Polygenic selection to a changing optimum under self-fertilisation

S2 File: Dynamics at a locus underlying a quantitative trait under partial selfing

Matthew Hartfield and Sylvain Glémin

In this notebook we analyse the dynamics of a single (focal) locus underlying a quantitative trait under partial selfing.

General model

We consider a quantitative trait affected by \( n \) bi-allelic loci with additive effect:

- \( A_i A_i : 2 \alpha_i \)
- \( A_i a_i : \alpha_i \)
- \( a_i a_i : 0 \)

We note \( x_i \) the frequency of allele \( A_i \)

The mean phenotypic effect of locus \( i \) is \( 2 \alpha_i x_i \)

In what follows we consider a single focal locus and the \( n - 1 \) other loci are considered as the background.

The mean phenotypic effect of the background is thus \( z_i = z - 2 \alpha_i x_i \)

If we neglect associations, the background is the same for the two alleles, hence the phenotype of the three genotypes are:

- \( A_i A_i : 2 \alpha_i + z - 2 \alpha_i x_i \)
- \( A_i a_i : \alpha_i + z - 2 \alpha_i x_i \)
- \( a_i a_i : z - 2 \alpha_i x_i \)

In the more general case we can note that allele \( A_i \) is associated with the background \( z_i + \beta_{A,i} \) and \( \alpha_i \) with the background \( z_i + \beta_{a,i} \)

As we have \( \beta_{A,i} x_i + \beta_{a,i}(1 - x_i) = 0 \).

We can thus write:

\[ \beta_{A,i} = \beta_i (1 - x_i) \]
\[ \beta_{a,i} = -\beta_i x_i \]
So the mean phenotypic effect of the three genotypes can be written as:

\[ A_i A_i: 2 \alpha_i - 2 \alpha_i x_i + 2 \beta_i (1 - x_i) \]
\[ A_i a_i: \alpha_i - 2 \alpha_i x_i + \beta_i (1 - x_i) - \beta_i x_i \]
\[ a_i a_i: -2 \alpha_i x_i - 2 \beta_i x_i \]

Then we can apply the classical gaussian selection model.

We note \( V_s = \omega^2 + V_e \), with \( \omega \) is the standard deviation of the gaussian fitness function and \( V_e \) the environmental variance.

In what follows we remove the subscript \( i \) for simplicity.

\[
\begin{align*}
\text{In[1]:} & \quad z_{AA} = 2 \alpha + (z - 2 \alpha x) + 2 \beta (1 - x) ; \\
z_{Aa} & = \alpha + (z - 2 \alpha x) + \beta (1 - x) - \beta x ; \\
z_{aa} & = 0 + (z - 2 \alpha x) - 2 \beta x ; \\
z_m & = \text{Simplify} \left[ x^2 + F x (1 - x) \right] z_{AA} + 2 x (1 - x) (1 - F) z_{Aa} + \left( (1 - x)^2 + F x (1 - x) \right) z_{aa} ; \\
w[z_] & = e^{\frac{z - z_{AA}}{2 V_s}} ; \\
W_m & = \text{Simplify} \left[ \left( x^2 + F x (1 - x) \right) w[z_{AA}] + 2 x (1 - x) (1 - F) w[z_{Aa}] + \left( (1 - x)^2 + F x (1 - x) \right) w[z_{aa}] \right] ; 
\end{align*}
\]

The change in allele frequency due to selection is given by:

\[
\begin{align*}
\text{In[1]:} & \quad dx = \text{Simplify} \left[ \left( x^2 + F x (1 - x) \right) w[z_{AA}] + x (1 - x) (1 - F) w[z_{Aa}] \right] \left( \frac{-x}{W_m} \right) \\
\text{Out[1]:} & \quad \frac{-x + \left( e^{\frac{(z - z_{AA}) (2 \alpha_i - 2 \alpha_i x_i) + (1 - x) (1 - F) x + e^{\frac{(z - z_{AA}) (2 \alpha_i - 2 \alpha_i x_i) + (1 - x) (1 - F) x}}{2 V_s}} (F + x - F x) \right)}{2 e^{\frac{(z - z_{AA}) (2 \alpha_i - 2 \alpha_i x_i) + (1 - x) (1 - F) x - e^{\frac{(z - z_{AA}) (2 \alpha_i - 2 \alpha_i x_i) + (1 - x) (1 - F) x}}{2 V_s}}} (1 + (1 - F) x) + e^{\frac{(z - z_{AA}) (2 \alpha_i - 2 \alpha_i x_i) + (1 - x) (1 - F) x}} (F + x - F x) \right)}{2 e^{\frac{(z - z_{AA}) (2 \alpha_i - 2 \alpha_i x_i) + (1 - x) (1 - F) x - e^{\frac{(z - z_{AA}) (2 \alpha_i - 2 \alpha_i x_i) + (1 - x) (1 - F) x}}{2 V_s}}} (1 + (1 - F) x) + e^{\frac{(z - z_{AA}) (2 \alpha_i - 2 \alpha_i x_i) + (1 - x) (1 - F) x}} (F + x - F x) \right)}{1} \\
\text{Out[1]:} & \quad \Delta x[F_m, D^0, -\alpha, -\beta, -\alpha, -\beta, z, x] = \text{Simplify} \left[ dx / . (z \to z_0 - D) . (D \to D^0 \text{Sqrt}[Vs], \beta \to \beta \text{Sqrt}[Vs], \alpha \to \alpha \text{Sqrt}[Vs]) , Vs > 0 \right] \\
\text{Out[1]:} & \quad \frac{-x + \left( e^{\frac{1}{2} (D^0 + (1 - 2 \alpha) (\alpha^2 + \beta^2))} (1 - F) x + e^{\frac{1}{2} (D^0 + (1 - 2 \alpha) (\alpha^2 + \beta^2))} (F + x - F x) \right)}{2 e^{\frac{1}{2} (D^0 + (1 - 2 \alpha) (\alpha^2 + \beta^2))} (1 - F) x - e^{\frac{1}{2} (D^0 + (1 - 2 \alpha) (\alpha^2 + \beta^2))} (F + x - F x) \right)}}{2 e^{\frac{1}{2} (D^0 + (1 - 2 \alpha) (\alpha^2 + \beta^2))} (1 - F) x - e^{\frac{1}{2} (D^0 + (1 - 2 \alpha) (\alpha^2 + \beta^2))} (F + x - F x) \right)}}{1} 
\end{align*}
\]

**Equilibrium (\( D = 0 \))**

**Allele frequency**

We assume weak selection, so \( \alpha^0, \beta^0 \ll 1 \). We thus obtain a simple expression for the change in allele frequency:
\textbf{Out[1]} = Δselθ[F\_\_\_, α\_\_, β\_\_, x\_\_] = 
Simplify[Normal[Series[Δ\[Delta][F, θ, α, β, ζ, x], {ζ, 0, 2}]] / . ζ -> 1]

\textbf{Out[2]} = \frac{1}{2} \left(1 + \frac{3 F}{2} \right) (1 - 3 x + 2 x^2) (α + β)^2

We also consider symmetrical mutation

\textbf{Out[3]} = Δmut[u\_\_, x\_\_] = Simplify[-u x + u (1 - x)];

So we can get the frequency at equilibrium:

\textbf{Out[4]} = xeq[F\_\_\_, α\_\_, β\_\_, u\_\_] = Simplify[x /. \text{Solve}[Δselθ[F, α, β, x] + Δmut[u, x] = 0, x]]

\textbf{Out[5]} = \left\{ \begin{array}{l}
\frac{1}{2} \frac{α^2 + 3 F α + β + 3 F β - √{1 + 3 F} √{-8 u + (1 + 3 F) (α + β)^2}}{2 (1 + 3 F) (α + β)^2} \\
\frac{α^2 + 3 F α + β + 3 F β + √{1 + 3 F} √{-8 u + (1 + 3 F) (α + β)^2}}{2 (1 + 3 F) (α + β)^2}
\end{array} \right.

And for low mutation rate

\textbf{Out[7]} = \text{Simplify}[Series[\text{seq}[F, α, β, u], \{u, 0, 1\}], \{α + β > 0\}] \\
\text{Simplify}[Series[\text{seq}[F, α, β, u], \{u, 0, 1\}], \{α + β < 0\}]

\textbf{Out[8]} = \left\{ \begin{array}{l}
\frac{1}{2} \frac{2 u}{(1 + 3 F) (α + β)^2} + O[u]^2, 1 - \frac{2 u}{(1 + 3 F) (α + β)^2} + O[u]^2
\end{array} \right.

\textbf{Out[9]} = \left\{ \begin{array}{l}
\frac{1}{2} \frac{2 u}{(1 + 3 F) (α + β)^2} + O[u]^2, \frac{2 u}{(1 + 3 F) (α + β)^2} + O[u]^2
\end{array} \right.

Which we can write with the initial, unscaled, parameters.

\textbf{Out[11]} = \text{FullSimplify}\left[\left\{ \frac{2 u}{(1 + 3 F) (α + β)^2}, 1 - \frac{2 u}{(1 + 3 F) (α + β)^2} \right\} / . \left\{ \begin{array}{l}
α^\Rightarrow α / \text{Sqrt}[Vs], β^\Rightarrow β / \text{Sqrt}[Vs], Vs > 0
\end{array} \right.\right]

\textbf{Out[12]} = \left\{ \begin{array}{l}
\frac{2 u Vs}{(1 + 3 F) (α + β)^2}, 1 - \frac{2 u Vs}{(1 + 3 F) (α + β)^2}
\end{array} \right.

There are two symmetrical equilibria, close to 0 or 1, and one central equilibrium at 1/2. The stability of equilibria depends on the sign of the derivative of Δx:

\textbf{Out[14]} = \text{FullSimplify}[\text{D}[Δselθ[F, α, β, x] + Δmut[u, x], x] / . x -> xeq[F, α, β, u]]

\textbf{Out[15]} = \left\{ \begin{array}{l}
-2 u + \frac{1}{4} (1 + 3 F) (α + β)^2, 4 u - \frac{1}{2} (1 + 3 F) (α + β)^2, 4 u - \frac{1}{2} (1 + 3 F) (α + β)^2
\end{array} \right.

The 1/2 equilibrium is stable only if selection is weak compared to mutation, that is:

(β + α)^2 < \frac{8 u}{1 + 3 F}

Or with initial parameters

(β + α)^2 < \frac{8 u Vs}{1 + 3 F}
When $\beta = 0$ we retrieve the result of [1].
As stabilizing selection generates negative LD (see also the two-locus model), $\alpha$ and $\beta$ have opposite sign, so genetic associations leads to higher equilibrium frequencies than predicted by single locus theory.

### Variance components

#### Derivations

We can also obtain the different variance component at equilibrium.
According to [2], the total genetic variance can be decomposed into:
- The genic variance, $V_g$, corresponding to Hardy–Weinberg expectations and no linkage
- The inbreeding variance, $V_I$, corresponding to the departure from Hardy-Weinberg expectations at each locus
- The linkage covariance, $C_{LD}$, which takes interactions among loci into account

The genic and inbreeding variance are readily obtained from the sum of each locus contribution:

\[
V_g = \sum_{i} \left( 2 \alpha_i^2 + 2 \alpha_i (1-x_i) \right)
\]

\[
V_I = \sum_{i} \left( 2 \alpha_i (1-x_i) \right)
\]

Which can be evaluated at equilibrium and assuming mutation rate is small.

For low frequency equilibrium:

\[
V_g(0) = \frac{4 u V_s \alpha^2}{(1 + 3 F)(\alpha + \beta)^2}
\]

\[
V_I(0) = \frac{4 F u V_s \alpha^2}{(1 + 3 F)(\alpha + \beta)^2}
\]

For intermediate frequency equilibrium ($1/2$):

\[
V_g(1/2) = \frac{4 u V_s \alpha^2}{(1 + 3 F)(\alpha + \beta)^2}
\]

\[
V_I(1/2) = \frac{4 F u V_s \alpha^2}{(1 + 3 F)(\alpha + \beta)^2}
\]
$$\begin{align*}
&\text{Cld} \{F_{\alpha}, \alpha, \beta_1, x_1\} = \text{Simplify}\left[\left(x^2 + F \times (1-x)\right) (2 \alpha) (2 \beta (1-x)) + 2 x (1-x) (1-F) (\alpha) (\beta (1-x) - \beta x) + (1-x)^2 + F \times (1-x)\right] (0) (-2 \beta x) \\
&\text{Out}\{1\} = -2 (1 + F) (-1 + x) \times \alpha \beta
\end{align*}$$

If there is no locus at intermediate equilibrium:

$$
\begin{align*}
V_g &= \sum_{i=1}^{n} \frac{4 U V_i}{1 + 3 F} \frac{\alpha_i}{(\alpha_i + \beta_i)^2} \\
V_l &= \sum_{i=1}^{n} \frac{4 F U V_i}{1 + 3 F} \frac{\alpha_i}{(\alpha_i + \beta_i)^2}
\end{align*}
$$

Further assuming that mutation rates and phenotypic effects are independent (or constant mutation rate) then:

$$
\begin{align*}
V_g &= \frac{4 UV_s}{1 + 3 F} \frac{1}{n} \sum_{i=1}^{n} \frac{\alpha_i}{(\alpha_i + \beta_i)^2} \\
V_l &= \frac{4 FU V_s}{1 + 3 F} \frac{1}{n} \sum_{i=1}^{n} \frac{\alpha_i}{(\alpha_i + \beta_i)^2}
\end{align*}
$$

The covariance due to LD can also be computed as the covariance of phenotypic effect of the locus and the background.

Adapting [2], we have:

$$C_{LD} = \sum_{i=1}^{n} \sum_{j=1}^{n} \text{Cov}(i, j) \alpha_i \alpha_j$$

where $\text{Cov}(i, j)$ is the covariance between the number of allele $A$ at locus $i$ and $j$. Note that [2] considered that alleles contribute either 0 or 1 to the trait.

Note that $\text{Cov}(i, j) \alpha_i \alpha_j$ can be written as:

$$
\begin{align*}
2 \alpha_i f(A_i A_j) (f(A_i A_j | A_i A_j) 2 \alpha_i + f(A_i A_j | A_i A_j) \alpha_j + f(a_i a_j | A_i A_j) 0) + \\
\alpha_i f(A_i) (f(A_i A_j | A_i A_j) 2 \alpha_i + f(A_i A_j | A_i A_j) \alpha_j + f(a_i a_j | A_i A_j) 0) + \\
0 f(a_j a_j) (f(A_i A_j | a_i a_j) 2 \alpha_i + f(A_i A_j | a_i a_j) \alpha_j + f(a_i a_j | a_i a_j) 0)
\end{align*}
$$

where $f(G_j | G_i)$ means the frequency of genotype $G_j$ knowing genotype $G_i$. [1] expressed these quantities with a set of genetic associations, for which they seek for a set of recursion equations.

With our formalism, they are encapsulated in the $\beta$ parameters such that:

$$
\begin{align*}
2 \beta_{A_i} &= 2 \beta (1-x) = \sum_{j=1}^{n} (f(A_i A_j | A_i A_j) 2 \alpha_j + f(A_i A_j | A_i A_j) \alpha_j + f(a_i a_j | A_i A_j) 0) \\
\beta_{A_i} + \beta_{A_j} &= \beta (1-2 x) = \sum_{j=1}^{n} (f(A_i A_j | A_i a_j) 2 \alpha_j + f(A_i A_j | A_i a_j) \alpha_j + f(a_i a_j | A_i a_j) 0) \\
2 \beta_{A_j} &= -2 \beta x = \sum_{j=1}^{n} (f(A_i A_j | a_i a_j) 2 \alpha_j + f(A_i a_j | a_i a_j) \alpha_j + f(a_i a_j | a_i a_j) 0)
\end{align*}
$$

So we can write:

$$C_{LD} = \sum_{i=1}^{n} \sum_{j=1}^{n} (f(A_i A_j) 4 \alpha_i \beta_j (1-x_i) + f(A_i a_j) 2 \alpha_i \beta_j (1-2x_i) + f(a_i a_j) 0 \ast (-2 \beta x_i))$$

For one locus:

$$\begin{align*}
&\text{Cl}_i \{F, \alpha, \beta, x\} = \text{Simplify}\left[\left(x^2 + F \times (1-x)\right) (2 \alpha) (2 \beta (1-x)) + 2 x (1-x) (1-F) (\alpha) (\beta (1-x) - \beta x) + (1-x)^2 + F \times (1-x)\right] (0) (-2 \beta x)
\end{align*}$$

$$\begin{align*}
&\text{Out}\{1\} = -2 (1 + F) (-1 + x) \times \alpha \beta
\end{align*}$$
In[1]:= \textbf{Simplify}\left[\textbf{Normal}\left[\textbf{Series}\left[\textbf{Cld}\left[F, \alpha, \beta, \frac{2 \ u \ Vs}{(1 + 3 \ F) \ (\alpha + \beta)^2}\right], \{u, 0, 1\}\right]\right]\right]

Out[1]= \frac{4 \ (1 + F) \ u \ Vs \ \alpha \ \beta}{(1 + 3 \ F) \ (\alpha + \beta)^2}

Summing over the $n$ loci:

$$C_{LD} = \frac{4(1 + F) \ U \ V_s}{1 + 3F} \frac{1}{n} \sum_{i=1}^{n} \frac{\alpha_i \ \beta_i}{(\alpha_i + \beta_i)^2}$$

Finally, by summing the three components we obtain the total genetic variance. For one locus:

In[1]:= \textbf{VG} = \frac{4 \ u \ Vs}{(1 + 3 \ F) \ (\alpha + \beta)^2}

Out[1]= \frac{4 \ (1 + F) \ u \ Vs \ \alpha}{(1 + 3 \ F) \ (\alpha + \beta)}

And summed over all loci:

$$V_G = \frac{4(1 + F) \ U \ V_s}{1 + 3F} \frac{1}{n} \sum_{i=1}^{n} \frac{\alpha_i}{\alpha_i + \beta_i}$$

Note that because the total variance must be positive, $\frac{\alpha_i}{\alpha_i + \beta_i} > 0$. As $\alpha_i$ and $\beta_i$ are opposite sign it implies that $|\beta_i| < |\alpha_i|$

**Interpretation**

The expressions for variance components are not closed but they provide useful insight on the effect of genetic associations, encapsulated in $\beta$s, on variance decomposition.

- When $\beta_i = 0$ we retrieve the house of card expression with selfing as in [1] with $C_{LD} = 0$
- As $\alpha_i$ and $\beta_i$ are of opposite sign, associations increase the genetic and inbreeding variance (because selection is reduced) and make $C_{LD}$ more negative. The overall effect is to increase genetic variance compared to no associations
- Selfing has two opposite effects: increasing purging, which reduces the genetic variance but increasing associations, which increased the genetic variance. However, we can’t determine the net effect, which would require to express $\sum_{i=1}^{n} \frac{\alpha_i}{\alpha_i + \beta_i}$ as a function of $F$

---

**Shift to a new optimum**

In this part we assume that the population is away from equilibrium. Without loss of generality we assume that $d_0 > 0$.

**Change in allele frequency**

As above we assume weak selection, so $\gamma^*, \beta^* \ll 1$. We also assume that $D^* \ll 1$, which means that
the shift to the new optimum didn’t generate a strong drop in fitness.

\[
\Delta x_{sel}[F_-, D^o_-, \alpha^o_-, \beta^o_-, x_n] =
\text{FullSimplify}[\text{Normal}[\text{Series}[\Delta x[F, D^o, \alpha^o, \beta^o, x], \{x, 0, 2\}]] / \xi \to 1]
\]

\[
\frac{1}{2} (-1 + x) \times (\alpha^o + \beta^o)

\bigl(-2 (1 + F) D^o - (1 + 3 F) (-1 + 2 x) (\alpha^o + \beta^o) + (1 + 3 F) (-1 + 2 x) D^{o2} (\alpha^o + \beta^o)\bigr)
\]

It can be re-written as the sum of two terms. The first term corresponds to pure directional selection and the second to stabilising selection

\[
\text{Simplify}\left[\frac{1}{2} (-1 + x) \times (\alpha^o + \beta^o) (-2 D^o (1 + F)) \right] + \text{Simplify}\left[\frac{1}{2} (-1 + x) \times (\beta^o + \alpha^o) \left((-1 + 3 F) (-1 + 2 x) (\beta^o + \alpha^o) + D^{o2} (1 + 3 F) (-1 + 2 x) (\beta^o + \alpha^o)\right)\right]
\]

\[
\frac{1}{2} (-1 + x) \times (\beta^o + \alpha^o) \left((-1 + 3 F) (-1 + 2 x) (\beta^o + \alpha^o) + D^{o2} (1 + 3 F) (-1 + 2 x) (\beta^o + \alpha^o)\right)

\bigl(-1 + F\bigr) \times D^o (\alpha^o + \beta^o) + \frac{1}{2} \left(1 + 3 F\right) \times (\alpha^o + \beta^o)^2
\]

This can be rewritten with the unscaled parameters

\[
\bigl\{\alpha^o \to \frac{\alpha}{\text{Sqrt}[Vs]}, \beta^o \to \frac{\beta}{\text{Sqrt}[Vs]}, D^o \to \frac{D}{\text{Sqrt}[Vs]}\bigr\}, \text{Vs} > 0\}

\text{Simplify}\left[\frac{1}{2} (-1 + x) \times (\beta^o + \alpha^o) \left((-1 + 3 F) (-1 + 2 x) (\beta^o + \alpha^o) + D^{o2} (1 + 3 F) (-1 + 2 x) (\beta^o + \alpha^o)\right)\right] / \text{Vs}
\]

\[
\text{Simplify}\left[\frac{1}{2} (1 + F) \times D (\alpha + \beta)\right] - \frac{\left(1 + 3 F\right) (1 - x) \times (\alpha^o + \beta^o) + D^{o2} (1 + 3 F) (1 - x) (\alpha^o + \beta^o)^2}{2 \text{Vs}^2}
\]

If we neglect LD we obtain an expression equivalent to the one of Hayward and Sella 2022 with the additional effect of selling

\[
\% / \beta \to 0
\]

\[
\frac{(1 + F) (-1 + x) \times D \alpha}{\text{Vs}} - \frac{(1 + 3 F) (-1 + x) \times (\text{Vs} - D^2) (\alpha + \beta)^2}{2 \text{Vs}^2}
\]

Change in the phenotype

The phenotypic change can be expressed in term of the distance to the optimum. Following Hayward and Sella 2022, we can write:

\[
E[\Delta D] = - \sum_{i=1}^n E[2 \alpha_i \Delta x_i]
\]

\[
= - \frac{D}{\text{Vs}} \sum_{i=1}^n \left(2 \alpha_i^2 (1 + F) x_i (1 - x_i) + 2 \alpha_i \beta_i (1 + F) (1 - x_i) x_i + (\text{Vs} - D^2) (1 + 3 F) \sum_{i=1}^n x_i (1 - x_i) \left(\frac{1}{2} - x_i\right) \alpha_i (\alpha_i + \beta_i)^2\right)
\]

The first term we recognise the expressions for the variance components. In the second term, we
need to introduce notations for third order moments

\[ \mu_3 = \sum_{i=1}^{n} x_i (1 - x_i) \left( \frac{1}{2} - x_i \right) \alpha_i^3 \]

\[ v_{2,1} = \sum_{i=1}^{n} x_i (1 - x_i) \left( \frac{1}{2} - x_i \right) \alpha_i^2 \beta_i \]

\[ v_{1,2} = \sum_{i=1}^{n} x_i (1 - x_i) \left( \frac{1}{2} - x_i \right) \alpha_i \beta_i \]

This leads to the expression in the main text

\[ E[\Delta \mathcal{D}] = - \sum_{i=1}^{n} E[2 \alpha_i \Delta x_i] \]

\[ = - \frac{\mathcal{D}}{Vs} \sum_{i=1}^{n} (1 + F) V_g + C_{LD} \frac{\left( Vs - \mathcal{D}^2 \right)}{Vs^2} (1 + 3 F) (\mu_3 + 2 v_{2,1} + v_{1,2}) \]

References
