HARDY WEINBERG PRINCIPLE OF GENETIC EQUILIBRIUM AND ITS MATHEMATICAL DERIVATION

Dr Poonam kumari
Dept Of Zoology
(PG SEMESTER II CC-08)

The Hardy–Weinberg principle, also known as the Hardy–Weinberg equilibrium, model, theorem, or law, states that allele and genotype frequencies in a population will remain constant from generation to generation in the absence of other evolutionary influences. These influences include genetic drift, mate choice, assortative mating, natural selection, sexual selection, mutation, gene flow, meiotic drive, genetic hitchhiking, population bottleneck, founder effect and inbreeding.
The mathematical proof of invariance of gene frequency under given assumptions, require:

a) simple knowledge of school algebra and
b) basic concepts of Mendelian genetics

The Hardy-Weinberg Theorem deals with Mendelian genetics in the context of populations of diploid, sexually reproducing individuals. Given a set of assumptions (discussed below), this theorem states that:

1. Allele frequencies in a population will not change from generation to generation.
2. If the allele frequencies in a population with two alleles at a locus are p and q, then the expected genotype frequencies are p^2, $2pq$, and q^2. This frequency distribution will not change from generation to generation once a population is in Hardy-Weinberg equilibrium. For example, if the frequency of allele A in the population is p and the frequency of allele a in the population is q, then the frequency of genotype $AA = p^2$, the frequency of genotype $Aa = 2pq$, and the frequency of genotype $aa = q^2$. If there are only two alleles at a locus, then $p + q$, by mathematical necessity, equals one. The Hardy-Weinberg genotype frequencies, $p^2 + 2pq + q^2$, represent the binomial expansion of $(p + q)^2$, and also sum to one (as must the frequencies of all genotypes in any population, whether it is in Hardy-Weinberg equilibrium). It is possible to apply the Hardy-Weinberg Theorem to loci with more than

two alleles, in which case the expected genotype frequencies are given by the multinomial expansion for all k alleles segregating in the population: $(p_1 + p_2 + p_3 + \ldots + p\mathrm{k})^2$.

The conclusions of the Hardy-Weinberg Theorem apply only when the population conforms to the following assumptions:

1. Natural selection is not acting on the locus in question (i.e., there are no consistent differences in probabilities of survival or reproduction among genotypes).
2. Neither mutation (the origin of new alleles) nor migration (the movement of individuals and their genes into or out of the population) is introducing new alleles into the population.
3. Population size is infinite, which means that genetic drift is not causing random changes in allele frequencies due to sampling error from one generation to the next. Of course, all natural populations are finite and thus subject to drift, but we expect the effects of drift to be more pronounced in small than in large populations.
4. Individuals in the population mate randomly with respect to the locus in question. Although nonrandom mating does not change allele frequencies from one generation to the next if the other assumptions hold, it can generate deviations from expected genotype frequencies, and it can set the stage for natural selection to cause evolutionary change.

If the genotype frequencies in a population deviate from Hardy-Weinberg expectations, it takes only one generation of random mating to bring them into the equilibrium proportions, provided that the above assumptions hold, that allele frequencies are equal in males and females (or else that individuals are hermaphrodites), and that the locus is autosomal. If allele frequencies differ between the sexes, it takes two generations of random mating to attain Hardy-Weinberg equilibrium. Sex-linked loci require multiple generations to attain equilibrium because one sex has two copies of the gene and the other sex has only one.

It is important to recognize that the Hardy-Weinberg equilibrium is a neutral equilibrium, which means that a population perturbed from its Hardy-Weinberg genotype frequencies will indeed reach equilibrium after a single generation of

random mating (if it conforms to the other assumptions of the theorem), but it will be a *new* equilibrium if allele frequencies have changed. This property distinguishes a neutral equilibrium from a stable equilibrium, in which a perturbed system returns to the same equilibrium state. It makes sense that the Hardy-Weinberg equilibrium is not stable, since a change from the equilibrium genotype frequencies will generally be associated with a change in allele frequencies (p and q), which will in turn lead to new values of p^2, $2pq$ and q^2. Thereafter, a population that meets Hardy-Weinberg assumptions will remain at the new equilibrium until perturbed again.

Although statistical deviation from Hardy-Weinberg expectations generally indicates violation of the assumptions of the theorem, the converse is not necessarily true. Some forms of natural selection (e.g., balancing selection, which maintains multiple alleles in a population) can generate genotypic frequency distributions that conform to Hardy-Weinberg expectations. It may also be true that migration or mutation is occurring, but at such low rates as to be undetectable using available statistical methods. And, of course, all real populations are finite and thus susceptible to at least some evolution via genetic drift.

Evolutionary Implications of the Hardy-Weinberg Theorem

The Hardy-Weinberg Theorem demonstrates that Mendelian loci segregating for multiple alleles in diploid populations will retain predictable levels of genetic variation in the absence of forces that change allele frequencies. A common way of visualizing these expectations is to plot p^2, $2pq$ and q^2 as a function of allele frequencies (Figure 01). This graphical presentation emphasizes two important consequences of the Hardy-Weinberg principle:

1. Population heterozygosity (the frequency of heterozygotes) is highest when $p = q = 0.5$.
2. Rare alleles are found primarily in heterozygotes, as they must be, given that q^2 is much smaller than $2pq$ when q is near zero, and p^2 is much smaller than $2pq$ when p is near zero.

The second point takes on particular significance if we consider the potential for natural selection to influence the frequencies of new mutations. If a population conforms to all other Hardy-Weinberg assumptions, selection will eventually fix an

advantageous allele in the population such that all individuals are homozygous for that allele. The initial increase in frequency of a rare, advantageous, dominant allele is more rapid than that of a rare, advantageous, recessive allele. This is because, as we have seen, rare alleles are found mostly in heterozygotes, such that a new recessive mutation can't be "seen" by natural selection until it reaches a high enough frequency (perhaps by drift in a real, finite population) to start appearing in homozygotes. A new dominant mutation, however, is immediately visible to natural selection because its effect on fitness is seen in heterozygotes. Thus, although Hardy (1908) demonstrated that dominance alone does not change allele frequencies at a locus, the dominance relationships among alleles can have substantial influence on evolutionary trajectories.

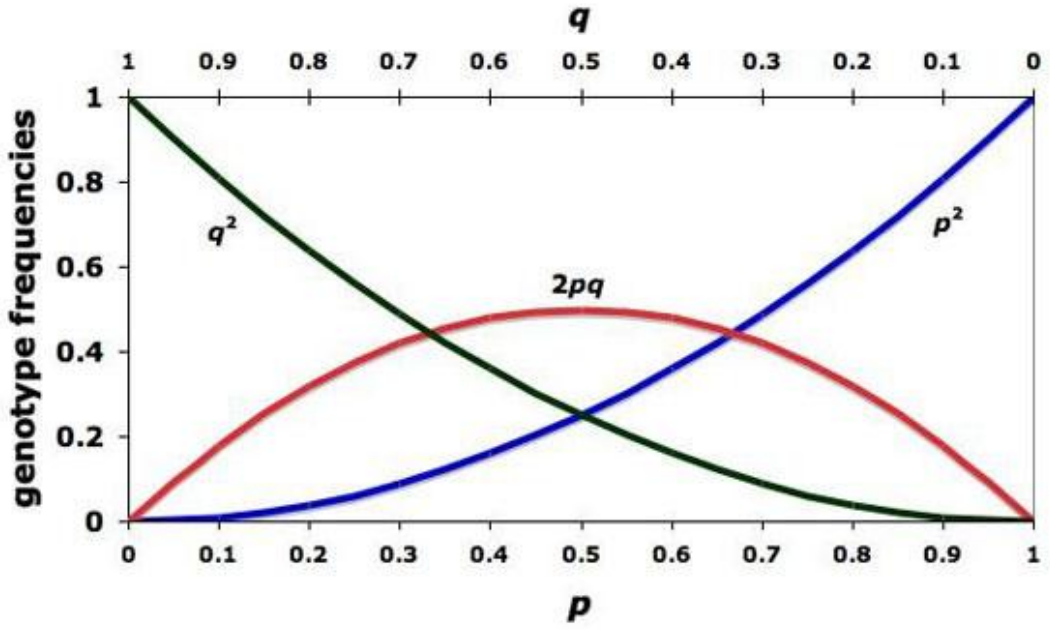

Figure 01: A plot of Hardy-Weinberg equilibrium genotype frequencies (p to the 2, 2pq, q to the 2) as a function of allele frequencies (p and q).

Selection, mutation, migration, and genetic drift are the mechanisms that effect changes in allele frequencies, and when one or more of these forces are acting, the population violates Hardy-Weinberg assumptions, and evolution occurs. The Hardy-Weinberg Theorem thus constitutes a null model for the discipline of population genetics, and is fundamental to the study of evolution.

Importance and implications of Hardy-Weinberg Theorem

- **Methodology:** Tells us how to calculate (or estimate) the allele frequency or genotype frequency from observed phenotypes in an empirical situation. It can help us to investigate how many alleles are governed by a phenotypic trait.
- **Evolution:** It is quantitative way of understanding the mechanism of evolutionary factors and its influences. Evolution is a dynamic and complex phenomenon and it is hardly possible to study evolution in the laboratory conditions. It gives insights into the inter-relationship between the forces and how to study the effects of each of these forces and the gene frequency.
- It is the **benchmark criterion** to test whether a new trait is in equilibrium or if not how to test the reasons for the deviations.
- It helps us in **genetic counselling** to expect the likelihood of a child being homozygous for a recessive deleterious trait given the parental genotype. It helps in forensic science in cases like identification of suspects, parent-offspring disputes etc.
- **Quantitative Genetics**: HWE helps us to investigate complex genetic traits, to estimate the role of environment and genetic components, spatial distribution of gene frequency etc

Deviations from Hardy–Weinberg equilibrium

The seven assumptions underlying Hardy–Weinberg equilibrium are as follows

- Organisms Are Diploid
- Only Sexual Reproduction Occurs
- Generations Are Nonoverlapping
- Mating Is Random
- Population Size Is Infinitely Large
- Allele Frequencies Are Equal In The Sexes
- There Is No Migration, Gene Flow, Admixture, Mutation Or Selection

Summary

1. Understanding of Population genetics principles, requires the basic concepts of Mendelian genetics: the result of segregation, the concept ‘gene’, ‘phenotype’, ‘genotype’, ‘dominant’ , ‘recessive’ traits, ‘allele’ etc.

Parental mating types and expected distribution of genotypes among the offspring.

2. Hardy-Weinberg equilibrium is the solution to an intriguing question: what happens to gene frequency of a dominant character over generations in a population. With three times more frequent than normal does this will increase over generations?
3. HWE law states that under the absence of intervening factors, especially in a large population, given random mating, no selection of any sort, no mutation and absence of demographic factors like migration, differential fertility and mortality etc., the allele frequency remain constant over generations. This can be proved theoretically, easily, for a 'biallelic locus and it can be extended to multilocus as well.
4. The importance of HWE: it gives a methodology to estimate the allele frequency in a population based on phenotypic/genotypic information of the parental mating types. It helps us to investigate the relationship between change in gene frequency with respect to mutation, migration, selection, genetic drift etc. The entire investigation is the kernel of a branch of biomathematics or the new field: 'population genetics' and 'quantitative genetics'.
5. HWE is the bench mark of qualitative test to check whether a trait, an allele, SNP, is in equilibrium. It tells how to distinguish between the effects of evolutionary forces from the demographic factors.
6. Mutation is a non-systematic and random, but rate of mutation is site specific. Mutations are more frequent at hot-spots and are rare at the 'conserved region'. The mitochondrial non-coding genome has a higher frequency of mutations than the nuclear genome.
7. Genetic drift is a non-systematic force which can lead to significant changes in gene frequency in a small population. If an allele is rare in a small population, it can get lost or get fixed in the population over generations.
8. Founder effect is one form of genetic drift. The founders are a sample (represent a fraction of the genetic diversity) of original populations. The descendents of a few founders have the gene frequency that is dependent on the genetic composition and genetic structure of the founders. It can also happen as bottleneck effect, especially as a result of sudden

population size reduction in a population, due to reasons such as natural causes or man-made causes or socio-cultural regulations. There could be serial founder effect as a result of waves of migration at different times. The mitochondrial investigation of human origins suggests that the human origins and migration to other continents appears as a result of serial founder effect from Africa.

9. Natural selection is one of the complex systematic forces that can influence significant changes in gene frequency. Selection can operate in multitude ways and it is a slow process than to the effect of migration or admixture etc.
10. Selection basically operates at differential fertility and mortality levels. It is measured as 'fitness' the ability to leave offspring and refers to 'relative rate of survival'. It is measured by 'selection coefficient' ('s') which is a function of fitness (W). The fitness or selection coefficient differs with respect to the type of dominance: complete, partial, over etc.
11. The effect of 'directional selection' to shift the mean allele frequency towards its extremes. Or it could be stabilizing selection that shifts the allele frequency of extreme alleles as a result the heterozygote frequency will increase. Or it could be disruptive selection where the extreme allele frequency increases as against the heterozygote frequency.
12. Selection can also be measured based on demographic factors of fertility and mortality trends. Crow's Index of opportunity for selection measures total selection intensity that a population can experience which depend on two components, fertility and mortality.
13. Gene flow (migration/admixture) is a systematic factor which can bring rapid changes in gene frequency within a short period. In general, human populations follow a variety of restrictions or regulations that restrict gene flow between and within populations. The barriers for gene flow could be because of culture or due to geographical, political, religious and linguistic etc.
14. There are theoretical models to investigate the effect of spatial gene flow or population structure between populations. Island model, stepping stone model, neighbourhood model help us to investigate the spatial gene flow in different situations of population structure.