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

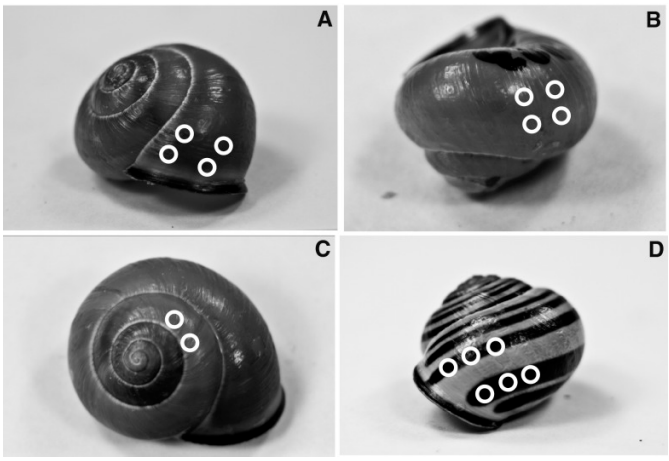
Cepaea nemoralis, the European land snail, is the classic example of the interplay of ***deterministic*** (natural selection) and ***stochastic*** (genetic drift) factors in evolution. EB Fords' studies on this species are the origin of **Ecological Genetics**.

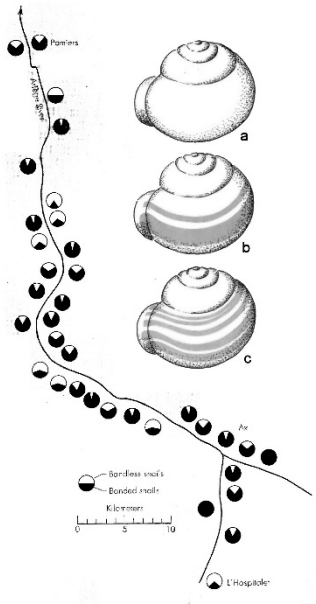
Cepaea is widespread in western Europe and southern Great Britain. Related species occur in North America, including Newfoundland. *C. nemoralis* is phenotypically polymorphic: individuals have dark or light shells, and from zero to five bands [below, left]. The polymorphism is controlled genetically: *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 and allows the birds to 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 show up less 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 variation (measured as phenotypic variance) in shell patterns among *smaller* colonies than *larger*, consistent with the expectations of genetic drift. Some small colonies may even '*fix*' or '*lose*' the '*wrong*', non-cryptic shell pattern.

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





Biol4250 – Evolutionary Genetics **Fall 2024**
MatLab simulation of **Natural Selection** / **Genetic Drift** in **finite populations**

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_0 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 single *finite* population **N**, *stochastic* change of **q** over time will depart from the *deterministic* Hardy-Weinberg expectation for an *infinite* population. The variance of allele frequency change **Δq** between generations is $\sigma^2_q = (\mathbf{q})(1-\mathbf{q})/2\mathbf{N}$, thus *smaller* populations have *larger* variance: they "*drift*" more than larger populations. Multiple populations drifting independently will also depart from each other stochastically over time, with variance of **q** among them is σ^2_t.

Model

The **MatLab** program **WriFish** generates **two** '**0**'s and '**1**'s corresponding to alleles **A** or **B**, according to **q = f(B)**. The two alleles define genotypes **AA**, **AB**, and **BB**. The appropriate fitness **W0**, **W1**, or **W2** is applied as a *probability of survival* of the newly-generated individual genotype. 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. **N** remains constant between pre- & post selection populations: **WriFish** thus corresponds to a "***soft selection***" model of Natural Selection.

Instructions

1) Start the program **WriFish** in **MatLab**. A screen will ask for the following parameters:

Fitness values **W0**, **W1**, & **W2** for **AA**, **AB**, & **BB** according to the scenario:
If **B** is *recessive* to **A**, **W0 = W1 ≠ W2**. If **B** is *dominant* to **A**, **W0 ≠ W1 = W2**
Initial value of **q** = f(**B**) in the range $0.0 < \mathbf{q_{init}} < 1.0$
Population size **N** for any value **≤ 100,000**
Number of generations **Ngen** , up to 10,000 (see **Suggestions**)
Number of populations **Npop**, **2** ~ 1,000 (see **Suggestions**)

2) Run the program by clicking on "wdzy 9Ø2fdzúλ2y".

The program will plot f(**B**) = **q**, for the input parameters.
For each scenario, run the suggested # of **Npop** simulations for **Ngen** generations**.**
The trajectories of f(**B**) are plotted for all populations in the simulation.
Counts of populations that go to **q = 1** or **0** (*fixation* or *loss* of **B**) are reported

Scenarios

3) Run the scenarios 2y the last page of this handout. Run *successive rows* in each scenario.

wŠŏ2ŇŘ úKŜ Ŏ2dzyű 2T LJ2LJdzfl·úλ2yã that go to **fixation** (**q = 1**) or **loss** (**q = 0**)
Note the number of generations required for fixation / loss. **Highlight** *unexpected* results.
Copy & Save JPG graphs as needed to a drawing program or WORD for your discussion:
Label graphs with starting parameters: suggested format **q.1 N10000 W2.9.jpg**
Suggestions: Best results are obtained when **Npop = 100** or **1,000** , over **Ngen = 1000 ~ 10,000**.
This may run very slowly, on slower PCs. Adjust **Ngen** accordingly, then **Npop**.

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 **W** and **N** do the effects of selection and drift achieve an approximate equilibrium of **Δq** ?

W0=W1=X ; W2 = 1	Rare Advantageous Recessive						Neutral	
q =	N = W0=W1 =	0.9000	0.9900	0.9990	0.9999	1.0000	Suggested npop	ngen
0.1000	5						1,000	1,000
0.0100	50						1,000	1,000
0.0010	500						1,000	2,000
0.0001	5,000						1,000	1,000

W0=W1=1 ; W2 = X	q=0.5 Deleterious Recessive							
q = 0.5	N = W2 =	0.9000	0.9900	0.9990	0.9999	1.0000	Suggested npop	ngen
	5						100	1,000
	50						100	1,000
	500						100	10,000
	5,000						10	1,000

W0 = X ; W1=W2 = 1	Common Advantageous Dominant							
q =	N = W0 =	0.9000	0.9900	0.9990	0.9999	1.0000	Suggested npop	ngen
0.9000	5						1,000	1,000
0.9900	50						1,000	1,000
0.9990	500						1,000	2,000
0.9999	5,000						1,000	100

W0 = X ; W1=W2 = 1	Rare Advantageous Dominant						Neutral	
q =	N = W0 =	0.9000	0.9900	0.9990	0.9999	1.0000	Suggested npop	ngen
0.1000	5						1,000	1,000
0.0100	50						1,000	500
0.0010	500						1,000	2,000
0.0001	5,000						1000	1,000