Population Genetics

A Fishy Process

Alan R.P. Journet

Alan R.P. Journet is an associate professor and coordinator of General Biology at Southeast Missouri State University, Cape Girardeau, MO 63701. He holds a B.Sc. from the Univ. of Wales and a Ph.D. from McGill Univ. He was awarded a post-doctoral fellowship at The Australian National Univ. and has taught at the Univ. of New South Wales and Texas A&M Univ. Journet's research interests and publications include plant/insect ecology, science teaching and cognitive development, technical writing and the application of multi-variate analyses to manuscripts.

In introductory courses, the process of evolution is often introduced through a mathematical treatment of the Hardy-Weinberg (H-W) equilibrium. Students have difficulty with this because of their lack of understanding of the mathematics of probability and the symbolism of binomial expansion. Nonetheless, population genetics is a logical introduction to evolution, and a mathematical treatment is appropriate since it underlines the inevitability of the mechanism and the firm foundation of evolutionary theory in mathematical proof. Prompted by this student confusion, I have developed a treatment of population genetics which guides students to a mathematical appreciation of the process. The section follows basic genetics, dealing with protein synthesis, DNA replication, mitosis and sexual life cycles (Journet 1984) and Mendelian genetics.

An initial student confusion develops from the textbook discussions, which start population genetics with H-W, then present (p + q) = 1 and derive $p^2 + 2pq + q^2 = 1$ through binomial expansion. Ritchie and Carola (1983, p. 531), for example, introduce the topic with the sentence, "The Hardy-Weinberg formula is usually expressed as $p^2 + 2pq + q^2 = 1$." Similar statements can be found in Wallace (1981) and Curtis and Barnes (1985). If introductory biology students are representative of all beginning college students, 50 percent are still concrete operational thinkers (Inhelder & Piaget 1958; McKinnon 1971; Stanley, Journet & Scheibe 1986). Such students are unable to deal with abstract algebraic symbolism (Sanders 1971) and will respond to such a discussion with confusion, followed by memorization without understanding (Renner 1977). In short, nonmathematical students are bewildered by the mathematics, and all students are bewildered by the significance of H-W., never seeing that it is the null against which evolution is measured. If our goal in science courses is understanding (Journet 1985), treatment of biological concepts should be designed to promote that end.

Students who do not develop H-W from initial nonequilibrium population conditions, tend to overlook the essential assumptions that must be met before the H-W genotype equilibrium occurs. Indeed, many students assume that all populations are automatically at H-W equilibrium for both allele and genotype frequencies. This error is frequently compounded by the use of the binomial expansion $(p + q)2 = 1$ to calculate the frequencies of unknown genotypes in the class (e.g. PTC tasters) of students. The average class of students, however, is not at H-W equilibrium with respect to genotype frequencies because it fails to satisfy H-W assumptions; it is equivalent to a founding population in which some genetic drift from H-W equilibrium has occurred. After many generations of random inbreeding, H-W genotype frequencies can be expected (again, if the population meets all H-W assumptions).

I develop population genetics through an arbitrarily defined population of distinguishable phenotypes. The H-W equilibrium is then used to demonstrate how allele frequencies may change. Evolution is defined simply as a change in allele frequency. (The abundant misuse of the terms gene and allele in most texts is a major cause of student confusion throughout the genetics sections.)

The discussion is based on questions and problems rather than terms and definitions, emphasizing that biology is the scientific study of principles, processes and their consequences, not memorization and vocabulary definitions (Journet 1985).

The (Castle) Hardy Weinberg Equilibrium

We start with a population of fish in a pond. The fish were collected, I argue, by the students themselves from a lake and placed in a newly made pond. In a system of incomplete dominance, we have: long tails (TT); medium tails (TT'); and short tails (T'T'). (T' is used rather than t for the sake of consistency with the Mendelian genetics treatment where T/t is used for complete dominant/recessive systems and

T/T' for incomplete dominance.) Suppose we start with 300 TT, 500 TT' and 200 T'T'.

"What are the genotype and allele frequencies?" Incomplete dominance is employed precisely because genotype frequencies are the same as phenotype frequencies which (by definition) add to 1: freq TT + freq TT' + freq T'T' = 1. It should not be assumed that students find this a simple proposition; I use several examples to illustrate frequency. This can be calculated as freq TT = 300/1000 = .3, and similarly freq TT' = .5 and freq T'T' = .2. From here there are two ways to calculate allele frequencies:

1. TT individuals contain only T alleles, T'T' individuals only T' and TT' contain half of each allele. Thus, freq T = freq TT + freq TT'/2 = .55 and freq T' = freq T'T' + freq TT'/2 = .45. We can see freq T (.55) + freq T' (.45) = 1 suggesting the shortcut method of calculating the frequency of the second allele as freq T' = 1 − freq T.

2. All organisms contain two segments of DNA coding for tail length, one on each chromosome, so the total number is 2,000. TT individuals contain two Ts, T'T' contain two T's and TT' contain one of each. Thus the T total is 2 * number of TT individuals + the number of TT' individuals = 1,100 while the total number of T's similarly calculated is 900. The frequency of allele T is thus 1,100/2,000 = .55, and the frequency of T' is 900/2,000 = .45.

While the second method is intuitively easier for many students to understand, it is much better to use the first method in subsequent calculations where genotype frequencies through future generations are calculated.

Now, we ask, "What will be the genotype frequencies in the next generation?" The two important assumptions are: egg and sperm production and fertilization are purely probabalistic; and subsequently, eggs hatch and juvenile fish grow to maturity as the parents die. The only fish present in the next season will be members of the next generation (G2) from these parents. Fish have the advantage, as students are aware, that they are relatively promiscuous; not only does random mating occur, but the liberation of sperm and egg into the water underlines the random nature of the fusion of eggs and sperm. As a result, it is possible to show the achievement of H-W genotype equilibrium very clearly and very quickly without asking students to accept, as a matter of faith, that the achievement of H-W genotype equilibrium will occur in any population, given random mating, with time and a large number of generations.

So, we must first ask "What are the sperm and egg allele frequencies in G1?" Students are often so locked into the simple Mendelian cross that they as-

sume 50 percent T and 50 percent T' regardless of the above genotype frequencies and calculations. (This is why it is inappropriate to start H-W discussions with an allele frequency of .5 T and .5 T' in a population; it confirms a common student misconception.) Most students will not see intuitively that allele frequencies among sperm and egg are identical to those in the entire population of adults. It will be necessary, therefore, to show the calculation, assuming equal numbers of males and females and equivalent genotype frequencies in each sex to those in the entire population. Thus, there are 150 TT males, 250 TT' males and 100 T'T' males. Meiosis produces four gametes per initial gonad cell, with each genotype equally as efficient at producing gametes. From the 500 male fish, we can argue simply, (each fish produces four gametes) there will be 500 * 4 = 2,000 gametes. The TT fish produce 150 * 4 = 600 T gametes, and the TT' fish produce 250 * 2 = 500 T gametes for a total of 1,100 out of the total 2,000 = 1,100/2,000 = .55. The total for T' is thus (100 * 4) + (250 * 2) and the frequency is 900/2,000 = .45. Egg allele frequencies may be similarly calculated.

Now we return to ask again about the genotype frequencies in the next generation. We can address this through the familiar Punnett Square, using the egg and sperm allele frequencies calculated above and remembering that the Punnett represents the fertilization stage in the sexual life cycle. Thus:

		Sperm	
		freq T = .55	freq T' = .45
Egg	freq T = .55	freq TT = .3025	freq TT' = .2475
	freq T' = .45	freq TT' = .2475	freq T'T' = .2025

So, the next generation (G2) genotype (and phenotype) frequencies are: freq TT = .3025, freq TT' = .495 and freq T'T' = .2025.

The genotype frequencies *have changed* from those in the initial arbitrarily defined population we introduced into the pond. This is precisely why we cannot use the binomial expansion of $(p + q) = 1$, $p^2 + 2pq + q^2 = 1$ to calculate (say) the frequency of homozygous dominants and heterozygotes given the homozygous recessive frequency (or number) in a population not *known* to be at H-W equilibrium.

The next question is, "What are the allele frequencies in this new (G2) generation?" Using the same techniques as initially used above:

freq T = freq TT + TT'/2 = .3025 + .2475 = .55 and
freq T' = freq T'T' + freq TT'/2 = .2025 + .2475 = .45.

Thus, we can now see, the allele frequencies *have not changed*.

"What about the genotype frequencies in the G3?" We can, again, use the Punnett Square with the allele = gamete frequencies just calculated, where freq T = .55 and freq T' = .45. Clearly, (though not to students without demonstration) this is identical to the Punnett Square shown above.

This shows that genotype frequencies in G3 are identical to those in G2. But again, we might ask, "What are the allele frequencies in this G3 generation?" By this time students begin to see that allele frequencies will remain unchanged indefinitely, while genotype frequencies initially change until equilibrium is reached, then they too remain unchanged. While the mating behavior of the fish leads to the rapid (one generation) achievement of Hardy Weinberg equilibrium, organisms with less random mating may take several generations to achieve genotype equilibrium.

We now ask, "What might prevent the achievement of H-W equilibrium, or disturb it once it has been achieved?" This leads to an introduction to and identification of the assumptions of Hardy Weinberg and, by corollary, the causes of evolution: unbalanced mutation rates, migration, chance (genetic drift in small populations and the founder effect), nonrandom mating and natural (also artificial) selection.

Only once we have achieved H-W genotype equilibrium and clearly identified its assumptions, is it reasonable to introduce p as the frequency of T and q as the frequency of T' where p + q = 1. We can now show from the last Punnett Square that:

Sperm

		p freq T = .55	q freq T' = .45
Egg	p freq T = .55	p^2 freq TT = .3025	pq freq TT' = .2475
	q freq T' = .45	pq freq TT' = .2475	q^2 freq T'T' = .2025

From this we can derive the H-W equilibrium genotype frequencies as $p^2 + 2pq + q^2 = 1$. But, q (freq T') is only $\sqrt{\text{freq T'T'}}$ when H-W genotype equilibrium frequency has been achieved. Given an equilibrium population, this equation can be used to calculate the frequency of unknown genotypes as usual. For example, if the recessive frequency is .04, q = .2 and p = 1 − q = .8 and the frequency of homozygote = .64 and heterozygote = .32.

If the point of introducing the mathematical treat-

ment of population genetics is to illustrate evolution, I have some doubts about introducing p and q at all. Once we start exploring evolutionary events, where H-W doesn't apply, these terms are not necessary anyway. The time it takes to convince students of the formulae and their application might be better spent on our main objective: evolution.

The Process of Evolution

Of the range of processes that leads to a change in allele frequency, I feel it appropriate to illustrate but a few through the mathematical principles developed above. The influence of *mutation rate* requires a little algebra. The discussion above shows that in the absence of disturbance, allele frequencies will remain the same from generation to generation. Certainly, if no mutation occurs this is also true. But in the presence of mutations, where T → T' (= u) or T' → T (= v), a change in allele frequency will occur. Thus, if T → T' at a frequency of (say) 1 in 100 per generation, then u = .01. After one generation, freq T will drop by .55 * .01 = .0055 and freq T' will increase by this amount.

In G2, then, freq T = .55 − .0055 = .5445 and
freq T' = .45 + .0055 = .4555.

In the absence of mutation T' → T (v = 0), freq T' increases through generations as freq T drops.

But, there will also be some reverse mutation T' → T. "What must be the value of the reverse mutation rate v that will maintain H-W equilibrium?" Clearly, the result of this reverse mutation T' → T must exactly replace the loss of T to T' = .0055 per generation. So, the mutations T' → T, in particular the value of v, must satisfy the equation:

freq T * u = .0055 = freq T' * v.

Since we know freq T' = .45, we can calculate the required value for v as .0055 = freq T' (.45) * v. So, for a balanced allele frequency, v must be .0055/.45 = .012.

The message, then is that if freq T * u =/= freq T' * v, then allele frequencies will change. It is possible now for students to predict the consequences if freq T * u > freq T' * v and the reverse.

To introduce the *Founder Principle* we return to the equilibrium fish population and argue that one wall of the pond is a dam which breaks, allowing 150 fish to escape and form a new colony downstream. There were 50 TT, 25 TT' and 75 T'T' in this new colony. The frequency of phenotypes and genotypes in this colony is purely a consequence of chance; those that happened to be swimming near the dam were washed over when it broke. No phenotypes, because of their genotypes, were more likely to be washed into the new pond than any others.

"What will be the H-W equilibria in the new colony and the old population left behind in the original pond?" Once again, we note, neither population is now at H-W equilibrium for genotype frequency, so we cannot use the binomial expectation (q = sq. root freq. T'T') to calculate allele frequencies. The genotype frequencies are: TT = 50/150 = .3333, TT' = .1667 and T'T' = .5. Allele frequencies, therefore, are:

$$\text{freq T} = \text{freq TT} (.3333) + \text{freq TT'}/2 (.0833) = .4166$$
and
$$\text{freq T'} = \text{freq T'T'} (.5) + \text{freq TT'}/2 (.0833) = .5833.$$

With rounding, small errors will occur.

Arguing again that sperm and egg are in the same frequencies as the alleles:

		Sperm	
		freq T = .4166	freq T' = .5833
Egg	freq T = .4166	freq TT = .1736	freq TT' = .2430
	freq T' = .5833	freq TT' = .2430	freq T'T' = .3402

So, in this new colony, H-W equilibrium genotype frequencies will be TT = .1736, TT' = .4860, and T'T' = .3402.

With random mating, this Punnett Square could be eliminated, and replaced with the H-W binomial expansion $p^2 + 2pq + q^2 = 1$. But this is likely to add to the confusion students already have in determining if a population is at H-W equilibrium or not. Maybe this is an illustration for the honors students.

A similar approach can also be taken to determining the H-W equilibrium in the population left behind. This entire example, clearly, can also be worked as the influence of *emmigration* on allele frequency. By simple addition of some random migrants, *immigration* likewise can be demonstrated. *Genetic Drift*, meanwhile, can be shown by arguing that a random event (a lightning bolt) kills some of the fish, again not as a consequence of their genotype. Comparison of such chance events in small as opposed to large populations also can show the influence of *population size* on chance changes in allele frequency.

Natural Selection now can be introduced by returning to the initial population of fish. Students readily identify predation as an agent capable of acting on the fish, with small tails more likely prey than medium or long tails. "If long and medium tails escape predation, but 20 percent of the short tails are eaten, what will happen to allele and genotype frequencies?"

Equilibrium genotype frequencies were: TT .3025, TT' .495 and T'T' .2025. Out of 1,000 fish, there are 302.5 TT, 495 TT' and 202.5 T'T'. These, then, are also the numbers for juvenile fish in the next generation immediately following parental death. Now, predators consume 20 percent of the short tails (T'T') so at maturity, there are only 202.5 * .8 T'T' = 162. With 302.55 TT and 495 TT', this makes a total of 959.5 fish. So, at maturity, genotype frequencies are TT = 302.5/959.5 = .315, TT' = 495/959.5 = .516 and T'T' = .169. Post-selection allele frequencies, therefore, are:

$$\text{freq T} = \text{freq TT} (.315) + \text{freq TT'}/2 (.258) = .573$$
and
$$\text{freq T'} = \text{freq T'T'} (.169) + \text{freq TT'}/2 (.258) = .427.$$

Predation, then, has altered both genotype and allele frequencies. These allele frequencies may now be cast on a Punnett Square to generate the juvenile frequencies in the next generation, selection again can be exerted through another cycle, and the long term effects of continued selection may be shown as it reduces freq T'.

An alternative means of calculation uses genotype frequencies rather than numbers. This is preferable mathematically, but more difficult for students conceptually. Equilibrium frequencies were TT .3025, TT' .495 and T'T' .2025. Suppose predators catch all fish, but the small tails more efficiently (a more realistic condition); for every 1 TT and 1 TT' that escape, only .8 T'T' escape. After predation, freq T'T' now becomes .8 * .2025 = .162. But since freq T'T' has dropped, the total is no longer 1, it is .3025 + .495 + .162 = .9595. The actual genotype frequencies are TT = .3025/.9596, TT' = .495/.9595 and T'T' = .162/.9596. The solution is identical to that above, but the use of frequencies eliminates the need to refer to and calculate the fish numbers. The value of .9596 in this example represents the reproductive success of the generation of fish; selection acts to maximize this value.

The bold instructor might introduce here frequency dependent selection where selection pressure changes such that the most abundant phenotype is always preyed upon the most. But, I don't recommend it.

Following the discussion of chance events and natural selection, it is appropriate to emphasize the difference. Chance events, such as drift, lead to changes in allele frequency that may or may not be adaptive; certainly they are not deterministic or directed, while natural selection is deterministic and directed.

Conclusion

The point of this discussion is to show students how chance events and natural selection influence allele frequencies and, therefore, cause evolution.

They see that the process of evolution is a mathematical inevitability consequent upon circumstances that disturb the Hardy-Weinberg equilibrium. The process of natural selection leading to adaptation is seen, also, as a consequence of the presence at a particular locus of two or more alleles which confer upon their bearers (the fish) different probabilities or reproductive success. Those which confer upon their bearers a greater probability of reproductive success will inevitably increase in frequency in the next generation. In an era when evolution is thought by many to be a matter of faith, equivalent to belief in creation, this is an important realization for students. Though all students may not fully comprehend the mathematics or be able to solve complex mathematical problems in population genetics, they will all be able to see more clearly how orderly and predictable is the process of evolution.

This section of the course then is followed by a similar discussion of the consequences of natural selection, including treatment of mimicry, camouflage, warning coloration, etc. with reference to an extensive slide program. Students are required, through this presentation, to explain evolutionary events from their origins in random DNA changes, through the spread of alleles by reproduction and gamete recombination, to the increase of some alleles at the expense of others resulting from their differential reproductive success. The fossil record, speciation (Journet 1986), convergence and divergence, and the evidence from comparative studies are introduced as the consequence of evolution rather than as evidence for the process. While I include some simple mathematical problems on tests, I do not emphasize them extensively. There is ample opportunity in this and other sections of the course to address the higher learning objectives of Bloom (1956) (application, analysis, synthesis and evaluation) through word rather than mathematical problems.
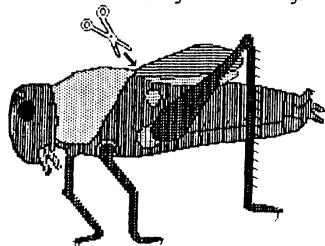
Acknowledgements

References

Bloom, B.S. (1956). *Taxonomy of educational objectives*. NY: Longman's.

Curtis, H. & Barnes, N.S. (1985). (4th ed.). *Invitation to biology*. NY: Worth Publishers, Inc.

Inhelder, B. & Piaget, J. (1958). *The growth of logical thinking from childhood to adolescence*. NY: Basic Books.

Journet, A.R.P. (1984). Pac-ing the most into genetics. *The Science Teacher, 51*(7), 5-53.

———. (1985). Are we teaching science? *Journal of College Science Teaching, 14*(2), 236-238.

———. (1986). A frog model of geographic speciation. *American Biology Teacher, 48*(1), 17-19.

McKinnon, J.W. (1971). Earth science, density and the college freshman. *Journal of Geological Education, 19*, 218-220.

Renner, J.W. (1977). *Evaluating intellectual development using written responses to selected science problems*. NSF Grant Report EPP75-19596. University of Oklahoma, Normal, OK.

Ritchie, D.D. & Carola, R. (1982). (2nd ed.). *Biology*. NY: Addison-Wesley Publishing Co.

Sanders, S.G. (1978). The importance of the Piaget learning theory for the college teacher. *Journal of College Science Teaching, 7*, 283-287.

Stanley, C.M., Journet, A.R.P., & Scheibe, J.S. (1986). *Cognitive development and the investigative biology laboratory*. In prep.

Wallace, R.A. (1981). (3rd ed.). *Biology: The world of life*. NY: Goodyear Publishers, Inc.