

Population Genetics

Theory of Allele Frequencies · Genetic Equilibrium · Changes in Gene Frequencies

Population genetics studies how the frequencies of alleles and genotypes in a population's **gene pool** behave over time. The Hardy-Weinberg principle first describes what happens when nothing is changing them - a baseline of genetic equilibrium - and this in turn makes it possible to detect and measure the four real forces that do change gene frequencies from one generation to the next: natural selection, mutation, migration, and genetic drift.

1. Theory of Allele Frequencies and Genetic Equilibrium

1.1 Allele and genotype frequencies

The **gene pool** of a population is the total collection of alleles carried by all its members at a given locus. For a locus with two alleles, A and a, the **allele frequency** p is the proportion of all alleles at that locus that are A, and $q = 1 - p$ is the proportion that are a. The **genotype frequency** is simply the proportion of individuals with a given genotype (AA, Aa, or aa). Allele frequencies are calculated directly from genotype counts: each homozygote contributes two copies of its allele and each heterozygote contributes one copy of each, a procedure often called **gene counting**.

1.2 The Hardy-Weinberg principle

The **Hardy-Weinberg principle** states that, in an idealised population, allele and genotype frequencies remain constant from generation to generation, with genotype frequencies given by $p^2 + 2pq + q^2 = 1$ (the expected frequencies of AA, Aa and aa respectively). This equilibrium requires five simplifying assumptions: no natural selection, no mutation, no migration, an infinitely large population (so no genetic drift), and random mating. Because real populations rarely meet all five conditions perfectly, the principle functions less as a description of nature and more as a **null model**: by comparing a real population's genotype frequencies against the Hardy-Weinberg expectation, geneticists can detect that evolution is occurring and start narrowing down which force is responsible.

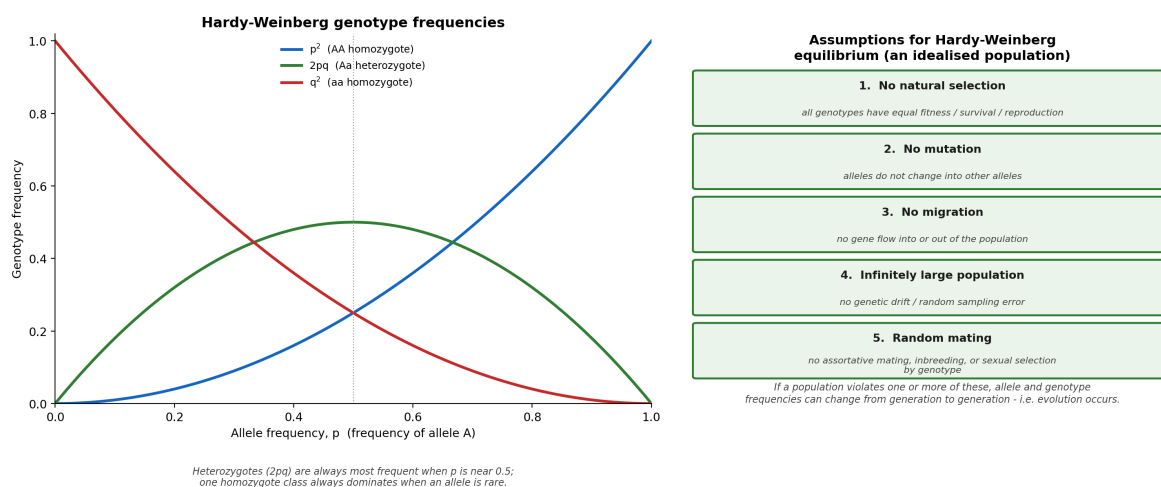


Fig. 1 - Hardy-Weinberg genotype frequencies as a function of allele frequency p , and the five assumptions required for equilibrium.

1.3 Estimating equilibrium frequencies in natural populations

Testing whether a real population is in Hardy-Weinberg equilibrium follows a standard procedure. A sample is genotyped at a codominant marker locus, allele frequencies p and q are estimated by gene counting, and these are used to predict the genotype counts a truly equilibrium population of the same size would show. These **expected** counts are then compared against the **observed** counts with a **chi-square goodness-of-fit test** (1 degree of freedom for a two-allele locus); a chi-square value at or above the critical value of 3.84 ($\alpha = 0.05$) indicates a statistically significant deviation from equilibrium, implying that selection, drift, migration, or non-random mating is acting on that locus in that population.

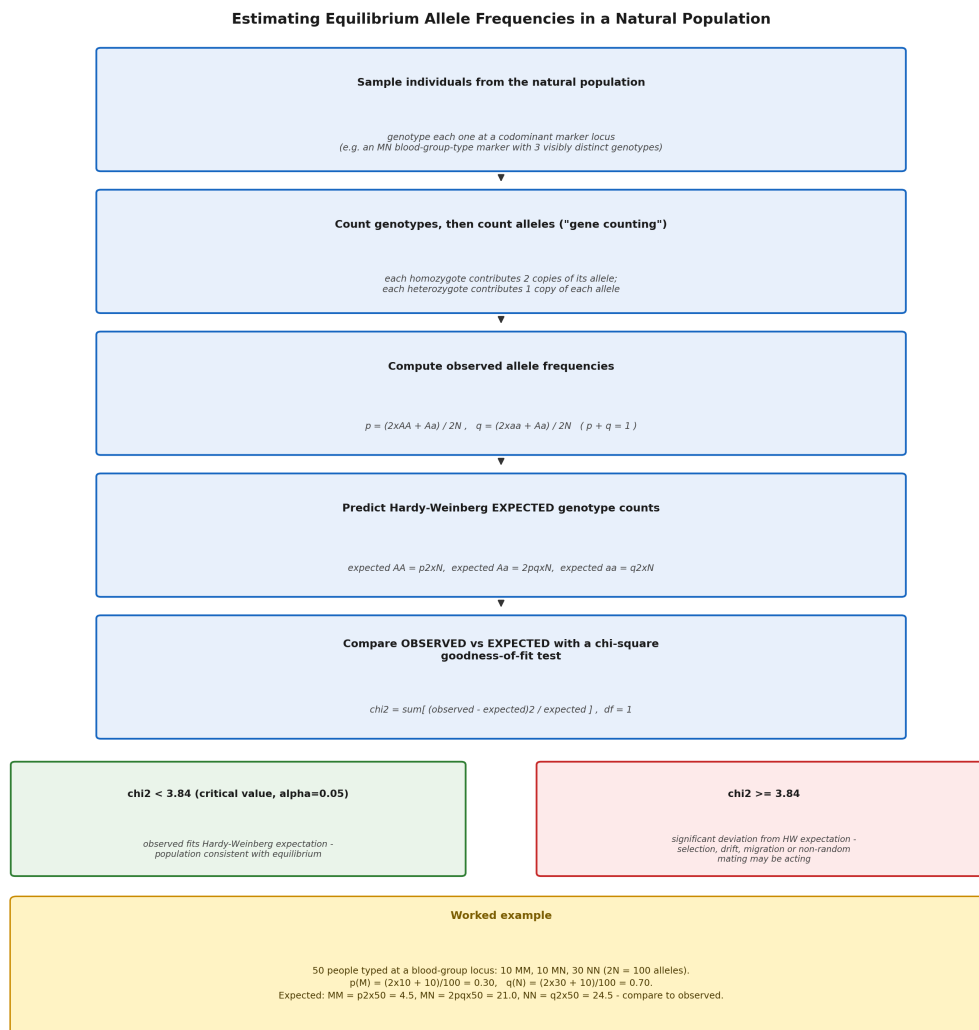


Fig. 2 - The standard workflow for estimating allele frequencies in a sampled population and testing them against Hardy-Weinberg expectation.

Key points

- p and q are frequencies (proportions, 0-1), never counts; $p + q$ always equals 1 for a two-allele locus.
- Hardy-Weinberg equilibrium is reached in a single generation of random mating regardless of starting genotype frequencies, provided allele frequencies themselves do not change - equilibrium concerns genotype frequencies relative to the allele frequencies, not the allele frequencies staying at any particular value.
- A significant chi-square result tells you equilibrium is violated, but not which of the four forces below is responsible - that requires further study of the specific locus and population.

2. Changes in Gene Frequencies

Four processes can shift allele frequencies away from Hardy-Weinberg expectation and drive real evolutionary change: natural selection, mutation, migration, and genetic drift.

2.1 Natural selection

Natural selection changes allele frequencies whenever genotypes differ in **relative fitness** - their average survival and reproductive success compared to other genotypes in the population; the strength of selection against a genotype is often summarised as a **selection coefficient**, s . Selection can act on a continuous phenotype in three characteristic modes: **directional selection** shifts the population toward one phenotypic extreme (favouring, for example, darker moths on soot-darkened trees), **stabilizing selection** favours the average phenotype and selects against both extremes (reducing variance), and **disruptive (diversifying) selection** favours both extremes over the intermediate phenotype (increasing variance, and potentially promoting speciation).

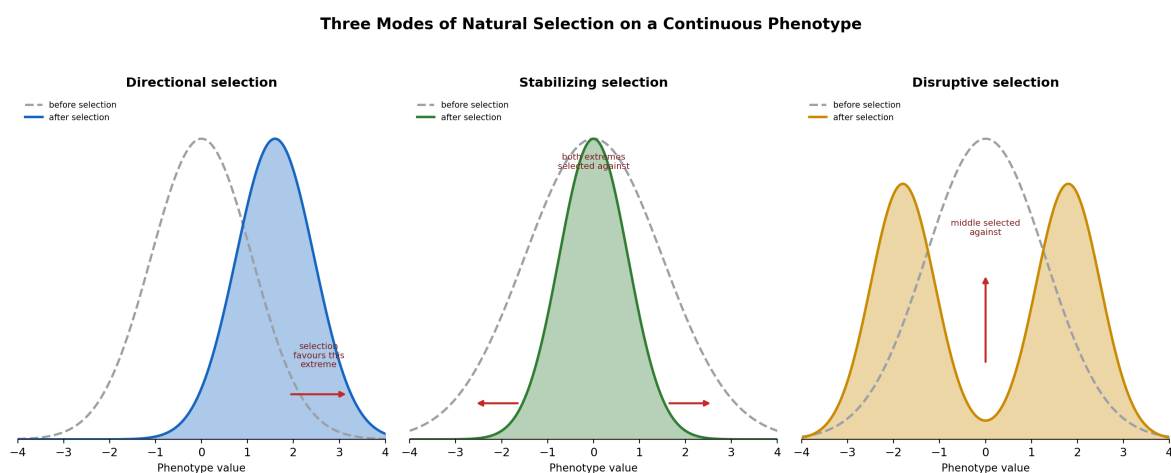


Fig. 3 - Directional, stabilizing, and disruptive selection each reshape the phenotype distribution of a population differently.

2.2 Mutation

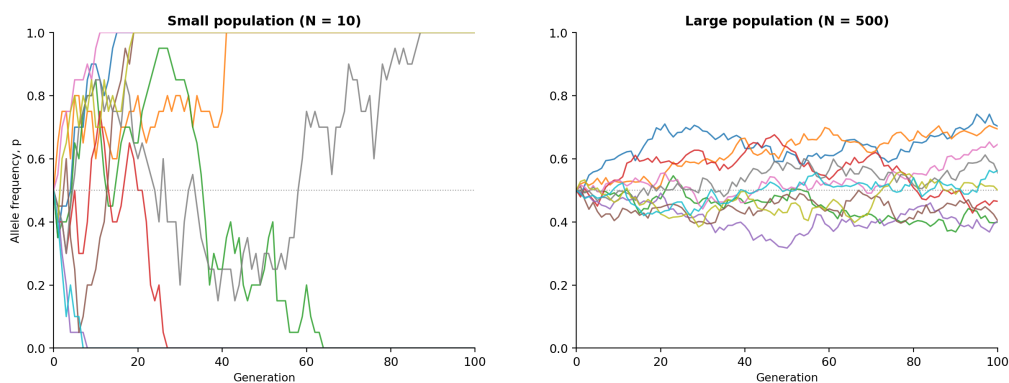
Mutation is the ultimate source of all new alleles, but because mutation rates at any single locus are typically very low (often around 10^{-5} to 10^{-6} per gene per generation), mutation alone is a very weak force for changing allele frequencies within a human timescale. Its evolutionary importance instead lies in continually resupplying variation for selection to act on, and in setting up a predictable equilibrium when opposed by selection: at **mutation-selection balance**, the rate at which mutation creates a harmful allele exactly offsets the rate at which selection removes it, giving an equilibrium frequency for a recessive deleterious allele of approximately $q = \sqrt{\mu/s}$, where μ is the mutation rate and s is the selection coefficient against the homozygote.

2.3 Migration (gene flow)

Migration moves alleles between populations that differ in allele frequency, and its effect is to make interbreeding populations more genetically similar to each other over time. In a simple one-island model, if a fraction m of a recipient population's next generation are migrants from a source population, the recipient's allele frequency shifts according to $p'_i = (1 - m)p_i + m p_m$: it moves a fraction m of the way toward the source population's frequency each generation. Even a small, steady migration rate can be a powerful homogenising force, and gene flow often opposes local adaptation by continually reintroducing alleles that selection is trying to remove from a population.

2.4 Genetic drift

Genetic drift is the random change in allele frequency that arises simply because each new generation is a finite sample of gametes drawn from the previous one - not because any allele is better or worse. Its effect is inversely related to population size: in a small population, chance sampling error can shift allele frequencies substantially in a single generation, eventually driving an allele to **fixation** (frequency 1) or **loss** (frequency 0) regardless of its fitness; in a large population, sampling error averages out and drift is negligible. Two special, historically important cases of drift are the **bottleneck effect**, in which a population crashes to a small size (e.g. by a disaster or over-hunting) and the survivors carry only a random subset of the original genetic diversity, and the **founder effect**, in which a new population is established by a small number of individuals who by chance carry a non-representative sample of alleles from their source population.



Genetic drift: 10 independently simulated populations starting at $p = 0.5$. In small populations, random sampling of gametes causes large, rapid, unpredictable swings - often ending in fixation ($p = 1$) or loss ($p = 0$) of an allele purely by chance, regardless of its fitness.

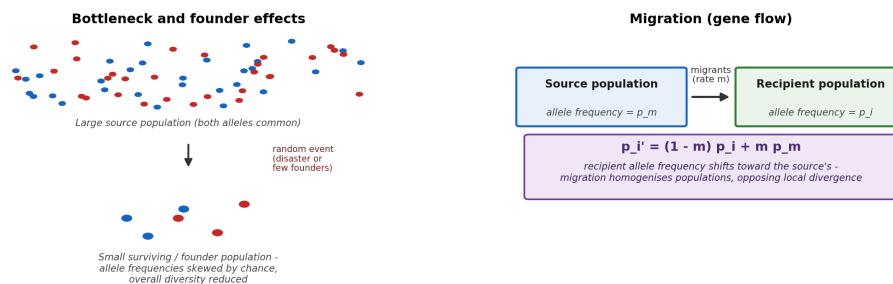


Fig. 4 - Genetic drift simulated in a small versus a large population, the bottleneck and founder effects, and the homogenising action of migration between populations.

Worth noting

- The northern elephant seal was hunted down to roughly 20 individuals in the 1890s; even after rebounding to tens of thousands, the population still shows far less genetic variation than the less-hunted southern elephant seal - a textbook bottleneck.
- Elevated frequencies of certain Mendelian disease alleles in some founder populations (e.g. Huntington disease among the Afrikaner population, descended from a small group of Dutch colonists) illustrate the founder effect in human populations.

2.5 Comparing the four forces

Force	Direction of change	Depends on population size?
Natural selection	Systematic - toward higher-fitness genotypes	No (though drift can overwhelm weak selection in small populations)
Mutation	Systematic but very slow; introduces new variation	No
Migration	Systematic - toward the source population's frequencies	No
Genetic drift	Random - direction unpredictable each generation	Yes - much stronger in small populations

Key points

- Selection, mutation and migration are directional/systematic forces (their expected effect on allele frequency can be predicted); genetic drift is random (its direction cannot be predicted, only its magnitude, which shrinks as population size grows).
- These forces do not act in isolation in real populations - for example, migration can continually reintroduce alleles that local selection is removing, and drift can fix a mildly deleterious allele in a small population despite weak selection against it.
- Detecting deviation from Hardy-Weinberg equilibrium is the starting point for identifying that one or more of these four forces is acting on a locus in a given population.

3. Quick Revision Summary

Concept	Essence
Gene pool	All alleles at a locus carried by every member of a population
Allele frequency (p, q)	Proportion of all copies of a locus that are a given allele; $p + q = 1$
Hardy-Weinberg equation	$p^2 + 2pq + q^2 = 1$; expected genotype frequencies at equilibrium
HW assumptions	No selection, no mutation, no migration, infinite population, random mating
Gene counting	Estimating p and q directly from observed genotype counts in a sample
Chi-square HW test	Compares observed vs HW-expected genotype counts; df=1, critical value 3.84
Natural selection	Differential fitness changes allele frequency; directional/stabilizing/disruptive modes
Mutation	Ultimate source of new alleles; weak alone; mutation-selection balance $q = \sqrt{\mu/s}$
Migration (gene flow)	Moves alleles between populations; homogenises frequencies toward the source
Genetic drift	Random sampling error changes allele frequency; strongest in small populations
Bottleneck effect	Sharp population crash leaves survivors with reduced, randomly skewed diversity
Founder effect	New population founded by few individuals carries a non-representative allele sample

Notes prepared for study purposes. Consult current primary literature for details required in formal citation or research use.