Evolution in ecological context

IICII Summer School in Quantitative Biology
September 2024 - IFOM, Milan, Italy

Onofrio Mazzarisi and William R. Shoemaker

September 2024

Contents

1 Single-locus evolutionary dynamics

In this lecture, we build a single-locus model of evolution inspired by a typical laboratory setting to introduce basic concepts of evolutionary dynamics: genetic drift and natural selection. We conclude by deriving the diffusion limit of population genetics for the model and discussing its scope and properties.

This lecture is inspired by Prof. Benjamin H. Good's course on Quantitative Evolutionary Dynamics and Genomics at Stanford University [1].

1.1 An experimental setting to study evolution

The ideas we will develop in this lecture can be, in principle, applied to populations of organisms at any scale. However, we would like to base our theoretical exploration on experiments. Therefore, something small that reproduces fast would be the perfect subject. Bacteria check both boxes.

1.1.1 Batch culture

Take a test tube with ~ 1 ml of growth media containing water, a carbon source (e.g., some sugar), salts, vitamins, and other ingredients. Then, inoculate with N_0 cells of a laboratory strain of, say, E. Coli. Let it cook (i.e., wait) for a given time Δt (e.g., 24 hours), and you will find a number N_f of cells. We can also measure time in number of generations as $\Delta t = \log_2 (N_f/N_0)$.

To measure N_0 and N_f, we can look at the population's growth curve. One way to track the population is through optical density (OD) measurement. OD is a measure of turbidity of the test tube, usually achieved by light scattering using a laser at a wavelength of 600 nm, which does no or little damage to the cells.

We will also need to track different strains, eventually, a mutated one and the wild type, for example. OD will not be able to differentiate between the two; we need other techniques. For example, flow cytometry allows one to count cells one by one. We can insert a gene producing a fluorescent protein in one strain and differentiate between strains in this way.

1.1.2 Serial dilution

In order to track evolution, we would like to repeat the procedure in the previous section many times.

Start with N_0 cells and let them grow for a fixed time Δt. For our purpose, we will ignore the lag phase and consider Δt much smaller than the time at which the population exits the exponential growth phase (i.e., before the environment changes so much it affects growth). This is not necessarily possible in practice, but we can already get plenty of insights in this simplified scenario. At time t, the number of cells will be

$$N(t) = N_0 e^{rt}, \tag{1}$$

where r is the growth rate. The final number of cells will, therefore, be

$$N_f = N_0 e^{r\Delta t}. \tag{2}$$

The growth rate can be measured as

$$r = \frac{1}{\Delta t} \ln\left(\frac{N_f}{N_0}\right). \tag{3}$$

Now, we dilute the test tube with a dilution factor D. The dilution factor is a dimensionless number that represents the ratio of the initial volume V of the test tube's content over the volume collected V_D and combined with fresh media to fill a new test tube: $D = V/V_D$. A ten-fold dilution, for example, consists of collecting a tenth of old media and cells and combining it with nine parts of fresh media. We also have $D = N_f/N_0$.

We want to set the dilution factor in such a way that the *expected* number of cells in the fresh tube is $\bar{N_0}$. If we label dilution cycles by $k \equiv t/\Delta t$, we can write that, for every k,

$$N_0(k+1) \sim \text{Poisson}(\bar{N_0}), \tag{4}$$

i.e., the number of cells at each new cycle is Poisson distributed with average $\bar{N_0}$. The Poisson distribution emerges because we are sampling a volume V_D from a volume V where the cells are well mixed, i.e., uniformly distributed. This is the only element of randomness in this lecture, and it comes from sampling.

Serial dilution consists of repeating this growth and dilution procedure as many times as we need.

Chemostat Another classical experimental setting worth mentioning is the chemostat. This device allows cells to be maintained at a fixed growth rate continuously. Bacteria are contained in a vessel with a constant influx of nutrients, and cells and media are diluted at a fixed rate. Bacteria will reach a population size where the growth rate is precisely balanced by dilution. We will consider this setting in the next lecture.

1.2 A single-locus model of evolution

Imagine, now, introducing a mutated strain of E. Coli that differs from the wild type because of a missing gene. Let us call strain 1 the wild type and strain 2 the mutant. The total population at time t is

$$N(t) \equiv N_1(t) + N_2(t), \tag{5}$$

and we define the relative frequency of the mutant as

$$f(t) \equiv N_2(t)/N(t). \tag{6}$$

Suppose that the missing gene in the mutant allows the processing of a fancy sugar that is absent from the medium and, therefore, frees up resources for growth; then we have that strain 2 will grow faster

$$N_1(t) = N_1(0)e^{rt}, \tag{7}$$

$$N_2(t) = N_2(0)e^{(r+s)t}, \tag{8}$$

where $s > 0$ encodes the growth advantage of the mutant.

1.2.1 Allele frequency dynamics

How does the frequency evolve over time? If, at the beginning of the experiment, the frequency of the mutated allele is $f(0)$, we have

$$f(\Delta t) \equiv \frac{N_2(\Delta t)}{N_1(\Delta t) + N_2(\Delta t)} \tag{9}$$

$$= \frac{N_0 f(0) e^{(r+s)\Delta t}}{N_0(1 - f(0))e^{r\Delta t} + N_0 f(0) e^{(r+s)\Delta t}} \tag{10}$$

$$= \frac{f(0)e^{s\Delta t}}{1 - f(0) + f(0)e^{s\Delta t}}, \tag{11}$$

therefore, the number of cells transferred to the fresh tube is distributed as

$$N_2(k+1) \sim \text{Poisson}\left(\bar{N}_0 \frac{f(k)e^{s\Delta t}}{1 - f(k) + f(k)e^{s\Delta t}}\right), \tag{12}$$

$$N_1(k+1) \sim \text{Poisson}\left(\bar{N}_0 \frac{1 - f(k)}{1 - f(k) + f(k)e^{s\Delta t}}\right). \tag{13}$$

We can finally write the frequency at the beginning of a cycle in terms of the frequency at the beginning of the previous one

$$f(k+1) = \frac{N_2(k+1)}{N_1(k+1) + N_2(k+1)}. \tag{14}$$

We generate a sequence of frequencies at the beginning of each cycle that constitutes a Markov Process in this way (the probability of $f(k+1)$ only depends on $f(k)$).

1.2.2 Genetic drift

Let us consider the simple scenario in which the mutation is *neutral*, meaning $s = 0$.

In this case, the dynamics for the two populations are reduced to

$$N_2(k+1) \sim \text{Poisson}\left(\bar{N}_0 f(k)\right), \tag{15}$$

$$N_1(k+1) \sim \text{Poisson}\left(\bar{N}_0 [1 - f(k)]\right), \tag{16}$$

and we can derive some system properties. If the frequency at the beginning of cycle k is $f(k)$, we will have the expected frequency at $k+1$ being

$$\mathbb{E}[f(k+1)] = f(k), \tag{17}$$

because of the symmetry of the problem [1]. This implies that if we start with an initial mutant frequency $f(0)$,

$$\mathbb{E}[f(k)] = f(0), \quad \forall k, \tag{18}$$

Now, consider the case in which $\bar{N}_0 f(k) \gg 1$ and $\bar{N}_0[1-f(k)] \gg 1$ (so that this is valid also for the wild type), which can be combined in $\bar{N}_0 f(k)[1-f(k)] \gg 1$. The Poisson distribution approximates a normal distribution when the mean is $\gg 1$; therefore, we can write Eq. (14) as

$$f(k+1) = \frac{N_2(k+1)}{N_1(k+1)+N_2(k+1)} = \frac{N_2(k+1)}{N(k+1)} \approx \frac{f(k) + \sqrt{\frac{f(k)}{\bar{N}_0}}\xi_{N_2}}{1+\sqrt{\frac{1}{\bar{N}_0}}\xi_N}, \tag{19}$$

where we used the property that the sum of two Poisson-distributed random variables is a Poisson-distributed random variable with the mean equal to the sum of the means of the initial variables and ξ_{N_2} and ξ_N are Gaussian variables with zero mean and unit variance. Now we Taylor-expand to the first order for small $\sqrt{1/\bar{N}_0}$ and obtain

$$f(k+1) \approx f(k) + \sqrt{\frac{f(k)}{\bar{N}_0}}\xi_{N_2} - \sqrt{\frac{f(k)^2}{\bar{N}_0}}\xi_N = f(k) + \sqrt{\frac{f(k)(1-f(k))}{\bar{N}_0}}\xi, \tag{20}$$

where ξ is a new Gaussian variable with zero mean and unit variance, and we used the property that the sum of Gaussian variables is a Gaussian variable.

We found that $f(k+1)$ can be approximated by a Gaussian with mean $f(k)$ (consistent with our previous heuristic argument) and standard deviation of order $\sqrt{1/\bar{N}_0}$

$$f(k+1) \approx f(k) + \mathcal{O}\left(\sqrt{\frac{1}{\bar{N}_0}}\right)\xi. \tag{21}$$

The fluctuations around the mean constitute the genetic drift. They are small! For example, for $\bar{N}_0 \sim 10^5$ we have $\sqrt{1/\bar{N}_0} \sim 0.3\%$. However, they can accumulate and lead to drastic scenarios. Indeed, $f=1$ and $f=0$ correspond to absorbing states, respectively, associated with the fixation and extinction of the mutant. After waiting long enough, the system will end in one of the two states, and the probability of fixation, $\mathcal{P}(f=1)$, can be computed by exploiting the fact that the average frequency remains constant

$$\mathbb{E}[f(k\to\infty)] = 0\mathcal{P}(f=0) + 1\mathcal{P}(f=1) = f(0), \tag{22}$$

[1] Notice that, formally, $f(k+1)$ is the ratio of two Poisson-distributed random variables, see Eq. (14), and therefore, we cannot obtain an analytical expression for it or even explicitly compute the average frequency.

therefore

$$\mathcal{P}(f=1) = f(0). \tag{23}$$

But how long is long enough? We have, under specific conditions that we will discuss better in the last section of this lecture,

$$f(k) \approx f(0) + \sum_{i=1}^{k} \mathcal{O}\left(\sqrt{\frac{1}{\bar{N}_0}}\right)\xi = f(0) + \mathcal{O}\left(\sqrt{\frac{k}{\bar{N}_0}}\right)\xi, \tag{24}$$

therefore, for $f(0) \sim 1/2$, we need $k \sim \bar{N}_0$. If a k is measured in days and we consider $\bar{N}_0 \sim 10^5$, this means that fixation and extinction start becoming possibilities after ~ 300 years! We will see that natural selection plays a more relevant role at lab scales in these conditions.

1.2.3 Natural selection

Consider now the case in which $s > 0$ and, for simplicity, $\bar{N}_0 \to \infty$, in order to ignore drift. We have in this case

$$f(k) = \frac{f(k-1)e^{s\Delta t}}{1 - f(k-1) + f(k-1)e^{s\Delta t}} \tag{25}$$

$$= \frac{f(k-2)e^{s2\Delta t}}{1 - f(k-2) + f(k-2)e^{s2\Delta t}} \tag{26}$$

$$= \frac{f(0)e^{sk\Delta t}}{1 - f(0) + f(0)e^{sk\Delta t}}. \tag{27}$$

In terms of generations $t \equiv k\Delta t$

$$f(t) = \frac{f(0)e^{st}}{1 - f(0) + f(0)e^{st}}, \tag{28}$$

which is the solution of the logistic equation

$$\frac{\partial f}{\partial t} = sf(1-f). \tag{29}$$

The time scale of natural selection is given by $t = 1/s$. For example, for $s = 0.01$, the time scale is $t = 100$ generations. Considering a 100-fold dilution factor, which implies $\Delta t = \log_2(100) \sim 6.6$ gen/day, this means that we see big changes after $k = t/(\Delta t) \sim 2$ weeks!

We can also use the definition of s to estimate it by measuring the initial frequency $f(0)$ and the frequency at time $f(t)$

$$s = \frac{1}{t}\ln\left(\frac{f(t)}{1-f(t)}\frac{1-f(0)}{f(0)}\right). \tag{30}$$

We can call s the *fitness difference*. Of course, in the case in which $s < 0$, the mutant will go extinct.

1.2.4 Spontaneous mutations

So far, we assumed to introduce a fraction of cells with mutations in a wild-type population and observe the effects of genetic drift and natural selection on allele frequency. We introduce now spontaneous, single-target mutation, which causes the loss of the gene and, therefore, causes the wild-type to mutate.

Let us assume that mutations appear with probability μ per division, with $\mu \ll 1$. Let us focus on the cycle in which the mutation first appears. To simplify the problem, let us assume that the mutation does not bring fitness benefits before the next cycle (reasonable). By definition, there are $\Delta t = \log_2(N_f/N_0)$ divisions in a cycle, so the probability that a cell at the previous cycle has acquired a mutation is

$$\mathcal{P}[\text{mutation}] = \mu \Delta t. \tag{31}$$

If we assume that each cell can acquire a mutation independently from the others, we have that

$$N_2(k+1) \sim \text{Poisson}(\bar{N_0}\mu\Delta t), \tag{32}$$

$$N_1(k+1) \sim \text{Poisson}(\bar{N_0}(1-\mu\Delta t)). \tag{33}$$

At this point, we can combine the dynamics we discussed for $f(0) > 0$ to obtain the full "microscopic" model of serial dilution, where we compute

$$N_2(k+1) \sim \text{Poisson}\left(\bar{N_0}\frac{f(k)e^{s\Delta t}}{1-f(k)+f(k)e^{s\Delta t}}\right) \tag{34}$$

$$+ \text{Poisson}\left(\bar{N_0}\mu\Delta t\frac{1-f(k)}{1-f(k)+f(k)e^{s\Delta t}}\right), \tag{35}$$

$$N_1(k+1) \sim \text{Poisson}\left(\bar{N_0}(1-\mu\Delta t)\frac{1-f(k)}{1-f(k)+f(k)e^{s\Delta t}}\right). \tag{36}$$

and then update the frequency

$$f(k+1) = \frac{N_2(k+1)}{N_1(k+1)+N_2(k+1)}. \tag{37}$$

1.3 A glimpse of universality, the diffusion limit of population genetics

In the previous sections, we were able to write down the explicit dynamics for the frequency in terms of time (generations) t only in the case of pure natural selection with no genetic drift, resulting in the logistic equation (29). The specific form of the microscopic model of serial dilution makes it so that the stochasticity cannot easily be dealt with.

In this section, we will derive the diffusion limit of population dynamics for the microscopic model, highlighting its limits of validity. It will consist of a Langevin equation which describes the dynamics of the frequency accounting for both natural selection and genetic drift.

1.3.1 Coarsening

1.3.2 Derivation

$$\frac{\partial f}{\partial t} = s_e f(1-f) + \sqrt{\frac{f(1-f)}{N_e}}\eta(t). \tag{38}$$

$$s_e \equiv s, \qquad N_e \equiv \bar{N}_0 \Delta t. \tag{39}$$

1.3.3 Universality

2 Single-locus eco-evolutionary dynamics

Single-locus models of evolution are useful models that can describe certain observations in experiments and natural populations. But not all observations. A notable exception is the Long-Term Evolution Experiment (Good, McDonald, Nature, 2017). "Simple" experiment with *E. coli* lasting >70k generations, but multiple co-existing lineages emerged in 75% (9/12) of replicate populations.

Whatsmore, these lineages are **stable**. Prior experiments have found that these strains will return to an intermediate frequency f^* if perturbed (Plucain et al., Science, 2014). So, in an evolutionary sense, the **sign** of the selection coefficient depends on the frequency of the lineage (frequency-dependent selection). This is a clear qualitative deviation from our model of evolution.

As a step towards understanding how ecology influences evolution, we are going to

- Incorporate ecology into a model of evolution via the mechanism of competition for substitutable resources.
- Derive a single-locus model of evolution where ecology gives rise to frequency-dependent fitness
- Investigate (some) eco-evolutionary consequences of resource competition.

$$\frac{\partial n_\mu}{\partial t} = n_\mu(g_\mu(\vec{c}) - D) + \sqrt{n_\mu \cdot D}\cdot \eta_\mu(t) \tag{40a}$$

$$\frac{\partial c_i}{\partial t} = S_i - Dc_i - \sum_\mu \frac{d_{\mu,i}(\vec{c}) n_\mu}{V} \tag{40b}$$

$$\langle \eta_\mu(t) \rangle = 0 \tag{41a}$$

$$\langle \eta_\mu(t)\eta_v(t') \rangle = \delta_{\mu,v}\delta(t-t') \tag{41b}$$

Assume growth and depletion have the form

$$g_\mu(\vec{c}) = \sum_i b_{\mu,i} d_{\mu,i}(\vec{c}) \tag{42a}$$

$$d_{\mu,i}(\vec{c}) = r_{\mu,i} \lambda_i(\vec{c}) \tag{42b}$$

- V vessel volume
- D dilution rate
- n absolute number of individuals
- $g_\mu(\vec{c})$ function of strain-specific per-capita growth rate
- $d_{\mu,i}(\vec{c})$ per-capita depletion/consumption rate
- $b_{\mu,i}^{-1}$ yield
- $r_{\mu,i}$ species and resource-specific factor (but concentration independent)
- $\lambda_i(\vec{c})$ species-independent but resource and concentration specific function.

Splitting growth between $r_{\mu,i}$ and $\lambda_i(\vec{c})$ can be viewed as allowing strains to vary their expression of a pathway, but not substantially change core biochemical properties (i.e., mutations alter expression of a transporter rather than altering the pump itself).

Goal: get a Langevin for allele frequencies that incorporates consumer-resource dynamics. Following steps.

- Sufficiently large large resource fluxed and concentrations: $Dc_i \approx 0$
- Separation of timescales s.t. resource concentrations reach quasi-equilibrium before abundances substantially change: $S_i C \approx \sum_\mu d_{\mu,i}(\vec{c}) n_\mu$
- $b_{\mu,i} \approx b_i$ (necessary for $\sum_\mu n_\mu(t)$ to close), giving us $N \equiv \sum_i S_i b_i V/D$

$$\frac{\partial f_\mu}{\partial t} = f_\mu \left[\sum_i \frac{\beta_i \alpha_{\mu,i} e^{X_\mu}}{\sum_v \alpha_{v,i} e^{X_v} f_v} - 1 \right] + \sum_v [\delta_{\mu,v} - f_\mu] \sqrt{\frac{f_v}{N}} \eta_v(t) \tag{43}$$

where we define

$$\xi_\mu(t) \equiv \sum_v [\delta_{\mu,v} - f_\mu] \sqrt{f_v} \eta_v(t) \tag{44}$$

Where time is in units of D^{-1}
with the following normalized parameters

$$\beta_i = \frac{S_i b_i}{\sum_j S_j b_j}, \qquad \text{Environmental supply vector (Fraction biomass supplied by resource } i) \tag{45a}$$

$$\alpha_{\mu,i} = \frac{r_{\mu,i}}{\sum_j r_{\mu,i}}, \qquad \text{Resource strategy (fraction energy spent importing resource } i) \tag{45b}$$

$$X_\mu = \log \sum_i r_{\mu,i}, \qquad \text{General fitness (max. growth rate)} \tag{45c}$$

We can define resource-specific mean fitness

$$\overline{X}_i(t) = \log\left(\frac{\sum_\mu \alpha_{\mu,i} e^{X_\mu} f_\mu}{\beta_i}\right) \tag{46}$$

Allowing us to re-arrange terms in the Langevin

$$\frac{\partial f_\mu}{\partial t} = f_\mu \sum_{i=1}^{\mathcal{R}} \alpha_{\mu,i}\left[e^{X_\mu - \overline{X}_i} - 1\right] + \frac{\xi_\mu(t)}{\sqrt{N}} \tag{47}$$

which has the following lowest-order expansion

$$\frac{\partial f_\mu}{\partial t} \approx f_\mu \sum_{i=1}^{\mathcal{R}} \alpha_{\mu,i}[X_\mu - \overline{X}_i(t)] + \frac{\xi_\mu(t)}{\sqrt{N}} \tag{48}$$

2.1 Competition for $\mathcal{R} = 2$ resources

2.1.1 Competition between $\mathcal{S} = 2$ strains

Set $f \equiv f_2$ and fitness difference $\Delta X - X_2 - X_1$, we can obtain an effective selection coefficient as follows

$$s_e(f) \equiv \frac{1}{f(1-f)}\left(\frac{\partial f}{\partial t}\right)_{\text{deterministic}} \tag{49a}$$

$$= \frac{\beta[\alpha_2 e^{\Delta X} - \alpha_1]}{\alpha_1 + [\alpha_2 e^{\Delta X} - \alpha_1]f} + \frac{(1-\beta)[(1-\alpha_2)e^{\Delta x} - (1-\alpha_1)]}{1 - \alpha_1 + f[(1-\alpha_2)e^{\Delta X} - (1-\alpha_1)]} \tag{49b}$$

for the single-locus Langevin

$$\frac{\partial f}{\partial t} = s_e(f) f(1-f) + \sqrt{\frac{f(1-f)}{N}}\eta(t) \tag{50}$$

We now have a model of frequency-dependent selection that is determined by consumer-resource parameters!

First, let's look at the limiting case of invasion of a rare mutant ($f \to 0$). The invasion fitness reduces to

$$S_{\text{inv}} \equiv \lim_{f\to 0} s_e(f) \tag{51a}$$

$$= e^{\Delta x} - 1 + e^{\Delta x}\left[\frac{(\beta-\alpha_1)(\alpha_2-\alpha_1)}{\alpha_1(1-\alpha_1)}\right] \tag{51b}$$

and the linear Langevin

$$\frac{\partial f}{\partial t} = S_{\text{inv}} f + \sqrt{\frac{f}{N}}\eta(t) \tag{52}$$

Which can be used to derive the standard branching process statistics covered by Onofrio for the regime $f \ll 1$. But there are three criteria for the branching process approximation to apply

$$NS_{\text{inv}} \gg 1 \tag{53a}$$

$$\frac{NS_{\text{inv}}\alpha_1}{\alpha_2 e^{\Delta X} - \alpha_1} \gg 1 \tag{53b}$$

$$\frac{NS_{\text{inv}}}{(1-\alpha_2)e^{\Delta X} - (1-\alpha_1)} \gg 1 \tag{53c}$$

Which are met for large N. The invasion criteria $S_{\text{inv}} > 0$ can be used to investigate fixation ($f = 1$) or co-existence at intermediate frequency f^*. Stable co-existence requires that the reciprocal of the invasion fitness is also positive

$$S^R_{\text{inv}} \equiv \lim_{f\to 1} - s_e(f) \tag{54a}$$

$$= (e^{-\Delta X} - 1) + e^{-\Delta X}\left[\frac{(\beta-\alpha_2)(\alpha_1-\alpha_2)}{\alpha_2(1-\alpha_2)}\right] \tag{54b}$$

at $S^R_{\text{inv}} = 0$ we get the critical fitness threshold

$$\Delta X_{\max} = \log\left(1 + \frac{(\alpha_1-\alpha_2)(\beta-\alpha_2)}{\alpha_2(1-\alpha_2)}\right) \tag{55}$$

when $S_{\text{inv}} > 0$ and $S^R_{\text{inv}} > 0$, we can calculate the equilibrium frequency from $s_e(f^*) = 0$. $\alpha_2 e^{\Delta X} - \alpha_1$ and $(1-\alpha_2)e^{\Delta X} - (1-\alpha_1)$ **must have different signs** for non-trivial frequency. We can then solve for f^*. In absence of fitness differences that frequency is $f^*_0 = \frac{(\beta-\alpha_1)}{\Delta\alpha}$. With fitness differences we obtain

$$f^* = \frac{f^*_0 + \left[f^*_0 + \frac{\alpha_1(1-\alpha_1)}{\Delta\alpha^2}\right]\left(e^{\Delta X} - 1\right)}{\left[1 + \frac{\alpha_2}{\Delta\alpha}\left(e^{\Delta X} - 1\right)\right]\left[1 - \frac{(1-\alpha_2)}{\Delta\alpha}\left(e^{\Delta X} - 1\right)\right]} \tag{56}$$

We can linearize this equation to gain some intuition. As $\Delta X \to 0$ and $\Delta\alpha \to 0$ we get

$$f^*(\Delta X) \approx f_0^* + \frac{\beta(1-\beta)}{\Delta\alpha^2} \cdot \Delta X \tag{57}$$

Fitness sensitivity determined by distance between resource strategies ($\Delta\alpha$).

We can then determine fluctuations around equilibrium frequencies $\delta f \equiv f - f^*$

At $f = f^*$ resource-specific mean fitness $\overline{X}_i$ are independent of β, meaning that they are independent of environmental conditions!

$$\overline{X}_1 = -\log\left[1 - (1 - e^{-\Delta X})(\frac{1-\alpha_2}{\Delta\alpha})\right] \tag{58a}$$

$$\overline{X}_2 = -\log\left[1 + (1 - e^{-\Delta X})(\frac{\alpha_2}{\Delta\alpha})\right] \tag{58b}$$

when ΔX is small, fitness functions can be linearized

$$\overline{X}_1 = \frac{(1-\alpha_2)}{\Delta\alpha}\Delta X \tag{59a}$$

$$\overline{X}_2 = -\frac{\alpha_2}{\Delta\alpha}\Delta X \tag{59b}$$

Using linearized fitness and linearized f^*, we can obtain a fitness scale over which $f^*(\Delta X)$ changes

$$X_f \equiv f^*(1-f^*)\left(\frac{\partial f^*}{\partial \Delta X}\right)^{-1} \tag{60a}$$

$$= \frac{(\beta-\alpha_1)(\alpha_2-\beta)}{\beta(1-\beta)+(\beta-\alpha_1)(\alpha_2-\beta)} \tag{60b}$$

Linearizing around $f \approx f^*$, we use our effective single-locus model to get a Langevin for the fluctuations around the equilibrium frequency

$$\frac{\partial \delta f}{\partial t} = -X_{eq} f^*(1-f^*)\delta f + \sqrt{\frac{f^*(1-f^*)}{N}}\eta(t) \tag{61}$$

where

$$X_{\text{eq}}(\Delta X) \equiv -\frac{\partial s_e(f)}{\partial f}\Big|_{f=f^*} \tag{62}$$

when $\Delta X \ll 1$ the Langevin can be solved, obtaining the stationary distribution $P(\delta f)$ $\mathcal{N}(0, (2NX_{eq}))^{-1/2}$. So the width of fluctuations decrease with $\sqrt{X_{eq}}$.

2.1.2 Competition between $\mathcal{S} = 3$ strains

Assume a mutant enters the community, with fitness $X_3 = \Delta X + s$ and strategy vector α_3. Using the observation that the mutant should **never invade if it is identical to its parent**, we can obtain the following invasion fitness.

$$S_{inv} = (e^s - 1) + (\alpha_3 - \alpha_2)\left(e^{-\overline{X}_1(t)} - e^{-\overline{X}_2(t)}\right)e^{\Delta X + s} \tag{63}$$

and examine special cases

1) Neutral. Fluctuations insufficient for strain replacement

$$\Delta X = s = 0 \tag{64}$$

2) Wright-Fisher model. Fixation then competition b/w strains 1 and 2

$$\alpha_3 = \alpha_2 \tag{65}$$

3) Strategy mutant (ecology!)

$$s = 0 \rightarrow S_{inv} = \frac{\alpha_3 - \alpha_2]}{\alpha_1 - \alpha_2}(e^{\Delta X} - 1) \tag{66}$$

Last scenario most "interesting". Direction of selection determined by sign of ΔX. On the fitter lineage ($\Delta X > 0$), selection favors mutations that push the strategy **towards** β (generalists). In the less-fit background selection favors mutations that push the strategy **away** from β (specialists). Commonality: selection favors mutations that lead push a lineage **towards** consuming resources consumed by less-fit lineages, minimizing effective competition.

We can understand the outcome of beneficial mutations on the more or less fit background using our linearized equilibrium frequency equation. Successful mutations in less-fit lineage always sweep through the lineage, increasing f^*. Successful mutations in the more-fit lineage have two possibilities: 1) $\alpha_3 < \beta \rightarrow$ outcompetes α_1 and stabily co-exists with parent α_2. 2) $\beta < \alpha_3 < \alpha_2$ sweeps the parent lineage.

There are multiple possibilities **within** option 2. If α_2 close to α_3 then f^* increases. BUT if α_3 is close enough to β that $\Delta X_{max}(\alpha_3)$ becomes less than ΔX then the mutant sweeps both lineages and the ecosystem collapses!

2.2 Many resources: $\mathcal{R} > 2$

In two-resource systems the ecological equilibria are either monocultures ($\mathcal{S} = 1$) or have the maximum number of strains ($\mathcal{S} = 2$). When $\mathcal{S} > 2$ you can have stable coexistence for values of $1 < \mathcal{S} < \mathcal{R}$ or $\mathcal{R} = \mathcal{S}$. Qualitatively different properties!

2.2.1 Saturated: $\mathcal{S} = \mathcal{R}$

Analogous to the $\mathcal{R} = 2$ case. If strains have dissimilar resource strategies, then $\overline{X}_i$ is independent of resource supply β_i

2.2.2 Unsaturated: $\mathcal{S} = \mathcal{R}$

Resource specific mean fitness is underdetermined $\overline{X}_i$, so $\overline{X}_i$ and f_μ^* must be jointly solved using non-linear constraints. Convex optimization problem, where $h_i = e^{-\overline{X}_i}$

$$\vec{h}^* = \text{argmax}_{\vec{h}} \left\{ \sum_i \beta_i \log h_i : \sum_i \alpha_{\mu,i} h_i = e^{-X_\mu} \forall \mu \right\} \tag{67}$$

depends on β_i! Turns out to be an important difference, as the ecosystem can no longer dynamically adjust, so internal selection pressures will change as the environment changes. This detail permits new opportunities for evolutionary adaptation.

3 References

[1] Benjamin H. Good, APHYS 237/BIO251: Quantitative evolutionary dynamics and genomics, https://bgoodlab.github.io/courses/apphys237/