z) &= \mathbf{E} \left( \left( \begin{array}{ccc} k_1(k_1-1) & k_1 k_2 & \dots \\ k_2 k_1 & k_2(k_2-1) & \dots \\ \vdots & \vdots & \ddots \end{array} \middle| j \right) \right) \\ &= \text{Cov}(\underline{K} | j) - \text{diag}(L_{m,j}) + L_{m,j} L_{m,j}^T \end{aligned} \quad (\text{B.8})$$

with

$$g''_j := \begin{pmatrix} \frac{\partial^2 f_j}{\partial z_1^2} & \frac{\partial^2 f_j}{\partial z_1 \partial z_2} & \dots \\ \frac{\partial^2 f_j}{\partial z_2 \partial z_1} & \frac{\partial^2 f_j}{\partial z_2^2} & \dots \\ \vdots & \vdots & \ddots \end{pmatrix}$$

and  $L_{m,j}$  the  $j^{\text{th}}$  column of  $\mathbf{L}_m$ .

Let  $P$  denote the vector of invasion probabilities depending on the birth state of the newly arrived mutant. By a similar argument as before,  $P$  can be shown to satisfy

$$\mathbf{1} - P = G(\mathbf{1} - P, Z) \quad (\text{B.9})$$

[33]. Expanding and making the Ansatz that  $P = \mathbf{C}Z + O(\|Z\|^2)$  gives

$$G(\mathbf{1} - P, Z) = \mathbf{1} - (\mathbf{L}_r^T + \Delta_L^T)P + \frac{1}{2}G''(\mathbf{1}, 0)[P, P] + O(\|Z\|^3) \quad (\text{B.10})$$

with  $\mathbf{L}_r := \mathbf{L}(X | X)$ ,  $\Delta_L := \mathbf{L}_m - \mathbf{L}_r = O(\|Z\|)$  and  $g''_j[P, P] := P^T g''_j P$ . Collecting the first order terms gives

$$P = \mathbf{L}_r^T P. \quad (\text{B.11})$$

Hence

$$P = \kappa V^T, \quad (\text{B.12})$$

with  $\kappa = O(\|Z\|)$ . Substituting this in the equation that results from collecting the second order terms gives

$$0 = \kappa V \Delta_L - \kappa^2 \frac{1}{2} G''(\mathbf{1}, 0)[V, V], \quad (\text{B.13})$$



which, on dividing by  $\kappa$  and right multiplying with  $U$  to get the projection on the dominant left eigenspace, gives

$$0 = V\Delta_L U - \kappa G''^T(\mathbf{1}, 0)[V, V]U + O(\|Z\|^2). \quad (\text{B.14})$$

Finally use that  $V\Delta_L U = \ln(R_0) + O(\|Z\|^2)$  to arrive at

$$\begin{aligned} P &= \frac{2 \ln(R_0)}{G''^T(\mathbf{1}, 0)[V, V]U} V^T + O(\|Z\|^2) = \frac{2 \ln(R_0)}{\sum_j \text{var}_T \left( \sum_i v_i k_i | j \right) u_j} V^T + O(\|Z\|^2) \\ &= \frac{2 \ln(R_0)}{\sigma_e^2} V^T + O(\|Z\|^2). \end{aligned} \quad (\text{B.15})$$

(The step from the factorial moments in  $G''^T(\mathbf{1}, 0)[V, V]U$  to variances is made by observing that  $\sum_{hij} v_h l_{r,hj} v_i l_{r,ij} u_j - \sum_{hj} v_h^2 l_{r,hj} u_j = \sum_j v_j^2 u_j - \sum_h v_h^2 u_h = 0$ .)

Now suppose that there is no relation between birth state and mutation propensity. Then the birth state distribution of a newly appeared mutant is  $U$ , and the probability that a random mutant invades can be approximated as

$$p = P^T U = \frac{2 \ln(R_0)}{\sigma_e^2} + O(\|Z\|^2). \quad (\text{B.16})$$

Like in the invasion fitness case, (B.16), with  $O(\|Z\|^2)$  replaced by  $O(\|Z\|^3)$ , also applies near evolutionarily singular strategies where  $\ln(R_0) = O(\|Z\|^2)$ .

### Appendix C: Calculating $R_0$ for a population in two patches with separate sexes

The proper next generation matrix for this situation is

$$\mathbf{L} = \frac{1}{2} \begin{pmatrix} \ell_{ff,11} & \ell_{ff,12} & \ell_{fm,11} & \ell_{fm,12} \\ \ell_{ff,21} & \ell_{ff,22} & \ell_{fm,21} & \ell_{fm,22} \\ \ell_{mf,11} & \ell_{mf,12} & \ell_{mm,11} & \ell_{mm,12} \\ \ell_{mf,21} & \ell_{mf,22} & \ell_{mm,21} & \ell_{mm,22} \end{pmatrix} \quad (\text{C.1})$$

with  $\ell_{fm,12}$  the average lifetime number of female offspring produced in patch 1 by a mutant male born in patch 2 (through fertilising resident type females), etc.

Now assume that an individual's sex is at most dependent on its birth patch, as is e.g. the case when sex determination is fully environmental or when fathers have no influence on the sex of their offspring, mothers let the sex of their offspring depend at most on where those offspring are born, and the traits are not connected with sex allocation. Then  $\mathbf{L}$  can be written as

$$\mathbf{L} = \frac{1}{2} \begin{pmatrix} p_{f,1} \ell_{\cdot f,11} & p_{f,1} \ell_{\cdot f,12} & p_{f,1} \ell_{\cdot m,11} & p_{f,1} \ell_{\cdot m,12} \\ p_{f,2} \ell_{\cdot f,21} & p_{f,2} \ell_{\cdot f,22} & p_{f,2} \ell_{\cdot m,21} & p_{f,2} \ell_{\cdot m,22} \\ p_{m,1} \ell_{\cdot f,11} & p_{m,1} \ell_{\cdot f,12} & p_{m,1} \ell_{\cdot m,11} & p_{m,1} \ell_{\cdot m,12} \\ p_{m,2} \ell_{\cdot f,21} & p_{m,2} \ell_{\cdot f,22} & p_{m,2} \ell_{\cdot m,21} & p_{m,2} \ell_{\cdot m,22} \end{pmatrix}, \quad (\text{C.2})$$

with  $\ell_{\cdot f,11} = \ell_{ff,11} + \ell_{mf,11}$ , etc., and  $p_{f,1} = \ell_{ff,11}/\ell_{\cdot f,11} = \ell_{ff,12}/\ell_{\cdot f,12}$ , etc., and thus has rank 2. After one generation the births can be written as  $(p_{f,1}b_1 \quad p_{f,2}b_2 \quad p_{m,1}b_1 \quad p_{m,2}b_2)^\top$ . From then on multiplication with  $\mathbf{L}$  gives

$$\begin{pmatrix} b_1 \\ b_2 \end{pmatrix} \mapsto \frac{1}{2} \begin{pmatrix} (\ell_{\cdot f,11}p_{f,1} + \ell_{\cdot m,11}p_{m,1})b_1 + (\ell_{\cdot f,12}p_{f,2} + \ell_{\cdot m,12}p_{m,2})b_2 \\ (\ell_{\cdot f,21}p_{f,1} + \ell_{\cdot m,21}p_{m,1})b_1 + (\ell_{\cdot f,22}p_{f,2} + \ell_{\cdot m,22}p_{m,2})b_2 \end{pmatrix}. \quad (\text{C.3})$$

Hence  $R_0$  equals the dominant eigenvalue of

$$\frac{1}{2} \begin{pmatrix} \ell_{\cdot f,11}p_{f,1} + \ell_{\cdot m,11}p_{m,1} & \ell_{\cdot f,12}p_{f,2} + \ell_{\cdot m,12}p_{m,2} \\ \ell_{\cdot f,21}p_{f,1} + \ell_{\cdot m,21}p_{m,1} & \ell_{\cdot f,22}p_{f,2} + \ell_{\cdot m,22}p_{m,2} \end{pmatrix}. \quad (\text{C.4})$$

When in the general case we write the dominant eigenvector of  $\mathbf{L}$  as  $U = (p_{f,1}b_1 \quad p_{f,2}b_2 \quad p_{m,1}b_1 \quad p_{m,2}b_2)^\top$ , with  $p_{f,i} := u_{f,i}/b_i$ ,  $p_{m,i} := u_{m,i}/b_i$ ,  $b_i := u_{f,i} + u_{m,i}$ , multiplying  $U$  with  $\mathbf{L}$  also gives (C.3). Hence, in this case  $R_0$  also equals the dominant eigenvalue of a matrix constructed by adding the average numbers of kids that individuals father or mother over their life and dividing by 2. However, in general these quantities are no more than population averages that can be determined only after establishing the stable birth state distribution of the full next generation operator.

#### Appendix D. The effective offspring variance for sexual diploid residents

We only consider the case where except for genetic and possibly sex differences all individuals are born equal. In sexually reproducing individuals an allele is born in either of two states, with potentially different futures, to wit in a macro-gamete or micro-gamete. In the simplest case these states are randomly allotted, independent of which route put the parental allele in the parental body. This happens e.g. in hermaphrodites or when sex is determined environmentally or by alleles on a different chromosome. We concentrate on that simplest case. Then  $\sigma_e^2$  is the variance of the number of allelic copies reaching the next generation of zygotes from a copy that has just ended up in a zygote. Let  $k_m$  be the number of alleles that does so micro-gametetically, and  $k_f$  the number that does so macro-gametetically. Generally  $k_m$  and  $k_f$  are dependent. For example, for a seed, and hence for the two alleles in it, the size of the plant it engenders is a random variable, with larger plants usually producing more ovules as well as pollen so that  $k_m$  and  $k_f$  are positively correlated. As the extreme opposite, in dioecious organisms a new zygote will go on to produce either micro- or macro-gametes, i.e.,  $k_m k_f = 0$ . Hence the starting point of the calculation is the generating function of the pair  $(k_f, k_m)$

$$g(w_f, w_m) := \mathbf{E} w_f^{k_f} w_m^{k_m}. \quad (\text{D.1})$$

The generating function of the number  $\underline{k}$  of  $aA$  offspring of an  $aA$  individual reproducing in an  $aa$  population background is then

$$g_A(w) = g\left(\frac{1}{2} + \frac{1}{2}w, \frac{1}{2} + \frac{1}{2}w\right). \quad (\text{D.2})$$

with  $\frac{1}{2} + \frac{1}{2}w$  the generating function of the number of  $A$  alleles, 0 or 1, in a gamete. For the calculation of  $\sigma_e^2$  the allele  $A$  is supposed to have no phenotypic effect. Hence, at population dynamical equilibrium both  $\partial_1 g(1,1) = \partial_2 g(1,1) = 1$  since the number of individuals stays constant over the generations and everybody has one father and one mother. Therefore

$$\mathbf{E}\underline{k} = \partial g_A(1) = \frac{1}{2}(\partial_1 g(1,1) + \partial_2 g(1,1)) = 1, \quad (\text{D.3})$$

and

$$\begin{aligned} \sigma_e^2 &= \mathbf{E}\underline{k}^2 - (\mathbf{E}\underline{k})^2 = \mathbf{E}\underline{k}^2 - \mathbf{E}\underline{k} \\ &= \left(\frac{1}{2}\right)^2 \left[ \partial_1^2 g(1,1) + 2\partial_1 \partial_2 g(1,1) + \partial_2^2 g(1,1) \right] \\ &= \frac{1}{4} \left[ \sigma_f^2 + 2(c+1) + \sigma_m^2 \right] \end{aligned} \quad (\text{D.4})$$

with

$$c = \mathbf{E}(k_f - 1)(k_m - 1) = \mathbf{E}k_f k_m - 1 \quad (\text{D.5})$$

the covariance between macro- and micro-gametically produced offspring.

#### Appendix E: The effective offspring variance for haplo-diploid residents

Following the pattern from Appendix D we first express the variances and covariances of the allelic offspring numbers in terms of the variances and covariances of the numbers of children of the resident individuals. Let  $g_a$  denote the generating function of the lifetime offspring numbers of  $a$  residing in a female and ending up in female and male children of that female and let  $g$  denote the generating function of those numbers of children  $k_f$  and  $k_m$  irrespective of their genotype, then

$$g_a(w_f, w_m) = g\left(\frac{1}{2} + \frac{1}{2}w_f, \frac{1}{2} + \frac{1}{2}w_m\right). \quad (\text{E.1})$$

Differentiating gives, with  $i, j = f, m$ ,

$$\begin{aligned} \partial_i g_a(w_f, w_m) &= \frac{1}{2} \partial_i g\left(\frac{1}{2} + \frac{1}{2}w_f, \frac{1}{2} + \frac{1}{2}w_m\right), \\ \partial_{ij} g_a(w_f, w_m) &= \frac{1}{4} \partial_{ij} g\left(\frac{1}{2} + \frac{1}{2}w_f, \frac{1}{2} + \frac{1}{2}w_m\right). \end{aligned} \quad (\text{E.2})$$

Hence

$$\partial_1 g_a(1,1) = \frac{1}{2} f_f = \frac{1}{2}, \quad \partial_2 g_a(1,1) = \frac{1}{2} f_m = \frac{1}{2} r, \quad (\text{E.3})$$

and

$$\begin{aligned} \partial_{11} g_a(1,1) &= \frac{1}{4} (\mathbf{E}k_f^2 - \mathbf{E}k_f) = \frac{1}{4} ((\sigma_{ff}^2 + 1) - 1) = \frac{1}{4} \sigma_{ff}^2, \\ \partial_{12} g_a(1,1) &= \frac{1}{4} \mathbf{E}k_f k_m = \frac{1}{4} (c_{ff} + r), \\ \partial_{22} g_a(1,1) &= \frac{1}{4} (\mathbf{E}k_m^2 - \mathbf{E}k_m) = \frac{1}{4} ((\sigma_{mm}^2 + r^2) - r). \end{aligned} \quad (\text{E.4})$$

Therefore

$$\begin{aligned}
\sigma_{a,\text{flf}}^2 &= \left(\frac{1}{4}\sigma_{\text{flf}}^2 + \frac{1}{2}\right) - \frac{1}{4} = \frac{1}{4}(\sigma_{\text{flf}}^2 + 1), \\
c_a &= \frac{1}{4}(c_{\text{lf}} + r) - \frac{1}{4}r = \frac{1}{4}c_{\text{lf}}, \\
\sigma_{a,\text{mlf}}^2 &= \frac{1}{4}((\sigma_{\text{mlf}}^2 + r^2) - r) + \frac{1}{2}r - \frac{1}{4}r^2 = \frac{1}{4}(\sigma_{\text{mlf}}^2 + r),
\end{aligned} \tag{E.5}$$

where the lf in the indices of  $\sigma$  and  $c$  indicates that the quantities are calculated conditional on the parent being a female. This gives all the ingredients for calculating  $\sigma_c^2$ . First calculate

$$\begin{aligned}
\text{var}\left(\sum_i v_i k_{a,i} \middle| \text{f}\right) &= \frac{1}{4}\left(\frac{2+r}{3r}\right)^2 \left[r^2(\sigma_{\text{flf}}^2 + 1) + 2rc_{\text{lf}} + (\sigma_{\text{mlf}}^2 + r)\right] \\
&= \frac{1}{4}\left(\frac{2+r}{3r}\right)^2 \left[r^2\sigma_{\text{flf}}^2 + r(2c_{\text{lf}} + 1 + r) + \sigma_{\text{mlf}}^2\right], \\
\text{var}\left(\sum_i v_i k_{a,i} \middle| \text{m}\right) &= \left(\frac{2+r}{3r}\right)^2 r^2\sigma_{\text{flm}}^2.
\end{aligned} \tag{E.6}$$

Substituting (4.3) and (E.6) in (3.9) gives

$$\begin{aligned}
\sigma_c^2 &= \frac{1}{2+r}\left(\frac{2+r}{3r}\right)^2 \left[\frac{1}{2}(r^2\sigma_{\text{flf}}^2 + r(2c_{\text{lf}} + 1 + r) + \sigma_{\text{mlf}}^2) + r^3\sigma_{\text{flm}}^2\right] \\
&= \frac{2+r}{18r} \left[2r^2\sigma_{\text{flm}}^2 + r\sigma_{\text{flf}}^2 + \sigma_{\text{mlf}}^2 + (2c_{\text{lf}} + 1 + r)\right].
\end{aligned} \tag{E.8}$$

For good measure we add the formula for the average age at reproduction

$$T_r = V \begin{pmatrix} T_{r,\text{ff}} & T_{r,\text{fm}} \\ T_{r,\text{mf}} & 0 \end{pmatrix} U = \frac{1}{3} \left[ 2(T_{r,\text{ff}} + r^{-1}T_{r,\text{mf}}) + rT_{r,\text{fm}} \right]. \tag{E.9}$$

## References

- [1] Dieckmann, U. & Law, R. 1996 The dynamical theory of coevolution: a derivation from stochastic ecological processes. *J. Math. Biol.* **34**, 579–612.
- [2] Champagnat, N. 2003 Convergence of adaptive dynamics n-morphic jump processes to the canonical equation and degenerate diffusion approximation. *Prépublication de l'Université de Nanterre (Paris X)* 03/7.
- [3] Champagnat, N. & Méléard, S. 2011 Polymorphic evolution sequence and evolutionary branching. *Probab. Theory. Relat. Fields* **151**, 45–94. (DOI 10.1007/s00440-010-0292-9)
- [4] Durinx, M., Metz, J.A.J. & Meszéna, G. 2008 Adaptive dynamics for physiologically structured models. *J. Math. Biol.* **56**, 673–742. (DOI 10.1007/s00285-007-0134-2)
- [5] Méléard, S. & Tran, V.C. 2009 Trait substitution sequence process and canonical equation for age-structured populations. *J. Math. Biol.* **58**, 881–921. (DOI 10.1007/s00285-008-0202-2)
- [6] Collet, P., Méléard, S. & Metz, J.A.J. 2013 A rigorous model study of the adaptative dynamics of Mendelian diploids. *J. Math. Biol.* (online first) (DOI 10.1007/s00285-012-0562-5)

- [7] Metz, J.A.J., Nisbet, R.M. & Geritz, S.A.H. 1992 How should we define "fitness" for general ecological scenarios? *TREE* **7**, 198-202.
- [8] Metz, J.A.J. 2008 Fitness. In *Evolutionary ecology*. Vol. 2 of *Encyclopedia of ecology*. (eds. S.E. Jørgensen & B.D. Fath), pp. 1599-1612. Oxford: Elsevier.
- [9] Diekmann, O., Heesterbeek, J.A.P. & Metz, J.A.J. 1990 On the definition and the computation of the basic reproduction ratio  $R_0$  in models for infectious diseases in heterogeneous populations. *J. Math. Biol.* **28**, 365-382.
- [10] Haldane, J.B.S. 1927 A mathematical theory of natural and artificial selection. v. selection and mutation. *Proc. Camb. Philos. Soc.* **23**, 838-844.
- [11] Ewens, W. 1969 *Population genetics*. London: Methuen.
- [12] Eshel, I. 1981 On the survival probability of a slightly advantageous mutant gene with a general distribution of progeny size—a branching process model. *J. Math. Biol.* **12**, 355-362.
- [13] Hoppe, F. 1992 Asymptotic rates of growth of the extinction probability of a mutant gene. *J. Math. Biol.* **30**, 547-566.
- [14] Athreya, K. 1992 Rates of decay for the survival probability of a mutant gene. *J. Math. Biol.* **30**, 577-581.
- [15] Eshel, I. 1984 On the survival probability of a slightly advantageous mutant gene in a multitype population: a multidimensional branching process model. *J. Math. Biol.* **19**, 201-209.
- [16] Hoppe, F. 1992 The survival probability of a mutant in a multidimensional population. *J. Math. Biol.* **30**, 567-575.
- [17] Pollak, E. 1992 Survival probabilities for some multitype branching processes in genetics. *J. Math. Biol.* **30**, 583-596.
- [18] Athreya, K. 1993 Rates of decay for the survival probability of a mutant gene. II. The multitype case. *J. Math. Biol.* **32**, 45-53.
- [19] Lotka, A.J. 1939 *Théorie analytique des associations biologiques*. Part II. *Analyse démographique avec application particulière à l'espèce humaine*. (*Actualités Scientifiques et Industrielles* 740). Paris: Hermann & Cie.
- [20] Metz, J.A.J. & Diekmann, O. eds. 1986 *The dynamics of physiologically structured populations*. *Lecture Notes in Biomathematics*, **68**. Berlin: Springer Verlag.
- [21] Geritz, S.A.H., Gyllenberg, M., Jacobs, F.J.A. & Parvinen, K. 2002 Invasion dynamics and attractor inheritance. *J. Math. Biol.* **44**, 548-560. (DOI 10.1007/s002850100136)
- [22] Dercole, F. & Rinaldi, S. 2008 *Analysis of evolutionary processes: the adaptive dynamics approach and its applications*. Princeton NJ: Princeton U.P.
- [23] Geritz, S.A.H. 2005 Resident-invader dynamics and the coexistence of similar strategies. *J. Math. Biol.* **50**, 67-82. (DOI 10.1007/s00285-004-0280-8)
- [24] Greiner, G., Heesterbeek J.A.P., & Metz, J.A.J. 1994 A singular perturbation theorem for evolution equations and time-scale arguments for structured population models. *Can. Appl. Math. Quart.* **2**, 435-459.
- [25] Diekmann, O., Getto, P. & Gyllenberg, M. 2007 Stability and bifurcation analysis of Volterra functional equations in the light of suns and stars. *SIAM J. Math. Anal.* **39**, 1023-1069. (DOI 10.1137/060659211)
- [26] Diekmann, O. & Gyllenberg, M. 2012 Equations with infinite delay: blending the abstract and the concrete. *J. Diff. Equ.* **252**, 819-851. (DOI 10.1016/j.jde.2011.09.038)
- [27] Carroll, S.B. 2005 *Endless forms most beautiful: the new science of Evo Devo*. New York: W. W. Norton & Cy

- [28] Davidson, E.H. 2006 *The regulatory genome: gene regulatory networks in development and evolution*. Amsterdam: Academic Press.
- [29] Van Dooren, T.J.M. 2006 Protected polymorphism and evolutionary stability in pleiotropic models with trait-specific dominance. *Evolution* **60**, 1991–2003.
- [30] Metz, J.A.J. & Jansen V.A.A. in prep. Adaptive dynamics for Mendelian genetics: relating the speed of evolution to the effective population size.
- [31] Diekmann, O., Gyllenberg, M. & Metz J.A.J. 2003 Steady state analysis of structured population models. *Theor. Pop. Biol.* **63**, 309-338.
- [32] Jagers, P. 1975 *Branching processes with biological applications*. London: Wiley.
- [33] Haccou, P., Jagers, P. & Vatutin, V.A. 2005 *Branching processes: variation, growth, and extinction of populations*. Cambridge: Cambridge U.P.
- [34] Ripa, J. & Dieckmann, U. 2013 Mutant invasions and adaptive dynamics in variable environments. *Evolution* **67**, 1279-1290. (DOI 10.1111/evo.12046)
- [35] Metz, J.A.J., Stanková, K., Johansson, J. in prep. The adaptive dynamics of life-histories: from fitness-returns to selection gradients and Pontryagin's maximum principle.
- [36] Metz, J.A.J. 2013 On the concept of individual in ecology and evolution. *J. Math. Biol.* **66**, 635-647. (DOI 10.1007/s00285-012-0610-1)
- [37] Otto, S.P. & Jarne, P. 2001 Evolution: haploids hapless or happening? *Science* **292**, 2441-2443. (DOI 10.1126/science.1062890)
- [38] Taylor, P.D. & Bulmer, M.G. 1980 Local mate competition and the sex ratio. *J. theor. Biol.* **86**, 409-419.
- [39] Courteau, J. & Lessard, S. 2000 Optimal sex ratios in structured populations. *J. theor. Biol.* **207**, 159-175. (DOI 10.1006/jtbi.2000.2160)
- [40] Metz, J.A.J. & Leimar, O. 2011 A simple fitness proxy for ESS calculations in structured populations with continuous traits, with applications to the evolution of haplo-diploids and genetic dimorphisms. *J. Biol. Dyn.* **5**, 163-190. (DOI 10.1080/17513758.2010.502256)
- [41] Meszéna, G., Gyllenberg, M., Jacobs, F.J. & Metz, J.A.J. 2005 Link between population dynamics and dynamics of Darwinian evolution. *Phys. Rev. Lett.* **95**, 078105(4). (DOI 10.1103/PhysRevLett.95.078105)
- [42] Lande, R. 1979 Quantitative genetic analysis of multivariate evolution, applied to brain:body size allometry. *Evolution* **33**, 402-416.
- [43] Lande, R. 1982 A quantitative genetic theory of life history evolution. *Ecology* **63**, 607-615.
- [44] Bulmer, M.G. 1980 *The mathematical theory of quantitative genetics*. Oxford: Clarendon Press.
- [45] Lande, R. 1976 Natural selection and random genetic drift in phenotypic evolution. *Evolution* **30**, 314-334.
- [46] Lande, R. 1980 The genetic covariance between characters maintained by pleiotropic mutations. *Genetics* **94**, 203-215.
- [47] Fisher, R.A. 1958 *The genetical theory of natural selection*. 2<sup>nd</sup> ed. New York: Dover.
- [48] Leimar O. 2009 Multidimensional convergence stability. *Evol. Ecol. Res.* **11**, 191-208
- [49] Metz, J.A.J., Geritz, S.A.H., Meszéna, G., Jacobs, F.J.A. & van Heerwaarden, J.S. 1996 Adaptive dynamics, a geometrical study of the consequences of nearly faithful reproduction. In *Stochastic and spatial structures of dynamical systems*. ( eds. S.J. van Strien & S.M. Verduyn Lunel), pp. 183-231. Amsterdam: North-

- Holland.
- [50] Geritz, S.A.H., Metz, J.A.J. Kisdi, É & Meszéna, G. 1997 Dynamics of adaptation and evolutionary branching. *Phys. Rev. Lett.* **78**, 2024-2027.
  - [51] Geritz S.A.H., Kisdi, É., Meszéna, G. & Metz, J.A.J. 1998 Evolutionarily singular strategies and the adaptive growth and branching of the evolutionary tree. *Evol. Ecol.* **12**, 35-57.
  - [52] Van Dooren, T.J.M., Durinx, M. & Demon, I. 2004 Sexual dimorphism or evolutionary branching? *Evol. Ecol. Res.* **6**, 857-871.
  - [53] Durinx, M., & van Dooren, T.J.M. 2009 Assortative mate choice and dominance modification: alternative ways of removing heterozygote disadvantage. *Evolution* **63**, 334-352. (DOI 10.1111/j.1558-5646.2008.0578.x)
  - [54] Geritz, S.A.H. & Metz, J.A.J. in prep. Evolutionary branching of multi-dimensional strategies.
  - [55] Caswell, H. 2001 *Matrix population models*. 2nd ed. Sunderland MA: Sinauer.