

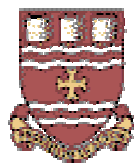
Biology 4250

Evolutionary

Genetics

October 2017

© 2017 by Steven M. Carr



Memorial
University of Newfoundland

Introduction

The laboratory portion of this course emphasizes hands-on experience with the measurement and experimental analysis of variation as it has evolved over time and within and among species. This handout includes the complete instructions for the major laboratory exercises in the course. I

Please consult the web page (<http://www.mun.ca/biology/scarr/Bio4250.html>) for information on the following aspects of the laboratory portion of the course.

The Home Page contains links to the supplementary material on each of the labs, including instructions for computation and interpretation of various statistical procedures (including **morphological distance**, **UPGMA**, **PCA**, and **phylogenetic analyses**)

Instructions for Lab Writeups describes the so-called **IMRAD** format (Introduction - Methods - Results - And - Discussion) to be used in the preparation of lab reports. This is described in more detail in Lobban & Schefter (1992). One of the main goals of this course is to introduce students to the idea of collecting, analyzing, and preparing data for publication in scientific journals: proper preparation of scientific MSS is a major emphasis for these labs.

Fundamentals of Statistical Hypothesis Testing reviews some basic statistical concepts, including the **null hypothesis** and **significance levels**. You are expected to be familiar with basic statistical concepts, which will be reviewed as needed in this course. Several of the exercises make use of **Excel** spreadsheets: ability to manipulate these is assumed.

Proper use of Reference Material describes correct (& incorrect) uses of outside materials in your lab reports. *In particular, use of material from laboratory reports from previous years is **plagiarism**.* You are responsible for knowing the University regulations governing **plagiarism** and academic dishonesty.

The laboratory exercises every year in response to student comments. I encourage you to let me know how I can make any of these exercises more relevant to your training as biologists, and where I can make these instructions clearer.

Steven M. Carr

St. John's, Newfoundland
October 2017

Laboratory #1: Computer Simulation of Natural Selection

This laboratory uses a computer program to simulate natural selection. You will use it to examine the consequences of directional selection on dominant and recessive alleles (Exercise 1), and to examine the conditions under which balancing selection maintains polymorphism in a population (Exercise 2). Review the lecture discussion of the **General Selection Equation** for complete dominance (Exercise 1) and heterozygote superiority ("**overdominance**") (Exercise 2).

The emphasis in this lab is how the pattern of genetic dominance affects the behaviour of alleles subject to negative selection. "Real" examples have been chosen to make the population genetics easier to visualize and discuss; the biology of these examples will be discussed in lecture.

Before coming to lab, read through each scenario and write down the initial conditions of \mathbf{q} , \mathbf{W}_0 , \mathbf{W}_1 , and \mathbf{W}_2 for each part of each exercise. Before running each simulation, check with the instructors to make sure that these values are correct.

Description of Computer Program NatSel

NatSel (v5.0) simulates natural selection on a one-locus, two-allele (**A** & **B**) model in a monoecious population with random union of gametes (the 'tide pool' model). A seven-digit random number is required when the program is first started. If the first digit is '0', allele frequencies will be called for, otherwise genotype counts will be called for. Required input parameters are the initial allele frequency of **B** ($\mathbf{q} = f(\mathbf{B})$), population size (**N**), and the fitness values of the phenotypes corresponding to each of the three genotypes (**AA**, **AB**, and **BB**): \mathbf{W}_0 , \mathbf{W}_1 , and \mathbf{W}_2 , respectively. Input fitness values are normalized to the maximum fitness, which is set as 1.00.

Exactly **N** individuals are born each generation. Each individual is generated by choosing two alleles at random to form one genotype, where the Expectation of a **B** allele is \mathbf{q} . Selection is applied by removal of that individual with an expectation of $(1 - \mathbf{W}_n)$, where \mathbf{W}_n is the normalized fitness of the appropriate phenotype. (This is called Viability Selection). Genotype proportions and phenotype survival are determined with a pseudo-random number generator, and are therefore not exactly equal to the expected values. **NatSel** thus includes a stochastic element alongside the deterministic equations. The allele frequency \mathbf{q} for the next generation is determined from the genotypes of all survivors. Each individual has a life-span of only one generation: it is convenient to think of the parental generation as dying immediately after gamete production.

Output is given in two lines per generation. Allele frequencies, genotype frequencies, and phenotype counts are given both "before" [at birth] and "after" [after selection]. The number of 'selective deaths' is $d(\mathbf{N}) = \mathbf{N}_{\text{after}} - \mathbf{N}_{\text{before}}$, where **N** is the sum of phenotype classes. The change in allele frequency in one generation ($\Delta\mathbf{q}$) is $d(\mathbf{q}) = \mathbf{q}_t - \mathbf{q}_{t-1}$, the change in \mathbf{q} ('before') from the previous to the current generation. Allele frequency (\mathbf{q}) may be displayed graphically at intervals of 10 generations, and are displayed cumulatively for up to 60 generations. Fitness values may be changed every 10 generations.

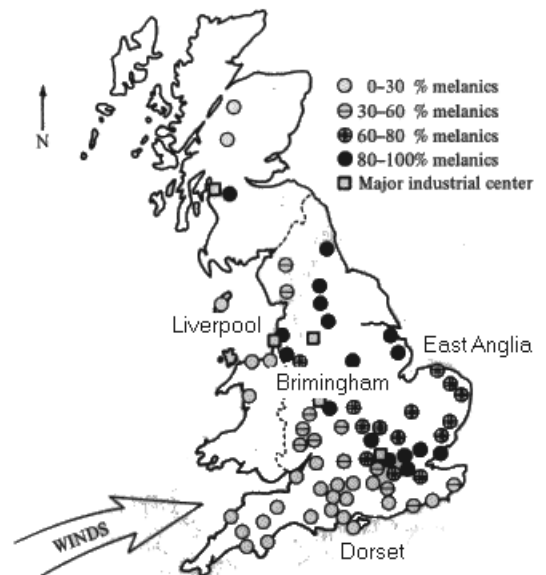
NatSel is written in FORTRAN77 and compiled with the Microsoft FORTRAN compiler (v.5.0) for PC-DOS machines. It has been supplemented by an Excel spreadsheet **GSM** that models natural selection according to the deterministic equations.

Lab #1, Exercise 1 - Directional & stabilizing selection in *Biston betularia*

The phenomenon of **industrial melanism** in *Biston betularia*, the common British pepper moth, is perhaps the best studied example of natural selection in the wild. Research on *B. betularia* has been carried out by H. B. D. Kettlewell and co-workers, and is the classical example of **Ecological Genetics**.

Pepper moths are nocturnal animals that during the day remain motionless on the surface of trees. Their major predators are diurnal predators, such as thrushes and other birds, that detect the moths visually and consume them in large numbers. The species is polymorphic for several colour phases. The common form is light grey and speckled, and is highly **cryptic** (camouflaged) when seen against the bark of the lichen-covered trees common in the rural English countryside. A second colour phase, a **melanic** (black pigmented) form called "*carbonaria*", is quite conspicuous on the light-coloured trees of the English countryside [left figure, below]. It has been demonstrated experimentally that the two colour phases are subject to differential predation by birds, according to degree of crypsis. The polymorphism has been determined to be under genetic control of a single locus with two alleles. The dominant allele corresponds to the melanic form and the recessive allele to the lighter form

Prior to the middle of the 19th century, new melanic mutants were rapidly eliminated whenever they arose, and were always rare. The melanic colour phase first began to appear in large numbers in the Manchester area about 1850, and is found in the insect collections of amateur naturalists from that period. This sudden increase coincided with the onset of the Industrial Revolution, which generated extensive air pollution in the form of vast quantities of black soot from the chimneys of coal-burning factories. In heavily industrialized areas (such as Manchester and Birmingham in the southeast of England)[right figure, below], forests of previously white, lichen-covered trees became completely blackened. Under these circumstances, the melanic form became more cryptic than the lighter form [middle figure, below]. By 1900, the proportion of melanic moths in the Manchester area exceeded 90%. Outside of these industrial areas, trees remained relatively uncontaminated and the lighter form continued to be more prevalent. In the latter twentieth century, the death of older trees and pollution control measures, such as installation of "scrubbers" on smokestacks, led to a partial restoration of pre-industrial environmental conditions, so that forests in some previously heavily polluted areas again became "white." The frequency of the melanic form has declined in these areas.



Lab #1, Exercise 1 - Directional & stabilizing Selection in *Biston betularia* (cont'd)

Directions

Start the program by typing "natsel" at the computer prompt. You will be prompted for a seven-digit random number seed [the first digit must be '0']. Then, enter the following values as instructed, in the proper format [confirm with the instructors that the values and format are correct]. Set the initial frequency of the 'light' allele [$q = f(\mathbf{B})$] as indicated. Set the population size $N = 9999$. Set the relative fitness values of the three genotypes **AA**, **AB**, and **BB** (**W₀**, **W₁**, **W₂**, respectively) as indicated in each of the scenarios below.

In each of the following five scenarios, **record $f(\mathbf{B})$ ('before') at the end of every ten generations**. Be prepared to explain to the instructors the changes between the 'before' and 'after' lines in any one generation.

(a) A rural area without pollution, pre-1850. [$q = 0.90$, 30% selection against the **dark** phenotype. Continue for up to 60 generations].

(b) A rural area undergoing the post-1850 Industrial Revolution. [$q = 0.90$, 30% selection against the **light** phenotype. Continue for 60 generations].

(c) A polluted industrial area, late 19th century. [$q = 0.10$, 30% selection against the **light** phenotype. Continue for 60 generations].

(d) An industrial area becoming a "smokeless zone," late 20th century. [$q = 0.10$, 30% selection against the **dark** phenotype. Continue for 60 generations].

(e) The British government has plans to establish Whittickshire Forest as a new wildlife preserve, to be located in a former industrial area in the north of England. The plan is to clean up the environment, and then to release a small number of light moths into the (predominantly dark) local population. What is the likely fate of such an experiment? [Re-start NatSel with a new random-number seed (enter '1' as the first digit). Start with 300, 400, or 500 homozygous light (**BB**) moths, **zero** heterozygous (**AB**) moths, and **9700, 9600, or 9500** homozygous dark (**AA**) moths, and 30% selection against the dark phenotype. Continue for 60 generations]. [Values for blanks will be supplied by the instructors]

Results

On the next page, RECORD $q = f(\mathbf{B})$ ('before') and $f(\mathbf{BB})$ at the beginning of the simulation, and at the end of every ten generations, for each scenario. Use these data to SKETCH a graph of the change in q and the frequency of the light phenotype over 60 generations. Use these data to answer the questions.

Lab #1, Exercise 1 - Results

RECORD $q = f(\mathbf{B})$ before [frequency of the recessive, 'light' allele] and $f(\mathbf{BB})$ [frequency of the 'light' phenotype] at 0, 10, 20, 30, 40, 50 & 60 generations. [Record all values to two decimals.]

	0	10	20	30	40	50	60
	$f(\mathbf{B})$ / $f(\mathbf{BB})$	$f(\mathbf{B})$ / $f(\mathbf{BB})$	$f(\mathbf{B})$ / $f(\mathbf{BB})$	$f(\mathbf{B})$ / $f(\mathbf{BB})$	$f(\mathbf{B})$ / $f(\mathbf{BB})$	$f(\mathbf{B})$ / $f(\mathbf{BB})$	$f(\mathbf{B})$ / $f(\mathbf{BB})$
a.							
b.							
c.							
d.							
e.							

SKETCH the graph of $f(\mathbf{B})$ and $f(\mathbf{BB})$ you obtained for scenarios **a, b, c, & d** (not e) (one graph for each scenario). Use these to answer the questions on the next page.

Lab #1, Exercise 1 - Questions (1.5 pts)

1. How rapidly does strong negative selection modify the frequency of a *rare dominant* (scenario **a**) allele as compared with a *rare recessive* (scenario **c**) allele? Can either sort of allele ever be completely eliminated from the population by selection? Why is there a difference?

2. How rapidly does strong negative selection modify the frequency of a *common recessive* (scenario **b**) allele as compared with a *common dominant* (scenario **d**) allele? How quickly does the **phenotype of the population** change (HINT: how many generations are required for the 'common' phenotype to become 'uncommon')? Why is there a difference?

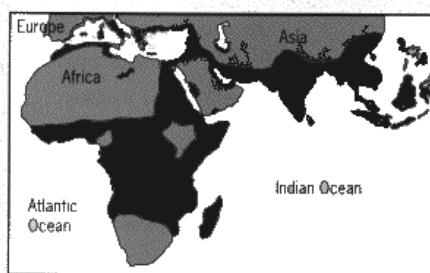
3. What is the difference in the behavior of a *very rare* (scenario **e**) as compared with a *rare* (scenario **d**) advantageous phenotype? What happens as the phenotype becomes less rare? Why?

Laboratory #1, Exercise 2 - Overdominant Selection of Hemoglobin A and S

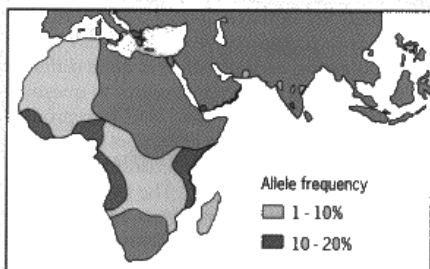
Standard human hemoglobin (Hemoglobin **A**, **HbA**) is a tetrameric protein composed of two identical alpha subunits and two identical beta subunits. Sickle-cell hemoglobin (Hemoglobin **S**, **HbS**) differs from HbA by a single DNA base substitution that leads to a single amino acid substitution in the beta subunit. Persons homozygous for the **S** beta chain (**SS**) show a severe blood disorder called **sickle-cell anemia**. Under conditions of reduced blood oxygen tension, HbS molecules form large, intracellular crystalline lattices that distort the red cells into crescent-shaped "sickles." The crystals impede blood flow through the capillaries (**infarction**) and the blockage results in episodes of acute, severe muscular pain (**infarctive crises**) as well as chronic hemolytic anemia. The fitness of **SS** homozygotes in the absence of adequate medical care is close to zero, since few survive to reproductive age. Persons heterozygous for the **S** beta chain (**AS**) show a much milder form of anemia, known as "**sickle cell trait**", which is seldom life threatening.

The sickle-cell allele occurs most commonly in human populations from West and Central Africa, in whom it reaches frequencies as high as 16%. These areas are also characterized by high incidence of malaria, and several lines of evidence indicate that the high frequency of this lethal allele is maintained by the increased resistance of **AS** heterozygotes to malaria. Besides the geographical correlation, (1) the frequency of the sickle-cell trait increases with age among African populations, (2) hospital records indicate increased morbidity due to malaria among **AA** homozygotes as compared with **AS** heterozygotes, and (3) clinical tests indicate that the malarial plasmodium parasite is less able to infect **AS** red cells. In malarial environments, the relative fitness of **AA** homozygotes seems to be substantially reduced with respect to the **AS** heterozygotes.

Sickle-cell anemia is a major health problem in the black community of North America, most of whose ancestors originated in West Africa. In the absence of malaria, the heterozygous advantage of the sickle-cell trait is absent, and the **S** allele is subject to uniform directional selection.



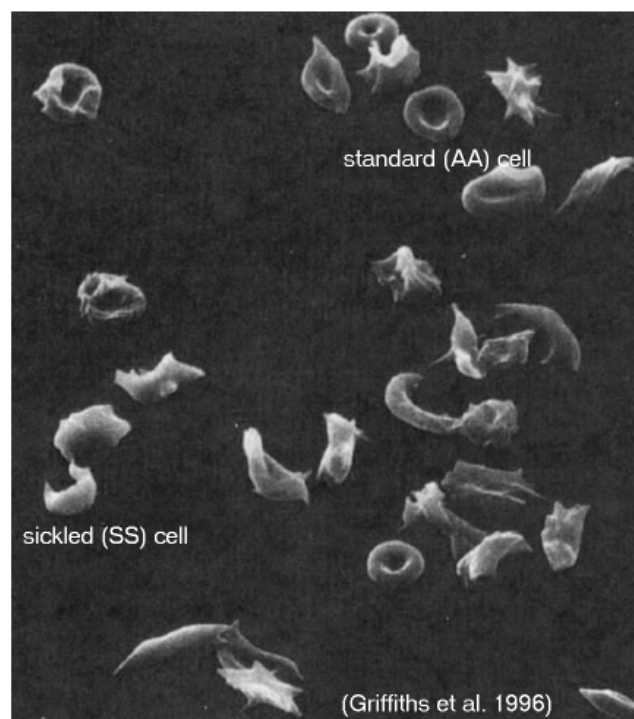
(a) Distribution of *Falciparum malaria*. ■



(b) Distribution of sickle-cell anemia allele (*HbS*).

Figure 27.7 Distribution of *Falciparum malaria* (a) and the distribution of the sickle-cell anemia allele (*HbS*) (b) in the Old World.

(Snustad et al. 1997)



Laboratory #1, Exercise 2 - Overdominant Selection between Hemoglobin A and S (cont'd)

Directions

Run NatSel with the following parameters. Let $q = f(S)$, where **S** is the allele for sickle-cell hemoglobin. Set the population size at $N = 9999$. The fitness of **AA** homozygotes is **1.00** in a non-malarial environment, or **0.50** in a malarial environment. In either environment, the fitness of **AS** heterozygotes is **1.0** and the fitness of **SS** homozygotes is **0.0**.

RECORD the values indicated in the table on the next page. The following scenario is continuous over 60 generations, and traces a human population as it moves between different selective environments.

(a) Consider a population of hunter-gatherers from a non-malarial environment in East Africa that initially carries the allele for sickle-cell hemoglobin at a frequency of $f(S) = 0.05$. [Enter **q**, **N**, and **W0**, **W1**, and **W2** as appropriate]. Continue for **20** generations.

(b) Suppose this population moves westward and begins farming in a high-malarial environment in West Africa. [Continue from part (a): change the fitness values as appropriate, and continue for **20** more generations.]

(c) West African blacks were brought to North America beginning about 400 years ago. Assume that North America is a non-malarial environment. [Continue from part (b): change the fitness values as appropriate, and continue for **20** more generations].

(d) How long would it take to eliminate the **S** allele if *all* persons with sickle-cell "trait" (**AS** individuals) voluntarily choose not to have children? If only 50% or 10% of carriers restrained? [Start with $q = 0.05$, set **W0**, **W1**, & **W2** as appropriate.] [**FOR FURTHER THOUGHT**: What are the social policy and ethical implications of such a solution?]

Results

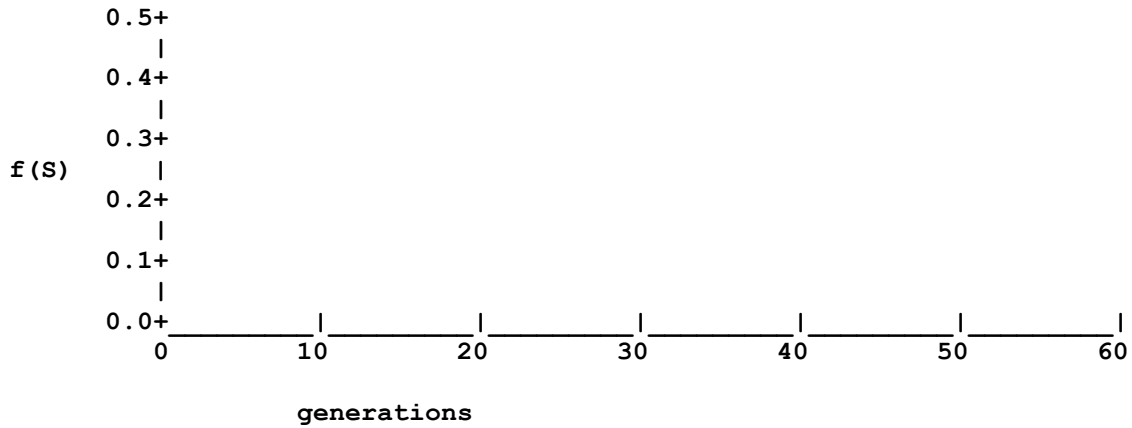
In the table on the next page, **RECORD** the following values at the beginning, and from the 'before' line at 5, 15, 25, 35, 45, and 55 generations. [Record $f(S)$ to three decimals.]

- i. $f(S)$ [= $f(B)$]
- ii. Total deaths [= $d(N)$]
- iii. Deaths due to sickle-cell [= # **BB**]
- iv. Deaths due to malaria [= $d(N) - \#BB$]
- v. **Mean fitness (W)** of the population
[= (total number of survivors after selection) / 10,000].

Laboratory #1, Exercise 2 - Overdominant Selection - Results

	generations					
	5	15	25	35	45	55
$f(S) [= f(B)]$						
Total deaths						
Sickle-cell deaths						
Malaria deaths						
Mean Fitness (W)						

SKETCH the graph of $f(S)$ from the computer screen. Use the graph and your data to answer the questions on the next page.



Laboratory #1, Exercise 2 - Questions (2.5 points)

1. In part (b), explain the patterns of deaths due to malaria or sickle-cell anemia after the population moves to the malarial environment (between generations 20 and 40).

2. CALCULATE the expected frequency of the **S** allele at equilibrium (between generations 30 & 40) (see your lecture notes or your text). Compare this with the observed frequency in this interval. Explain how $f(\mathbf{S})$ is maintained in this interval.

3. What happens to the mean fitness (**W**) of the population when the mode of selection changes at 20 generations (hint: what is the total population size *after* selection in generations 25 and 35)? In the malarial environment, at which value of $f(\mathbf{S})$ is the total population size after selection maximized? Explain.

4. Compare the rate of increase of the frequency of the **S** allele between generations 20 & 30 with the rate of decrease between generations 40 & 60 (compare the shape of the curves). Why is there a difference (hint: how does the fitness of the **AA** phenotype change with respect to **AS** in these two intervals)?

5. In part (c), approximately how many generations are required before $f(\mathbf{S})$ declines to the original value ($f(\mathbf{S}) = 0.05$)? Predict how long it would take to eliminate the **S** allele under these circumstances. How long would it take to eliminate the **S** allele if *all* persons with sickle-cell "trait" (**AS** individuals) voluntarily choose not to have children? If only 50% or 10% of carriers restrained? [**FOR FURTHER THOUGHT**: What are the social policy and ethical implications of such a solution?]

6. In this model of overdominant selection, selection against **AA** in malarial environments has been set unrealistically high in order to make the principles clearer. If the actual observed frequency of **S** in West African populations ($q = 0.16$) represents the true equilibrium frequency, calculate the selection coefficient associated with the **A** allele.

Computer Simulation #3: Evolutionary consequences of inbreeding (2 pts)

Individuals with a common ancestor are said to be **consanguineous** (the term means 'common blood'). The progeny of consanguineous parents are said to be **inbred**. As discussed in lecture, an inbred individual may have both alleles at a locus **identical by descent**, that is, both alleles are replicate copies of the exact same stretch of DNA in the common ancestor. Such an individual (or locus) is said to be **autozygous**. The probability that an individual will be autozygous at a particular locus is the **inbreeding coefficient (F)**. **F** increases according to the degree of relatedness of the consanguineous parents: for example, full sibs are related by 1/2 (they share half their alleles) and their offspring would have $F = (1/2)(1/2) = 1/4$. The offspring of a first-cousin marriage would have $F = 1/16 = 0.0625$: for second-cousins $F = 1/64$, and for third-cousins $F = 1/256$.

DIRECTIONS

Suppose that a genetic disease is caused by a rare recessive allele **B** that is lethal in the homozygous state. Consider two communities with different patterns of inbreeding. In one community, cousin marriages are considered unacceptable [**F = 0.00**]. In a second community, marriages between close relatives are more common [**F = 0.10**]. [In both communities: $N = 9999$, initial $f(B) = 0.05$, fitness of **BB** = 0.0].

Follow the change in $f(B)$ in each community over 60 generations. RECORD $f(B)$ before selection and the number of deaths [$d(N)$] after 1, 5, 10, 20, 40, and 60 generations. RECORD a running count of the **total** number of deaths after 5, 10, 20, 40, and 60 generations. DETERMINE the total number of deaths in the first 10 generations *versus* the number of deaths in the last 50 generations. [Record values to two significant places].

RESULTS

Community 1 (not inbred) F = 0.00	Generations					
	1	5	10	20	40	60
$f(B)$						
$d(N)$						
$\sum [d(N)]$						

Community 2 (Inbred) F = 0.10	Generations					
	1	5	10	20	40	60
$f(B)$						
$d(N)$						
$\sum [d(N)]$						

Briefly answer the following questions, based on the data you collected. Do not exceed the space provided.

1. Which community has a lower frequency of the condition in generations 1, 5, 10, 20, 40, and 60? [Which community is '**healthier**'?] Why does the change occur in each community, and why is there a difference between communities?

2. Which community show greater evolutionary '**fitness**' in generations 1, 5, 10, 20, 40, and 60? Why does the change occur in each community, and why is there a difference between communities?

3. Discuss the difference between the short-term 'genetic' consequences and the long-term 'evolutionary' consequences of inbreeding. Does inbreeding change allele frequencies?

Computer Simulation #4: Genetic Drift *versus* Selection in *Cepaea nemoralis* (2 pts)

Cepaea nemoralis, the European land snail, is the classic example of the interplay of deterministic (natural selection) and stochastic (genetic drift) forces in evolution.

Cepaea is widespread in western Europe and southern Great Britain, and 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. The polymorphism is controlled genetically: dark is dominant to light, and unbanded is dominant to banded.

Cepaea, like *Biston betularia*, is subject to differential predation by birds based on crypsis. The principle predators are song thrushes, which bring the snails to flat rocks (called "thrush anvils") where the birds break open the shell and eat the contents. Predation patterns and pressure can be measured by examining the phenotypes of shells at the "anvils." The incidence of the various shell types at the "anvils" varies according to habitat type. Broken light-colored shells are more common at anvils in dark "brown" backgrounds [e.g., bare soil], and are less common in light "green" backgrounds [e.g., grasslands]. Broken unbanded shells are more common at anvils in "high contrast" habitats [e.g., short grass fields], and are less common on "low contrast" [e.g., deep-shadow woodlands] habitats. Much of the variation among colonies seems to be correlated with these two environmental variables (see Fig. 13.6, p. 372 in Futuyma 1997). In other words, the shells that stand out visually are preferentially eaten by thrushes: those that blend in are not eaten.

However, **genetic drift** also plays a role. The snails occur in colonies of different sizes, from as few as ten to as many as several thousand individuals. Colonies are usually well-separated microgeographically, and gene flow between colonies is severely limited by the creatures' notoriously sedentary habits. Different shell forms typically co-exist in most colonies, and the frequency of each type varies greatly over short distances, not always in accordance with the predictions from crypsis. It has been shown that there is greater genetic variation (measured as variance) in shell patterns *among* smaller colonies than larger, an observation that is consistent with genetic drift. Some small colonies may even be 'fixed' for the 'wrong' (non-cryptic) shell pattern.

This exercise is a quantitative comparison of patterns of genetic drift in large and small populations, and the balance between **stochastic** and **deterministic** processes in determining genetic polymorphism in natural populations.

Genetic Drift *versus* Selection in *Cepaea nemoralis* (cont'd)

DIRECTIONS

In each of the following scenarios, set the initial frequency of the recessive (banded) allele $q = f(\mathbf{B}) = 0.50$. Do five replicate simulations with $N = 10$, and five simulations with $N = 200$ (representing five small and five large colonies). Set the selection coefficients as indicated. Follow each simulation for 20 generations or until fixation/loss of the banded allele. For each colony, record $q=f(\mathbf{B})$ at 0, 5, 10 and 20 generations (including cases of fixation or loss).

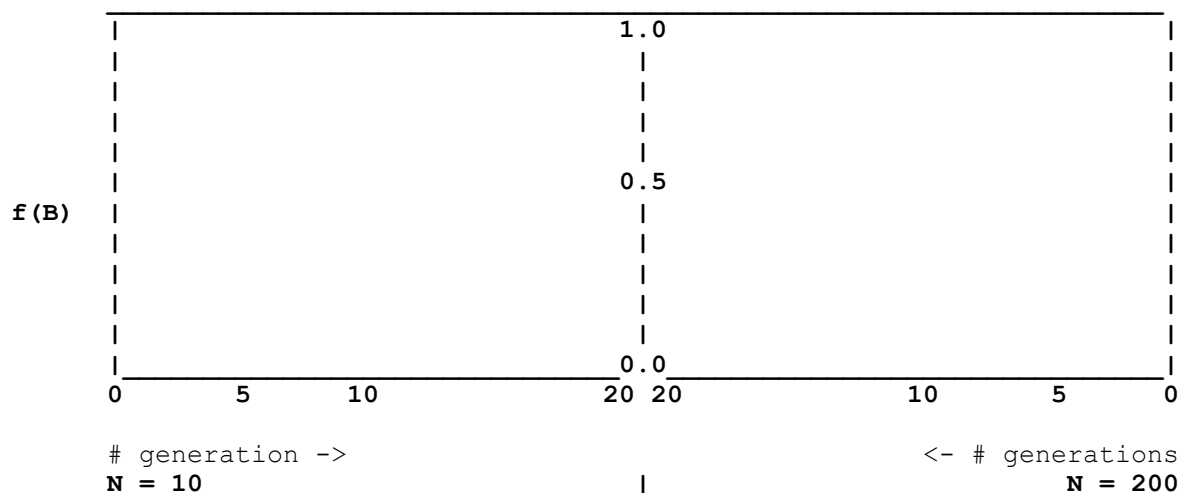
Scenarios

- (a) Drift with **no** selection: set all three fitness values at 1.00.
- (b) Drift with **moderate** selection: set 10% selection against the banded (**BB**) genotype.
- (c) Drift with **stronger** selection: set 15% selection against the **BB** genotype.
- (d) Drift with **weaker** selection: set 5% selection against the **BB** genotype.

FOR FURTHER UNDERSTANDING: Repeat scenario (a), with colonies of size $N = 200$ and $N = 1000$, and continue until loss or fixation is reached in each colony (the program limit is 1000 generations). Note how many generations are required to reach loss or fixation.

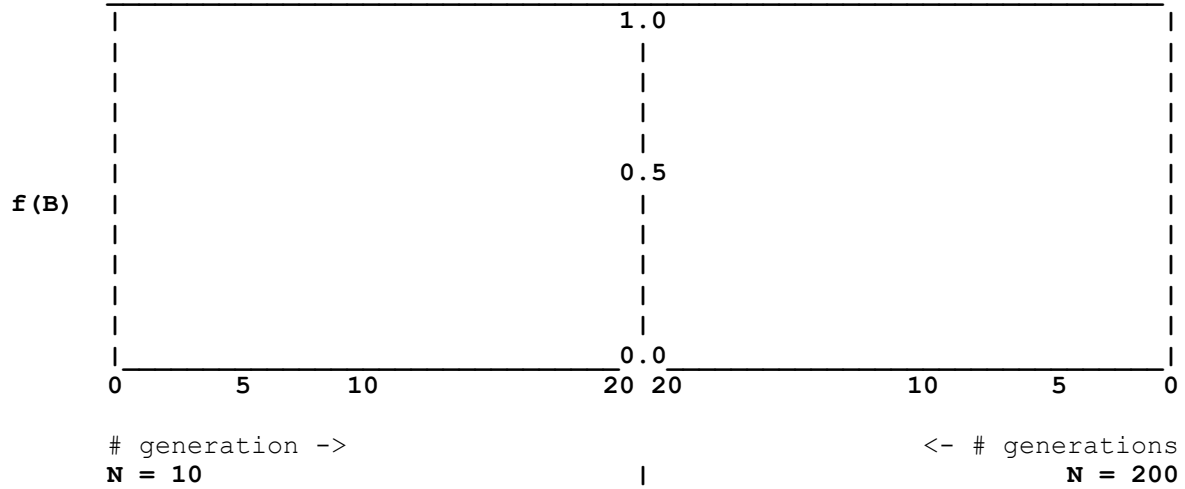
RESULTS

Sketch the graphs in parts (a) & (b) separately, using the format shown below. Plot the frequency of $q = f(\mathbf{B})$ for each colony at 0, 5, 10, and 20 generations; connect the dots from each colony. Plot all five colonies of the same size on one side of the graph, and the five colonies of the other size on the other side of the same graph. (See the examples in Fig. 11.3, p. 301, in Futuyma 1997) Note, for comparison with (b), the results in parts (c) & (d) *but don't sketch parts (c) & (d)*.

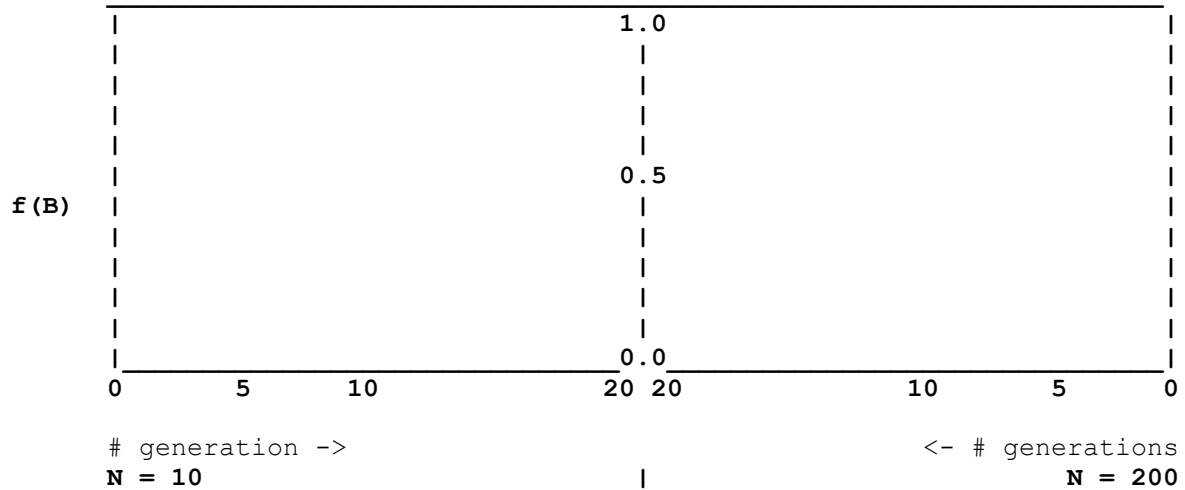


Laboratory exercise #4: Questions

Sketch your results in scenario (a)



Sketch your results in scenario (b)



1. Over 20 generations, in which size colony are allele frequencies more variable *among populations*?
 Why?

2. Can deleterious (unfavourable) alleles be fixed in a population? Under what circumstances? Explain.

3. After 20 generations of drift, is a species that is composed of *multiple small populations* 'more variable' than one composed of a *few large populations*? Explain.

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