**SUPPLEMENTAL MATERIAL: FILE S2**

**Diffusion approximations for Wright-Fisher models with selection**

We present the formulation of the diffusion approximation for the Wright-Fisher model with selection here and refer to Durrett (2008) for a rigorous proof, especially for one- and two-locus population dynamics. For simplicity of notation, we introduce a short-hand notation for the conditional expectation of a random variable given the population of the haplotype frequencies $x$ in generation $k$

$$E^{(k,x)}(\cdot) = E\left(\cdot \mid X^{(N)}(k) = x\right),$$

and a function $\rho_{uv}^i$ of three variables $u$, $v$ and $i$, defined as

$$\rho_{uv}^i = \frac{1}{2}(\delta_{ui} + \delta_{vi}),$$

where $\delta_{ui}$ and $\delta_{vi}$ are the Kronecker delta functions.

Let $\Delta X_i^{(N)}(k)$ denote the change in the frequency of haplotype $i$ from generation $k$ to the next, and we have

$$E^{(k,x)}(\Delta X_i^{(N)}(k)) = \sum_{u,v} \rho_{uv}^i E^{(k,y)}(\Delta Y_{uv}^{(N)}(k))$$

$$E^{(k,x)}(\Delta X_i^{(N)}(k)\Delta X_j^{(N)}(k)) = \sum_{u,v} \sum_{m,n} \rho_{uv}^i \rho_{mn}^j E^{(k,y)}(\Delta Y_{uv}^{(N)}(k)\Delta Y_{mn}^{(N)}(k)).$$

Based on the construction of the Wright-Fisher model with selection, we have

$$E^{(k,y)}(\Delta Y_{uv}^{(N)}(k)) = q_{uv} - y_{uv}$$

$$E^{(k,y)}(\Delta Y_{uv}^{(N)}(k)\Delta Y_{mn}^{(N)}(k)) = \frac{1}{N} q_{uv}(\delta_{um}\delta_{vn} - q_{mn}) + (q_{uv} - y_{uv})(q_{mn} - y_{mn}).$$

Substituting Eqs. (3) and (4) into Eqs. (1) and (2), respectively, we have

$$E^{(k,x)}(\Delta X_i^{(N)}(k)) = p_i - x_i$$

$$E^{(k,x)}(\Delta X_i^{(N)}(k)\Delta X_j^{(N)}(k)) = \frac{1}{2N} p_i(\delta_{ij} - p_j) + (p_i - x_i)(p_j - x_j),$$

where $q_{uv}$ is the frequency of genotype made up of haplotype $u$ and $v$ of an effectively infinite population after the possible mechanisms of evolutionary change (except population regulation) at intermediate stages of the life cycle such as natural selection, meiosis and random mating within generation $k$. Substituting Eqs. (3) and (4) into Eqs. (1) and (2), respectively, we have
where
\[ p_i = \sum_{u,v} p_{uv} q_{uv} \]  
(5)
is the frequency of haplotype \( i \) of an effectively infinite population after the possible mechanisms of evolutionary change (except population regulation) at intermediate stages of the life cycle such as natural selection, meiosis and random mating within generation \( k \).

Considering the limits as the population size \( N \) goes to infinity, we can formulate the infinitesimal mean vector \( \mu(t, x) \) as
\[ \mu_i(t, x) = \lim_{N \to \infty} 2N E^{(2Nt, x)} \left( \Delta X_i^{(N)}([2Nt]) \right) \]
= \[ 2N(p_i - x_i) \]
and the infinitesimal covariance matrix \( \Sigma(t, x) \) as
\[ \Sigma_{ij}(t, x) = \lim_{N \to \infty} 2N E^{(2Nt, x)} \left( \Delta X_i^{(N)}([2Nt]) \Delta X_j^{(N)}([2Nt]) \right) \]
= \[ \lim_{N \to \infty} p_i(\delta_{ij} - p_j) + 2N(p_i - x_i)(p_j - x_j), \]
where \([\cdot]\) is used to denote the integer part of the value in the brackets, according to standard techniques of diffusion theory (see, for example, Karlin and Taylor, 1981).

Therefore, the process \( X^{(N)} \) converges to a diffusion process, denoted by \( X = \{ X(t), t \geq 0 \} \) and referred to as the Wright-Fisher diffusion with selection, satisfying the stochastic differential equation of the form
\[ dX(t) = \mu(t, X(t)) dt + \sigma(t, X(t)) dW(t), \]
where the diffusion coefficient matrix \( \sigma(x) \) satisfies the relation that
\[ \sigma(t, x) \sigma^T(t, x) = \Sigma(t, x) \]
and \( W(t) \) is a multi-dimensional standard Brownian motion.

**One-locus population dynamics.** For the population evolving at a single locus, we use a common category of fitness values for a diploid population at a single locus (see Hamilton, 2011, for other categories of fitness values presented in terms of selection coefficients), which can be presented as follows: genotypes \( A_1A_1, A_1A_2 \) and \( A_2A_2 \) at a given locus \( A \) have fitness values 1, \( 1 - h_A s_A \) and \( 1 - s_A \), respectively, where \( s_A \) is the selection coefficient and \( h_A \) is the dominance parameter. Substituting the sampling probabilities \( q_{uv}^{(v)} \) and \( q_{uv}^{(f)} \) for \( u, v = 1, 2 \) into Eq. (5), respectively, we have
\[ p_1^{(v)} = \frac{x_1^2 + (1 - h_A s_A)x_1x_2}{1 - (2h_Ax_1x_2 - x_2^2)s_A} \]
\[ p_2^{(v)} = \frac{(1 - h_A s_A)x_1x_2 + (1 - s_A)x_2^2}{1 - (2h_Ax_1x_2 - x_2^2)s_A} \]
and

\[ p_1^{(f)} = \frac{x_1^2 + (1 - h_A s_A)x_1 x_2}{1 - (2h_A x_1 x_2 - x_2^2)s_A} \]

\[ p_2^{(f)} = \frac{(1 - h_A s_A)x_1 x_2 + (1 - s_A)x_2^2}{1 - (2h_A x_1 x_2 - x_2^2)s_A} \]

That is, we have

\[ \left| p_i^{(v)} - p_i^{(f)} \right| = 0 \]

for \( i = 1, 2 \), which implies that we have the same representation of the infinitesimal mean vectors and covariance matrices for the cases that natural selection takes the form of viability or fecundity selection,

\[ \mu_1(x) = \alpha_A x_1 x_2[x_1 h_A + x_2(1 - h_A)] \]

\[ \mu_2(x) = -\alpha_A x_1 x_2[x_1 h_A + x_2(1 - h_A)] \] (6)

and

\[ \Sigma_{ij}(x) = x_i(\delta_{ij} - x_j) \] (7)

for \( i, j = 1, 2 \). Eqs. (6) and (7) ensure the same SDE-representation of the one-locus Wright-Fisher diffusion with selection for the two types of natural selection, viability and fecundity selection.

**Two-locus population dynamics.** For the population evolving at two linked loci, we use a same category of fitness values for a diploid population at a single locus and assume that the fitness values of two-locus genotypes are determined multiplicatively from fitness values at individual loci, e.g., the fitness value of the \( \text{A}_1\text{B}_2/\text{A}_2\text{B}_2 \) genotype is \((1 - h_A s_A)(1 - s_B)\). We let \( r \) denote the recombination rate between the two loci (i.e., the rate that a recombinant gamete is produced at meiosis). Substituting the sampling probabilities \( q_u^{(v)} \) and \( q_u^{(f)} \) for \( u, v = 1, 2, 3, 4 \) into Eq. (5), respectively, and sing Taylor expansions with respect to the selection coefficients \( s_A \) and \( s_B \) and the recombination rate \( r \), we have

\[ p_1^{(v)} = x_1 - (x_1 x_4 - x_2 x_3)r + d_A(x_1(x_3 + x_4) - (x_3 + x_4)(x_1 x_4 - x_2 x_3))s_A \]

\[ + d_B(x_1(x_2 + x_4) - (x_2 + x_4)(x_1 x_4 - x_2 x_3))s_B + \mathcal{O}(s_A^2 + s_A s_B + s_B^2 + r^2) \]

\[ p_2^{(v)} = x_2 + (x_1 x_4 - x_2 x_3)r + d_A(x_2(x_3 + x_4) + (x_3 + x_4)(x_1 x_4 - x_2 x_3))s_A \]

\[ - d_B(x_2(x_1 + x_3) + (x_1 + x_3)(x_1 x_4 - x_2 x_3))s_B + \mathcal{O}(s_A^2 + s_A s_B + s_B^2 + r^2) \]

\[ p_3^{(v)} = x_3 + (x_1 x_4 - x_2 x_3)r - d_A(x_3(x_1 + x_2) + (x_1 + x_2)(x_1 x_4 - x_2 x_3))s_A \]

\[ + d_B(x_3(x_2 + x_4) + (x_2 + x_4)(x_1 x_4 - x_2 x_3))s_B + \mathcal{O}(s_A^2 + s_A s_B + s_B^2 + r^2) \]

\[ p_4^{(v)} = x_4 - (x_1 x_4 - x_2 x_3)r - d_A(x_4(x_1 + x_2) - (x_1 + x_2)(x_1 x_4 - x_2 x_3))s_A \]

\[ - d_B(x_4(x_1 + x_3) - (x_1 + x_3)(x_1 x_4 - x_2 x_3))s_B + \mathcal{O}(s_A^2 + s_A s_B + s_B^2 + r^2) \]
and

\[
p_1^{(f)} = x_1 - (x_1x_4 - x_2x_3)r + d_A(x_1(x_3 + x_4) + (x_1 + x_2 - x_3 - x_4)(x_1x_4 - x_2x_3)r)s_A \\
+ d_B(x_1(x_2 + x_4) + (x_1 - x_2 + x_3 - x_4)(x_1x_4 - x_2x_3)r)s_B + \mathcal{O}(s_A^2 + s_A s_B + s_B^2)
\]

\[
p_2^{(f)} = x_2 + (x_1x_4 - x_2x_3)r + d_A(x_2(x_3 + x_4) - (x_1 + x_2 - x_3 - x_4)(x_1x_4 - x_2x_3)r)s_A \\
- d_B(x_2(x_1 + x_3) + (x_1 - x_2 + x_3 - x_4)(x_1x_4 - x_2x_3)r)s_B + \mathcal{O}(s_A^2 + s_A s_B + s_B^2)
\]

\[
p_3^{(f)} = x_3 + (x_1x_4 - x_2x_3)r - d_A(x_3(x_1 + x_2) + (x_1 + x_2 - x_3 - x_4)(x_1x_4 - x_2x_3)r)s_A \\
+ d_B(x_3(x_2 + x_4) - (x_1 - x_2 + x_3 - x_4)(x_1x_4 - x_2x_3)r)s_B + \mathcal{O}(s_A^2 + s_A s_B + s_B^2)
\]

\[
p_4^{(f)} = x_4 - (x_1x_4 - x_2x_3)r - d_A(x_4(x_1 + x_2) - (x_1 + x_2 - x_3 - x_4)(x_1x_4 - x_2x_3)r)s_A \\
- d_B(x_4(x_1 + x_3) - (x_1 - x_2 + x_3 - x_4)(x_1x_4 - x_2x_3)r)s_B + \mathcal{O}(s_A^2 + s_A s_B + s_B^2),
\]

where

\[
d_A = (x_1 + x_2)h_A + (x_3 + x_4)(1 - h_A)
\]

\[
d_B = (x_1 + x_3)h_B + (x_2 + x_4)(1 - h_B).
\]

That is, we have

\[
\left| p_i^{(v)} - p_i^{(f)} \right| = \mathcal{O}\left( \frac{1}{N^2} \right)
\]

for \( i = 1, 2, 3, 4 \) under the assumption that the selection coefficients \( s_A \) and \( s_B \) and the recombination rate \( r \) are all of order \( \mathcal{O}(1/N) \), which implies that we have the same representation of the infinitesimal mean vectors and covariance matrices for the cases that natural selection takes the form of viability or fecundity selection,

\[
\mu_1(\mathbf{x}) = \alpha_A x_1(x_3 + x_4)((x_1 + x_2)h_A + (x_3 + x_4)(1 - h_A)) \\
+ \alpha_B x_1(x_2 + x_4)((x_1 + x_3)h_B + (x_2 + x_4)(1 - h_B)) - R(x_1x_4 - x_2x_3)
\]

\[
\mu_2(\mathbf{x}) = \alpha_A x_2(x_3 + x_4)((x_1 + x_2)h_A + (x_3 + x_4)(1 - h_A)) \\
- \alpha_B x_2(x_1 + x_3)((x_1 + x_3)h_B + (x_2 + x_4)(1 - h_B)) + R(x_1x_4 - x_2x_3)
\]

\[
\mu_3(\mathbf{x}) = -\alpha_A x_3(x_1 + x_2)((x_1 + x_2)h_A + (x_3 + x_4)(1 - h_A)) \\
+ \alpha_B x_3(x_2 + x_4)((x_1 + x_3)h_B + (x_2 + x_4)(1 - h_B)) + R(x_1x_4 - x_2x_3)
\]

\[
\mu_4(\mathbf{x}) = -\alpha_A x_4(x_1 + x_2)((x_1 + x_2)h_A + (x_3 + x_4)(1 - h_A)) \\
- \alpha_B x_4(x_1 + x_3)((x_1 + x_3)h_B + (x_2 + x_4)(1 - h_B)) - R(x_1x_4 - x_2x_3),
\]

and

\[
\Sigma_{ij}(\mathbf{x}) = x_i(\delta_{ij} - x_j)
\]

for \( i, j = 1, 2, 3, 4 \), where \( \alpha_A = 2Ns_A \), \( \alpha_B = 2Ns_B \) and \( R = 2Nr \) are the scaled selection coefficients and recombination rate, respectively. Eqs. (8) and (9) guarantee the same SDE-representation of
the two-locus Wright-Fisher diffusion with selection for the two types of natural selection, viability and fecundity selection.

Therefore, we can conclude that we have the same diffusion approximations for the Wright-Fisher models with selection for the population dynamics at a single locus or two linked loci no matter whether viability or fecundity selection is occurring.

REFERENCES

