Speciation induced by dormancy in a model with changing environment

by Jason Schweinsberg
University of California San Diego

Joint work with Fernando Cordero and Adrián González Casanova

Background

In many populations, individuals can enter a dormant state:

- Some mammals hibernate in the winter.
- Many plants produce seeds which germinate after a dormant period.
- Many bacteria can enter a dormant state.

Dormancy allows individuals to survive periods of unfavorable environmental conditions.

It has been suggested that dormancy could contribute to speciation:

- Blath, Paul, Tóbiás, and Wilke-Berenguer (2024) found that dormancy could increase evolutionary branching.
- Tellier, Laurent, Lanier, Pavlidis, and Stephan (2011) found that dormancy may contribute to a high speciation rate among South American tomatoes.

Dormancy and Seasonal Changes

We consider a model of dormancy that is similar to Blath, González Casanova, Kurt, and Wilke-Berenguer (2016), but includes seasonal changes.

Individuals can acquire mutations that are advantageous in the summer and disadvantageous in the winter, or vice versa.

We see two waves of adaptation. Some individuals repeatedly acquire mutations that benefit them in the summer. Others repeatedly acquire mutations that benefit them in the winter.

Individuals can survive the season during which they are less fit by entering a dormant state.

This may be a mechanism by which dormancy could lead to speciation.

The Model

In each generation, there are N active individuals and K_N dormant individuals.

The environment alternates between summer and winter. Each year lasts U_N generations. Summer lasts V_N generations, and winter lasts $U_N - V_N$ generations.

Each individual has a type in $\mathbb{Z}$. An individual of type k has fitness $(1 + s_N)^k$ during the summer and $(1 + s_N)^{-k}$ during the winter.

To obtain the next generation:

- Randomly choose c_N dormant individuals to become active, and c_N active individuals to become dormant.
- The other $K_N - c_N$ dormant individuals remain dormant.
- The other $N - c_N$ active individuals choose their parent at random from the previous generation, with probability proportional to fitness.
- If the parent has type k, the offspring has type k with probability $1 - 2\mu_N$, type $k + 1$ with probability μ_N, and type $k - 1$ with probability μ_N.

Assumptions on the parameters

1. Strong selection:
$$\lim_{N\to\infty} s_N = s \in (0,\infty).$$
2. Winter and summer have comparable length:
$$\lim_{N\to\infty} \frac{V_N}{U_N} = \beta \in (0,1).$$
3. Seasons are not too short:
$$U_N \gg \log N.$$
4. Transitions from active to dormant are not too fast:
$$c_N \ll N.$$
5. Mutation rate is not too slow:
$$\mu_N \gg N^{-b} \quad \text{for some } b > 0.$$

The population after a beneficial mutation

Initially, all individuals have type 0. Suppose one individual mutates to type 1 during the summer.

In the next generation, each individual has type 1 with probability

$$\frac{1+s_N}{(N-1)+(1+s_N)} \approx \frac{1+s_N}{N}.$$

There are $N - c_N \approx N$ new active individuals born in the next generation. The number of type 1 in the next generation is approximately Poisson$(1+s_N)$.

For a short time, the number of type 1 individuals is well approximated by a branching process with a Poisson$(1+s_N)$ offspring distribution.

Conditional on survival, most of the active population is type 1 after a time which is $O(\log N)$. (Kimura and Ohta, 1969)

Conditional on extinction, there is not sufficient time for a type 1 individual to become dormant because $c_N \ll N$.

The population after a seasonal change

Suppose type 1 spreads to most of the active population, then winter arrives.

The dormant population will contain both type 0 individuals (which have been dormant since early in the previous summer) and type 1 individuals (which became dormant late in the summer).

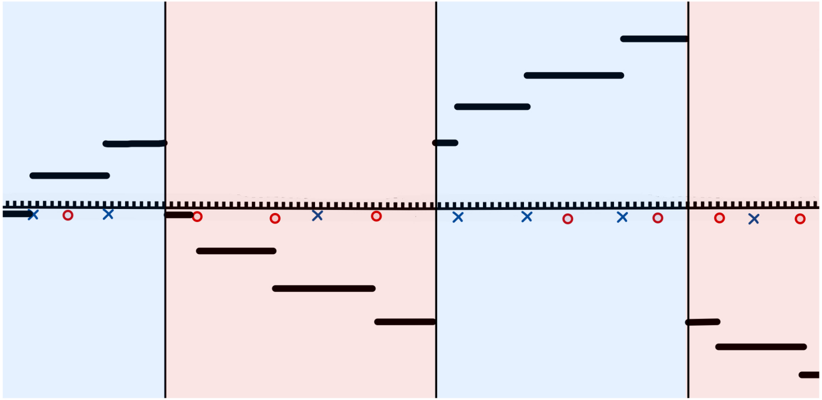
Some type 0 individuals will move from the dormant to the active population.

Because type 0 is fitter, type 0 will quickly spread to most of the active population, in a time which is $O(\log N)$. Recall that $\log N \ll U_N$.

Another mutation may cause type -1 to spread to most of the active population. However, type 1 individuals will remain in the dormant population and quickly take over the following summer.

The type 1 individuals will be overtaken again by type -1 individuals at the beginning of the following winter, and the process continues.

A graph of the dominant type in the active population over time



× represents a positive mutation

○ represents a negative mutation

Blue is summer, and red is winter

Some notation

$X_{k,N}(m) =$ fraction of active individuals in generation m having type k.
$Y_{k,N}(m) =$ fraction of dormant individuals in generation m having type k.

Let Δ be the space of sequences $\mathbf{x} = (\ldots, x_{-2}, x_{-1}, x_0, x_1, x_2)$ such that $x_k \geq 0$ for all $k \in \mathbb{Z}$ and $\sum_k x_k = 1$. We call x_k the kth coordinate of $\mathbf{x}$ for $k \in \mathbb{Z}$.

Let $\mathbf{X}_N(m)$ and $\mathbf{Y}_N(m)$ be the Δ-valued random variables whose kth coordinates are $X_{k,N}(m)$ and $Y_{k,N}(m)$ respectively:

$$\mathbf{X}_N(m) = (\ldots, X_{-2,N}(m), X_{-1,N}(m), X_{0,N}(m), X_{1,N}(m), X_{2,N}(m), \ldots),$$
$$\mathbf{Y}_N(m) = (\ldots, Y_{-2,N}(m), Y_{-1,N}(m), Y_{0,N}(m), Y_{1,N}(m), Y_{2,N}(m), \ldots).$$

Three time scales

1. The length of a year is U_N.
2. Let $h(u)$ be the survival probability for a branching process with a Poisson$(1 + u)$ offspring distribution. Then $\rho_N = N\mu_N h(s_N)$ is approximately the probability of a successful mutation in a generation. The expected number of generations until a successful mutation is ρ_N^{-1}.
3. Because a fraction c_N/K_N of the seed bank is replaced every generation, the number of generations required for the dominant type in the active population to spread to a substantial fraction of the seed bank is K_N/c_N.

We will scale time by ρ_N, so that one unit of time corresponds to ρ_N^{-1} generations.

Two parameter regimes

Regime 1: Suppose that

$$\lim_{N\to\infty} \frac{c_N}{K_N \rho_N} = \alpha \in (0, \infty), \qquad \lim_{N\to\infty} \rho_N U_N = \theta \in (0, \infty).$$

In Regime 1, all three time scales are comparable.

Regime 2: Suppose that

$$U_N \ll \frac{K_N}{c_N} \ll \rho_N^{-1}$$

and

$$\lim_{N\to\infty} \frac{c_N^2 U_N^2}{K_N \log N} = \infty.$$

In Regime 2, it takes longer for active individuals to enter the dormant population than it takes for the season to change, but the active and dormant populations will reach equilibrium before the next mutation.

Convergence of the mutation times to a Poisson process

Recall in Regime 1, $\rho_N U_N \to \theta$ and $\rho_N V_N \to \beta\theta$. Let

$$R(t) = \begin{cases} 1 & \text{if } j\theta \leq t < j\theta + \beta\theta \text{ for some nonnegative integer } j \\ 0 & \text{otherwise.} \end{cases}$$

in Regime 1, and $R(t) = \beta$ for all t in Regime 2.

Let $(N^+(t), t \geq 0)$ and $(N^-(t), t \geq 0)$ be independent inhomogeneous Poisson processes whose intensities are given by $R(t)$ and $1 - R(t)$ respectively.

Let $T_0 = 0$, and for positive integers k, define

$$T_k = \inf\{t : N^+(t) = k\}, \qquad T_{-k} = \inf\{t : N^-(t) = k\}.$$

Theorem: Fix $a \in (0,1)$. Let $S_{k,N}(a) = \min\{m : X_{k,N}(m) > a\}$. In both regimes, for all positive integers K and all $a \in (0,1)$, we have

$$\big(\rho_N S_{k,N}(a)\big)_{k=-K}^{K} \Rightarrow (T_k)_{k=-K}^{K} \qquad \text{as } N \to \infty.$$

Composition of the active population

Recall that $S_{k,N}(a) = \min\{m : X_{k,N}(m) > a\}$, for $a \in (0,1)$.

Typically one type dominates the active population, except for short periods when a beneficial mutation is spreading or the season is changing.

For $a \in (0,1)$ and nonnegative integers m, let

$$D_{N,a}(m) = \begin{cases} \max\{k : S_{k,N}(a) \leq m\} & \text{if } m \text{ is in the summer} \\ \min\{k : S_{k,N}(a) \leq m\} & \text{if } m \text{ is in the winter} \end{cases}$$

We expect $D_{N,a}(m)$ to be the dominant type in the active population in generation m.

Theorem: Fix $t_0 > 0$ and $a \in (0,1)$. In both regimes, for all $k \in \mathbb{Z}$, we have

$$\rho_N \sum_{m=0}^{\lfloor \rho_N^{-1} t_0 \rfloor} \left| X_{k,N}(m) - \mathbb{1}_{\{D_{N,a}(m)=k\}} \right| \to_p 0 \qquad \text{as } N \to \infty.$$

Convergence of stochastic processes in Regime 1

In Regime 1, let $D(t) = N^+(t)\mathbb{1}_{\{R(t)=1\}} - N^-(t)\mathbb{1}_{\{R(t)=0\}}$.

Let $\mathbf{D}(t)$ be the Δ-valued random variable whose kth coordinate is $\mathbb{1}_{\{D(t)=k\}}$.

Theorem: In Regime 1, we have

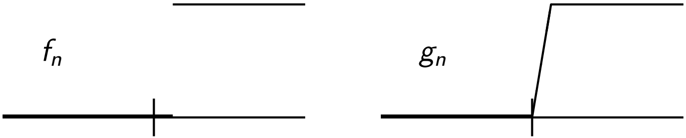
$$(\mathbf{X}_N(\lfloor \rho_N^{-1} t \rfloor), t \geq 0) \Rightarrow (\mathbf{D}(t), t \geq 0) \qquad \text{as } N \to \infty,$$

where $\Rightarrow$ represents weak convergence of stochastic processes with respect to the Meyer-Zheng topology.

The Meyer-Zheng topology (Meyer and Zheng, 1984)

Let (E, d) be a metric space. Càdlàg functions $f_n : [0, \infty) \to E$ converge to f in the Meyer-Zheng topology if for all $T > 0$ and $\varepsilon > 0$, the Lebesgue measure of $\{x \in [0, T] : d(f_n(x) - f(x)) > \varepsilon\}$ tends to zero as $n \to \infty$.

Let $f_n(t) = \begin{cases} 0 & \text{if } t < 1 + 1/n \\ 1 & \text{if } t \geq 1 + 1/n \end{cases}$ $\quad g_n(t) = \begin{cases} 0 & \text{if } t \leq 1 \\ nt & \text{if } 1 < t < 1 + 1/n \\ 1 & \text{if } t \geq 1 + 1/n \end{cases}$



In the Skorohod J_1 topology, $f_n \to f_\infty$ but $g_n \not\to f_\infty$.

In the Meyer-Zheng topology, $f_n \to f_\infty$ and $g_n \to f_\infty$.

See also Kurtz (1991), Gonzalez Casanova, Miró Pina, and Siri-Jégousse (2022).

Composition of the dormant population: Regime 1

Suppose $X_{k,N}(m) = x$ and $Y_{k,N}(m) = y$. Then

$$Y_{k,N}(m+1) \approx \frac{xc_N + y(K_N - c_N)}{K_N} = y + (x-y)\frac{c_N}{K_N}.$$

After scaling time by ρ_N, we obtain

$$\frac{dy}{dt} \approx (x-y)\frac{c_N}{K_N \rho_N} \approx \alpha(x-y).$$

Theorem: Define a Δ-valued stochastic process $(\mathbf{Y}(t), t \geq 0)$ such that $\mathbf{Y}(0)$ is the sequence whose 0th term is 1 and whose other terms are zero, and

$$\frac{d}{dt}\mathbf{Y}(t) = \alpha(\mathbf{D}(t) - \mathbf{Y}(t)).$$

In Regime 1, the processes $(\mathbf{Y}_N(\lfloor \rho_N^{-1} t \rfloor), t \geq 0)$ converge as $N \to \infty$ to $(\mathbf{Y}(t), t \geq 0)$, in the sense of weak convergence of stochastic processes with respect to Skorohod's J_1 topology.

Composition of the dormant population: Regime 2

In Regime 2, seasons change rapidly. It is summer a fraction β of the time.

Because $c_N/(K_N \rho_N) \to \infty$, the differential equation converges to an equilibrium before the next beneficial mutation.

Theorem: For all $k \in \mathbb{Z}$, define

$$Y_k(t) = \begin{cases} \beta \mathbb{1}_{\{T_k \leq t < T_{k+1}\}} & \text{if } k \geq 1 \\ (1-\beta)\mathbb{1}_{\{T_k \leq t < T_{k-1}\}} & \text{if } k \leq -1 \\ \beta \mathbb{1}_{\{t < T_1\}} + (1-\beta)\mathbb{1}_{\{t < T_{-1}\}} & \text{if } k = 0. \end{cases}$$

Let $\mathbf{Y}(t)$ be the Δ-valued random variable whose kth coordinate is $Y_k(t)$.
In Regime 2, the processes $(\mathbf{Y}_N(\lfloor \rho_N^{-1} t \rfloor), t \geq 0)$ converge as $N \to \infty$ to $(\mathbf{Y}(t), t \geq 0)$, in the sense of weak convergence of stochastic processes with respect to the Meyer-Zheng topology.

Times when individuals enter dormant population

Label the dormant individuals $1, \ldots, K_N$.

Let M_i denote the number of generations in the past that the ith dormant individual became dormant. Then $M_i \sim \text{Geometric}(c_N/K_N)$.

Let T be a finite subset of the positive integers. Let

$$Z_T = \sum_{i=1}^{K_N} \mathbb{1}_{\{M_i \in T\}}.$$

Then

$$E[Z_T] = K_N P(M_1 \in T) = K_N \sum_{m \in T} \frac{c_N}{K_N}\left(1 - \frac{c_N}{K_N}\right)^{m-1} = c_N \sum_{m \in T}\left(1 - \frac{c_N}{K_N}\right)^{m-1}.$$

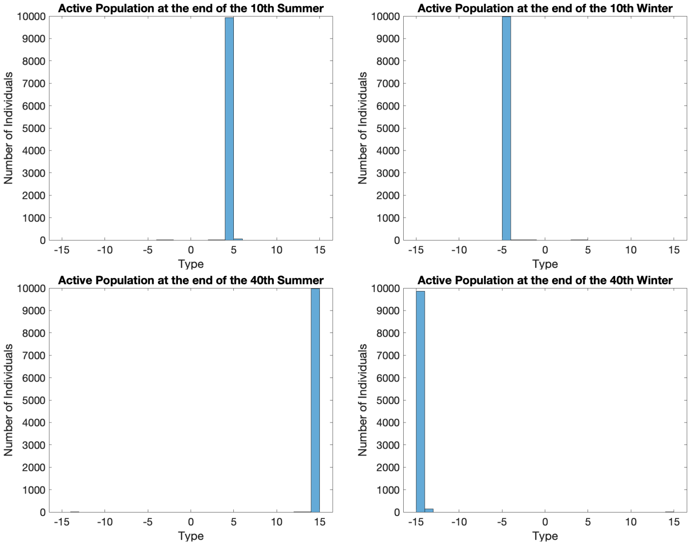
The events $\{M_i \in T\}$ are negatively correlated by Harris' Inequality, so

$$P(|Z_T - E[Z_T]| > \varepsilon E[Z_T]) \leq 2e^{-(\varepsilon^2 \wedge \varepsilon)E[Z_T]/3}$$

and we have good control on when individuals entered the dormant population.

Simulation Results for the Active Population

$$N = K_N = 10000, c_N = 10, U_N = 500, V_N = 250, s_N = 0.1, \mu_N = 0.000002$$


Active Population at the end of the 10th Summer

Active Population at the end of the 10th Winter

Active Population at the end of the 40th Summer

Active Population at the end of the 40th Winter

Simulation Results for the Dormant Population