H

Hardy–Weinberg Equilibrium and Random Mating

J Lachance, Georgia Institute of Technology, Atlanta, GA, USA

Glossary

Allele One of a number of alternative forms of a gene.
Assortative mating Non-random mating on the basis of phenotype.
Fitness The expected contribution of an individual or genotype to next generation's gene pool.
Genetic drift Changes in allele or genotype frequencies due to random sampling.
Genotype The pair of alleles present at a single locus.
Inbreeding Preferential mating between relatives.
Natural selection Differential survival or reproduction.
Phenotype The set of an organism's observable traits.
Ploidy The number of sets of chromosomes in a cell.
Single nucleotide polymorphism (SNP) DNA sequence variation at a single nucleotide position that differs between individuals.

History

Darwin's theory of evolution transformed biology in the mid-nineteenth century. However, Darwin had an inaccurate understanding of heredity and his belief in blending inheritance was problematic. Because blending inheritance ultimately results in homogenized populations full of intermediate genotypes, it is unable to explain how genetic variation can persist over evolutionary time (Charlesworth and Charlesworth, 2009). It wasn't until the rediscovery of Mendelian genetics in the early twentieth century that a potential solution could be found. Importantly, Mendel treated genes as discrete units, which enabled a mathematical formulation of population genetics. Despite the rediscovery of Mendelian genetics and the pioneering work of biologists and statisticians (Castle, 1903; Pearson, 1904) there was still some confusion, and during a 1908 meeting of the Royal Society of Medicine it was claimed that dominant mutations would increase in frequency until phenotypes reached 3:1 Mendelian ratios (Provine, 2001). Something about this claim seemed wrong to R.C. Punnett, who passed on the question to his cricket-playing friend G.H. Hardy. Hardy was one of the leading mathematicians in Britain and he quickly deduced a solution. In a 1908 letter to the journal *Science*, Hardy began by writing "I should have expected the very simple point that I wish to make to have been familiar to biologists" before using the binomial expansion to demonstrate that genotype frequencies will reach a stable equilibrium after one generation of random mating (Hardy, 1908). Unbeknownst to Hardy, a Germany physician named Wilhelm Weinberg had published a similar paper on the same topic six months earlier. Weinberg derived a general equilibrium principle for a single locus with two segregating alleles (Weinberg, 1908).

The Hardy–Weinberg Principle

The Hardy–Weinberg principle relates allele frequencies to genotype frequencies in a randomly mating population. Imagine that you have a population with two alleles (A and B) that segregate at a single locus. The frequency of allele A is denoted by p and the frequency of allele B is denoted by q. The Hardy–Weinberg principle states that after one generation of random mating genotype frequencies will be p^2, $2pq$, and q^2. In the absence of other evolutionary forces (such as natural selection), genotype frequencies are expected to remain constant and the population is said to be at *Hardy–Weinberg equilibrium*. The Hardy–Weinberg principle relies on a number of assumptions: (1) random mating (i.e, population structure is absent and matings occur in proportion to genotype frequencies), (2) the absence of natural selection, (3) a very large population size (i.e., genetic drift is negligible), (4) no gene flow or migration, (5) no mutation, and (6) the locus is autosomal. When these assumptions are violated, departures from Hardy–Weinberg proportions can result.

One useful way to think about the Hardy–Weinberg principle is to use the metaphor of a gene pool (Crow, 2001). Here, individuals contribute alleles to an infinitely large pool of gametes. In a randomly mating population without natural selection, offspring genotypes are found by randomly sampling two alleles from this gene pool (one from their mother and one from their father). Because the allele that an

 doi:10.1016/B978-0-12-800049-6.00022-6

individual receives from their mother is independent of the allele they receive from their father, the probability of observing a particular genotype is found by multiplying maternal and paternal allele frequencies. Mathematically this involves the binomial expansion: $(p + q)^2 = p^2 + 2pq + q^2$ (see the modified Punnett Square in Figure 1 for a graphical representation). Note that there are two ways that an individual can be an *AB* heterozygote: they can either inherit an *A* allele from their mother and a *B* allele from their father or they can inherit a *B* allele from their mother and an *A* allele from their father.

Additional insight can be found by considering an empirical example (Figure 2). Consider a population that initially contains 18 *AA* homozygotes, 4 *AB* heterozygotes, and 3 *BB* homozygotes. The alleles in the gene pool, 80% are *A* and 20% are *B*. After a single generation of random mating we observe Hardy–Weinberg proportions: 16 *AA* homozygotes, 8 *AB* heterozygotes, and 1 *BB* homozygote. Note that allele frequencies remain unchanged.

There are a number of evolutionary implications of the Hardy–Weinberg principle. Most importantly, genetic variation is conserved in large, randomly mating populations. A second implication is that the Hardy–Weinberg principle allows one to determine the proportion of individuals that are carriers for a recessive allele. Third, it is important to note that dominant alleles are not always the most common alleles in a population. Another implication of the Hardy–Weinberg principle is that rare alleles are more likely to be found in heterozygous individuals than in homozygous individuals. This occurs because q^2 is much smaller than $2pq$ when q is close to zero.

The Hardy–Weinberg principle can be generalized to include polyploid organisms and genes that have more than two segregating alleles. Equilibrium genotype frequencies are found by expanding the multinomial $(p_1 + \ldots + p_k)^n$, where n is the number of sets of chromosomes in a cell and k is the number of segregating alleles. For example, tetraploid organisms ($n = 4$) with two segregating alleles ($k = 2$) are expected to have genotype frequencies of: ${p_1}^4$ (*AAAA*), $4{p_1}^3p_2$ (*AAAB*), $6{p_1}^2{p_2}^2$ (*AABB*), $4p_1{p_2}^3$ (*ABBB*), and p^4 (*BBBB*). Similarly, diploid organisms ($n = 2$) with three segregating alleles ($k = 3$) are expected to have genotype frequencies of: ${p_1}^2$ (*AA*), ${p_2}^2$ (*BB*), ${p_3}^2$ (*CC*), $2p_1p_2$ (*AB*), $2p_1p_3$ (*AC*), and $2p_2p_3$ (*BC*). Genotype frequencies sum to one for each of the above scenarios. Although the Hardy–Weinberg principle can also be generalized to include genes located on sex chromosomes (e.g., X chromosomes in humans), it is important to note that it can take multiple generations for genotype frequencies at sex-linked loci to reach equilibrium values.

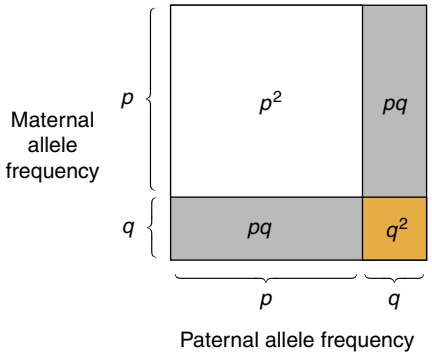

Figure 1 Graphical representation of the Hardy–Weinberg principle. The frequency of *A* alleles is denoted by *p* and the proportion of *B* alleles by *q*. *AA* homozygotes are represented by white, *AB* heterozygotes by gray, and *BB* homozygotes by gold. Shaded areas are proportional to the probability of observing each genotype.

Testing for Hardy–Weinberg Proportions

A Chi-square (χ^2) test with one degree of freedom can be used to determine whether a population is at Hardy–Weinberg equilibrium (Weir, 1996) (see Table 1 for an example). First, the observed numbers of individuals with each genotype are counted. These genotype frequencies are then used to obtain allele frequencies (p and q). The expected numbers of individuals with each genotype are calculated by multiplying the total number of individuals sampled by p^2, $2pq$, and q^2. Once observed and expected genotype counts are known, χ^2 statistics

Table 1 Chi-square test of Hardy–Weinberg proportions. Allele frequencies are $p = 0.7$ and $q = 0.3$. The expected number of individuals with each genotype are calculated by multiplying the total number of individuals by p^2 (*AA*), $2pq$ (*AB*), and q^2 (*BB*). A χ^2 of 8.724 with one degree of freedom yields a *p*-value of 0.00172, indicating that there is a statistically significant departure from Hardy–Weinberg proportions

Genotype	*Observed*	*Expected*	χ^2
AA	117	107.8	0.785
AB	74	92.4	3.664
BB	29	19.8	4.275
Total	220	220	8.724

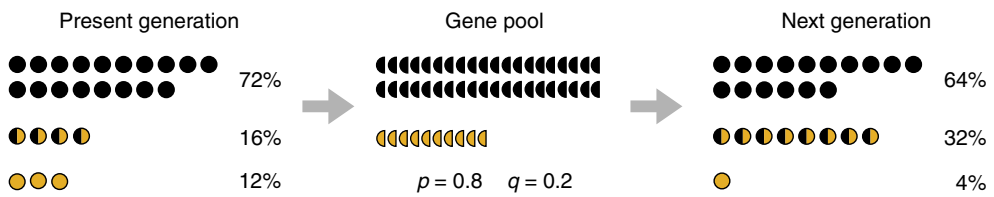

Figure 2 Hardy–Weinberg example. *AA* homozygotes (black circles), AB heterozygotes (black and gold circles), and *BB* homozygotes (gold circles) contribute to the gene pool. *A* alleles are shown as black half-circles and *B* alleles are shown as gold half-circles. After a single generation of random mating Hardy–Weinberg proportions are obtained.

can be calculated for each genotype using the equation $\chi^2 = (observed - expected)^2/expected$. These χ^2 values are summed, and if the overall χ^2 test statistic is greater than 3.84 the null hypothesis that the population is in Hardy–Weinberg equilibrium can be rejected (p-value <0.05). The power of this statistical test is directly related to sample size: departures from Hardy–Weinberg proportions are unlikely to be statistically significant if you have genotype frequency data from a small number of individuals.

Departures from Hardy–Weinberg Proportions

There are a number of reasons why one might fail to observe Hardy–Weinberg proportions, and these departures from Hardy–Weinberg equilibrium can involve either an excess of homozygotes or an excess of heterozygotes.

Genotyping Error

While it may not be the most interesting scenario from an evolutionary perspective, departures from Hardy–Weinberg proportions sometimes occur because of genotyping error. These errors can be due to low quality DNA samples, biochemical artifacts, or human error (Pompanon *et al.*, 2005). For example, low-coverage DNA sequencing can cause genotypes to be incorrectly inferred. To improve data quality, a common practice in genome-wide association studies is to filter out any single nucleotide polymorphisms (SNPs) that show significant departures from Hardy–Weinberg proportions (Turner *et al.*, 2011). However, this runs the risk of eliminating SNPs that have departures from Hardy–Weinberg proportions that are due to some other reason than genotyping error.

Non-Random Mating and Population Structure

Non-random mating leads to departures from Hardy–Weinberg proportions. For example, inbreeding and positive assortative mating (where individuals prefer to mate with phenotypically similar individuals) yield an excess of homozygotes. By contrast, negative assortative mating (where opposites attract and individuals prefer to mate with phenotypically different individuals) results in excess of heterozygotes. Population structure also causes departures from Hardy–Weinberg proportions. For example, consider what happens when samples are drawn from multiple populations instead of a single randomly mating population. If these samples are pooled together and there are allele frequency differences between source populations the resulting mixture will have an excess of homozygotes. This reduction in heterozygosity is known as the *Wahlund effect* (Wahlund, 1928). In practice, departures from Hardy–Weinberg proportions due to non-random mating and population structure tend to be genome-wide (i.e., their effects can be seen at multiple loci).

Natural Selection

Natural selection modifies allele and genotype frequencies and these effects depend on both the magnitude and type of selection present. Selection results in departures from Hardy–Weinberg proportions whenever genotypic fitnesses are non-multiplicative (i.e., $w_{AB}{}^2 \neq w_{AA} \times w_{BB}$) (Lachance, 2008). Not surprisingly, overdominance (heterozygote advantage) results in an excess of heterozygotes compared to Hardy–Weinberg expectations, and underdominance (heterozygote disadvantage) results in an excess of homozygotes. Strong directional selection, such as when one allele is a recessive lethal, leads to marked departures from Hardy–Weinberg expectations. However, weak directional selection has only modest effects on genotype frequencies, and detecting these effects can require sample sizes that are larger than 10 000 individuals (Lachance, 2009). Note that the effects of natural selection tend to be locus-specific rather than genome-wide.

Other Causes

Additional population genetic phenomena that can result in departures from Hardy–Weinberg proportions include genetic drift and mutation. In small populations the proportion of individuals with each genotype often differs from p^2, $2pq$, and q^2. The idea here is that sampling effects due to genetic drift yield a slight excess of homozygotes or heterozygotes. These effects are more pronounced in very small populations. This is akin to what happens when you flip a coin a small number of times and end up with an unequal number of heads and tails. Although mutation can lead to slight departures from Hardy–Weinberg proportions, these departures are difficult to detect. This is because mutation rates tend to be small (on the order of 10^{-8} mutations per base pair per generation in humans).

Ultimately, departures from Hardy–Weinberg proportions reveal the biological complexity of natural populations. By building upon the foundation of the Hardy–Weinberg principle, generations of population geneticists have been able to construct more complex models of evolution. These mathematical models have led to an increased understanding of how evolutionary forces shape natural patterns of genetic diversity.

See also: Genetic Variation in Populations. Linkage Disequilibrium: Population Genetics of Multiple Loci

References

Castle, W.E., 1903. The laws of heredity of Galton and Mendel: And some laws governing race improvement by selection. Proceedings of the American Academy of Arts and Sciences 39, 223–242.

Charlesworth, B., Charlesworth, D., 2009. Darwin and genetics. Genetics 183, 757–766.

Crow, J.F., 2001. The beanbag lives on. Nature 409, 771.

Hardy, G.H., 1908. Mendelian proportions in a mixed population. Science 28, 49–50.

Lachance, J., 2008. A fundamental relationship between genotype frequencies and fitnesses. Genetics 180, 1087–1093.

Lachance, J., 2009. Detecting selection-induced departures from Hardy–Weinberg proportions. Genetics Selection and Evolution 41, 15.

Pearson, K., 1904. Mathematical contributions to the theory of evolution. XII. On a generalised theory of alternative inheritance, with special reference to Mendel's laws. Philosophical Transactions of the Royal Society A. 53–86.

Pompanon, F., Bonin, A., Bellemain, E., Taberlet, P., 2005. Genotyping errors: Causes, consequences and solutions. Nature Reviews Genetics 6, 847–859.
Provine, W.B., 2001. The Origins of Theoretical Population Genetics. Chicago: University of Chicago Press.
Turner, S., Armstrong, L.L., Bradford, Y., *et al*., 2011 Quality control procedures for genome-wide association studies. Current Protocols in Human Genetics Chapter 1: Unit1 19.
Wahlund, S., 1928. Zusammensetzung von Populationen und Korrelationserscheinungen vom Standpunkt der Vererbungslehre aus betrachtet. Hereditas 11, 65–106.
Weinberg, W., 1908. Über vererbungsgesetze beim menschen. Molecular and General Genetics MGG 1, 440–460.
Weir, B.S., 1996. Genetic Data Analysis II: Methods for Discrete Population Genetic Data. Sunderland, MA: Sinauer Associates.