Computer Simulation #4: Genetic Drift *versus* Selection in *Cepaea nemoralis*

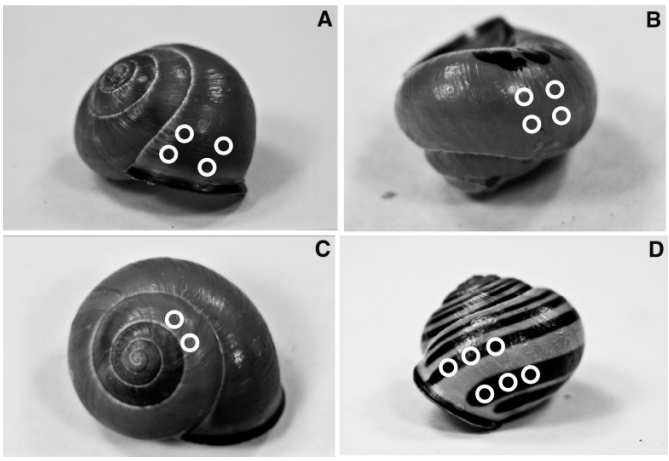
Cepaea nemoralis, the European land snail, is the classic example of the interplay of ***deterministic*** (natural selection) and ***stochastic*** (genetic drift) factors in evolution. EB Fords' studies on this species are the origin of Ecological Genetics.

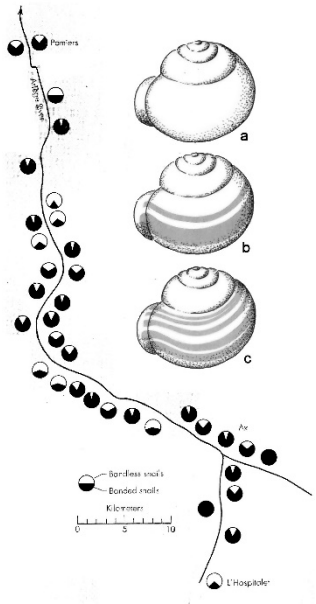
Cepaea is widespread in western Europe and southern Great Britain; related species occur in North America, including Newfoundland. *C. nemoralis* is phenotypically polymorphic: individuals have dark or light shells, and from zero to five bands [below, left]. The polymorphism is controlled genetically: *dark* alleles are dominant to *light*, and *un-banded* alleles are dominant to *banded*.

Cepaea, like *Biston betularia*, is subject to differential visual predation by birds based on **crypsis**. The principle predators are song thrushes, which drop the snails on flat rocks ("*thrush anvils*"), which breaks the shells open so that the birds can eat the contents. Predation patterns can be measured by examination of the phenotypes of *broken* shells at the anvils. The incidence of the various shell types varies *inversely* according to habitat type. Broken *light-colored* shells are more common at anvils in *dark* "brown" backgrounds [*e.g.*, bare soil], and less common in *light* "green" backgrounds [*e.g.*, grasslands]. Broken *un-banded* shells are more common in "high contrast" habitats [*e.g.*, short grass fields], and are less common on "low contrast" habitats [*e.g.*, deep-shadow woodlands]. Much of the variation among colonies seems to be correlated with these two environmental variables. Snails that stand out visually are preferentially eaten by thrushes and show up more frequently as broken shells at the anvils: the more cryptic do not show up as frequently.

Genetic drift also plays a role. The snails occur in colonies of different sizes, from tens to thousands of individuals. Colonies are well-separated micro-geographically, and gene flow between colonies is severely limited by the creatures' sedentary habits. Different shell forms typically co-exist in most colonies, and the frequency of each type varies greatly over short distances [below, right], not always in accordance with predictions from crypsis. It has been shown that there is greater genetic variation (measured as variance) in shell patterns among *smaller* colonies than *larger*, consistent with the expectations of genetic drift. Some small colonies may even become '*fixed*' for '*lost*' for the '*wrong*', non-cryptic shell pattern.

This exercise compares the behavior of alleles in large rand smaller populations, with greater or lesser degrees intensity of selection, and thus the interaction between **stochastic** Genetic Drift and **deterministic** Natural Selection in determining the degree of genetic polymorphism in natural populations.





Biol4250 – Evolutionary Genetics
MatLab simulation of **Natural Selection** / **Genetic Drift** in **finite populations** **Fall 2021**

In a "***tide pool***" model, consider a **finite population** of **N** individuals at t_0, each with two **alleles A** & **B** at a **gene locus**, for a total of **2N** alleles. Emphasis in the **MatLab** model is the **B** allele and its frequency **q** = f(**B**). Recall

If alleles unite at random in three **genotypes** **AA**, **AB**, & **BB**, the *expected* frequencies at time t_1 are

$f(\mathbf{AA}) = p^2 = (\mathbf{1 - q})^2$ $f(\mathbf{AB}) = 2pq = \mathbf{2(q)(1 - q)}$ $f(\mathbf{BB}) = \mathbf{q}^2$

In any *finite* population **N**, *stochastic* expectaions over time will depart from the *deterministic* expectation in an infinite population, where the variance between generations is $\sigma^2_q = (\mathbf{q})(1\text{-}\mathbf{q})/2\mathbf{N}$. Therefore, **Δq** in a set of such populations will also "drift" stochastically *among* populations over time, with variance of **q** *among* populations $\sigma^2_t = (\mathbf{q})(1\text{-}\mathbf{q})/2\mathbf{N}$. Finally, the magnitude of **Δq** will differ between two population of *different* size, again with the expectation $\sigma^2_N = (\mathbf{q})(1\text{-}\mathbf{q})/2\mathbf{N}$.

Model

The **MatLab** program **hardy_weinberg_2020** draws 2 alleles at random as '**0**' or '**1**', where '**0**' = **A** and '**1**' = **B**. The alleles are then added to give **0** for **AA**, **1** for **AB**, and **2** for **BB**. The appropriate fitness **WO**, **W1**, or **W2** is applied as a *probability of survival*: if the individual does not '*survive*', allele sampling and the fitness test are applied until one does, and this individual is added to the gene pool. This is repeated until the sample reaches **N** individuals, in each of $\mathbf{N_{pop}}$ populations. This version corresponds to a "***soft selection***" model of Natural Selection.

Instructions

1) Start the program. A screen will ask for the following parameters:

Initial value of **q**, in the range $0.0 < \mathbf{q_{init}} < 1.0$
Population size **N** for any value **≤ 100,000**
Number of populations **Npop**, typically **10,** up to 1000
Number of generations **Ngen** (see **Suggestion** below)
Fitness values **W0**, **W1**, & **W2** for **AA**, **AB**, & **BB**
B is **recessive** to **A**; set **W1** = **W2** = **1** for **AA** & **AB**, set **W2** per the scenario

2) Run the program by clicking on "Run Evolution.

The program will plot f(**B**) = **q**, for these input parameters.
For each of the following scenarios, run **10** [or more] **simulations simultaneously**.
The trajectories of f(**B**) are plotted for all populations in the simulation.
Counts of populations that fix (**q = 1**) or lose (**q = 0**) allele **B** are noted.
Copy & Save JPG graphs to a drawing program or WORD for your discussion:

Label graphs with starting parameters: suggested format ***q.1 N10000 s.1***.*jpg*

Scenarios

See p. 4 of this lab handout. Start the sets of simulations at **q = 0.1**, **0.5**, & **0.**
Record the count of populations that go to **fixation** (**q = 1**) or **loss** (**q = 0**)
Note the approximate range of number of generations required for loss / fixation
Suggestion: It may be useful to run **Npop = 100** or **1,000** to get a more accurate ratio,
or to *repeat* the sample of 10 or 100 populations 10 times to see the variance

Biol4250 Evolutionary Genetics – Natural Selection in Finite Populations **Dr Carr**

1. Over time, in which size colony are allele frequencies more variable ***among populations***? Why?

2. Is a species composed of ***multiple small populations*** 'more variable' than one composed of a ***few large populations***? Explain.

3. Can **deleterious** alleles be **fixed** in a population? Under what circumstances? Explain.

4. What is the ***ultimate fate*** of alleles, even in 'large' populations? Explain.

5. For what values of **s** and **N** do the effects of selection and drift achieve an approximate equilibrium of **Δq** ?

q = 0.1 N= / s=	0.100	0.010	0.001	0.000	0.001	0.010	0.001
10							
100							
1,000							
10,000							

q = 0.5 N= / s=	0.100	0.010	0.001	0.000	0.001	0.010	0.001
10							
100							
1,000							
10,000							

q = 0.9 N= / s=	0.100	0.010	0.001	0.000	0.001	0.010	0.001
10							
100							
1,000							
10,000							