The dynamics induced in a biological system by broken mutational symmetry and non-linear feedback

D. Waxman
Centre for the Study of Evolution
School of Life Sciences, University of Sussex
Brighton BN1 9QG, Sussex UK
D.Waxman@sussex.ac.uk

Abstract

Some biological systems involving mutation and selection are characterised by nonlinear equations that can have dynamically complex outcomes. We consider the genes affecting a quantitative trait, that have been previously found to have a distribution that persistently changes and does not equilibrate. This is the outcome of non-linear feedback arising from selection and mutational processes that vary from one genetic locus to another, thereby breaking the symmetry that all loci are equivalent. Here we adopt an approximate analytical approach to identify the evolutionary "force" that drives the persistent change.

Keywords: mutation, selection, biological dynamics, quantitative trait

1 Model

This work is concerned with a mathematical description of a population of living organisms. We investigate a model of a quantitative trait. Quantitative traits are measurable features of an organism. Their value is affected by many genes along with the environment in which the organism develops, and hence are continuously distributed. One example of such a trait is the height of an organism.
Consider a large population of randomly mating sexual organisms, with each individual characterised by a phenotypic trait. There are no systematic differences in the traits of the two sexes (no sexual dimorphism). The value of the phenotypic trait, Z, is the sum of an environmental component, ε, and the individual's "genotypic value", G, so $Z = \varepsilon + G$.

The environmental component, ε, is drawn for each individual at random from a normal distribution with mean zero and a fixed variance.

To determine the genotypic value, G, we take the population to be diploid and have n freely recombining loci. The effects of the maternally and paternally inherited alleles at locus i $(= 1, 2, ..., n)$ are X_i and Y_i, respectively. Following Crow and Kimura [1], we take allelic effects to be continuous and have an infinite range: $-\infty < X_i, Y_i < \infty$, and genotypic values to be additively determined from the effects of the alleles at the loci controlling the trait. That is, $G = \sum_{i=1}^{n}(X_i + Y_i)$. Maternally and paternally inherited alleles have identical distributions (apart from a short transient period) and henceforth we shall only refer to the distribution of alleles of maternal origin.

Generations are taken to be overlapping (time is continuous). Following Kimura [2], we take the population to be subject to stabilising viability selection which is characterised by a death rate that is a quadratic function of the phenotypic trait value. Thus an individual with phenotypic value z has death rate $s\,(z - z_{\rm opt})^2 +$ "z independent terms", where s is a positive parameter characterising the strength of selection on phenotypes towards the value $z_{\rm opt}$, which is termed the optimal phenotype. We take $s \ll 1$ as is often found in natural populations [3].

We derive the effect of selection on *genotypic* values, by focussing on a large number of individuals with identical genotypes, but different phenotypes (due to different environmental effects, ε). Averaging over ε yields a death rate for individuals with genotypic value G of $s\,(G - z_{\rm opt})^2 +$ "G independent terms".

Gamete formation is taken to result from Mendelian segregation and free recombination. Given that each of the $2n$ alleles in an individual is a copy of a parental allele, the effect of an allele in an offspring is identical to that of the parental allele, unless a mutation occurred during its production. The probability of a mutation of an allele at locus i is written μ_i.

Mutations to different alleles are taken to occur independently. Mutational changes of allelic effects are taken to have continuous values [1], [2] and have a Gaussian distribution. Thus if x is the effect of an allele in an individual and x^* is the effect of the parental allele, then if no mutation occurred, we have $x = x^*$. If a mutation did occur then x is drawn from a normal distribution with mean $x^* + b_i$ and variance m_i^2. The probability density of mutated allelic effects is thus

$$f_i(x - x^* - b_i) = \sqrt{\frac{1}{2\pi m_i^2}}\exp\left(-\frac{(x - x^* - b_i)^2}{2m_i^2}\right). \qquad (1)$$

If $b_i = 0$ we have the model of Lande [4]. If $b_i \neq 0$ the mutation distribution is not symmetric around the parental value and b_i is a bias introduced by mutation [6].

Lastly, we take birth rate to be independent of phenotype and by a choice of time units set the birth rate to unity and assume population numbers are regulated by an unspecified ecological process (see e.g., Appendix A of [7]).

2 Calculation

With $\mathbf{x} = (x_1, x_2, ...x_n)$ a realisation of the n allelic effects in gametes of maternal origin, let $\Phi(\mathbf{x}, t)$ denote the probability density of these at time t. Under the neglect of linkage disequilibria [9], allelic effects at different loci are statistically independent. Thus $\Phi(\mathbf{x}, t)$ can then be approximated by $\prod_{j=1}^{n} \varphi_j(x_j, t)$ where $\varphi_j(x_i, t)$ is the distribution of allelic effects of maternal origin at locus j at time t. From the description of the model, it then follows that

$$-\frac{\partial \varphi_i(x_i,t)}{\partial t} = s\left[(x_i + \overline{G} - z_{\rm opt} - \bar{x}_i)^2 - \overline{(x_i + \overline{G} - z_{\rm opt} - \bar{x}_i)^2}\right]\varphi_i(x_i,t) + \mu_i\varphi_i(x_i,t) - \mu_i \int f_i(x_i - y - b_i)\varphi_i(y,t)dy. \quad (2)$$

Here, overbars denote averaging at time t, hence $\overline{A(x_i)} = \int A(x_i)\,\varphi_i(x_i,t)dx$, so the mean trait value is

$$\overline{G} \equiv \overline{G}(t) = 2\sum\nolimits_{i=1}^{n} \bar{x}_i \quad (3)$$

and we use the convention that all integrals with unspecified limits cover the full $-\infty$ to ∞ range of the integration variable.

3 Gaussian approximation

For general values of the mutation rate, μ_i, the only known exact solution of Eq. (2) is when (i) the population is at equilibrium (ii) all b_i vanish and (iii) $f_i(x)$ has a particular symmetric form similar to a Gaussian [7]. For analytical progress, approximations need to be considered.

Equation (2) was, with an unspecified distribution of mutations, analysed by Kimura under a Gaussian approximation for $\varphi_i(x_i, t)$ [2]. Two aspects to Kimura's calculation require comment.

(i) Kimura's approximation was shown to have good validity at loci where $\mu_i/(sm_i^2) \gg 1$ [5]. This parameter region is not believed appropriate for a *typical* locus [3] but Welch and Waxman [8] suggested that there may be *some* loci where $\mu_i/(sm_i^2) \gg 1$. Perhaps the simplest situation where this arises is if there are loci whose values of m_i^2 are anomalously small.

(ii) Kimura implicitly assumed the population equilibrates. That is, at long times, he assumed $\varphi_i(x_i, t)$ becomes independent of time. Waxman and Peck [6] showed that when there is variation of the b_i across loci, as well, possibly, as variation of other mutational parameters (such as the m_i and the μ_i), the assumption that $\varphi_i(x_i, t)$ equilibrates is mathematically inconsistent. Numerical evidence [6] suggests that at long times, the only aspect of $\varphi_i(x_i, t)$ that remains time dependent is the mean allelic effect, $\bar{x}_i(t) = \int x\varphi_i(x,t)dx$. At long times, the $\bar{x}_i(t)$ were numerically found to change linearly with time, with some $\bar{x}_i(t)$ taking increasingly positive values, and

others takes increasingly negative values. It was observed that the sum of the mean allelic effects, which is proportional to the mean genotypic value $\overline{G}(t)$ (Eq. (3)), achieves a constant value at long times, indicating that $\overline{G}(t)$ equilibrates.
Here, we present a Gaussian approximation of Eq. (2) that does not make implicit assumptions about the $\varphi_i(x_i, t)$ at long times. A purpose of the present work is to provide some analytical evidence for the "running allele" phenomenon that was numerically seen by Waxman and Peck [6], where the $\bar{x}_i(t)$ asymptotically change linearly with time. Additionally, we have the aim of identifying the evolutionary "force" that drives the "running" of alleles. The calculations apply when $\mu_i/(sm_i^2) \gg 1$ and we would like to extend the analysis to the regime $\mu_i/(sm_i^2) \ll 1$ that is believed to be relevant to most loci, but this has not yet been possible.
The Gaussian approximation entails replacing $\varphi_i(x_i, t)$ by a Gaussian function of x_i. In such a case $\varphi_i(x_i, t)$ is completely characterised by its mean, $\bar{x}_i$, and its variance, V_i, and in general, these change with time. In Appendix A it is shown that $\bar{x}_i$ and V_i, obey

$$\frac{d}{dt}\bar{x}_i = -2sV_i\left(\overline{G} - z_{\rm opt}\right) + \mu_i b_i \tag{4}$$

$$\frac{d}{dt}V_i = -2sV_i^2 + \mu_i m_i^2 + \mu_i b_i^2. \tag{5}$$

We note that: (i) there is coupling between variables of the same locus, since Eq. (4) explicitly involves the variance, V_i, of the same locus, (ii) there is coupling between variables of the different loci, since $\overline{G}$ has the form in Eq. (3) and hence Eq. (4) involves mean allelic effects of loci other than just locus i.

4 Long-time solutions

Since Eq. (5) is independent of $\bar{x}_i$, it simply results in V_i approaching its single stable equilibrium value

$$\widehat{V}_i = \sqrt{\mu_i\left(m_i^2 + b_i^2\right)/(2s)}. \tag{6}$$

To determine the long time behaviour of the $\bar{x}_i$, we first determine the equilibrium mean trait value, $\widehat{\overline{G}}$, and use this in Eq. (4). We proceed by summing Eq. (4) over all i and find $\overline{G}$ equilibrates to

$$\widehat{\overline{G}} = z_{\rm opt} + \frac{\sum_{i=1}^n \mu_i b_i}{2s\sum_{i=1}^n \widehat{V}_i}. \tag{7}$$

Using Eqs. (6) and (7) in Eq. (4) leads to the $\bar{x}_i$ having a rate of change, $d\bar{x}_i/dt$, that approaches a constant value R_i given by

$$R_i = \lim_{t\to\infty}\frac{d\bar{x}_i}{dt} = \mu_i b_i - \widehat{V}_i\frac{\sum_{j=1}^n \mu_j b_j}{\sum_{j=1}^n \widehat{V}_j}. \tag{8}$$

In general, the R_i do not vanish, so the $\bar{x}_i$ do not come to equilibrium. Equation (8) signals that there is an asymptotically linear change of $\bar{x}_i$ with t: $\bar{x}_i(t) \sim R_i t$. This behaviour, previously described as a "running" of the mean allelic effects [6], corresponds to a constant rate of turnover of alleles at every locus. Alleles of larger *absolute* effect are substituted at each locus with $R_i \neq 0$. In Figure 1 the rates of change of the mean allelic effects are plotted against time t and may be observed to approach constant values.

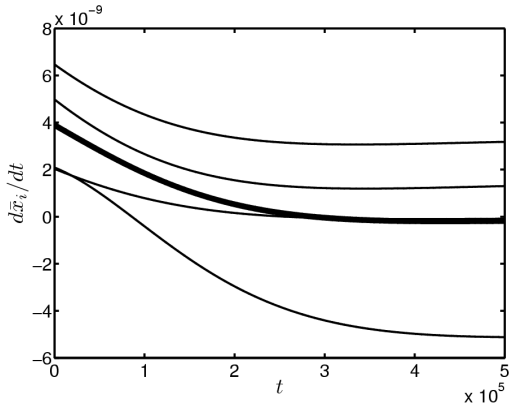

Figure 1: The rate of change of the mean allelic effects of different loci, $d\bar{x}_i/dt$, are plotted against time, t (thin lines). Also plotted is the rate of change, averaged over all loci, $n^{-1}\sum_{j=1}^{n} d\bar{x}_i/dt$ (thick line). The case illustrated is for $n = 4$ non-equivalent loci, where $\boldsymbol{\mu} = 10^{-4}\times$(0.0356,0.0719,0.0113,0.1044), $\mathbf{m} = 10^{-2}\times$(0.07,0.04,0.04, 0.18), $\mathbf{b} = 10^{-2}\times$(0.14,0.09,0.18,0.02), $s = 0.1$ and the initial conditions are $\mathbf{V}(0) = 10^{-5}\times$(0.4747,0.4418,0.2504,0.2425) and $\overline{G}(0) - z_{\rm opt} = 0$.

An automatic consequence of the form of Eq. (8) is that

$$\sum\nolimits_{i=1}^{n} R_i = 0 \tag{9}$$

i.e., at long times, $d\left(2\sum_{i=1}^{n}\bar{x}_i\right)/dt \equiv d\overline{G}/dt = 0$. Thus the running of the mean allelic effects and the equilibration of the mean trait value are exact features of a Gaussian approximation. In Figure 1, the thick line represents $d\left(n^{-1}\sum_{i=1}^{n}\bar{x}_i\right)/dt$ and approaches zero at long times, in accordance with Eq. (9).

The form of the R_i (Eq. (8)) has an interesting structure. When there are no mutational biases (all $b_i = 0$), all R_i vanish. Thus non-zero mutational biases are a *necessary condition* for the R_i to be non zero. That this is not sufficient can be seen in the case where all loci are mutationally equivalent, with all μ_i, b_i and m_i, and hence $\widehat{V}_i$, not exhibiting variation across loci. In this "symmetric" case, we have all $R_i = 0$. Thus variation of some mutational parameters across loci, combined with non-zero b_i, are

required for all R_i not to vanish. There appears to be no special reason why loci are mutationally equivalent or all b_i vanish, and since the slightest breaking of equivalence of loci leads to the R_i being non-zero, we infer the *generic* long time behaviour of the model is running of the $\bar{x}_i$.

5 Two loci

It is illuminating to consider properties of the rates of running of mean allelic effects (the R_i) when there are just $n = 2$ loci controlling the trait. In this case we can write $R_1 = \frac{\widehat{V}_1\widehat{V}_2}{\widehat{V}_1+\widehat{V}_2}\left(\frac{\mu_1 b_1}{\widehat{V}_1} - \frac{\mu_2 b_2}{\widehat{V}_2}\right)$ and $R_2 = -R_1$. It is clear that the direction of "running" of alleles at locus 1 (the sign of R_1) can depend sensitively on which of the two terms within the bracket determine its sign. If both b_1 and b_2 are positive then a small change in the sizes of the mutation rates at the two loci has the possibility of changing the sign of R_1. More generally, a small change in parameters has the possibility of making a very large change in the long time properties of the population, where $\bar{x}_1$ will have very different values at long times if $R_1 > 0$ compared with $R_1 < 0$. Such parameters may be sensitive to external environmental conditions, such as temperature, and small changes in the location of a population have the possibility of resulting in large genetic changes.

An interesting feature is that the direction of the rate of running of allelic effects at a locus can be in the *opposite* direction to the direction mutational bias pushes alleles at that locus. For example, if $\mu_1 = \mu_2 = \mu$, $m_1 = m_2 = m$ but $0 < b_1 \ll b_2 \ll m$ then the equilibrium allelic variances are similar: $\widehat{V}_1 \simeq \widehat{V}_2$, but the rates of running are of opposite sign: $R_1 \simeq -\mu b_2/2$ and $R_2 \simeq \mu b_2/2$. In this example, locus 1 is subject to a positive mutational bias but remarkably has a negative rate of change of its mean allelic effect at long times.

6 Feedback

The phenomenon that the $\bar{x}_i$ asymptotically increase with time is, in part, a result of feedback arising from selection. In the absence of selection ($s = 0$), Eq. (4) is simply $d\bar{x}_i/dt = \mu_i b_i$ and this accounts for the leading term on the right hand side of Eq. (8). The second term on the right hand side of Eq. (8) must arise from feedback arising from selection, even though in the Gaussian approximation this term is independent of the strength of selection, s.

To get an idea of the timescale over which such feedback occurs, we have $d\bar{x}_i/dt = \mu_i b_i - 2sV_i\left(\overline{G} - z_{\rm opt}\right)$ hence the second term on the right hand side achieves its long time form when both V_i and $\overline{G} - z_{\rm opt}$ have achieved their long time equilibrium values. It may be shown that V_i takes a time of order $\left(4s\widehat{V}_i\right)^{-1}$ to equilibrate, while

$\overline{G} - z_{\rm opt}$ takes a shorter time of order $\left(4s\sum_{j=1}^{n}\widehat{V}_j\right)^{-1}$. Thus we can say effectively all feedback has occurred at locus i after a time of order $\left(4s\widehat{V}_i\right)^{-1}$ and in a Gaussian approximation, $\widehat{V}_i \propto s^{-1/2}$ (Eq. (6)). Thus the timescale of selectively induced feedback scales with the strength of selection as $\left(s \times s^{-1/2}\right)^{-1} = s^{-1/2}$.

7 Summary

We have considered the dynamics of a quantitative trait. The full mathematical problem is undoubtedly complex and we have implemented a Gaussian approximation. This generally represents a significant departure from the full set of equations, but has the virtue that explicit results can be derived for quantities of direct interest. One such quantity is the rate at which mean allelic effects "run" with time. Results for these are in qualitative agreement with a numerical treatment of the full problem, suggesting that a Gaussian approximation does capture important features.

It is interesting to identify the aspects of the problem that lead to "running" alleles. It would appear that a necessary feature is the presence of the mean trait value, $\overline{G}(t)$, in the equations describing the alleles at *different* loci. The presence of this term accounts for *communication* or *feedback* between different loci. A second key factor is variation in mutational properties of the different loci. Because of this, different loci do not have identical dynamical equations. Of all of the possible differences of mutational properties of different loci, the mutational biases (the b_i) play a central role in "running" of alleles; if all $b_i = 0$ then irrespective of variation of other parameters there is no running of alleles.

It would also be interesting to find other models, perhaps simpler ones, where the dynamics depends on the mean value of a "central variable" analogous to $\overline{G}(t)$, and has a "broken symmetry" analogous to non equivalence of the mutational properties of the different loci, to see if persistent, non equilibrium behaviour arises.

Appendix A

In this Appendix, we derive equations for the mean allelic effect, $\bar{x}_i$ and the variance, V_i under a Gaussian approximation. We begin by defining $\delta x_i = x_i - \bar{x}_i$ so $V_i = \overline{(\delta x_i)^2}$ and Eq. (2) takes the form

$$\frac{\partial}{\partial t}\varphi_i(x_i,t) = -s\left[(\delta x_i)^2 + 2\delta x_i(\overline{G} - z_{\rm opt}) - V_i\right]\varphi_i(x_i,t) - \mu_i\varphi_i(x_i,t) + \mu_i\int f_i(x_i - y - b_i)\varphi_i(y,t)dy. \tag{10}$$

Multiplying Eq. (10) by x_i and integrating over all x_i yields

$$\frac{d}{dt}\bar{x}_i = -s\overline{(\delta x_i)^3} - 2sV_i\left(\overline{G} - z_{\rm opt}\right) + \mu_i b_i. \tag{11}$$

Multiplying Eq. (10) by x_i^2 and integrating over all x_i yields $\frac{d}{dt}\overline{x^2}_i = -2s\bar{x}_i\overline{(\delta x_i)^3} - s\overline{(\delta x_i)^4} - s\left[4\bar{x}_i V_i + 2\overline{(\delta x_i)^3}\right](\overline{G} - z_{\rm opt}) + sV_i^2 + \mu_i(m_i^2 + 2\bar{x}_i b_i + b_i^2)$. Using $\frac{d}{dt}V_i = \frac{d}{dt}\overline{x^2}_i - 2\bar{x}_i\frac{d}{dt}\bar{x}_i$ we find

$$\frac{d}{dt}V_i = sV_i^2 + \mu_i m_i^2 - s\overline{(\delta x_i)^4} - 2s\overline{(\delta x_i)^3}(\overline{G} - z_{\rm opt}) + \mu_i b_i^2. \qquad (12)$$

Some approximations of Eqs. (11) and (12) amount to "moment closure" relations (high order moments of δx_i are expressed in terms of low order moments). Here we employ a Gaussian approximation; Eqs. (11) and (12) then simplify, since $\overline{(\delta x_i)^3} = 0$ and $\overline{(\delta x_i)^4} = 3V_i^2$, to Eqs. (4) and (5).

References

[1] Crow, J. F. and Kimura M. *The theory of genetic loads*, in S. J. Geerts (Ed.) Proceedings of the XI International Congress of Genetics. Pergamon, Oxford, U.K. (1964) 495-505.

[2] Kimura, M. *A stochastic model concerning the maintenance of genetic variability in quantitative characters*. Proc. Natl. Acad. Sci. USA (1965) 54: 731-736.

[3] Lynch, M. and Walsh B. *Genetics and Analysis of Quantitative Traits. Sinauer, Sunderland, MA*. (1998)

[4] Lande R. *The maintenance of genetic variability by mutation in a polygenic character with linked loci*. Genet. Res. (1975) 26: 221-236.

[5] Turelli M. *Heritable genetic variation via mutation-selection-balance: Lerch's zeta meets the abdominal bristle.* Theor. Popul. Biol. (1984) 25: 138-193.

[6] Waxman D. and Peck J. R. The Anomalous Effects of Biased Mutation. Genetics (2003) 164: 1615–1626

[7] Waxman, D. , *Numerical and Exact Solutions for Continuum of Alleles Models*. J. Math. Biol. (2003) 46: 225-240

[8] Welch J. and Waxman D. Non-Equivalent Loci and the Distribution of Mutant Effects. Genetics (2002) 161: 897-904

[9] Turelli M., and Barton N. H. *Dynamics of polygenic characters under selection.* Theor. Popul. Biol. (1990) 38: 1-57.