INNOVATIONS

How to Teach the Hardy-Weinberg Principle Using Engaging, Non-trivial Evolutionary Scenarios

Michael J. Wise

Department of Biology, Roanoke College, Salem, VA 24153

Corresponding author**:** wise@roanoke.edu

Abstract: As a foundational evolutionary concept, the Hardy-Weinberg principle should be taught enthusiastically in introductory biology courses. In a companion Perspectives paper, I made the case that students are often given limited or incorrect information on the HW principle due to a lack of mastery or confidence on the part of their teachers. The purpose of this Innovations paper is to identify where errors are most-often made in the set-up and solutions of HW problems. The centerpiece of this paper is a set of six biological scenarios to which students need to apply the HW principle to answer interesting evolutionary questions. I provide explanations for solving the problems (including Excel instructions with formulas to perform chi-square tests), and I identify several teachable moments that are likely to arise in the discussion of the solutions. The use of this problem set and the pedagogical strategy described in this paper significantly improved students' performance on HW problems in my introductory biology class, and I expect that they can benefit other teachers at least as much.

Keywords: chi-square test, evolutionary mechanisms, Hardy-Weinberg principle, introductory biology, population genetics, problems

INTRODUCTION

The Hardy-Weinberg principle (HWP) is among the more challenging concepts taught in introductory biology courses. The probabilistic nature of the HWP makes analysis of evolutionary scenarios deceptively difficult for students, as well as teachers who are not specifically trained in quantitative analyses (Mertens, 1992; Masel, 2012; Brewer & Gardner, 2013). As a result, the exercises that students are given to apply the HWP tend to be over-simplified, and the students' understanding and appreciation of the HWP ends up being superficial and fleeting (Masel, 2012; Smith & Baldwin, 2015).

My goals in writing this and the companion Perspectives paper are to encourage introductory-biology instructors to teach the HWP enthusiastically and to provide them with the means to do so confidently. This paper begins with general guidance on teaching students to analyze evolutionary scenarios using the HWP, based on a strategy I have developed over many years of teaching introductory biology. This strategy includes the option of using chi-square tests to make statistically supported inferences. The rest of the paper focuses on a HW-problem set that includes six scenarios that I have used in my introductory biology courses. I explain the rationale for each scenario, the mathematics used to analyze the problems, and hypotheses for how evolution would most likely cause the patterns in the data. Throughout, I provide guidance on avoiding the most common mistakes made by students, as well as by teachers, in applying the HWP.

PROCEDURE

General Guidance on Teaching Hardy-Weinberg Problems

The strategy that I teach students to use in solving HW problems involves a four-step process (Fig. 1). The first step is to identify the relevant information given (generally about phenotypes, but maybe about genotypes or alleles) that can be used to calculate the actual (i.e., observed) allele frequencies and genotype frequencies in the population. It is very important to stress that the actual allele frequencies can always be calculated if one knows the genotype frequencies, but the reverse is not true: The actual genotype frequencies in a population cannot be calculated from the allele frequencies alone. (Note the one-way red arrow from the genotype-frequency box to the allele-frequency box in Step 1 of Fig. 1). A very common mistake in HW problems is that students plug the allele frequencies into the HW equilibrium equation (see below) to try to calculate actual genotype frequencies (Smith & Baldwin, 2015).

The second step is the calculation of the HW equilibrium (HWE) genotype frequencies from the actual allele frequencies (i.e., p^2, $2pq$, and q^2; Fig. 1). The third step is to compare the actual genotype frequencies with the HWE genotype frequencies. This comparison can be done with or without the aid of statistical analysis (e.g., chi-square test). Without statistical analysis, students can still make qualitative

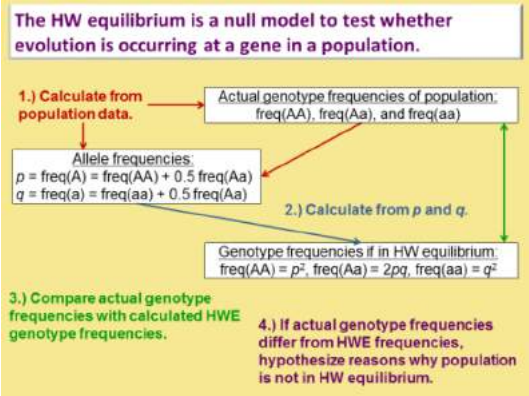

Fig. 1. Lecture slide used to overview strategy for solving Hardy-Weinberg problems.

statements such as whether there are more or fewer heterozygotes in the population than predicted by the HWE frequencies. The fourth step is to compare any discrepancies between the actual and HWE genotype frequencies with patterns expected if any of the assumptions of the HWP are violated. These patterns allow students to propose hypotheses about what mechanisms are the most likely to be causing the evolution observed in the population. Ecological information about the organisms given in the text of the problem should guide students toward explanations of how these mechanisms might act in terms of the biology of the system.

Group Activity on Hardy-Weinberg Scenarios

In the most-recent version of my introductory Biological Diversity course, I employed the six-scenario problem set using a group activity. Before presenting the problem set, I lectured briefly on the historical context and development of the HWP. I used a textbook example of the inheritance of coloration in cats (Raven et al., 2014) to illustrate the four steps shown in Fig. 1. I then had the class of 24 students break into six groups of four students, and I handed each student a sheet with six evolutionary scenarios (i.e., "HW problems"). These scenarios are based on actual biology, but the numbers were fabricated to achieve my pedagogical goals. Each student-group volunteered to be responsible for solving one of the problems and presenting it to the class during the next meeting. Students spent the rest of the class period solving their problem together. They were provided with colored markers and a flipboard to work on their presentations. While they were working, I rotated among groups to check their progress and give hints as needed so that they would not present wildly erroneous information to the rest of the class. Most groups finished preparing their presentations for homework, and I strongly encouraged students to work on the other five problems on their own so that they could get the most value out of watching the presentations of the other student groups during the next class meeting.

Use of Chi-Square Tests in Hardy-Weinberg Problems

Because my course was the second of a two-semester sequence of introductory biology, I knew that the majority of my students had been exposed to analyzing genotype frequencies (in the context of Mendelian genetics) in the prerequisite course. Therefore, I required students to analyze their data by performing a chi-square test using calculators and a chi-square probability table (Table 1), as they had done in the previous course. Alternatively, chi-square tests can be performed in a relatively straightforward manner using Excel. I have provided a generic Excel table (Fig. 2) that can be used to quickly calculate all of the information needed for Steps 1-3 in most HW problems. The orange-shaded cells show the formulas for the calculations using Excel syntax (i.e., exactly what would be typed into the cells). The pink-shaded cells (C5:C7) are where data on the number of individuals of each phenotype or genotype would be entered. (Be sure that students enter counts of individuals into these cells—rather than frequencies or proportions.) The generic information in the unshaded cells should be replaced with the information specific to the problem (e.g., the phenotype names in Column A and convenient abbreviations for the alleles for the genotypes in

Table 1. Table of chi-square (X^2) probabilities and associated *P*-values. Use this table to perform a chi-square test of the null hypothesis that observed genotype frequencies in a sample population did not differ from Hardy-Weinberg equilibrium (HWE) frequencies. First find the row for the number of degrees of freedom in your test. Then find the columns between which your calculated X^2 value falls. Drop to the bottom (shaded) row of these columns to find the range of *P*-values associated with your X^2 and df. If $P < 0.05$, then you should reject the null hypothesis, and you can infer that the genotype frequencies in your sample population are not at HWE.

df	Null hypothesis is supported ←							→ Null hyp. is rejected		
	Chi-square values (X^2)							Chi-square values (X^2)		
1	0.00016	0.0039	0.064	0.455	1.074	1.642	2.706	3.841	5.412	6.635
2	0.0201	0.103	0.446	1.386	2.408	3.219	4.605	5.991	7.824	9.210
***P* =**	**0.99**	**0.95**	**0.80**	**0.50**	**0.30**	**0.20**	**0.10**	**0.05**	**0.02**	**0.01**

Fig. 2. Excel table for performing a chi-square test of the null hypothesis that a population is in Hardy-Weinberg equilibrium for a given gene. Data on observed numbers of individuals are entered directly in the pink-shaded cells (C5-C7). The formulas in the orange-shaded cells calculate the information needed for the chi-square test.

	A	B	C	D	E	F	G	H	I
1									
2									
3			Number	Number		$\frac{(Obs. - Exp.)^2}{Expected}$		Frequencies	
4	Phenotype	Genotype	Observed	Expected	Observed - Expected			p =	=C5/C8+C6/C8/2
5	Homozygote 1	*AA*	N_{AA}	=C8*I6	=C5-D5	=(E5^2)/D5		q =	=(C7/C8+C6/C8/2)
6	Heterozyogote	*Aa*	N_{Aa}	=C8*I7	=C6-D6	=(E6^2)/D6		p^2 =	=I4*I4
7	Homozygote 2	*aa*	N_{aa}	=C8*I8	=C7-D7	=(E7^2)/D7		$2pq$ =	=2*I4*I5
8	Sums:		=SUM(C5:C7)	=SUM(D5:D7)	=ROUND(SUM(E5:E7),1)	=SUM(F5:F7)		q^2 =	=I5*I5
9	X^2 =	=F8	df =	=(COUNT(D5:D7)-1)*(2-1)-1	0	=CHISQ.DIST.RT(F8,D9)			

Column B).

There is substantial flexibility in how much guidance you choose to give your class depending on how much time you have to spend on these problems. For example, you could provide them with Excel tables with formulas already filled in, you could provide them with a printout of a blank table with only headings, or you could let them figure out for themselves how to tabulate the data. If you do not want to involve computers, you can have students make statistical inferences by hand-calculating chi-square (χ^2) values and degrees of freedom (df). Students can then compare these values with the values in a chi-square-probability table to find the *P*-value range for their test (e.g., Table 1).

The *P*-value range identified using the table (or the precise value calculated by Excel) can be interpreted as the likelihood of finding actual genotype frequencies as divergent from the HWE frequencies by random chance alone. By convention, if the *P*-value is < 0.05, then we can infer that the null hypothesis (viz., that the population is in HWE) can be rejected; that is, the actual genotype frequencies are "statistically significantly different" from HWE. Therefore, the population can be inferred to be undergoing evolution at the gene of interest. If the *P*-value is ≥ 0.05, then we cannot reject the null hypothesis, and there is not sufficient evidence to conclude that the population is evolving.

One common area of confusion in applying chi-square tests for HW problems is the appropriate number of degrees of freedom to use. In general, the degrees of freedom in a chi-square test are calculated as the number of columns (*c*) minus one, times the number of rows (*r*) minus 1, or df = (*c*-1) x (*r*-1). This is the formula that Excel uses to determine the df for a chi-square test when using the built-in function "=CHITEST( )" in which the range of cells containing the counts are entered as the argument in parentheses. For comparing actual to expected genotype frequencies, there will always be two columns—one for actual counts of individuals and one for expected counts of individuals under HWE. Thus, the df calculation reduces to just the number of rows - 1. For three genotypes, there would be three rows of data, and the df would thus be 2, or df = (2-1) x (3-1). Equivalently, a commonly taught approach to calculating df for chi-square tests in HW problems is simply to subtract 1 from the number of genotypes under consideration (cf. McMurran, 2010).

While this approach is convenient and may help avoid a class-wide discussion of the statistical/philosophical rationale behind degrees of freedom, it actually perpetuates a critical error. Simply put, the degrees of freedom available for any statistical test is reduced by the number of parameters that must be estimated from the data. For calculating the expected genotype frequencies, one must generally estimate allele frequencies (*p* or *q*) from the data. This estimation uses up one degree of freedom. Therefore, in a typical HW problem, the appropriate number of degrees of freedom for a chi-square test involving three genotypes would be 1 (or, df = 3-1-1) (Hartl & Clark, 1989; Freeman & Herron, 2004).

This statistical issue can be easily rectified by using a different Excel function: "=CHISQ.DIST.RT(_, _)" in which the calculated χ^2 and the df are entered as the arguments in parentheses. While conceptually a bit more difficult, this correct approach offers few additional practical difficulties in statistically analyzing HW problems than the typical, incorrect approach. An exception is when the problem involves only two genotypic categories instead of three. In such a case, the degrees of freedom will equal 0, which precludes a valid application of a chi-square test. The second scenario in the problem set below illustrates such a case in which a chi-square test cannot be applied to analyze the data. Instead, students are limited to making qualitative, rather than quantitative, inferences in that scenario.

Set of Six Hardy-Weinberg Scenarios

The question set that I used is displayed in Appendix 1. In this section, I overview the rationale behind the scenarios, explain the solutions, and provide hints and caveats for each of the six problems. A summary of the intermediate and final quantitative answers for each problem is displayed in Excel-table format in Fig. 3. (An Excel file with formulas entered into the cells is also available from

Fig. 3. Excel tables showing the intermediate calculations, X^2, df, and *P*-values for analyses of the six Hardy-Weinberg scenarios shown in Appendix 1.

1) Sickle-Cell Anemia

Phenotype	Genotype	Number Observed	Number Expected	Obs. - Exp.	$(Obs. - Exp.)^2$ / Expected	Frequencies	
						p =	0.6625
normal	*SS*	75	87.78125	-12.78125	1.8610	q =	0.3375
carrier	*Ss*	115	89.4375	25.5625	7.3061	p^2 =	0.4389
sickle-cell anemia	*ss*	10	22.78125	-12.78125	7.1708	$2pq$ =	0.4472
Sums:		200	200	0	16.3379	q^2 =	0.1139
X^2 =	16.3379	df =	1	P =	5.29934E-05		

2) Bitterness in Dandelions

Phenotpye	Genotype	Number Observed	Number Expected	Obs. - Exp.	$(Obs. - Exp.)^2$ / Expected	Frequencies	
						p =	0.5
mild	*bb*	5	31.25	-26.25	22.0500	q =	0.5
bitter	*BB + Bb*	120	93.75	26.25	7.3500	p^2 =	0.25
Sums:		125	125	0	29.4000	$2pq$ =	0.5
X^2 =	29.4000	df =	0	P =	#NUM!	q^2 =	0.25

3) Copper Tolerance in Bent Grass

Phenotype	Genotype	Number Observed	Number Expected	Obs. - Exp.	$(Obs. - Exp.)^2$ / Expected	Frequencies	
						p =	0.775
Cu-tolerant	*TT*	55	60.0625	-5.0625	0.4267	q =	0.225
Cu-tolerant	*Tt*	45	34.875	10.125	2.9395	p^2 =	0.6006
Cu-sensitive	*tt*	0	5.0625	-5.0625	5.0625	$2pq$ =	0.3488
Sums:		100	100	0	8.4287	q^2 =	0.0506
X^2 =	8.4287	df =	1	P =	0.003693443		

4) Flower Color in Morning Glory

Phenotype	Genotype	Number Observed	Number Expected	Obs. - Exp.	$(Obs. - Exp.)^2$ / Expected	Frequencies	
						p =	0.6227
white	*WW*	388	290.785333	97.21467	32.5006	q =	0.3773
purple center	*Ww*	158	352.429333	-194.4293	107.2634	p^2 =	0.3877
purple	*ww*	204	106.785333	97.21467	88.5018	$2pq$ =	0.4699
Sums:		750	750	0	228.2657	q^2 =	0.1424
X^2 =	228.2657	df =	1	P =	1.4242E-51		

5) Pigmentation in Water Boatmen

Phenotype	Genotype	Number Observed	Number Expected	Obs. - Exp.	$(Obs. - Exp.)^2$ / Expected	Frequencies	
						p =	0.35
pigmented	*AA*	20	12.25	7.75	4.9031	q =	0.65
pigmented	*Aa*	30	45.5	-15.5	5.2802	p^2 =	0.1225
albino	*aa*	50	42.25	7.75	1.4216	$2pq$ =	0.455
Sums:		100	100	0	11.6049	q^2 =	0.4225
X^2 =	11.6049	df =	1	P =	0.000657783		

6) *Delta-32* Mutation

Phenotype	Genotype	Number Observed	Number Expected	Obs. - Exp.	$(Obs. - Exp.)^2$ / Expected	Frequencies	
						p =	0.7
HIV-susceptible	*SS*	500	490	10	0.2041	q =	0.3
HIV-susceptible	*Ss*	400	420	-20	0.9524	p^2 =	0.49
HIV-resistant	*ss*	100	90	10	1.1111	$2pq$ =	0.42
Sums:		1000	1000	0	2.2676	q^2 =	0.09
X^2 =	2.2676	df =	1	P =	0.132104339		0

the author. This file has the data that I used in my class for the scenarios, but these data can be modified as desired, and the statistics will be automatically recalculated.)

Scenario 1: Sickle-Cell Anemia

Sickle-cell anemia is a classic example of a disease caused by a recessively expressed mutation that is found in higher-than-expected frequencies due

to a phenomenon called "heterozygote advantage." Specifically, carriers of the disease (i.e., heterozygous individuals) have an increased resistance to malaria, which is a deadly disease prevalent in tropical areas in which sickle-cell anemia is also common. Thus, while individuals homozygous for the sickle-cell gene are selected against (i.e., they are likely to die before reproducing), individuals heterozygous for the gene are favored by selection because they are less likely to die of malaria before reproducing.

The numbers of individuals possessing the different genotypes are given in the text of the problem. (If using the Excel table, enter these in Cells C5-C7, Fig. 2.) With these numbers of individuals, the genotype frequencies and then the allele frequencies can be calculated using the formulas in Step 1 of Fig. 1 (or Cells I4 and I5, Fig. 2). In Step 2 (Fig. 1), the HWE genotype frequencies are calculated from the allele frequencies (Cells I6-I8, Fig. 2). The numbers of individuals in the sample that would be expected if the population is at HWE are then calculated from these genotype frequencies (Cells D5-D7, Fig. 2). Then the actual numbers of individuals of each genotype are compared with the number of individuals we would expect from a population of the same sample size if the population is at HWE (Step 3, Fig. 1).

With the numbers of individuals given in this problem, there are fewer than half as many recessive homozygotes as expected at HWE (10 vs. 23, Fig. 3), slightly fewer dominant homozygotes than expected (75 vs. 88), and more heterozygotes than expected (115 vs. 89). Qualitatively, these results are consistent with strong selection against individuals with the sickle-cell phenotype, and selection in favor of heterozygous carriers.

An obvious question is whether these differences from the HWE expectations are substantial enough to be of biological interest. This is where a chi-square test proves to be very useful. Using Excel, the CHISQ.DIST.RT function (Cell F9 of Fig. 2) returns a *P*-value of 0.000053 (Fig. 3; or if you use the chi-square probabilities in Table 1, then $P < 0.01$), which is well below the traditional cutoff of < 0.05 for statistical significance. Thus, the genotype frequencies are highly significantly different from those predicted by the HWE null hypothesis, and we can reject this null hypothesis and infer that evolution is occurring at the sickle-cell gene in this population. The pattern of evolution found in the genotypes is indeed consistent with the hypothesis of heterozygote advantage.

In this scenario on sickle-cell anemia, some students might answer that genetic drift is responsible for the deviation of the genotypic frequencies from HWE. As detailed in the companion Perspectives paper, the biological and stochastic processes that result in genetic drift are always acting in every population of organisms. Therefore, students might argue that genetic drift is a reasonable answer for *any* of the scenarios in the problem set. After all, one cannot prove for certain that genetic drift alone did not cause a deviation in genotype frequencies of any magnitude from HWE expectations. The rebuttal to such arguments involves a rational appeal to likelihood (i.e., perhaps, common sense). Specifically, the employment of a chi-square test enables a quantitative and objective assessment of the likelihood that drift is solely responsible for a deviation of genotype frequencies from HWE expectations.

In general terms, the *P*-value from a statistical test indicates the probability of finding a deviation from the null hypothesis as large as you observed by random chance alone (what statisticians call "sampling error"). In terms specific to the context of population genetics, the evolutionary mechanism that causes random departures in genotype frequencies from HWE is genetic drift. The probability of genetic drift causing a deviation from HWE as large as was observed in a population is embodied in the *P*-value. Specifically, the lower the *P*-value, the less likely it is that genetic drift alone was responsible for the deviation, and thus the more likely another evolutionary mechanism was also at play. The identity of that mechanism should be hypothesized from the background information given in the text of the scenario.

For this scenario on sickle-cell anemia, the *P*-value for the chi-square test of the null hypothesis that the genotype frequencies were at HWE was 0.000053. Therefore, the chance that genetic drift alone would cause the observed deviation in genotype frequencies from HWE was only about one in 20,000. While a student could still argue that we cannot 100% prove that genetic drift was not solely responsible, the low probability of that outcome makes that answer relatively untenable. A much better answer to this question would include an explanation of the evolutionary mechanisms (e.g., natural selection and heterozygote advantage) that are consistent with the observed patterns in deviations of genotype frequencies from HWE predictions in the specific scenario.

In any given generation, genetic drift may act either to accentuate or obscure the influence of other evolutionary mechanisms on the deviations of genotype frequencies from HWE expectations. However, the smaller the *P*-value is, the more confidently we can infer the influence of another evolutionary mechanism through the noise that is caused by genetic drift.

Scenario 2: Bitterness in Dandelions

In this scenario, it is the allele frequencies that are given; these can be entered directly into Cells I4 and I5 of Fig. 2. We cannot calculate the actual frequencies for all three genotypes because we don't

know how many of the 120 bitter plants are heterozygous and how many are homozygous. Nevertheless, a careful reading of the scenario will show that we do not need to know all the genotype frequencies to address the specific question—at least qualitatively. We can still calculate HWE genotype frequencies from the allele frequencies (Step 2 of Fig. 1; Cells I6-I8 of Fig. 2). We know that the number of mild plants expected under HWE is the number of individuals homozygous for the mildness allele (Cell D5, Fig. 2), and that the number of bitter plants expected under HWE is the number of heterozygotes plus the number of individuals homozygous for the bitterness allele (Cells D6+D7 of Fig. 2).

In Step 3 (Fig. 1), we just need to compare the ratio of mild:bitter plants observed (5:120) with the much higher ratio expected under HWE (31:94). A chi-square value can be calculated from these data (χ^2 = 29.4; Fig. 3). This χ^2 value is much higher than all of the critical values for statistical significance shown in Table 1. However, a valid chi-square test cannot be performed on these data, because the df for such a test would equal zero. Note that Excel will return an error message (#NUM!) if one asks it to perform a chi-square test with df = 0. Therefore, students will be required to interpret the results without the benefit of a statistical test. The inclusion of a scenario in which a chi-square test is not an option can serve as a topic of discussion regarding students' confidence in statistical tests, as well as the difference between biological significance and statistical significance.

Even without a statistical test, students can see that there were 24 times as many bitter dandelion as mild dandelions in my yard, while if the population was in HWE, there would only be three times as many bitter dandelions plants as mild ones. The most reasonable explanation for the discrepancy in the actual phenotype ratio from that predicted by HWE is that natural selection acted against the mild phenotype. If students need hints to hypothesize a mechanism, ask them what kinds of animals one might find in yards. Among these animals, several are bound to be herbivorous (e.g., rabbits, slugs, and insects), and these animals may perceive the palatability of dandelion greens just as we do. Herbivores feed on (and thus reduce the fitness of) mild dandelions, while bitter dandelions survive and reproduce. In answering the question posed in the scenario text, one could conclude that rather than being unlucky, I was not thinking enough about evolution by natural selection.

Scenario 3: Copper Tolerance in Bent Grass

In this scenario, the numbers of individuals of each genotype in the sample are given, which allows for simple calculation of the allele frequencies (Cells I4-I5, Fig. 2). The HWE genotype frequencies are then calculated from the allele frequencies (Cells I6-I8, Fig. 2). The numbers of individuals expected for each genotype under HWE are then calculated by multiplying these HWE genotype frequencies by the number of individuals (100) in the sample (Cells D5-D7, Fig. 2).

A comparison of these expectations with the actual numbers of individuals shows that the number of copper-susceptible individuals at the abandoned mine site is lower than the HWE expectation (0 vs. 5, Fig. 3). This result is easily explained by natural selection in the abandoned copper mine acting against the individuals that are susceptible to copper poisoning. However, the number of copper-tolerant homozygotes is also a bit lower than the HWE expectation (55 vs. 60). This result is not consistent with natural selection for copper tolerance being the only evolutionary mechanism in action.

The more surprising result is that the number of heterozygotes observed was greater than HWE expectations (45 vs. 35). The clue to understanding this result is knowledge that grass is wind-pollinated. Some bent-grass plants in the abandoned mine site are likely to be fertilized by pollen that blows in from plants adjacent to the mine, where copper-sensitive individuals are more fit than copper-tolerant individuals. Such pollen would carry the susceptibility allele. Thus, gene flow from a non-mine site is a likely evolutionary mechanism responsible for the larger-than-expected number of heterozygotes in the abandoned-copper-mine site. Seeds originating from the low-copper populations adjacent to the mine are also likely to be dispersed by wind into the abandoned mine site. These seeds would likely be homozygous for the susceptibility allele. However, plants that grow from such seeds in high-copper soil are likely to die before maturity, and thus homozygous susceptible plants are not likely to be found in a sample of mature individuals from the abandoned mine site.

Scenario 4: Flower Color in Morning Glory

In this scenario, there is incomplete dominance for flower-color phenotype. Thus, the genotype frequencies are directly reflected in the phenotype frequencies, which are easily obtained from the numbers given in the text of the problem. I used *p* to represent the frequency of the white allele, and *q* for the purple allele—though the choice of letters is arbitrary. The allele frequencies are calculated from the observed numbers of individuals (Cells I4-I5, Fig. 2), and HW-equilibrium genotype frequencies are calculated from the allele frequencies (Cells I6-I8, Fig. 2). The numbers of individuals in the sample that would be expected if the population is in HWE are then calculated from these genotype frequencies (Cells D5-D7, Fig. 2).

A comparison of these expectations with the actual numbers of individuals shows that the observed numbers of both types of homozygotes were much greater than the expected numbers (Fig. 3). In contrast, the number of heterozygotes observed was much smaller than expected under HWE (158 vs.

352). With such a large sample (750 plants), these deviations from HWE were very highly statistically significant ($P = 1.4 \times 10^{-51}$, Fig. 3). This pattern of an excess of both homozygotes is a hallmark of assortative mating, wherein individuals with similar phenotypes mate with each other and avoid mating with individuals of different phenotypes.

Hypothesizing a mechanism that would cause this pattern requires a recognition of how morning glory plants "choose" mates. The hint to how this occurs is given in the first sentence of the text of the problem. Specifically, it is the bumblebees that transfer pollen among plants, and thus they are choosing which plants are mated. To forage efficiently, bumblebees may form a search image for one color or the other. For instance, an individual bee may find purple flowers to be rewarding, and thus may go from one purple flower to another. Another bee may focus on just white flowers. The result is that purple alleles tend to match up with other purple alleles, and white with white, which results in a preponderance of offspring homozygous for flower color.

Scenario 5: Pigmentation in Water Boatmen

This scenario is perhaps the most challenging of the set, as its setup is more complex than the others, and because the answer is not as simple as one evolutionary force acting consistently in one direction. The text of the problem clues students in to the fact that the pigmentation gene is indeed evolving in the tanks. Specifically, the phenotype ratios have changed over time in the tanks, going from extreme ratios to 50:50. Students will also probably realize that the predatory fish are the drivers of the evolution. However, the scenario is not as simple as fish always preferring to eat pigmented individuals or always preferring albino individuals. This scenario involves frequency-dependent selection, where the fish will favor whichever prey phenotype is the most common.

The actual frequencies at the end of the experiment are the important numbers to compare with the HWE frequencies. From the information given, it is a simple matter to figure out that each tank ended up with 50 albino (all homozygotes), 30 heterozygous pigmented, and 20 homozygous pigmented water boatmen. These values can be used to calculate the actual allele frequencies (Cells I4-I5, Fig. 2). The allele frequencies are then used to calculate the HWE genotype frequencies (Cells I6-I8, Fig. 2), which are then used to calculate the expected number of individuals of each genotype at HWE (Cells D5-D7, Fig. 2). A comparison of these expectations with the actual numbers of individuals shows that number of both homozygous-pigmented and homozygous-albino individuals were greater than expected (Fig. 3). This deviation from expectation was statistically significant ($P = 0.00066$), which indicates that evolution is indeed occurring on this gene, even when the two phenotypes are equally common in the tanks.

Natural selection by the predatory fish is the most likely explanation for the deviation from HWE in this problem. However, neither phenotype has a consistent advantage. That is, if fish always preferred pigmented insects, then the number of pigmented insects would be less than the HWE expectation. Likewise, if fish always preferred albinos, then the number of albino insects would be less than the HWE expectation. The fact that both homozygotes were more frequent than expected shows that they both experienced selective advantages. However, they experienced them at different times—when each was the minority phenotype. Thus natural selection always favors the rare phenotype, such that the two phenotypes become equally common over time—a hallmark of frequency-dependent selection. The text in the setup of the scenario should lead students to the conclusion that the fish prefer to eat the more-common phenotype, even though they may not yet have the language to explain it as "frequency-dependent selection."

This pattern of the both homozygotes being more common than expected, at the expense of the heterozygotes, is also consistent with assortative mating (as observed in the question on morning glories). Therefore, students might hypothesize that pigmented individuals prefer to mate with other pigmented individuals, and albino individuals with other albinos. However, assortative mating by itself does not explain why the two phenotypes always became equally common in the experimental tanks. The text of the scenario intentionally did not mention anything about mating. In fact, the time scale involved is likely so short that the fish ate the insects before the insects completed their life cycles. Nevertheless, these details were not specified in the text so that students would have more flexibility in hypothesizing about evolutionary mechanisms, which could then generate more critical thinking and discussion.

Scenario 6: Delta-32 Mutation

This scenario may be the most interesting to students because it is about human health. (Be aware, however, that it involves topics that may be uncomfortable for some students.) The genotype frequencies are given in this problem, so the allele frequencies can be readily calculated (Cells I4-I5, Fig. 2). From the allele frequencies, the HWE genotype frequencies can be calculated (Cells I6-I8), and the expected number of individuals with each genotype (Cells D5-D7) can then be compared to the observed numbers (Cells C5-C7).

Qualitatively, there were more individuals homozygous for the Delta-32 mutation than expected under HWE (100 vs. 90). However, this difference was not statistically significant, with a chi-square test *P*-value of 0.13 (Fig. 3). Thus, there is not definitive

evidence that this population is evolving at the *CCR5* gene due to natural selection. Instead, the discrepancy in the observed data with the HWE expectations could easily be due to genetic drift alone. Nevertheless, this failure to reject the null hypothesis does not mean that the Delta-32 allele is not advantageous in some fashion. Individuals homozygous for the mutation may be likely to survive longer than sexually active gay men without the mutation. However, fitness involves not just survival, but reproduction to pass along the genes. Gay men may be unlikely to pass on their genes, whether or not they have the Delta-32 allele. Or, people may reproduce before contracting HIV and thus may pass along the non-mutated allele to their offspring before natural selection for mutation has a chance to take place. Finally, the Delta-32 mutation may provide other immunity benefits, but natural selection was not strong enough to be statistically significant in this sample of 1000 individuals.

This example on the Delta-32 allele provides a good opportunity for discussing the limits of using the HWE test to detect the action of natural selection. Specifically, the strength of selection must be quite high, or the sample population must be very large, for selection to result in a statistically significant departure in genotype frequencies from HWE within a generation (Hartl & Clark, 1989). Nevertheless, even weak selection can have a substantial effect on the allele frequencies of a population over very long periods of time. Thus, methods other than testing for HWE are often more useful for finding evidence of evolution by natural selection.

CONCLUSION

In the first introductory biology course in which I used this problem set, the students' scores on a set of questions based on a HW scenario on the final exam improved by 20% compared to scores on the same question in the last biology course in which I did not use the problem set (one-way ANOVA: $F_{1,57} = 5.82$; $P = 0.02$). I expect that the improvement in the exam performance would have been even greater if I had done two things: 1) insist that every student turned in a written attempt at each of the six scenarios, and 2) remind students to review these problems prior to the final exam—rather than just studying material covered in professor-led lectures, which is something that several students admitted to doing on their course evaluations.

In addition to the objective benefit of higher exam scores, my use of this problem set provided several subjective benefits. For example, the relatively complex scenarios required students to employ quantitative reasoning and higher-order cognitive skills. The activity also provided students a chance to work cooperatively and make an oral presentation to the class—both important skills that are not generally practiced in the traditional lecture structure of most introductory biology courses (Gokhale, 1995; Prince, 2004). It is my hope that other instructors will be willing to use or modify this problem set for their own courses, including the Excel template for performing chi-square tests.

ACKNOWLEDGMENTS

I thank the students of my BIOL 125 course in 2015 for participating in the group activity on solving Hardy-Weinberg problems, and the Biology Department at Roanoke College for logistical support during the writing of this paper. I also thank an anonymous reviewer for insightful comments that substantially improved this manuscript as well as the companion Perspectives manuscript.

REFERENCES

BREWER, M.S., AND G.E. GARDNER. 2013. Teaching evolution through the Hardy-Weinberg principle: a real-time, active learning exercise using classroom response devices. *The American Biology Teacher* 75(7): 476-479.

FREEMAN, S., AND J.C. HERRON. 2004. *Evolutionary Analysis*. *third ed*. Pearson Education, Inc., Upper Saddle River, NJ. 802p.

GOKHALE, A.A. 1995. Collaborative learning enhances critical thinking. *Journal of Technology Education* 7(1): 22-30.

HARTL, D.L., AND A.G. CLARK. 1989. *Principles of Population Genetics*. *second ed*. Sinauer Associates, Inc., Sunderland, MA. 682p.

MASEL, J. 2012. Rethinking Hardy-Weinberg and genetic drift in undergraduate biology. *Bioessays* 34: 701-710.

MCMURRAN, S.L. 2010. The Hardy-Weinberg principle. *PRIMUS* 20(6): 529-549.

MERTENS, T.R. 1992. Introducing students to population genetics & the Hardy-Weinberg principle. *The American Biology Teacher* 54(2): 103-107.

PRINCE, M. 2004. Does active learning work? A review of the research. *Journal of Engineering Education* 93(3): 223-231.

RAVEN, P.H., G.B. JOHNSON, K.A. MASON, J.B. LOSOS, AND S.R. SINGER. 2014. *Biology*. *tenth ed*. McGraw-Hill, New York. 1279p.

SMITH, M.U., AND J.T. BALDWIN. 2015. Making sense of Hardy-Weinberg equilibrium. *The American Biology Teacher* 77(8): 577-582.

Appendix 1. Hardy-Weinberg problem set: six evolutionary scenarios.

1. In a population of humans in a village in central Africa, doctors took blood samples from 200 adolescent boys to study sickle-cell anemia, which is a recessively inherited disease caused by a mutation in a single gene coding for hemoglobin production. In the sample, 115 boys were found to be carriers (heterozygote) for sickle-cell anemia, and 75 were homozygous for the normal-hemoglobin allele. Is there evidence for evolution in the sickle-cell anemia gene in this population? If so, what might be causing the evolution?
2. Dandelion greens can be purchased in spring at farm markets and green grocers. The greens generally have a mild flavor, and they are considered a healthy addition to salads. However, dandelions in the wild can taste quite bitter, depending on the alleles present for one gene (the bitterness gene). Mildness is a recessive phenotype, such that leaves from heterozygotes are just as bitter (and unpleasant tasting) as leaves from plants that are homozygous for the bitterness allele. Out of curiosity, I had a genetic analysis done from a sample of the huge dandelion population in my yard. The genetics lab reported that the frequency for the mildness allele was 50% in my dandelion population. Instead of mowing, I decided that it was worth harvesting all of my dandelion plants one spring to sell the greens at the local market. Out of 125 dandelion plants, it turned out that 120 had bitter leaves, and thus only five produced leaves worth eating. Should I have been surprised at my misfortune, based on a population-genetics perspective? If so, what is a likely explanation for the preponderance of bitter-leaved dandelions in my yard?
3. A species of bent grass has a gene that controls whether it is tolerant of (or susceptible to) copper poisoning. The more copper that is present in soil, the greater the survival and reproductive advantage tolerant plants have over susceptible plants. When copper levels are particularly high, susceptible plants die as seedlings before reaching reproductive maturity. In contrast, where soil-copper levels are normal, susceptible plants grow and reproduce much better than copper-tolerant plants. Bent-grass reproduction does not rely on animals, as its pollen and seeds are both carried by wind. The tolerant phenotype is inherited in a completely dominant fashion over the susceptible phenotype. In an abandoned copper-mine site, a sample of 100 mature bent-grass individuals was taken to a genetics lab, and the genotypes for copper tolerance were identified: 55 homozygous tolerant and 45 heterozygous. Is there evidence for evolution of the copper-tolerance gene in this population of bent grass at the abandoned mine site? If so, what might be causing the evolution? Is there evidence for more than one evolutionary mechanism acting?
4. Many species of morning glories produce large, showy flowers that are attractive to bumblebees. Consider a species whose flowers are either entirely white, entirely purple, or mostly white but with purple just at the center of the flower. These colors are determined by one gene with two alleles, and heterozygotes have white flowers with purple centers. Any given plant may have many flowers, but all of its flowers are the same color phenotype. A graduate student sampled a population of 750 morning glory plants and found the following phenotypic frequencies: 388 white, 204 purple, and 158 white with purple centers. Is there evidence for evolution at the flower-color gene in this population? If so, what might be causing the evolution?
5. Surface-swimming aquatic insects called water boatmen occur in two color morphs: pigmented and albino. The albino phenotype is recessive to the pigmented phenotype. In a series of cattle-tank experiments, a predatory fish was placed into each tank along with 1000 water boatmen. The water boatmen were a mixture of different proportions of albino and pigmented individuals, ranging from a low of 10% albinos to a high of 90% albinos. The experiment was terminated for each tank when 100 water boatmen remained in the tank. At the end of the experiment, the water boatmen percentages consistently ended up at half pigmented and half albino, regardless of the starting percentages. Electrophoresis determined that roughly 30% of the water boatmen were heterozygous for color in each tank. Did evolution at the pigmentation gene occur in these tanks? Does evolution continue to occur after the phenotypic ratios reach 50:50?
6. Recently, a deletion mutation in the *CCR5* gene on chromosome 3 of humans has been identified and named the Delta-32 mutation. There is evidence that individuals who are homozygous for this mutation may be resistant to infection by certain strains of the HIV virus. In a study of 1000 sexually active gay men in England, 100 men were found to be homozygous for the Delta-32 allele, while 500 men had no copies of the Delta-32 allele. Is there evidence for evolution of the gene in question in England? If so, what might be causing the evolution?