Biology 4250 Evolutionary Genetics

16 September 2024

Biology 4250 - Principles of Evolution
Dr. Carr

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** ("*overdominance*") maintains polymorphism in a population (Exercise 2). Review the lecture discussion of modes of selection.

The emphasis in this lab is how patterns of genetic dominance affect 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 **q**, **W0**, **W1**, and **W2** for each part of each exercise. Before running each simulation, check with the instructors to make sure that these values are correct.

Description of Excel spreadsheet GSM

The **Excel** spreadsheet **GSM** 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). As shown, this accurately simulates selection for the *dioecious* species used here. Required input parameters are the initial **allele frequency q0** = f(**B**)) and the **fitness values W0**, **W1**, and **W2** of the *phenotypes* corresponding to each of the three *genotypes* **AA**, **AB**, and **BB**, respectively. Recall that ***fitness* is a *phenotype***, but may be assigned to a genotype *iff* each genotype has a different phenotype. Population size is infinite and is unchanged by selection. The model is therefore **deterministic**, and **q** approaches either fixation (**q=1**) or loss (**q=0**) **asymptotically**: **q** = 1.0000... or 0.0000... to four decimals can be thought of as such. [The program **NatSel** will explore selection with *finite* populations].

Three graphs of the tabled results are given. All three are plotted on the same **X** axis (time). (1) Change in allele frequency f**(B) = q**. (2) Change in genotype frequencies f(**AA**), f(**AB**), & f(**BB**), and for complete dominance models f("**A**") = f(**AA** + **AB**). (3) Change in population fitness, **W-bar**, calculated as defined, **W-bar** = **W0** x f(**AA**) + **W1** x f(**AB**) + **W2** x f(**BB**).

The spreadsheet runs as a single continuous scenario, with **W0**, **W1**, and **W2** changing periodically according to the environment. Changing the number of generations ("*as instructed*") for each scenario alters **q** at the end of that segment, which will affect behavior of the model in the next segment. For example, how does f(**B**) behave when **BB** is favored, but **q** is uncommon, rare, or very rare (**q = 0.1**, **q = 0.01** or **q = 0.001**)? As well, transitions between different values of **W** can be made smooth rather than abrupt. For example, a change from **W** = 1.0 to 0.70 in five 5% steps (0.95, 0.90, 0.85. 0.80, & 0.75) (or the reverse) will smooth the change in **q**.

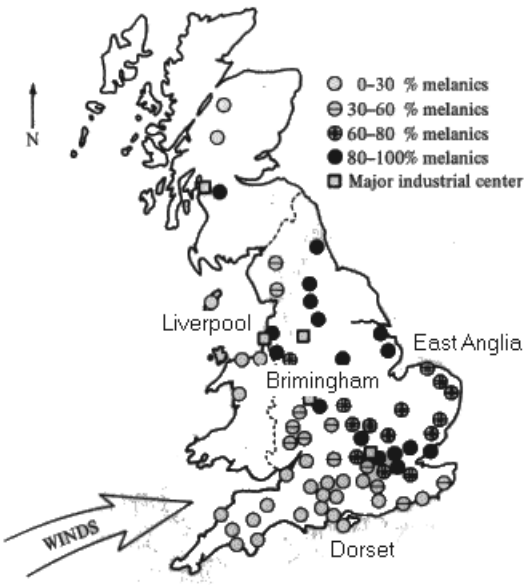
Biology 4250 - Principles of Evolution
Dr. Carr

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

The phenomenon of **industrial melanism** in *Biston betularia*, the common British pepper moth, is the classic 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 sit 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 English Midlands. A second colour phase, a **melanic** (black pigmented) form called "*carbonaria*", is quite conspicuous on the same light-coloured trees [compare forms in the 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 classes of alleles. The **dominant** alleles corresponds to the **melanic** form and the **recessive** alleles to the **lighter** form.

Knowledge of the frequency of the two forms has been obtained from amateur insect collections from the 19th century onward. Early on, melanic forms were rare, because the alleles were rapidly eliminated whenever they arose. The melanic form first began to appear in larger numbers in the English Midlands about 1850. This coincides 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 [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 forms in the Manchester area exceeded 90%. Outside of these industrial areas, trees remained relatively uncontaminated and the lighter form continued more prevalent. In the latter twentieth century, the death of older trees and institution of 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.



Biology 4250 - Principles of Evolution
Dr. Carr

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

Directions

Set the initial frequency of the recessive '***light***' allele [**q0** = f(**B**)] as indicated. Set the relative fitness values of the three genotypes **AA**, **AB**, and **BB** (**W0**, **W1**, **W2**, respectively) as indicated in the first scenario below.

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

(a) A rural area without pollution, pre-1850. [**q0** = 0.90, 30% selection against the **dark** phenotype]. Continue for 10 generations, to g = 10, ***OR*** as instructed. [What is the expected initial **f('A')**?]

(b) A rural area undergoing the post-1850 Industrial Revolution. [30% selection against the **light** phenotype]. Continue for an additional 60 generations, to g = 70, ***OR*** as instructed.

(c) A polluted industrial area, late 19th century. [30% selection against the **light** phenotype, as in (b)]. Continue for an additional 30 generations, to g = 100, **OR** as instructed.

(d) An industrial area becoming a "smokeless zone," late 20th to early 21st century. [30% selection against the **dark** phenotype. Continue for 100 generations, to g = 200].

Results

Save your Excel Spreadsheet file data: print out the graph at the end of all four scenarios. Label your axes: indicate the points where the scenario changed. The graphs will record

(1) **q = f(B)** over the course of the model,

(2) ***f*[AA+AB]** = *f*['**dark**' phenotype], *f*[**AA**], *f*[**AB**], *f*[**BB**] = *f*['**light**' phenotype], and

(3) **Wbar** = mean fitness of the population.

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*** (**q = 0.001**) as compared with a ***rare*** (**q = 0.01**, as in scenario **d**) advantageous phenotype? What happens as the phenotype becomes less rare? Why?

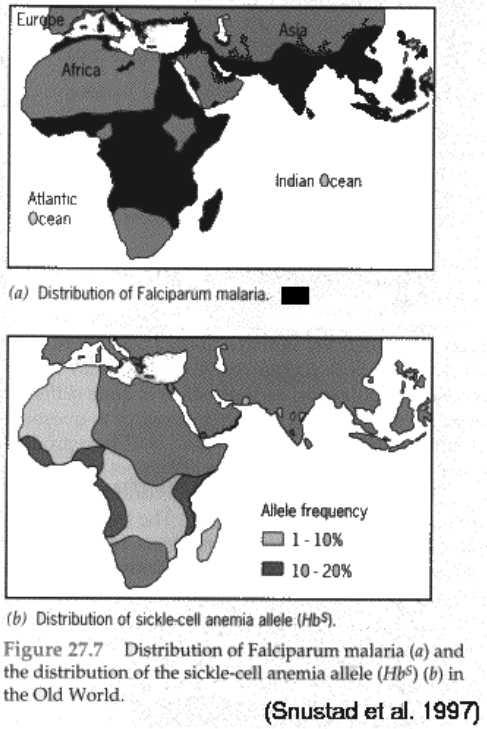
Biology 4250 - Evolutionary Genetics
Dr. Carr

Laboratory #1, Exercise 2 - Balancing 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 **SNP** that leads to a single amino acid substitution in the **beta** subunit. Persons homozygous for the **S** beta chain allele (**SS**) show a severe blood disorder called **sickle-cell anemia**. Under conditions of reduced blood oxygen tension, **HbS** molecules form large, crystalline lattices that distort red blood cells into crescent-shaped "*sickles*." The crystals impede blood flow through capillaries, which 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**) allele show a much milder form of anemia, known as "**sickle cell trait**", which is seldom life threatening. [Distinguish '*trait*' from '*anemia*'.]

The sickle-cell allele occurs most commonly in human populations from West and Central Africa, in whom it reaches frequencies as high as f(**S**) = 0.20. These areas are also characterized by high incidence of **malaria**, and several lines of evidence indicate that f(**S**) is maintained by the increased resistance of **AS** heterozygotes to malaria. Besides the geographical correlation, (1) the frequency of sickle-cell trait increases with age among African populations, (2) hospital records indicate increased morbidity due to malaria among **AA** homozygotes relative to **AS** heterozygotes, and (3) laboratory 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. Thus, **A** & **S** are subject to **balancing selection**.

Sickle-cell anemia is a major health and social problem in Africa and in black communities 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 lost, and the **S** allele is subject to **directional selection**.

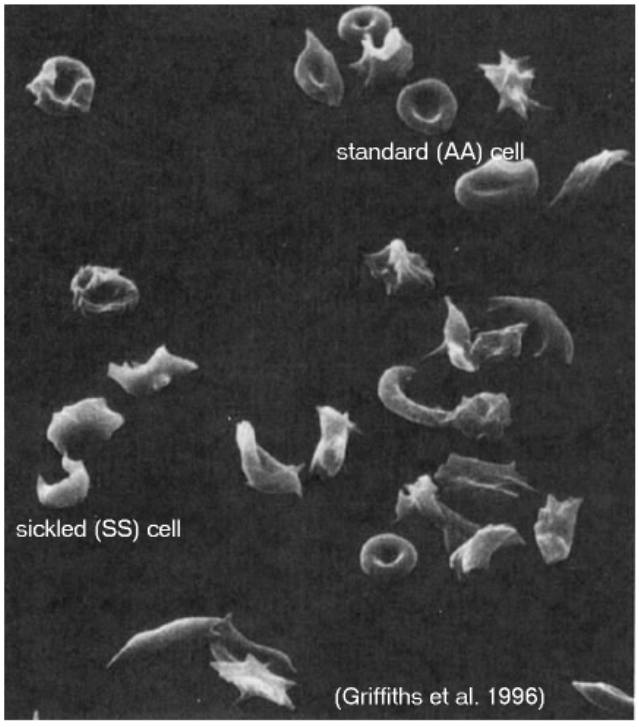


(a) Distribution of Falciparum malaria.

(b) Distribution of sickle-cell anemia allele (Hb^S).

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

(Snustad et al. 1997)



Biology 4250 - Evolutionary Genetics
Dr. Carr

Laboratory #1, Exercise 2 - Balancing Selection between Hemoglobin A & S alleles (cont'd)

Directions

Note that the column heads are re-labelled as **AA**, **AS**, & **SS**. Run **GSM** with the following parameters. Let **q** = f(**S**), where **S** is the allele for sickle-cell hemoglobin. 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**. [These values allow the model to run in a more easily analyzed manner. Realistic values for f(**S**) and **W** have been estimated from clinical data, see **Question #6** below. They may also be used in the **GSM** model]

The following scenario is continuous over many generations, and traces a human population as it moves between different selective environments. The number of generations in each environment may be adjusted upward

(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 indicated]. Continue for **20** generations, ***OR*** as instructed.

(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 indicated, and continue for **40** more generations, ***OR*** as instructed.]

(c) West African blacks were brought to North America beginning about 400 years ago [how many generations is this?]. Assume that North America is a non-malarial environment. [Continue from part (b): change the fitness values as appropriate, and continue for **40** more generations, ***OR*** as instructed.].

(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 **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? How do recent developments of **CRISPR** technology change the scenario presented here?]

Results

Save the Excel Spreadsheet(s). **Copy & Paste graphs** as necessary to answer the questions.

Remember to label you axes.

Biology 4250 - Evolutionary Genetics NAME _____ Group _____
Dr. Carr

Laboratory #1, Exercise 2 - Questions

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(**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(**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(**S**) declines to the original value (f(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? What 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 balancing selection, selection against **AA** in malarial environments has been set unrealistically high in order to make the principles clearer.

a. 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.
b. Repeat the scenarios with ***realistic selections coefficients*** as calculated in [https://www.mun.ca/biology/scarr/NS_07-Box7smc.html].
You will likely need to double the number of generations at each stage.

Biology 4250 - Evolutionary Genetics
Dr. Carr
Laboratory #1, Exercise 3 - Natural Selection in a Variable Environment