

A Project-Based Approach to Teaching Complex Population Genetics to Undergraduates.

Richard M. Kliman

Department of Biological Sciences
Kean University
1000 Morris Avenue
Union, NJ 07083
e-mail: rkliman@kean.edu

ABSTRACT: In today's cramped undergraduate biology curriculum, adequate coverage of population genetics is difficult. Yet, this field is becoming increasingly important in areas as diverse as anthropology and medical genetics. Educators also face the challenge of teaching students to develop and test hypotheses, and to understand stochastic processes. I describe an approach to teaching population genetics that allows students to gain an appreciation for the breadth of the field, while also developing and testing specific hypotheses. The keys to the approach are a computer simulation (available by download), group exercises, and individual projects followed by class presentations.

KEYWORDS: Population genetics, evolution, computer simulation, hypothesis testing.

INTRODUCTION.

As the emphasis in undergraduate general genetics courses shifts increasingly from classical genetics to molecular genetics, adequate coverage of population genetics becomes more difficult. This is unfortunate, because (i) a proper understanding of microevolutionary processes requires some understanding of population genetics beyond Hardy-Weinberg equilibrium; (ii) population genetics offers opportunities to generate and test hypotheses using quantitative methods; and (iii) population genetics is one of the few subjects that introduces students to stochastic processes. Also, with the arrival of the "post-genome" era, population genetics methods are increasingly important in medical genetics research (Reich *et al.*, 2001).

Misunderstanding of population genetics is widespread, and part of the problem can be attributed to a lack of appreciation of stochastic processes. For example, highly publicized work by Vigilant and co-workers (1991) indicated that extant human mitochondria arose from a common ancestor living about 200,000 years ago. However, the coupled processes of recombination and meiosis generate stochastic variance in gene histories. The time back to the most recent common ancestor (MRCA; see Box 1 for this and other abbreviations/symbols used in this paper) of the autosomal *PDHA* gene seems to be an order of magnitude greater (Harris and Hey, 1999). These findings are perfectly compatible. While gene histories are influenced by population histories, the

time to the MRCA of a given gene clearly does not tell us when a population or species arose.

Box 1. *Important abbreviations and symbols used in this paper.*

MRCA: Most Recent Common Ancestor.
N : Population size.
N_e : Effective population size.
f : Ploidy.
μ : mutation rate.
s : selection coefficient.

Consider, also, two ways in which a misunderstanding of population genetics affects our understanding of medical genetics. First, nonrandom associations (i.e., linkage disequilibrium) between alleles of different genetic loci are often used to map "disease" genes. However, linkage disequilibrium arises for reasons other than physical linkage, including nonrandom mating and natural selection on epistatically interacting genes, and proper interpretation of association studies must take this into consideration (Nordborg and Tavaré, 2002). Second, while the persistence of disease-causing alleles is often attributed to stabilizing selection (e.g., on β -globin for malaria resistance), some form of selection-mutation balance may be a better explanation in certain cases. This is particularly reasonable when selection is weak

and mutations can occur in many locations along the relevant gene(s).

MORE ON STOCHASTIC PROCESSES: GENETIC DRIFT AND COALESCENCE

In addition to addressing common misconceptions, teaching population genetics provides the opportunity to reinforce other skills important for biology majors: applied mathematics and appreciation for stochastic processes. The classic example is genetic drift. Consider a population in which the A allele is at a frequency of 0.6. If the population is closed and of finite size, and if all other Hardy-Weinberg assumptions are met, then the fixation probability for the A allele is 0.6. This can be shown mathematically and by computer simulation. A simulation must be run many times to demonstrate this; the estimate of the probability will approach 0.6, and the standard error will approach 0.0, as the number of iterations approaches infinity. The ramifications of this are two-fold. First, to accurately estimate probabilities, simulations must be run a sufficient number of times. A few years back, when doing some Monte Carlo statistics, I settled on 10,000,000 iterations. No one questions 10,000,000 iterations. Second, while simulation can provide insight into the often broad *distribution* of population histories under a given set of assumptions, the *actual* history, being analogous to a single iteration, can not be inferred with much confidence. In fact, ignoring this point will lead to overconfidence in the placement in time of an MRCA (see Templeton, 1993).

A slightly more advanced example of applied probability involves the fate of a new mutation. There are three possible scenarios, the one most applicable determined by the joint influence of natural selection and population size (Hartl and Clark, 1997; Li, 1997).

- If the mutation is selectively neutral and if all of the Hardy-Weinberg assumptions (other than infinite population size) are met, the mutant allele has a $1/fN$ chance of being fixed, where N is the census population size and f is a scaling factor associated with the effective ploidy of the gene. [For an autosomal gene, f is 2. For an X-linked gene, f is somewhere between 1 and 2, depending on the sex ratio of the population. For a Y-linked or organellar gene, f is between 0 and 1.] If the gene is autosomal, the probability of fixation assuming selective neutrality is $1/2N$; it will usually be lost to genetic drift.
- If the mutation is deleterious, its probability of fixation is above 0.0 and no more than $1/2N$. This depends on two parameters: the selection coefficient, s , and the population size. If selection is effective, the allele will likely be removed quickly from the population. If selection is ineffective (because the product of s and N falls

below 1.0), the mutation is *effectively* neutral and its probability of fixation is $1/2N$.

- If the mutation is beneficial, its probability of fixation is below 1.0 and no less than $1/2N$. Again, this depends on s and N . If s is small and selection is effective, the probability of fixation is approximately $2s$. At higher values of s , the probability of fixation drops below $2s$, asymptotically approaching 1.0. If the mutation is effectively neutral due to the joint values of s and N , the probability of fixation is $1/2N$.

While all of this can be derived mathematically, it is not reasonable to assume that an undergraduate biology major will have the requisite background.

Another advanced concept deals with the time to fixation of a single gene copy. "Coalescence theory" deals with the time depth of genealogies, particularly with respect to estimating the number of generations back to the single gene copy that serves as the MRCA of all extant gene copies (Hudson, 1990). While the mathematics of coalescence theory are often beyond undergraduates, predictions based on the theory can be confirmed by means of computer simulation. Similarly, we can assess how certain factors influence effective population size (N_e) and, by extension, the time to fixation of gene copies in a population. The results are often counterintuitive, all the more reason to encourage students to investigate this area.

We are starting to consider some pretty hairy issues in population genetics. Still, students can gain insight into these problems, and we can achieve this goal with a surprisingly small time investment. Central to this is (i) the use of a computer simulation developed specifically for the instructional goals, and (ii) class presentations of individual research projects.

METHODS.

A. AN EMPIRICAL APPROACH TO COMPLEX CONCEPTS IN POPULATION GENETICS.

Even if a student has had no prior exposure to population genetics, it takes little time to explain Hardy-Weinberg equilibrium. I use a simple gamete pool approach that gets the point across in about 15-20 minutes.

Following the math, we proceed to a computer simulation I specifically designed for this laboratory exercise. *EvolGenius* can be downloaded along with an HTML-based manual from <http://www.kean.edu/~rkliman/evolgenius.htm>. [*EvolGenius* is currently compiled to run in a MS-DOS window, and is not currently designed to run efficiently on a Macintosh computer.] The students perform three group exercises: (1) genetic drift; (2) fixation of a new beneficial mutation; and (3) mutation-selection balance assuming no dominance. These exercises are described later. The third exercise requires that students import the *EvolGenius* output into a spreadsheet program, thus ensuring that each student has had the opportunity to

work with the data in this manner before proceeding with an independent project.

Over the next several weeks, individual students (or pairs) choose a more sophisticated project. In general, I supply several suggestions (see Box 2), but students are free to come up with their own ideas. One rule is that no student work on a project too similar to

that of another. The students consult relevant literature and present their findings in standard format (introduction, methods, results and discussion) to the class. Thus, while individual students carry out just one study, they are exposed to a variety of population genetics problems through the presentations of their peers.

Box 2. Some suggested independent projects using *EvolGenius*.

- **Influence of population size on the effectiveness of natural selection.** Population genetics theory shows that the probability of fixation of a beneficial mutation depends on population size (N) and the selection coefficient (s). Selection will not be effective if the product of N and s is below 1.0. The student should begin with one copy of the beneficial allele (see Exercise 2 in text). A few different values of N can be chosen (e.g., 20, 50, 100 and 500) and each used with a few different values of s (e.g., 0.005, 0.01, 0.05 and 0.1). Note that the neutral probability of fixation is $1/2N$.
- **Mutation-selection balance with complete dominance.** The equilibrium deleterious allele frequency for complete dominance (where the deleterious allele is recessive) is $(\mu/s)^{0.5}$. The student can essentially follow the methods used in Group Exercise 3 (see text).
- **Number of generations needed to achieve equilibrium allele frequencies under selection-mutation balance with no dominance.** This could depend on (a) the starting allele frequency, (b) the equilibrium deleterious allele frequency and (c) the selection coefficient. Following Exercise 3, the equilibrium deleterious allele frequency can be calculated from μ and s . For a given initial deleterious allele frequency and a given combination of μ and s , the simulation can be stopped at a set of arbitrary times (e.g., 25, 50, 75, 100 generations, etc.). Deleterious allele frequency can be plotted against time, providing some insight into the time required to reach equilibrium. A more objective measure might be the time it takes for 50% of the iterations to reach equilibrium.
- **Influence of selection on one gene on the fixation rate of another gene (2 different projects).** 1. For a given selection coefficient for gene A , how does the fixation rate of gene B respond to changing linkage map distance? 2. For a given linkage map distance, how does fixation rate of gene B respond to changing the selection coefficient on gene A ? The problem can be addressed in a few different ways, depending on whether or not mutation is allowed. If mutation is allowed, the problem can be simplified slightly by starting at the expected equilibrium allele frequencies of the gene subject to selection. Regardless, the deleterious allele (e.g., a) should be present at fairly high frequency at the start of the simulation. Note that for a given value of N , the average $Beta$ gene copy fixation time is expected to be $4N$. Thus, the effective population size of the $Beta$ gene can be estimated from the average fixation time (i.e., fixation time divided by four). The results of this exercise can be counterintuitive. For example, weaker selection will allow deleterious alleles to reach a higher frequency; this decreases the number of individuals lacking the deleterious allele and, by extension, effective population size. For this reason, the fixation time of $Beta$ gene copies may be shorter when selection on $Alpha$ is weaker, but still effective.
- **Evolutionary advantage of sexual reproduction.** In the absence of recombination, novel genotypes can arise only by mutation. In the presence of recombination, novel genotypes can arise without new mutation. Start with a population 50% $AAbb$ and 50% $aaBB$. Relative fitnesses can be set such that, for a given value of s , relative fitness is 1.0 minus s times the number of lower-case alleles. Thus, the $AABB$ genotype, which is not present at the start, confers maximum fitness. In one set of simulations, mutation rates can be set to zero and recombination rates can vary (e.g., from 0.001 to 0.01). In another set of simulations, recombination rates can be set to zero and mutations rates varied (perhaps using the same set of values). It should be noted that, for a given value, mutation often "beats" selection; that is, the time to fixation of the $AABB$ genotype takes less time with mutation. However, it should also be noted that a mutation rate of 0.001 mutations/gene/generation is several orders of magnitude higher than occurs in nature, while a corresponding recombination map distance of 0.001 (0.1 cM) is quite low. For most randomly chosen genes, the recombination map distance will be 50 cM, since it is unlikely that the genes are syntenic. With this in mind, it is easy to show that adaptive evolution is facilitated better by recombination than by recurrent mutation.
- **What is the probability that an individual is the MRCA for gene B if it carries a deleterious allele at gene A ?** Under strict neutrality, all individuals have an equal chance of being the MRCA for a gene. Obviously, if an individual has a deleterious allele for that gene, it is less likely to be the MRCA for that gene. However, does this affect its ability to be the MRCA for another gene? One approach to this problem is to start with a population 50% $AABB$ and 50% $aaBB$. Thus, the only haplotypes present at the start are AB and aB , and the expectation from neutrality is that, for i iterations of the simulation, the MRCA for the $Beta$ gene should be AB $i/2$ times and aB $i/2$ times. Relative fitnesses can be set to 1.0 ($AABB$), $1.0-s$ ($AaBB$) and $1.0-2s$ ($aaBB$). A range of values for s , each used with recombination map distances of 0.0, 25.0 and 50.0 cM, can be tested. The result of interest is the fraction of times the MRCA of the $Beta$ gene had the aB haplotype.

The point of this exercise is not to make the students population genetics experts. Rather, the goal is to teach some basic population genetics and to foster an appreciation for more complex principles

B. A DESCRIPTION OF THE COMPUTER SIMULATION.

The program allows the student to run thousands of iterations of a given simulation in a fairly short

period of time. [The program, like others, can be set to show allele and genotype frequencies at fixed intervals, but this is not its strength.] It tracks the histories of two autosomal genes, *Alpha* and *Beta*, in a diploid, hermaphroditic population of constant size N ; self-fertilization is allowed. Thus, for a population of N individuals, there will be $2N$ copies of each gene in the population. Each gene can have one or two distinct alleles. Each gene copy gets a unique identifier at the start of the simulation, so the program can track the fate of each starting gene copy and determine when all copies of a gene are descended from a single copy that existed at generation 0. The student sets the number of individuals with each of the nine possible genotypes up to a combined total of 1000 individuals. The student can also set the following:

- **The relative fitness of each genotype**, ranging from 0.0 to 1.0.
- **The mutation rates/generation** of A to a , a to A , B to b and b to B .
- **The linkage map distance** between the *Alpha* and *Beta* genes, from 0.0 to 50.0 cM.
- **The maximum number of generations** that the simulation will run. Fixation of a neutral gene copy is expected to take $4N$ generations on average, and it is extremely unlikely that the simulation will ever require 30,000 generations (the default maximum) to complete, given a maximum N of 1000. An exception, and the motivation for a limit, is when the user creates some form of mutation-selection balance, such that neither allele of a particular gene is fixed.

- **The number of iterations** using the same starting conditions and parameters. The student can run the same simulation thousands of times, with the output of each iteration listed sequentially in a file suitable for import into a spreadsheet program.

Figure 1 shows the main menu page of *EvolGenius*. The program output (see Figure 2 for an example) will include a list of the parameters at the top of the file. It will also report the following in table format:

- the iteration number;
- the allele (A or a) of gene *Alpha* fixed (if any);
- the final frequency of the A allele;
- the number of generations required to fix an allele of gene *Alpha*;
- the number of generations required to fix an ancestral gene copy of gene *Alpha*, as well as the following information: the ID number of the MRCA, the haplotype to which the MRCA was initially linked (AB , Ab , aB or ab) and the diploid genotype in which it was initially found; and
- the analogous data for gene *Beta*.

Means and standard deviations are given for the fixation times. The program also reports the number of times that a given allele is fixed. While some of this output is usually unnecessary, there are a few conceivable exercises when the data would be required. The program keeps track of the information regardless, so there is no point in discarding it.

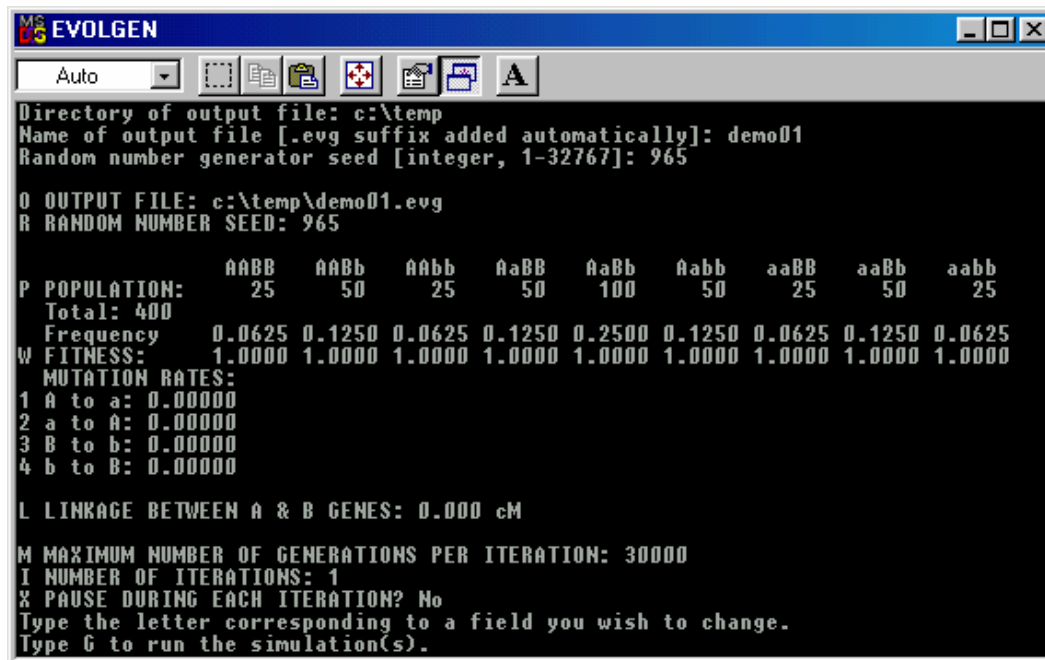


Figure 1. Main menu of *EvolGenius*.

```

OUTPUT FILE: c:\temp\demo02.evg
RANDOM NUMBER SEED: 965

      AABB  AABb  AAbb  AaBB  AaBb  Aabb  aaBB  aaBb  aabb
POPULATION:      24    1    0    50    0    0    25    0    0
Total: 100
Frequency  0.2400 0.0100 0.0000 0.5000 0.0000 0.0000 0.2500 0.0000 0.0000
FITNESS:    0.9000 0.9500 1.0000 0.9000 0.9500 1.0000 0.9000 0.9500 1.0000

MUTATION RATES:
A to a: 0.00000
a to A: 0.00000
B to b: 0.00000
b to B: 0.00000

LINKAGE BETWEEN A & B GENES: 50.000 cM

MAXIMUM NUMBER OF GENERATIONS PER ITERATION: 30000

..... Haplo- Geno- ..... Haplo- Geno-
..... type.. type. .... type.. type.
..... ID#.of from.. from. .... ID#.of from.. from.
..... Allele Gens.. Gens.. which. which Allele Gens.. Gens.. which. which
..... of.... to.fix Alpha. to.fix fixed. fixed of.... to fix Beta.. to fix fixed. fixed
..... Gene.. Gene.. copy.. Gene.. Alpha. Alpha. Gene.. Gene.. copy.. Gene.. Beta.. Beta.
Iter- Alpha. Alpha. Final. fixed. Alpha. copy.. copy. Beta.. Beta.. Final. fixed. Beta.. copy.. copy.
ation fixed. Allele f(A).. (MRCA) copy.. arose. arose fixed. Allele f(B).. (MRCA) copy.. arose. arose
-----
0001 A      54 1.0000      49 146 Ab  AaBB b      152 0.0000      49 152 Ab  AaBB
0002 a     329 0.0000     183 329 aB  aaBB B      1 1.0000      58 493 AB  AaBB
0003 a      94 0.0000     131 156 aB  AaBB B      1 1.0000     174 176 aB  aaBB
0004 a     223 0.0000     91 501 aB  AaBB B      1 1.0000     186 381 aB  aaBB
0005 a     439 0.0000    111 439 aB  AaBB B      6 1.0000     183 242 aB  aaBB
0006 a     677 0.0000     89 677 aB  AaBB B      6 1.0000     48 224 AB  AaBB
0007 A      43 1.0000     11 202 AB  AaBB B      7 1.0000     72 371 AB  AaBB
0008 a     258 0.0000    170 258 aB  aaBB B      1 1.0000     52 636 aB  AaBb
0009 A     167 1.0000     42 336 AB  AaBB B      5 1.0000     56 292 AB  AaBB
0010 a      89 0.0000     81 269 aB  AaBB B      2 1.0000    184 845 aB  aaBB

NUMBER          10              10              10              10
MEAN           237.3          331.3          18.2           381.2
STD. DEV.      200.5          167.0          47.1           220.9

A allele fixed 3 times.
a allele fixed 7 times.
Neither allele fixed 0 times.

B allele fixed 9 times.
b allele fixed 1 times.
Neither allele fixed 0 times.

```

Figure 2. Sample *EvolGenius* output file.

C. GROUP EXERCISES FOR WEEK ONE.

Following derivation of Hardy-Weinberg equilibrium, each student participates in three group "experiments," one of which requires that the *EvolGenius* output be exported to a spreadsheet program for further analysis. To be efficient, I provide a handout listing the values that each student will enter in the program.

1. Genetic drift using a constant *N* neutral model with no mutation.

Students estimate the probability that an allele will be fixed as a function of its initial frequency. The only exception to the assumptions of Hardy-Weinberg equilibrium is finite population size. Selective neutrality is obtained by setting the relative fitnesses of all nine genotypes to 1.0 (the default). Mutation rates are also left at 0.0; if this were not the case, allele fixation probabilities would also be a function of

mutation rates. Linkage map distance is set to 50 cM; that is, the *Alpha* and *Beta* genes are independently assorting. The number of iterations is set arbitrarily high (e.g., 200) to allow for the calculation of standard deviations and standard errors of fixation times. While *N* is the same for each student (100 in this case), the genotype frequencies differ (see Table 1). Note that the allele frequencies for students 1 through 5 differ. The initial allele frequencies for student 3 and students 6-9 are the same (50% each allele), but arranged into different initial genotype frequencies. This exercise shows students that, under this neutral model, (a) the probability of an allele's fixation is its initial frequency and (b) it doesn't matter how the alleles are initially apportioned into genotypes. Of course, the estimates of the probabilities are not likely to be perfect. This is an additional take-home message from the exercise: computer simulations give *estimates*, not precise values.

Table 1. Genotypes for Exercise 1.

Student	<i>AABB</i>	<i>AABb</i>	<i>AAbb</i>	<i>AaBB</i>	<i>AaBb</i>	<i>Aabb</i>	<i>aaBB</i>	<i>aaBb</i>	<i>aabb</i>
1	90	0	0	0	0	0	0	0	10
2	70	0	0	0	0	0	0	0	30
3	50	0	0	0	0	0	0	0	50
4	30	0	0	0	0	0	0	0	70
5	10	0	0	0	0	0	0	0	90
6	0	0	0	0	100	0	0	0	0
7	0	0	50	0	0	0	50	0	0
8	25	0	25	0	0	0	25	0	25
9	0	25	0	25	0	25	0	25	0

This exercise also shows that, regardless of initial allele frequencies, the time to fixation of single gene copies is approximately $4N$ generations for an autosomal gene in a diploid organism (assuming a constant N neutral model). Since all nine simulations have N set to 100 individuals, gene copy fixation takes, on average, approximately 400 generations. There will be considerable variation, but in most cases the mean will be within the standard error of 400. This provides an opportunity to discuss means, variance and statistical confidence.

2. Probability of fixing a new, beneficial mutation.

Population genetics theory predicts that the probability of fixing a new, beneficial mutation is approximately twice its selective advantage (s), at least when s is low. All students begin with the same genotype frequencies: 99 *AABB* and 1 *AABb*; all other genotype frequencies are set to 0. The single b allele represents the new beneficial mutation. Notice that gene *Alpha* is irrelevant if the A allele is fixed in the population and the mutation rate from A to a is set to 0.0. The mutation rates for the *Beta* gene must be set to zero, so that the novel allele enters the population only at the start of the simulation. Again, an arbitrarily high number of iterations is chosen.

The simplest way to model this problem is to assume no dominance; that is, the heterozygote has a fitness intermediate to either homozygote. For a selection coefficient of s , the fitnesses of the three genotypes should be set to 1.0 (*AAbb*), $1.0-s$ (*AABb*) and $1.0-2s$ (*AABB*). [In higher-level population genetics, the relative fitnesses would be 1.0, $1.0-0.5s$ and $1.0-s$, respectively. However, defining s as we have done here works just as well and simplifies matters somewhat.] The settings shown in Table 2 show values of s ranging from 0.00 to 0.16 in increments of 0.02. The last six genotypes never occur in the simulation, so their relative fitnesses are irrelevant. The linkage distance between the two genes is also irrelevant, though setting this to 0.0 cM will speed the simulation.

With a population size of 200 gene copies, population genetics theory predicts that selection should be effective with all non-zero selection coefficients used in this example. That is, $2Ns$ ranges from 4.0 to 32.0, and selection should be effective if $2Ns$ is sufficiently greater than 1.0 (Ohta, 1996). Because there is no selection, the probability of fixation for student 1 should be $1/2N$, or 0.005 in this case.

Table 2. Relative fitnesses for Exercise 2.

Student	<i>AABB</i>	<i>AABb</i>	<i>AAbb</i>	<i>AaBB</i>	<i>AaBb</i>	<i>Aabb</i>	<i>aaBB</i>	<i>aaBb</i>	<i>aabb</i>
1	1	1	1	1	1	1	1	1	1
2	0.96	0.98	1	1	1	1	1	1	1
3	0.92	0.96	1	1	1	1	1	1	1
4	0.88	0.94	1	1	1	1	1	1	1
5	0.84	0.92	1	1	1	1	1	1	1
6	0.80	0.90	1	1	1	1	1	1	1
7	0.76	0.88	1	1	1	1	1	1	1
8	0.72	0.86	1	1	1	1	1	1	1
9	0.68	0.84	1	1	1	1	1	1	1

3. Selection-mutation balance with no dominance.

Here, students examine an equilibrium situation, where deleterious alleles enter the population by mutation at rate μ and are removed by selection of intensity s . The equilibrium frequency of the deleterious allele is μ/s . [For one independent project, a student might study the analogous problem assuming complete recessiveness of the deleterious allele. In this case, the equilibrium frequency is $(\mu/s)^{0.5}$.]

Every student begins with a population of 50 *AABB*, 100 *AABb* and 50 *AAbb*. That is, the *B* and *b* alleles are present at initial frequencies of 0.5 each. The relative fitnesses of the three genotypes are set to 1.0 (*AABB*), 0.95 (*AABb*) and 0.90 (*AAbb*); that is, $s = 0.05$. Students will vary the mutation rate of *B* to *b*, as shown in Table 3. Since μ ranges from 0.005 to 0.045, the estimates of equilibrium *b* allele frequency range from 0.1 through 0.9. I find that 100 iterations is sufficient to make the point.

To make the problem a little more interesting, the linkage map distance between genes *Alpha* and *Beta* can be set to 10 cM. By doing so, one can see how selection on one gene affects the fixation rate of gene copies of a partially linked neutral gene. The results can be surprising.

In most cases, the students' initial allele frequencies differ from the expected equilibrium frequencies, so the simulation must run through several generations to reach equilibrium. However, if too many generations pass, allele fixation will still occur by drift. A compromise is necessary, and it depends on the population size. For the values described above and in Table 3, stopping each iteration at 250 generations seems to alleviate the problem.

Table 3. Mutations rates for Exercise 3.

Student	Mutation Rate
1	0.005
2	0.01
3	0.015
4	0.02
5	0.025
6	0.03
7	0.035
8	0.04
9	0.045

Unlike the first two exercises, the program output must be entered into a spreadsheet program to obtain estimates of equilibrium *b* allele frequencies. The *EvolGenius* output includes the final frequency of the *B* allele. The spreadsheet program can be used to calculate the mean, variance and standard deviation of this value. Note, however, that this gives the equilibrium frequency of the *favored* allele, not the deleterious allele. One solution is to reverse the relative fitnesses in the simulation, such that *B* is the

deleterious allele. Alternatively, it can be noted that, since $f(b)$ is simply $1-f(B)$, the variances, standard deviations and standard errors are the same. All that must be calculated is mean $f(b)$, which is trivial.

D. INDEPENDENT STUDENT PROJECTS.

Over the next few weeks, students carry out independent projects to explore somewhat more complicated problems. I provide a list of ideas from which the students may choose (see Box 2), though more confident students might develop a different project altogether.

Two of the goals of the overall exercise, better appreciation of hypothesis development/testing and stochastic processes, could be satisfied without independent projects. However, the third goal, increased appreciation of population genetics, requires greater exposure to the breadth of the field. While performing an independent project works toward the first two goals, it does not on its own do much to address the third goal. However, I require students to give oral presentations on their projects. They must describe the hypotheses they are testing, explain why the particular approach is valid, and carefully describe and discuss their results. Thus, all of the students are exposed to a broad range of projects. Even if they do not ultimately understand everything to which they are exposed, they should gain an appreciation for the breadth and complexity of the topic. The goal is not to make the students experts in population genetics. Rather, it is to drive home the point that there is much more to know than Hardy-Weinberg equilibrium.

DISCUSSION

This exercise is not a substitute for a population genetics course, but such a course is rarely offered at the undergraduate level. The problem of insufficient coverage is, of course, not limited to the teaching of population genetics. And cases can be made for increased emphasis on other areas of genetics, particularly with respect to gene regulation and development. However, genetics as a field has progressed to the point where formerly distinct subdisciplines are clearly integrated.

Many human disease states reflect a complex interaction of multiple genes and the environment, a problem historically in the territory of quantitative genetics. Today's "association analyses," whereby the presence/absence of a disease state is correlated with a particular mapped genetic marker, are an extension of quantitative genetics. The difference is that the markers can, in principle, lead to the identification of actual genes, variants of which contribute to the disease state. However, a particular variant may only be associated with the disease state in certain genetic backgrounds and in certain environments, and the likelihood of the allele appearing in a certain genetic

background is historically the concern of population genetics.

In addition to familiarizing students with an important, and often under-appreciated area of genetics, the exercises described here help students develop other strengths associated with a well-rounded scientist. The project requires students to develop and test hypotheses, and to seek out relevant information in the primary and secondary scientific literature. Students are asked to test hypotheses throughout their course work, but we don't often ask them to develop hypotheses. Because there is a large literature in population genetics, with theory relevant to most projects developed to some extent, students may often be able to compare their results to those expected from theory. If the findings are consistent with theory, this can be reported. If the findings appear to be inconsistent with theory, then the student can be challenged to account for the discrepancy. Usually, the fault is not with the theory, nor even the hypothesis, but with the predictions. It is useful, in fact, to have some students tackle a project that produces seemingly incongruous results, in order to foster discussion about

the actual nature of hypothesis testing: namely, that a mismatch between observed results and expected results may reflect error in determining the latter, and not error in the hypothesis itself.

Finally, all of the exercises provide insight into stochastic processes. All too often, material in biology courses becomes a series of related facts, and it is easy for students to think that the scientist is seeking to describe, at increasingly high resolution, a static, deterministic world. It only takes a few runs of a population genetics simulation to show that, for purely stochastic reasons, the outcome can vary dramatically. We may say we "expect" fixation of a gene copy to take $2fN$ generations, but this isn't really what we anticipate. If we understand stochastic processes, we anticipate that our result will probably differ from the expected result. If the student learns nothing else, let it be this.

ACKNOWLEDGMENTS

I thank Reyna Favis for helpful comments on the manuscript.

LITERATURE CITED

- Hartl, D. L. and A. G. Clark (1997) *Principles of Population Genetics. Third Edition.* Sinauer Associates, Sunderland, MA.
- Harris, E. E. and J. Hey (1999) X chromosome evidence for ancient human histories. *Proc. Natl. Acad. Sci. USA* 96: 3320-3324.
- Li, W.-H. (1997) *Molecular Evolution.* Sinauer Associates, Sunderland, MA.
- Nordborg, M. and S. Tavaré (2002) Linkage disequilibrium: what history has to tell us. *Trends in Genetics* 18: 83-90.
- Ohta, T. (1996) The current significance and standing of neutral and nearly neutral theories. *Bioessays* 18: 673-677.
- Reich, D. E., M. Cargill, S. Bolk, J. Ireland, P.C. Sabeti, D. J. Richter, T. Lavery, R. Kouyoumjian, S. F. Farhadian, R. Ward and E. S. Lander (2001) Linkage disequilibrium in the human genome. *Nature* 411: 199-204.
- Templeton, A. R. (1993) The "Eve" hypotheses: a genetic critique and reanalysis. *American Anthropologist* 95: 51-72.
- Vigilant, L., M. Stoneking, H. Harpending, K. Hawkes and A. C. Wilson (1991) African populations and the evolution of human mitochondrial DNA. *Science* 253: 1503-1507.

ACUBE

46th Annual Meeting

Columbia College

Chicago, IL

September 12-14, 2002

