Hardy Weinberg Equilibrium

John T. Baldwin// University of Illinois at Chicago

Mar 2, 2022

Sample problem for the revised AP biology exam, 2010

The data below demonstrate the frequency of tasters and non-tasters in an isolated population at Hardy-Weinberg equilibrium. The allele for non-tasters is recessive. How many of the tasters in the population are heterozygous for tasting?

tasters	non-tasters
8235	4325

What do these words mean?

Sample problem for the revised AP biology exam, 2010

The data below demonstrate the frequency of tasters and non-tasters in an isolated population at Hardy-Weinberg equilibrium. The allele for non-tasters is recessive. How many of the tasters in the population are heterozygous for tasting?

tasters	non-tasters
8235	4325

What do these words mean?

Solution provided:

An acceptable answer would be any number in the range of 6030-6156, depending on how the students rounded the variables in the Hardy-Weinberg equation.

We will come back to this.

Hardy Weinberg Equilibrium

Definition

A population is in Hardy-Weinberg equilibrium if the genotype frequencies are the same in each generation.

What do these words mean?
[BS15]
https://online.ucpress.edu/abt/article/77/8/577/18809/Making-Sense-of-Hardy-Weinberg-Equilibrium

Brief History of Genetics

Chronology

1. Mendel: forgotten or ignored?
   i. Mendel G. Versuche über Pflanzen-Hybriden. Verh. Naturfsch Ver in Brünn 1866;4: 3-47
   ii. 1866- 1892 (10 mostly minor references)
   iii. Liberty Hyde Bailey 1892/1895 Cross-breeding and hybridizing; book on Plant breeding
   iv. 1901 Bateson W. Problems of heredity as a subject for horticultural investigation.
2. The stats/genetics wars in UK 1900 - 1930
   i. Pearson and Yale vs
   ii. Bateson and Punnett
   iii. 1908 Hardy/ Weinberg
   iv. The rest is another story. Modern Synthesis early 1930's.

1908

Weinberg: January

Weinberg lectured on January 13, 1908, in Württemberg. published in German in September.
'derived the general equilibrium principle for a single locus with two alleles'

Punnett lectures on genetics (February 28)

1. Punnett gives clear account of Mendel's theory *Mendelism in Relation to Disease*
2. Yale and other 'statisticians' objected that Mendelians were poor experimentalists and over-emphasized the importance of inheritance, especially in regard to disease.

1908: Punnett February; Hardy: April

Question period and Hardy

1. 'Punnett says Mr. Yule wondered why the nation was not slowly becoming brown-eyed and brachydactylous, since these characters were both dominant.'
2. That is, Punnett wondered dominant gene did not continually increase in frequency.
3. Yale/Pearson knew better. But thought $p = 1/2$ was essential to arguments about fixation.
4. Punnett asked Hardy
5. Hardy gave the one page answer overnight. (submitted April 5 appeared July 1908 in Science
6. Not in Nature because it was about Genetics.

Mendelian traits: Allele Frequencies

monogenic Mendelian trait

The gene has two alleles:
one dominant A and the other recessive a.

examples?

Mendelian traits: Allele Frequencies

monogenic Mendelian trait

The gene has two alleles:
one dominant A and the other recessive a.

examples?
tasting PTC, cleft chin, freckles, widow peak, sickle cell anemia, cystic fibrosis, Brachydactyly, Huntington disease ?, and Duchenne muscular dystrophy.

Allele frequency in one generation

In a population of n individuals. The frequency of the *A* allele is the number of A alleles divided by total number of alleles at this locus (for this gene) within the population ($2n$).

If n is 4000 and 2000 of the alleles of this locus in the population are *a*, how many alleles are *A*?

Allele frequency in one generation

In a population of n individuals. The frequency of the *A* allele is the number of A alleles divided by total number of alleles at this locus (for this gene) within the population ($2n$).

If n is 4000 and 2000 of the alleles of this locus in the population are *a*, how many alleles are *A*?
the frequency of the *A* allele is $\frac{3}{4}$ and that of the *a* allele is $\frac{1}{4}$

Mendelian traits: Genotype Frequencies

Genotypes: An individual can be AA, Aa or aa
The same allele frequency can be attained by many different genotype distributions:

Genotype frequency

1. 3000 AA, 1000 aa;
2. 500 aa, 1000 Aa, and 2500 AA;
3. 250 aa, 1500 Aa and ? AA.

Mendelian traits: Genotype Frequencies

Genotypes: An individual can be AA, Aa or aa
The same allele frequency can be attained by many different genotype distributions:

Genotype frequency

1. 3000 AA, 1000 aa;
2. 500 aa, 1000 Aa, and 2500 AA;
3. 250 aa, 1500 Aa and ? AA.

2250 AA

Genotype frequency

Variants of a Punnett Square

	A	a
A	AA	Aa
a	aA	aa

Suppose there are 4000 individuals - all *homozygous*.

	A	a
A	3000	0
a	0	1000

	A $\frac{3}{4}$	a $\frac{1}{4}$
A $\frac{3}{4}$	3000	0
a $\frac{1}{4}$	0	1000

Requirements for Hardy-Weinberg Equilibrium

Definition

A population is in Hardy-Weinberg equilibrium if the genotype distribution remains constant from generation to generation.

Theorem

A population is in Hardy-Weinberg equilibrium (with respect to a particular gene) if the following conditions hold :

1. There is no migration (gene flow) in or out of the population.
2. Natural selection is not occurring.
3. Mutation is not occurring.
4. Each member of the population is equally likely to breed.
5. The population is infinite.
6. (Full random mating) Each pair from the population is equally likely to breed

Preserving Allele frequency: The observation

Observation 1

As long as a population satisfies conditions 1-5 the allele frequencies (p and q) are the same in each generation.

Proof Conditions 1 and 2 guarantee that there is no change in the allele frequencies between the birth and maturity of the next generation; there are no unaccounted forces that would change the allele frequencies (i.e., one phenotype is not more fit than the other).

Conditions 3 and 4 guarantee that at birth the pool of alleles in the next generation is the same as in the current generation; mating just reshuffles the alleles; the allele frequencies remain the same.

Why does ∞ matter?

Why does ∞ matter?

The population needs to be infinite to guarantee that the frequencies remain exactly p and q. The probabilities p and q represent the averages over many trials and so it will only be approximate in a particular trial on a finite population.

Preserving Genotype frequency

Let A and a represent the two possible alleles of a simple Mendelian trait; let p and q represent the frequencies of A and a, respectively in the current generation.

the Crucial H-W Insight:

Under fully random mating and (the other biological conditions), the frequency of AA homozygotes in the next generation is given by:

Genotype

	A p	a q
A p	p^2	pq
a q	pq	q^2

Infinite population is now mathematically necessary, not just empirically. Why?

Hardy Weinberg equilibrium

The Hardy Weinberg Theorem

Consider a population satisfying conditions 1-6.
If the allele frequencies in one generation are p and q then in the next:

1. the allele frequencies remain p and q
2. and the genotype frequencies in the next generation are $p^2, 2pq, q^2$
3. and genotype and allele frequencies remain the same for as many generations as conditions 1-6 continue to hold.

Proof

1. 1) The observation: allele frequency constant
2. 2) the insight: Under condition 1-6 allele frequency determines genotype frequency.

A fact which does not depend on HW-equilibrium

genotype distribution determines allele distribution in the same generation

$p = 2P(AA) + P(Aa)$ and $q = 2P(aa) + P(Aa)$
Where $P(X)$ is the probability of genotype X.

Proof: just count.

Hardy Weinberg equations

The 9th grade bio distraction

Hardy and Weinberg NEVER mentioned them!
These two equations are widely used in biology teaching, but all too often they are used as a mathematics exercise that obfuscates the problem.

Equation 1: $p + q = 1$ (allele frequencies) is true for any probabilities of complementary outcomes. E.g.for any monogenic Mendelian trait

Squaring equation 1 yields:
Equation 2: $p^2 + 2pq + q^2 = 1$ (genotype frequencies)

These observations have no biological content.
Equation 2 simply follows mathematically from Equation 1. There is no assumption about random mating and no biological assumption in the step from Equation 1 to Equation 2.

Two problems: Compare and contrast

Sample problem for the revised AP biology exam, 2010

The data below demonstrate the frequency of tasters and non-tasters in an isolated population at Hardy-Weinberg equilibrium. The allele for non-tasters is recessive. How many of the tasters in the population are heterozygous for tasting?

tasters	non-tasters
8235	4325

POGIL Activities for AP Biology

The ability to taste PTC is due to a single dominant allele T. You sampled 215 individuals and determined that 150 could detect the bitter taste of PTC and 65 could not. Calculate the following:

(a) The frequency of the recessive allele.

(b) The frequency of the dominant allele.

(c) The frequency of the heterozygous individuals.

Compare and contrast solutions

what do you think of those problems? How would you solve them?s

Compare and contrast solutions

what do you think of those problems? How would you solve them?s

AP Solution provided:

An acceptable answer would be any number in the range of 6030-6156, depending on how the students rounded the variables in the Hardy-Weinberg equation.

Pogil Solution provided:

$$q^2 = \sqrt{(\frac{65}{115} = \sqrt{.3} = .55$$

$$p = 1 - q = .45$$

$$2pq = .495$$

HW-equilibrium ?

(K-State Parasitology Laboratory, 2000: Sickle cell anemia

1. Normal homozygous individuals (SS) are easily infected with the malaria
2. Individuals homozygous for the sickle-cell trait (ss) have red blood cells that readily collapse when deoxygenated. Although malaria cannot grow in these red blood cells, individuals often die because of the genetic defect.
3. Individuals with the heterozygous condition (Ss) have tend to survive better than either of the homozygous conditions.

If 9 % of an African population is born with a severe form of sickle-cell anemia (ss), what percentage of the population will be more resistant to malaria because they are heterozygous (Ss) for the sickle-cell gene?

Can you solve this problem?

Genetic Drift

Suppose condition 1-4 and 6 are satisfied but the population is $N < \infty$

.

1. There is no reason to think p^2N is an integer.
2. Moreover, from one generation to the next is a random sample; the number of A alleles in the next generation is only approximately $2pN$.

Genetic drift in the change in the genotype frequencies as a finite population satisfying conditions 1)-5) evolves.

Rate of Genetic drift

Given a population of size N with $P(A) = p$ and $P(a) = q$

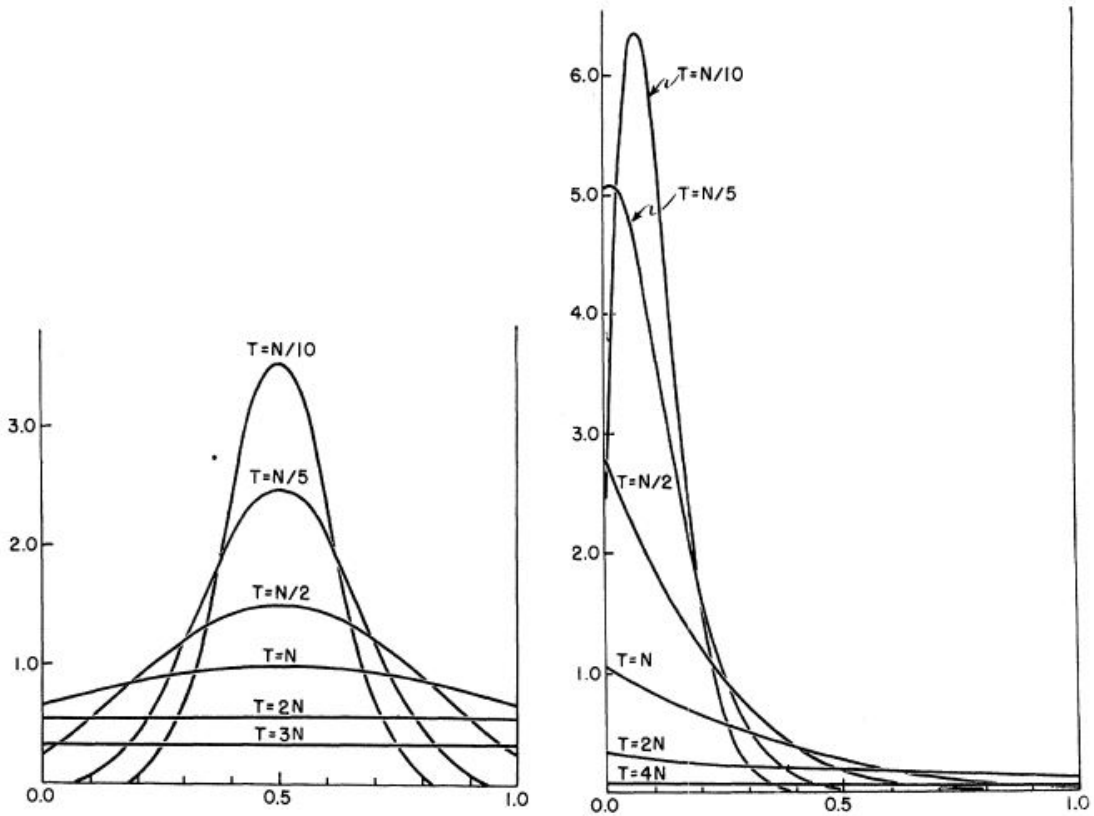
Wright 1931 With probability p the a will die out and the population will all be A; With probability q?

Kimura 1955 the expected time to fixation (in one of the two) is $2N$ generations.

[Hol12] Let $\bar{t}$ denote the average time until one the alleles disappears

Kimura Graph

The graph below summarizes the epic study of genetic drift in [Kim55].



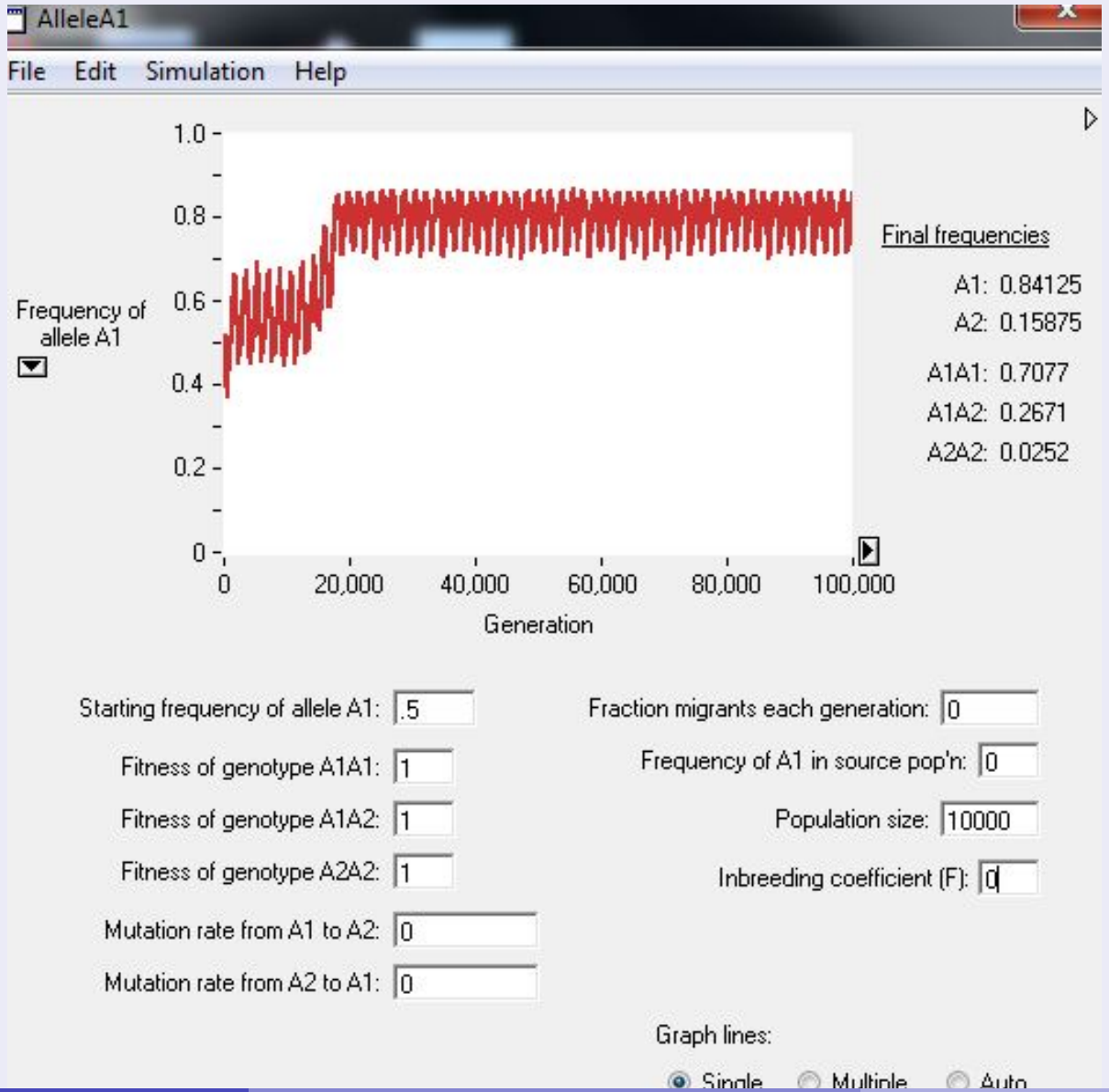
FIGS. 1–2.—The processes of the change in the probability distribution of heterallelic classes, due to random sampling of gametes in reproduction. It is assumed that the population starts from the gene frequency 0.5 in Fig. 1 (left) and 0.1 in Fig. 2 (right). t = time in generation; N = effective size of the population; abscissa is gene frequency; ordinate is probability density.

Kimura graph explanation

Each curve represents the situation after T generations, where T is given in terms of N, the size of the original population. The x-axis represent the frequency of a given allele A. The y-axis is a probability density function.
Note that for some arguments the value of the probability density can be much greater than 1. The requirement for be a probability density function is that the area under the curve is 1.
Thus, in figure 1, where the original gene frequency is .5, after a time only N/10 generation, the gene frequency is bunched between .4 and .6; perhaps 70% still have a frequency in this interval ($3.5 \times .2$). But after 2N generations about 25% of the population is fixed at each of A and a and the other 50% are equally like to have any allele frequency p with $0 < p < 1$. But after 3N generations about 1/3 of the population is fixed at each of A and a and the other 1/3 are equally like to have any allele frequency p with $0 < p < 1$.

Simulations

Allele A1 program on windows 7



Hardy could be wrong

No discovery of mine has made, or is likely to make, directly or indirectly, for good or ill, the least difference to the amenity of the world.
G.H Hardy, A mathematicians apology

References I

John T. Baldwin and M.U. Smith.
Making sense of hardy-weinberg equilibrium.
American Biology Teacher, 77:577–582, 2015.
https://doi.org/10.1525/abt.2015.77.8.3.

Kent Holsinger.
Lecture notes in Population Genetics.
self-published, 2012.

Motoo Kimura.
Solution of a process of random genetic drift with a continuous model.
Proceedings of the National Academy of Sciences of the United States of America, 41:11–20, 1955.
http://www.jstor.org/stable/89306.