Population Genetics

By William Beavis, Laura Merrick, Kendra Meade, Arden Campbell, Deborah Muenchrath, Shui-Zhang Fei (ISU)

Introduction

Introduction

Population genetics is a sub-discipline of genetics that characterizes the structure of breeding populations. The forces of mutation, migration, selection and **genetic drift** will alter the structure of populations. In this introductory module we will focus on characterizing population structure at a single locus. In more advanced modules you will learn how to characterize populations based on the multi-dimensional space determined by multiple loci throughout the genome.

Objectives

- Understand the importance of a reference population.
- Become familiar with modeling and estimation of genetic variation.
- Understand the principles of allele frequency, genotype frequency, and genetic equilibrium in populations.
- Be aware of the conditions required for Hardy-Weinberg Equilibrium (HWE).
- Examine the forces that cause deviations from HWE.

Fig. 1 A bed of tulips at Pecherskyi Landscape Park in Ukraine. Photo by OZ-OK. Licensed under CC-SA 3.0 via Wikimedia Commons.

Two possible challenges are described in the following scenarios:

Scenario 1—Fate of a Transgene
Imagine a community of small farms in a valley located in the highlands of Central America. The farmers of this community produce grain from an open pollinated maize variety that is adapted to their preferred cultural practices. They also select partial ears from about 5% of their better performing plants to be used for seed in their next growing season. One day a truck filled with seed of a transgenic insect resistant hybrid overturns on the highway while passing through the valley. 99.999% of the seed is recovered, but about 500 kernels remain in a farmer's 10 acre field adjacent to the highway. The transgenic seeds germinate and grow to maturity alongside the planted open pollinated variety. You are asked to determine the fate of an insect resistant transgene in this valley.

Fig. 2 Maize and bean plants in a field in Quiché, Guatemala. Photo by Fabian Hanneforth. Licensed under CC-SA 3.0 via Wikimedia Commons.

Scenario 2

Two possible challenges are described in the following scenarios:

Scenario 2—Fixation of an Allele
Imagine a naturally occurring allele at a locus that regulates the structure of carbohydrates in the wheat kernel; with the allele the carbohydrates in the kernel have low glycemic indices. For the last 100 years hard-red winter wheat varieties have not been selected for low glycemic indices, but with the emergence of a Type II diabetes epidemic, there is a demand for low glycemic carbohydrates in hard-red winter wheat varieties. How will you develop a breeding population in which this allele is fixed, that is the frequency of this allele = 1.0?

Fig. 3 Hard red winter wheat kernels. Photo by U.S. Department of Agriculture.

Fields of Genetics

These challenges are fundamentally about population genetics. In this section you have the opportunity to successfully address these types of challenges by learning how to model and estimate allelic frequencies and the forces that affect population structures. In the study of population genetics, the focus shifts away from the individual (which is the focus for **transmission genetics**) and the cell (which is the focus for **molecular genetics**) to emphasis on a large group of individuals—a Mendelian population—that is defined as a group of interbreeding individuals who share a common set of genes.

This module will include discussion of **inbreeding**, which is one type of mating of individuals that is often of particular significance to plant breeders. Inbreeding is the mating of individuals that are more closely related than individuals mated at random in a population. Self-pollination (mating of an individual to itself) represents the most extreme form of inbreeding.

Fig. 5 Molecular genetics focuses on DNA operating within the cell.

Fig. 4 Transmission genetics studies how traits are passed from an individual to its progeny

Fig. 6 Population genetics studies the forces that affect groups of

Reference Population

Goals

Population genetics has three major goals, all of which are interrelated (Conner and Hartl, 2004):

- Explain the origin and maintenance of genetic variation.
- Describe the genetic structure of populations, i.e., the patterns and organization of genetic variation.
- Recognize the mechanisms that cause changes in allele and genotypic frequencies.

Similar to **quantitative genetics**, population genetics is concerned with application of Mendelian principles and is amenable to mathematical treatment. Understanding population genetics will require you to apply concepts from high school algebra.

Fig. 7 Analysis of rice crop populations. Photo by Iowa State University.

Description

In order to understand the genetic structure of a population it is necessary to establish a standard reference population so that the breeding population can be characterized relative to the standard.

Consider an 'ideal' population that is infinitely large. Further consider development of sub-populations as in Figure 6, described in Falconer and Mackay (1996).
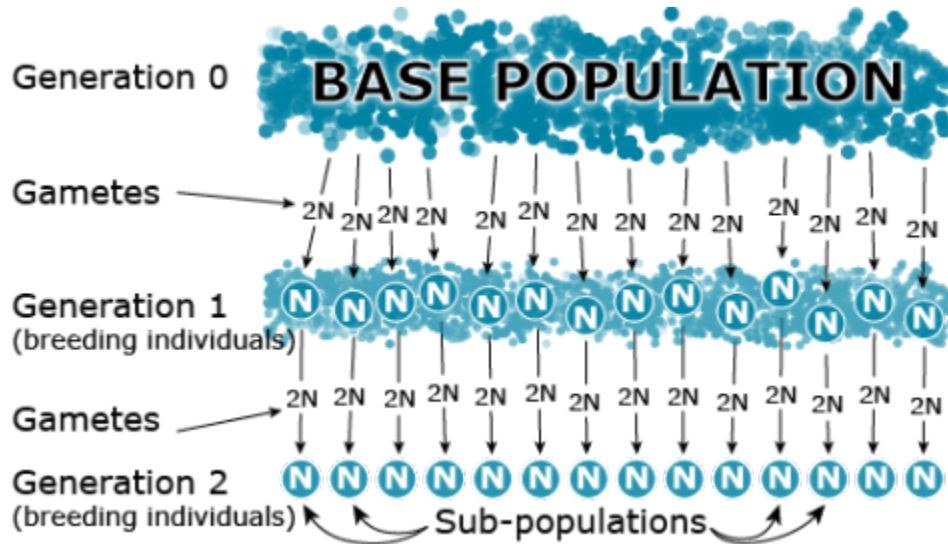

Fig. 8 Diagram of the subdivision of a single large base population into a number of sub-populations.

Note that the sub-populations depicted in the figure above are based on a genetic sampling process that is affected by reproductive biology of the species. The reproductive mode of most plant species can be classified as **sexual** or **asexual**. Species that reproduce sexually are generally categorized into three **types of mating systems** – primarily cross-pollinated, primarily self-pollinated, or a mixture of self- and cross-pollinated. Asexual modes of reproduction include three main categories: **vegetative** or clonal propagation, and **apomixis**. Under different mating systems (e.g., random vs. inbreeding) different genotypic frequencies will be generated from the same allele frequencies. With sexually reproducing individuals, mating combines alleles in the pool of haploid gametes produced by meiosis into genotypes in the diploid individuals.

In the ideal **model population** depicted in Fig. 8, we make the following assumptions:

- The base population is extremely large (too large to count)
- No migration between sub-populations
- Non-overlapping generations
- Number of breeding individuals is the same in each sub-population
- Random mating within a sub-population
- No selection

- No mutation

Models such as that shown above are theoretical abstractions. Models provide methods to simulate real-life situations and they are used for two principal reasons: 1) to reduce complexity, allowing underlying patterns to become more visible and 2) to make specific predictions to test with experiments or observations (Connor and Hartl 2004).

Discussion

Discuss the two challenges described earlier with respect to each reference population:

For Scenario 1—Fate of a Transgene, characterize the breeding population. Assume that there are 100 10-acre farms in the Central American valley, where farmers plant about 10,000 maize kernels per acre.

Photo by the U.S. Department of Agriculture.

For Scenario 2—Fixation of an Allele, determine how many hard red winter wheat varieties exist for the Southern Great Plains region. The number can include all historical varieties grown in the region. Assume that you have identified one additional ancient accession of hard red winter wheat that has the desirable allele for low glycemic carbohydrates. Assume that these varieties represent the lines you will use for your basic breeding population. Characterize this breeding population.

Photo by Fabian Hanneforth. Licensed under CC-SA 3.0 via Wikimedia Commons.

Allele and Genotypic Frequencies

Model

We first model a single locus with only two alleles (e.g., presence or absence of a transgene) in an ideal breeding population of diploid individuals. Define the following:

N = Number of breeding individuals in a sub-population (population size)

t = Time in generations with base population at t_0

q = Frequency of a particular allele at a locus within a sub-population

p = 1 - q = Frequency of other allele at a locus within a sub-population

$\bar{p}$ = Frequency in the whole population (the mean of p)

p_0 = Frequency of p in the base population

q_0 = Frequency of q in the base population

Because of the assumptions associated with an ideal reference population, $\bar{q}$ = q_0 at any stage or generation of the sampling process, so q_0 can be used interchangeably with $\bar{q}$.

Equations

The alleles, allele frequencies, genotypes and genotypic frequencies can be represented as follows:

	Alleles		Genotypes		
	A	a	AA	Aa	aa
Frequency	p	q	P_{AA}	P_{Aa}	P_{aa}

Where,

$$p + q = 1$$

and

$$P_{AA} + P_{Aa} + P_{aa} = 1$$

The relationship between allele frequencies and genotype frequencies can be expressed as follows:

$$p = P_{AA} + \frac{1}{2}P_{Aa}$$

and

$$q = P_{aa} + \frac{1}{2}P_{Aa}$$

Hardy-Weinberg Equilibrium

Concept of Genetic Equilibrium

Concept of Genetic Equilibrium

Plant breeders recombine and select the alleles present in the **gene pool**. The gene pool of a population is the total of all alleles within a population, and consists of all of the genes shared by individuals in the population. Gene pools are described in terms of allele and genotype frequencies. Knowing the frequency with which desired (or undesirable) alleles occur in the gene pool of the population influences the choice of breeding population(s), breeding method, and likelihood of progress. The breeding population must contain not only sufficient genetic variability to allow selection, but also have favorable alleles present in high enough frequencies to facilitate their selection and allow efficient breeding progress to occur.

- **Allele frequency** (often also called gene frequency) – the proportion of contrasting alleles present in the gene pool of a population.
- **Genotype frequency** – the proportion of various genotypes present in a population.

Concept of Genetic Equilibrium

The frequencies of specific alleles and genotypes in a large, **random-mating** population will reach equilibrium and will remain in equilibrium with continued random mating. This tendency toward equilibrium is the foundation of a model called the Hardy-Weinberg Law or Hardy-Weinberg Equilibrium (HWE). This law states that

The probability of two alleles uniting in a zygote is the product of the frequency of the allles in the population

The law makes several assumptions.

- There are two alleles at a gene locus.
- The population is large (that is, the number of breeding individuals is in the hundreds, rather than in the tens).
- The population is random-mating.

Concept of Genetic Equilibrium

Hardy-Weinberg Equilibrium mathematically describes the relationship between allele frequency and genotype frequency. According to the Hardy-Weinberg law, if the frequencies of two contrasting alleles at a locus in the parent population are p and q, respectively, then

- $p + q = 1$, always; and

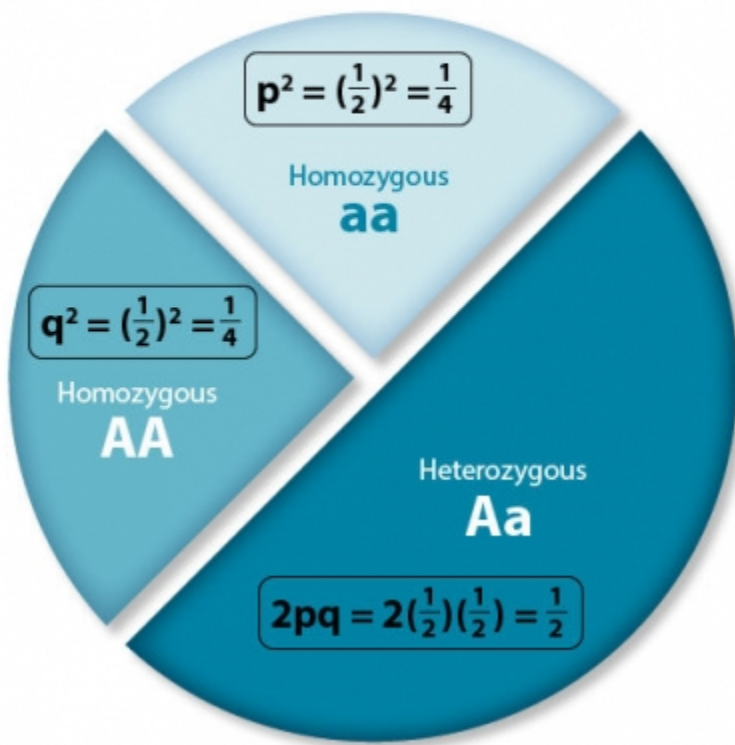- genotype frequency in the progeny is $(p + q)^2$ or $p^2 + 2qp + q^2 = 1$.

Fig. 9 Genotype frequencies in a hypothetical population in Hardy-Weinberg equilibrium.

For each of the following populations, indicate whether the Hardy-Weinberg Law would apply.

Population	Does the Hardy-Weinberg Law apply?	
Naturally self-pollinating species	NO ⌄	
Naturally cross-pollinating species	NO ⌄	
Limited population size	NO ⌄	

✔ Check

Study Question 2

Locus Alpha has two contrasting gene forms or alleles (**A** and **a**) in a large, random-mating population. The population is at equilibrium.

Study Question 2 Explanation

The correct frequency of aa genotype following selection and random mating is 0.17. Selection for the A_ phenotype (or against the aa phenotype), shifts the allele and genotype frequencies. Here's how the answer is determined:

- Initial population is 0.09 AA + 0.42 Aa + 0.49 aa
- Selection removes aa genotypes, so the unselected portion of the population is 0.09 AA + 0.42 Aa and the remaining individuals are all A_.
- Thus, setting p equal to the frequency of the **A** allele, and q equal to frequency of the **a** allele, the resulting allelic frequencies are now

$$p = \frac{\text{frequency of A in the AA genotype} + \text{frequency of A in the Aa genotype}}{\text{total allele frequencies of A and a}}$$

$$p = \frac{0.09 \times 2 + 0.42 \times 1}{0.09 \times 2 + 0.42 \times 2}$$

$$p = \frac{0.60}{1.02} = 0.59 \qquad q = 1 - p = 0.41$$

- So, the frequency of the **A** allele is 0.59 and the frequency of the **a** allele is 0.41.
- Now, we can calculate the frequency of the aa genotype in the population after one generation of selection and subsequent random mating.

$p^2(AA) + 2pq(Aa) + q^2(aa) = 1$
$(0.59)^2 + 2 \cdot 0.59 \cdot 0.41 + (0.41)^2 = 1$
$0.35\ AA + 0.48\ Aa + 0.17\ aa = 1$

Thus, the correct frequency of the aa genotype is 0.17.

Concept of Genetic Equilibrium

Several factors may disturb the genetic equilibrium of a population.

- mutation of an allele at the locus of interest
- natural or human selection may favor one allele over the other
- migration of alleles into or out of the population (for example, via an introduction of a different allele from another population, or loss of an allele through selection)
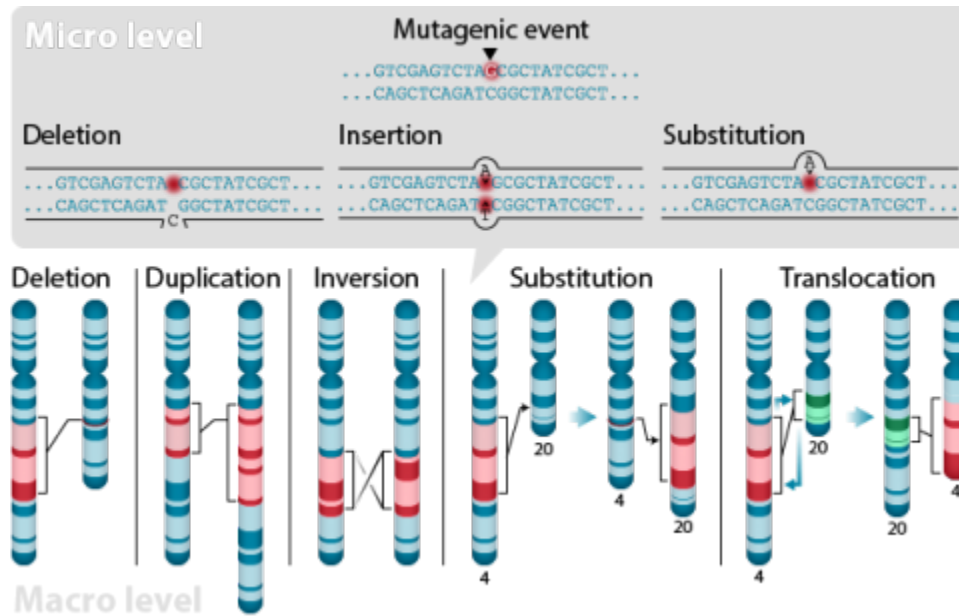

Fig. 10 Types of mutational events. Illustration by NIH/NHGRI, 2003.

Generally, a population not in genetic equilibrium, but retaining two contrasting alleles at a single, independently-segregating (non-linked) locus, will be restored to equilibrium at that locus after just one generation of random mating.

Random-Mating Interference

What is the significance of the Hardy-Weinberg Law to plant breeders? The random-mating assumption is often violated in breeding populations because breeding populations are smaller than natural plant populations. Thus, a mating design that minimizes gamete (allele) sampling errors is an important consideration. The breeder must be aware of several factors:

- **Self-pollinated population** — allele frequency will remain in equilibrium (assuming sufficiently large population, no selection, or other factors that disturb equilibrium). However, with each successive generation of self-pollination, the genotype frequency of homozygous loci will increase and the frequency of heterozygous loci will decrease. Ultimately, the heterozygous genotype will be eliminated from the population with continued selfing.

- **Cross-pollinated population** — sampling errors occur if plants in the population differ in their vigor, time of flowering, or mate more frequently with plants in close proximity.

- **Selection for or against a particular allele will alter the allele and genotype frequencies of the population.** Selection against a dominant allele (i.e., selection for homozygous recessive) will remove the dominant allele from the population in a single generation. Selection against a recessive allele will require more than a few generations to remove the recessive allele from the population because the homozygous dominant and heterozygous genotypes have indistinguishable phenotypes.

Concept of Genetic Equilibrium

In addition to being able to estimate allele and genotype frequencies, the breeder also needs to understand the gene action affecting the character of interest.

The breeding of cross-pollinated crops differs from self-pollinated species because of differences in the structures of their gene pools and opportunity for genetic recombination.

Table 1 Natural genetic structure of self- vs. cross-pollinated species.

Reproductive mode	Individuals	Population
Self-pollinated	Homozygous	**Homogeneous or heterogeneous**
Cross-pollinated	Heterozygous	Heterogeneous

Wheat, the crop found in the second challenge, **Scenario 2—Fixation of an Allele**, has **bisexual flowers** and is normally a self-pollinated crop.

Maize, the crop found in the first challenge, **Scenario 1—Fate of a Transgene**, is **monoecious and is cross-pollinated**.

Let's examine the genetic structure of populations of self- and cross-pollinated species.

Scenario 2—Fixation of an Allele Imagine a naturally occurring allele at a locus that regulates the structure of carbohydrates in the wheat kernel; with the allele the carbohydrates in the kernel have low glycemic indices. For the last 100 years hard-red winter wheat varieties have not been selected for low glycemic indices, but with the emergence of a Type II diabetes epidemic, there is a demand for low glycemic carbohydrates in hard-red winter wheat varieties. How will you develop a breeding population in which this allele is fixed, that is the frequency of this allele = 1.0?

Scenario 1—Fate of a Transgene Imagine a community of small farms in a valley located in the highlands of Central America. The farmers of this community produce grain from an open pollinated maize variety that is adapted to their preferred cultural practices. They also select partial ears from about 5% of their better performing plants to be used for seed in their next growing season. One day a truck filled with seed of a transgenic hybrid overturns on the highway while passing through the valley. 99.999% of the seed is recovered,

but about 500 kernels remain in a farmer's 10 acre field adjacent to the highway. The transgenic seeds germinate and grow to maturity alongside the planted open pollinated variety. You are asked to determine the fate of an insect resistant transgene in this valley.

Genetics of Cross-Pollinated Species

Because cross-pollinated species have evolved to outcross, individuals tend to be heterozygous at many loci and they usually perform best when that heterozygosity is maintained. This is a characteristic referred to as heterosis or hybrid vigor. When repeated self-pollination occurs in cross-pollinated species, homozygosity increases and plant vigor is reduced, a phenomenon called inbreeding depression. Heterosis and inbreeding depression will be further discussed in Lesson 6.

Several morphological and physiological features of cross-pollinated species promote cross-pollination. Let's briefly review these.

- **Monoecy** — pistillate and staminate flowers occur on different sections of the same plant.

- **Dioecy** — pistillate and staminate flowers occur on different plants.

- **Protandry or protogyny** — pistillate and staminate flowers mature at different times.

- **Self-incompatibility** — pollen from the same plant cannot effect fertilization or seed set.

- **Male or female sterility** — pollen or ovule does not function normally.

Genetics of Self-Pollinated Species

Self-pollinated species rarely hybridize naturally. Although cross-pollinating may occasionally occur, ovules of a self-pollinated plant are normally fertilized by pollen produced on that same plant. The result of repeated generations of selfing is that homozygosity is increased or maintained.

- Homozygous loci will remain homozygous.

- Heterozygous loci will segregate such that the **frequency of homozygotes** will increase at the expense of the frequency of heterozygotes with each generation of selfing.

How does a locus become heterozygous? A contrasting allele can be acquired when a plant out-crosses or when a mutation occurs. Each successive self-pollination thereafter will reduce heterozygosity by half. Breeders rely on the natural tendency of self-pollinated crops to become homozygous to obtain lines that exhibit uniformity in characters that affect appearance and performance.

Notice how rapidly populations lose heterozygosity with selfing. For self-pollinated crops, one of the breeder's objectives is usually to develop pure lines. Since pure lines are homozygous, their rapid loss of heterozygosity speeds cultivar development. Some background heterozygosity may remain in a pure line, but the line is sufficiently homozygous to provide the uniformity in characters required for reliable and predictable appearance and performance.

Concept of Genetic Equilibrium

The tendency of a species to self-pollinate or outcross influences allelic and genotypic frequencies in the population. In a self-pollinated homozygous population, the effect of a gene (allele) is determined by the gene's effect in combination with itself and with alleles at other loci. What determines the effect of a gene in a cross-pollinated population?

Effect or fate of an allele in a cross-pollinated population is determined by its effect

- in combination with other alleles at the same locus
  - additive effects
  - dominance effects
  - overdominance effects
- in combination with alleles at other independent loci (epistatic effects)
- in combination with alleles at closely linked loci

One difference between a self-pollinated and a cross-pollinated population is that in the cross-pollinated population there is constant inter-crossing. Thus, recombination and rearrangement of alleles and expression of dominance and epistatic effects occurs.

Review **gene action** or **gene interactions**, such as epistasis in the next screens.

Gene Action

There are several general types of gene action. The type of gene action and the alleles present for a given gene affect the phenotype. Let's consider the gene action as indicated by the phenotype of a diploid individual heterozygous at the given single locus compared to the phenotype of its parents.

Additive gene action (no dominance)

The progeny's phenotypic value is at the midpoint between both parents. In this example, each A allele adds 20 units.

Parents	→	Progeny
AA × aa		Aa

Genotype		Phenotypic Value
Parent 1 AA	—	80
Progeny Aa	—	60
Parent 2 aa	—	40
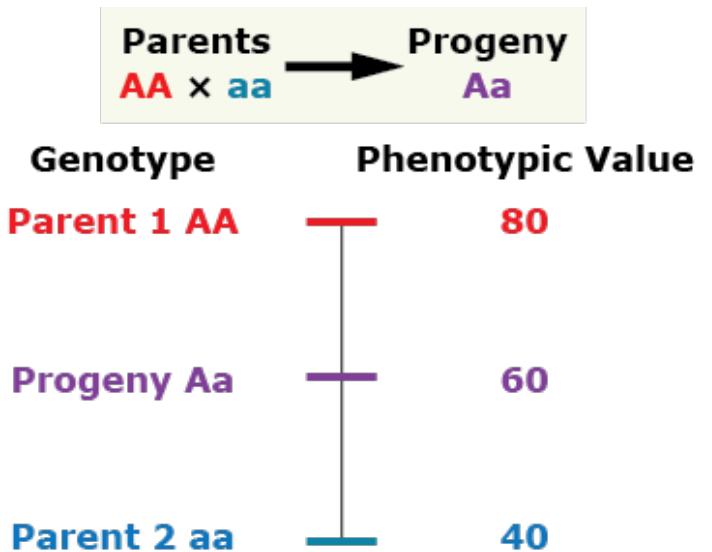
Complete Dominance

The phenotype of the heterozygous progeny equals the phenotype of the homozygous dominant parent.

Parents	→	Progeny
AA × aa		Aa

Genotype		Phenotypic Value
Parent 1 AA Progeny Aa	—	80 80
Parent 2 aa	—	40
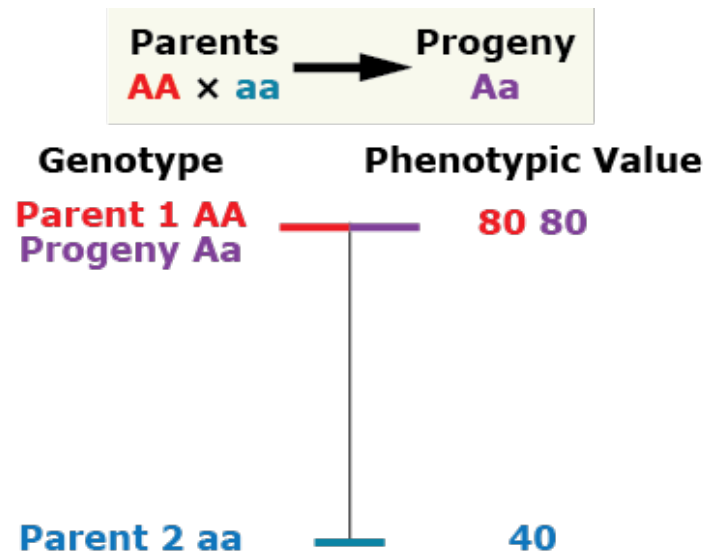
Partial (incomplete) dominance

The heterozygous progeny has a phenotypic value greater than that of the midparent value, but less than that of the homozygous dominant parent.
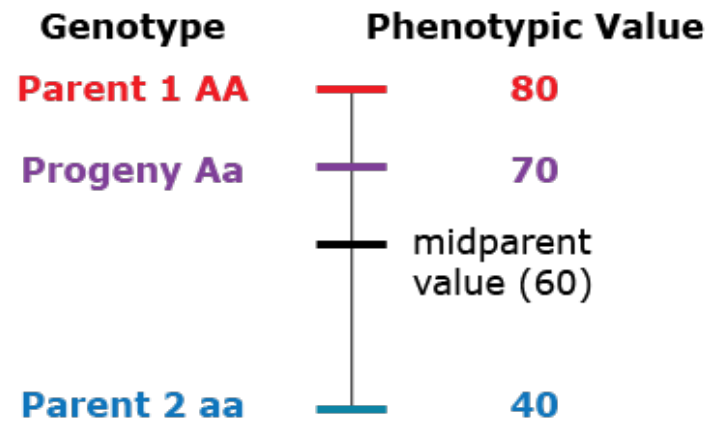
Fig. 11

Over-Dominance

The phenotype of the heterozygous progeny is greater than either parent.
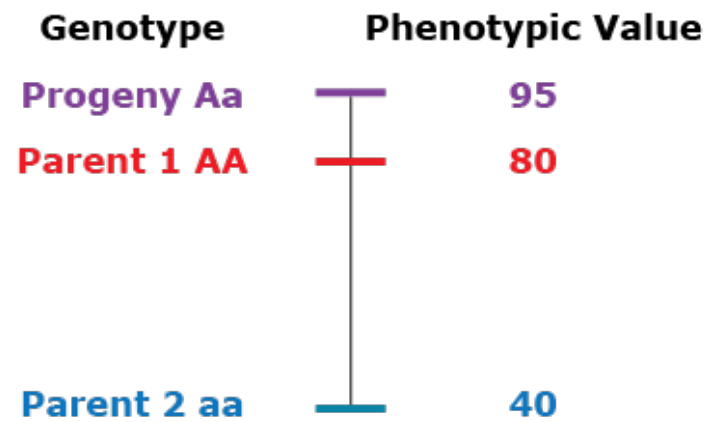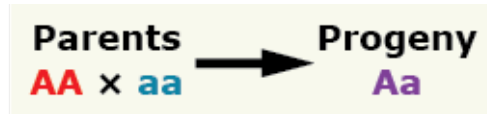

Fig. 12

Gene Interactions

When multiple genes control a particular trait or set of traits, gene interactions can occur. Generally, such interactions are detected when genetic ratios deviate from common phenotypic or genotypic proportions.

Pleiotropy – Genes that affect the expression of more than one character.
Epistasis – Genes at different loci interact, affecting the same phenotypic trait. Epistasis occurs whenever two or more loci **interact to create new phenotypes**. Epistasis also occurs whenever an allele at one locus either **masks** the effects of alleles at one or more other loci or if an allele at one locus **modifies** the effects of alleles at one or more other loci. There are numerous types of epistatic interactions.

Epistasis is expressed at the **phenotypic** level. It is important to note that genes that are involved in an epistatic interaction may still exhibit independent assortment at the **genotypic** level. In the case of two completely dominant, non-interacting (i.e., no linkage) genes, all of the deviations observed in results involving epistatic interactions are modifications of the expected 9:3:3:1 ratio.

Study Question 3

Describe a natural cross-pollinated population as to its heterozygosity, heterogeneity, and effect of inbreeding. For each of the following, select the best terms to complete the statement.

A natural cross-pollinated population is a [heterozygous ▾] mixture of [heterogeneous ▾] plants. Deleterious recessive alleles can remain in the population at a [low ▾] frequency because of [heterozygosity ▾]. Inbreeding in a cross-pollinated population will result in [improved ▾] vigor.

[✓ Check]

Proof

The proof of Hardy-Weinberg Equilibrium (HWE) requires the following assumptions (Falconer and Mackay, 1996):

1. Allele frequency in the parents is equal to the allele frequency in the gametes
    1. Assumes normal gene segregation
    2. Assumes equal fertility of parents
2. Allele frequency in gametes is equal to the allele frequency in gametes forming zygotes
    1. Assumes equal fertilizing capacity of gametes
    2. Assumes large population
3. Allele frequency in gametes forming zygotes is equal to allele frequencies in zygotes
4. Genotype frequency in zygotes is equal to genotype frequency in progeny
    1. Assumes random mating
    2. Assumes equal gene frequencies in male and female parents
5. Genotype frequencies in progeny do not alter gene (allele) frequencies in progeny.
    1. Assumes equal viability

For a two allele locus in a population in HWE:

$$P_{AA} = p^2 \quad P_{Aa} = 2_{pq} \quad P_{aa} = q^2$$

Proof

HWE at a given genetic locus is achieved in one generation of random mating. Genotype frequencies in the progeny depend only on the gene (allele) frequencies in the parents and not on the genotype frequencies of the parents.

If a population is in HWE, relationships between frequencies of alleles and genotypes may be derived as depicted in figure 13.

As shown in the figure, in HWE:

- frequency of heterozygotes does not exceed 0.5
- heterozygotes are most frequent genotype when p or q are between 0.33 and 0.66
- very low allele frequency should result in very low frequency of homozygotes for that allele
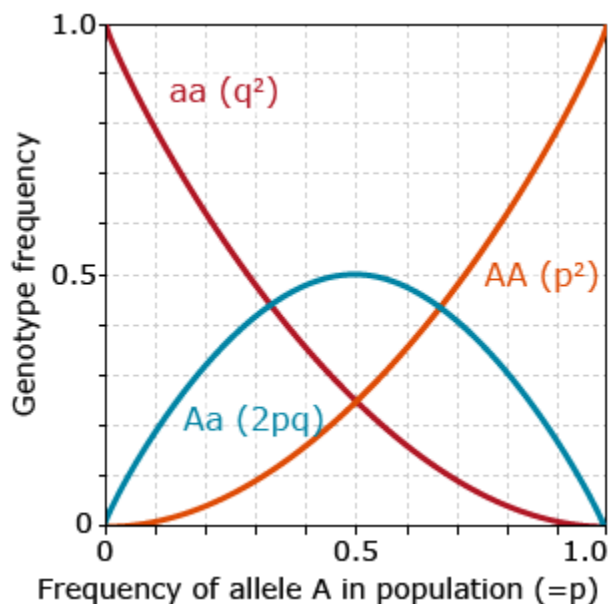- if there are only two alleles at a locus in the population, p+q=1.

Fig. 13 Relationship between genotype frequencies and gene frequency for two alleles in a population in Hardy-Weinberg equilibrium.

A **chi-square test** is typically used to determine whether or not a population varies significantly from Hardy-Weinberg expectations. The Hardy-Weinberg formula is useful in describing situations where mating is completely randomized. But more commonly, mating is not at random and populations are subjected to other forces, such as mutation, migration, genetic drift, and selection. Linkage can also have a significant effect on gene frequencies.

Forces Affecting Population Structures

Descriptions

Non-Random Mating

Two methods of non-random mating that are important in plant breeding are **assortative mating** and **disassortative mating**.

Assortative mating occurs when similar phenotypes mate more frequently than they would by chance. One example would be the tendency to mate early x early maturing plants and late x late maturing plants. The effect of assortative mating is to increase the frequency of homozygotes and decrease the frequency of heterozygotes in a population relative to what would be expected in a randomly mating population. Assortative mating effectively divides the population into two or more groups where matings are more frequent within groups than between groups.

Disassortative mating occurs when unlike or dissimilar phenotypes mate more frequently than would be expected under random mating. Its consequences are in general opposite those of assortative mating in that disassortative mating leads to an excess of heterozygotes and a deficiency of homozygotes relative to random mating. Disassortative mating can also lead to the maintenance of rare alleles in a population. For example, in self-incompatible species, an individual will only mate with another individual that differs in the self-incompatibility loci. This is a type of disassortative mating, resulting in a great allelic diversity in the self-incompatibility loci. It is an effective mechanism to maintain heterozygosity and prevent inbreeding.

Study Question 4

Scenario 2—Fixation of an Allele

Imagine a naturally occurring allele at a locus that regulates the structure of carbohydrates in the wheat kernel; with the allele the carbohydrates in the kernel have low glycemic indices. For the last 100 years hard-red winter wheat varieties have not been selected for low glycemic indices, but with the emergence of a Type II diabetes epidemic, there is a demand for low glycemic carbohydrates in hard-red winter wheat varieties. How will you develop a breeding population in which this allele is fixed, that is the frequency of this allele = 1.0? In order to form a breeding population with the low glycemic allele fixed among all of the lines, what sort of mating will need to be pursued?

Enter your answer here.

Show Answer

Forces Affecting Allele Frequency

Factor Categories

The factors affecting changes in allele frequency can be divided into two categories: **systematic processes**, which are predicable in both magnitude and direction and **dispersive processes**, which are predictable in magnitude but not direction. The three systematic processes are migration, mutation, and selection. Dispersive processes are a result of sampling in small populations.

Table 2

Systematic Processes	Dispersive Processes
Migration	Small Population Size
Mutation	
Selection	

Forces Affecting Allele Frequency

Migration

Clearly, the first challenge described in the introduction represents a case of migration. A new set of genes in a developed transgenic hybrid have been introduced into an open pollinated variety of maize. When discussing population genetics, migration is also sometimes referred to as **gene flow**, a concept that is often used interchangeably with migration by population geneticists. However, the term migration means the movement of individuals between populations, whereas gene flow is the movement of genes between populations. New genes would be established in the population if the immigrant successfully reproduces in its new environment, but if it doesn't reproduce migration would still have occurred while gene flow would not.

Fig. 14 Maize grows along with beans in a traditional milpa field in Quiché, Guatemala. Photo by Fabian Hanneforth. Licensed under CC-SA 3.0 via Wikimedia Commons.

Migration

Assume a population has a frequency of m new immigrants each generation, with 1- m being the frequency of natives. Let q_m be the frequency of a gene in the immigrant population and q_0 the frequency of that gene in the native population. Then the frequency in the mixed population will be:

$$q_1 = mq_m + (1 - m)q_0$$

$$q_1 = m(q_m - q_0) + q_0$$

The change in gene frequency brought about by migration is the difference between the allele frequency before and after migration

$$\Delta q = q_1 - q_0$$

$$\Delta q = m(q_m - q_0)$$

Thus the change in gene frequency from migration is dependent on the rate of migration and the difference in allele frequency between the native and immigrant population. Migration or gene flow can introduce new alleles into a population at a rate and at more loci than expected from mutation. It can also alter allele frequencies if the populations involved have the same alleles but not in the same proportions. Thus the effect of migration on changes in allele frequency depends on differences in allele frequencies (migrants vs. residents) and proportion of migrants in the population.

Study Question 5

Scenario 1—Fate of a Transgene

Imagine a community of small farms in a valley located in the highlands of Central America. The farmers of this community produce grain from an open pollinated maize variety that is adapted to their preferred cultural practices. They also select partial ears from about 5% of their better performing plants to be used for seed in their next growing season. One day a truck filled with seed of a transgenic hybrid overturns on the highway while passing through the valley. 99.999% of the seed is recovered, but about 500 kernels remain in a farmer's 10 acre field adjacent to the highway. The transgenic seeds germinate and grow to maturity alongside the planted open pollinated variety. You are asked to determine the fate of an insect resistant transgene in this valley. Determine Δq for the transgenic allele assuming that the allele is homozygous in the hybrid seed. This farmer's field is **5 acres** with **10,000 plants per acre.**

Δq = [   ]

🔘 Check

Mutation

Mutations are the source of all genetic variation. Loci with only one allelic variant in a breeding population have no effect on phenotypic variability. While all allelic variants originated from a mutational event, we tend to group mutational events in two classes: rare mutations and recurrent mutations where the mutation occurs repeatedly.

Rare Mutations

By definition a rare mutation only occurs very infrequently in a population. Therefore, the mutant allele is carried only in a heterozygous condition and since mutations are usually recessive, will not have an observable phenotype. Rare mutations will usually be lost, although theory indicates rare mutations can increase in frequency if they have a selective advantage.

Fate of a Single Mutation

Consider a population of only **AA** individuals. Suppose that one **A** allele in the population mutates to **a**. Then there would only be one **Aa** individual in a population of **AA** individuals. So the **Aa** individual must mate with a **AA** individual.

AA x Aa → 1AA:1Aa

From Li (1976; pp 388), this mating has the following outcomes:

1. **No offspring are produced** in which case the mutation is lost.
2. **One offspring is produced:** the probability of that offspring being **AA** is 1/2 so the probability of losing the mutation is 1/2.
3. **Two offspring are produced:** the probability of them both being **AA** is 1/4 so the probability of losing the mutation is 1/4.

If k is the number of offspring from the above mating then the probability of losing the mutation among the first generation of progeny is $(1/2)^k$.

The probability of losing the gene in the second generation can be calculated by making the following assumptions:

- Number of offspring per mating is distributed as a Poisson process (which means that they follow a stochastic distribution in which events occur continuously and independently of one another).
- With the average number of offspring per mating = 2.
- New mutations are selectively neutral.

With these assumptions, the probabilities of extinction are:

Table 3 Probability of extinction in different generations.

Generation	Probability of Loss
1	0.37
7	0.79

Generation	Probability of Loss
15	0.89
31	0.94
63	0.97
127	0.98

Recurrent Mutations

Let the mutation frequencies be:

Mutation rate: $A \xrightarrow{u} a$

Frequency: $p_0 \xleftarrow{v} q_0$

Then the change in gene frequency in one generation is:

⚠ Invalid Equation

at equilibrium

⚠ Invalid Equation

⚠ Invalid Equation

Conclusions:

- Mutations alone produce very slow changes in allele frequency
- Since reverse mutations are generally rare, the general absence of mutations in a population is due to selection

Selection

Selection is one of the primary forces that will alter allele frequencies in populations. **Selection** is essentially the differential reproduction of genotypes. In population genetics, this concept is referred to as **fitness** and is measured by the reproductive contribution of an individual (or genotype) to the next generation. Individuals that have more progeny are more fit than those who have less progeny because they contribute more of their genes to the population.

The change in allele frequency following selection is more complicated than for mutation and migration, because selection is based on phenotype. Thus, calculating the change in allele frequency from selection requires knowledge of genotypes and the degree of dominance with respect to fitness. Selection affects only the gene loci that affect the phenotype under selection—rather than all loci in the entire genome—but it also would affect any genes that are linked to the genes under selection.

Effects of Selection

Change in allele frequency

The strength of selection is expressed as a **coefficient of selection**, **s**, which is the proportionate reduction in gametic output of a genotype compared to a standard genotype, usually the most favored. **Fitness** (relative fitness) is the proportionate contribution of offspring to the next generation.

Partial selection against a completely recessive allele

To see how the change in allele frequency following selection is calculated consider the case of selection against a recessive allele:

Table 4

	Genotypes			
	AA	**Aa**	**aa**	**Total**
Initial Frequencies	p^2	$2pq$	q^2	1
Coefficient of Selection	0	0	s	
Fitness	1	1	1-s	
Gametic Contribution	p^2	$2pq$	$q^2(1\text{-}s)$	$1 - sq^2$

Frequency Equations

The frequency of allele **a** after selection is:

$$q_1 = \frac{q^2(1-s) + pq}{1 - sq^2} \qquad q_1 = \frac{q - sq^2}{1 - sq^2}$$

The change in allele frequency is then:

$$\Delta q = q_1 - q$$

$$\Delta q = \frac{q - sq^2}{1 - sq^2} - q$$

In general, you can show that the number of generations, t, required to reduce a recessive from a frequency of q_0 to a frequency of q_t, assuming complete elimination of the recessive (s = 1) is:

$$t = \frac{1}{q_t} - \frac{1}{q_0}$$

Discussion

Review the two challenges at the beginning of the lesson and then answer these questions:

- **Allele Frequency**—For Scenario 1, **calculate the frequency** of the insect resistant transgene in the Central American maize farmer's 10 acre field assuming that it is a) hemizygous and b) homozygous in the spilled hybrid seed. Remember that hemizygous means that the individual has only one single homologous chromosome, and therefore is neither homozygous nor heterozygous; in contrast homozygous means that there are two homologues.
- **Allele Frequency**—For Scenario 2, **calculate the frequency** of the allele responsible for low glycemic carbohydrates in the wheat breeding population, assuming the allele is not present in any wheat variety except one.
- **Mutation**—For Scenario 2, assume the mutation that produced the low glycemic allele was selectively neutral in the hard red-winter wheat breeding population. **Why was that allele lost from all varieties that were developed over the last 100 years?**
- **Selection**—In Scenario 1 a transgene (and likely other genes) is introduced into an open pollinated variety in one farmer's field. **Determine Δq for the transgenic allele** assuming that the allele is homozygous in the hybrid seed, the insect resistant allele is completely dominant and the selective advantage of the allele is a) two to one (2:1) when the insect is present and b) one to one (1:1) when it is absent.

Scenario 1: Fate of a Transgene Imagine a community of small farms in a valley located in the highlands of Central America. The farmers of this community produce grain from an open pollinated maize variety that is adapted to their preferred cultural practices. They also select partial ears from about 5% of their better performing plants to be used for seed in their next growing season. One day a truck filled with seed of a transgenic hybrid overturns on the highway while passing through the valley. 99.999% of the seed is recovered, but about 500 kernels remain in a farmer's 10 acre field adjacent to the highway.
The transgenic seeds germinate and grow to maturity alongside the planted open pollinated variety. You are asked to determine the fate of an insect resistant transgene in this valley.

Scenario 2: Fixation of an Allele Imagine a naturally occurring allele at a locus that regulates the structure of carbohydrates in the wheat kernel; with the allele the carbohydrates in the kernel have low glycemic indices. For the last 100 years hard-red winter wheat varieties have not been selected for low glycemic indices, but with the emergence of a Type II diabetes epidemic, there is a demand for low glycemic carbohydrates in hard-red winter wheat varieties. How will you develop a breeding population in which this allele is fixed, that is the frequency of this allele = 1.0?

Small Population Size

Unlike the three systematic forces that are predictable in both amount and direction, changes due to small population size are predictable only in amount and are random in direction.

The effects of small population size can be understood from two different perspectives. It can be considered a sampling process and it can be considered from the point of view of inbreeding. The inbreeding perspective is more interesting, but looking at it from a sampling perspective lets us understand how the process works.

A particular sub-population is a random sample of N individuals or 2N gametes (for a diploid) from the base population. Therefore, the expected gene frequency of a particular allele in the sub-populations is q_0 and the variance of q is ⚠ **Invalid Equation**

Since q_0 is a constant, the variance of the change in allele frequency ($q_1 - q_0$) is also: ⚠ **Invalid Equation**

Examples

Example 1: Let q = 0.5 and N = 50, then $\sigma_q^2 = \dfrac{(0.5)(0.5)}{100} = 0.0025$

Example 2: Let q = 0.5 and N = 4, then $\sigma_q^2 = \dfrac{(0.5)(0.5)}{8} = 0.03125$

Consequences of small population size

1. Random genetic drift: random changes in allele frequency within a subpopulation

2. Differentiation between subpopulations

3. Uniformity within subpopulations

4. Increased homozygosity

Small Population Size

Random genetic drift refers to allelic frequencies that change through time (generations) due to errors and other random factors (i.e., not selection or mutation). When sample sizes are small, all genotypes may not be produced and then mate at expected frequency. The **effective population** size (N_e) of a population is a term used to describe the number of parents that actually contribute gametes to the next generation; not all individuals may contribute equally, thus resulting in genetic drift. Small populations are susceptible to **genetic bottlenecks**, which are sudden decreases in breeding population due to deaths, migration, or other factors. Small populations can be subject to so-called **founder effects**, which occur when a breeding population is small when initially founded, then increases in size but the gene pool is largely determined by the genes present in the original founders.

Random Genetic Drift Simulation

PARENTS ←

PROGENY →

Plants remaining:

Blue:

Yellow:

Green:

Purple:

Red:

Brown:

Genotypes lost:

Small Population Size

The rate of change due to random genetic drift depends on population size and allele frequency. As illustrated in the figure below, the more frequent the allele, the higher chances of being fixed and the smaller the population, the faster it will either move towards fixation or loss. In the absence of other forces:

- genetic drift leads to loss or fixation of alleles
- frequency of rare alleles would be expected to go to zero
- lower frequency of heterozygotes in later generations
- less genetic variation within subpopulations
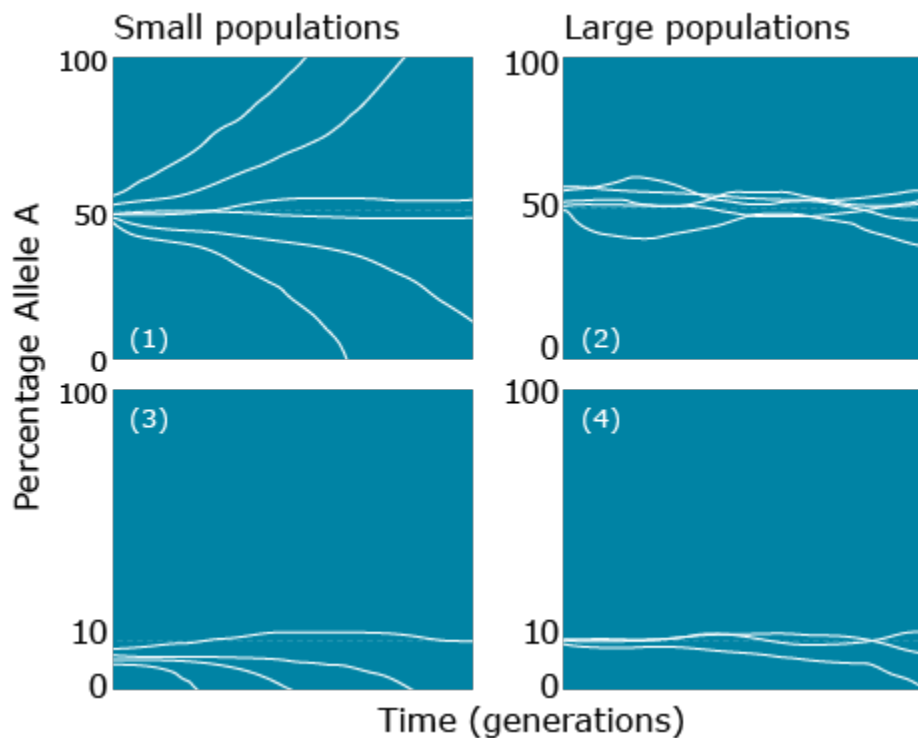- more genetic variation among subpopulations

Fig. 15 Effect of population size and gene frequency on rate of fixation due to drift. Each line represents a different population. In small populations, allele frequency for A showed greater differences among populations (1) whereas in larger populations, allele frequencies are similar over generations (2). When allele frequency is very low, the allele is more likely to be lost in small populations (3) than in larger populations (4).

Inbreeding and Small Populations

Inbreeding and Small Populations

Inbreeding and Small Populations Inbreeding is the mating together of individuals that are related by ancestry. The degree of relationship among individuals in a population is determined by the size of the population. This can be seen by examining the number of ancestors that a single individual has:

Just 50 generations ago note that a single individual would have more ancestors than the number of people that have existed or could exist on earth.

Therefore, in small populations individuals are necessarily related to one another. Pairs mating at random in a small population are more closely related than pairs mating together in a large population. Small population size has the effect of forcing relatives to mate even under random mating, thus with small population sizes inbreeding is inevitable.

Generation	Ancestors
0	1
1	2
2	4
3	8
4	16
5	32
6	64
10	1,024
50	1,125,899,906,842,620
100	1,267,650,600,228,230,000,000,000,000,000
t	2^t

Identical Types

In finite populations there are two sorts of homozygotes: Those that arose as a consequence of the replication of a single ancestral gene — these genes are said to be identical **by descent** (Bernardo, 1996). If the two genes have the same function, but did not arise from replication of a single ancestral gene, they are said to be alike **in state**. It is the production of homozygotes that are identical by descent that gives rise to inbreeding in a small population.

Scenario 2—Fixation of an Allele

Imagine a naturally occurring allele at a locus that regulates the structure of carbohydrates in the wheat kernel; with the allele the carbohydrates in the kernel have low glycemic indices. For the last 100 years hard-red winter wheat varieties have not been selected for low glycemic indices, but with the emergence of a Type II diabetes epidemic, there is a demand for low glycemic carbohydrates in hard-red winter wheat varieties. How will you develop a breeding population in which this allele is fixed, that is the frequency of this allele = 1.0?

Assume the mutation that produced the low glycemic allele was selectively neutral in the hard red-winter wheat breeding population. Why was it lost from all varieties that were developed over the last 100 years?

- Check

Summary of Factors

Hardy and Weinberg discovered mathematically that genotype frequencies will reach an equilibrium in one generation of random mating in the absence of any other evolutionary force. If the conditions of equilibrium are met, the frequencies of different genotypes in the progeny will depend only upon the allele frequencies of the previous generation. If allele frequencies do not accurately predict genotype frequencies, then plants are mating in a non-random way or another evolutionary force is operating.

	Effect of level of variation		
	Within subpopulations	Among subpopulations	Affect all loci equally?
Mutation	▲	▲	No
Migration (Gene flow)	▲	▼	Yes
Random Genetic Drift	▼	▲	Yes
Selection	▲ ▼	▲ ▼	Yes

Within subpopulations the degree of genetic variation can be assessed by heterozygosity, while variation among subpopulations is measured by population differentiation. Mutation is the ultimate source of all genetic variation and it tends to increase variation both within and among subpopulations. But because most mutations are rare, the effect of mutation is slow relative to the change the other forces can effect. Migration or gene flow and random genetic drift are opposite in their effects: migration tends to increase variation within subpopulations but decrease it among subpopulations, and random drift does the opposite. In contrast, the effects of selection vary both within and between populations. For example, variation can decrease if one homozygote is favored, or may increase or be maintained if heterozygosity is advantageous. Selection acts on the phenotype so it will affect only those genes that control the trait under selection, as well as genes linked to those loci.
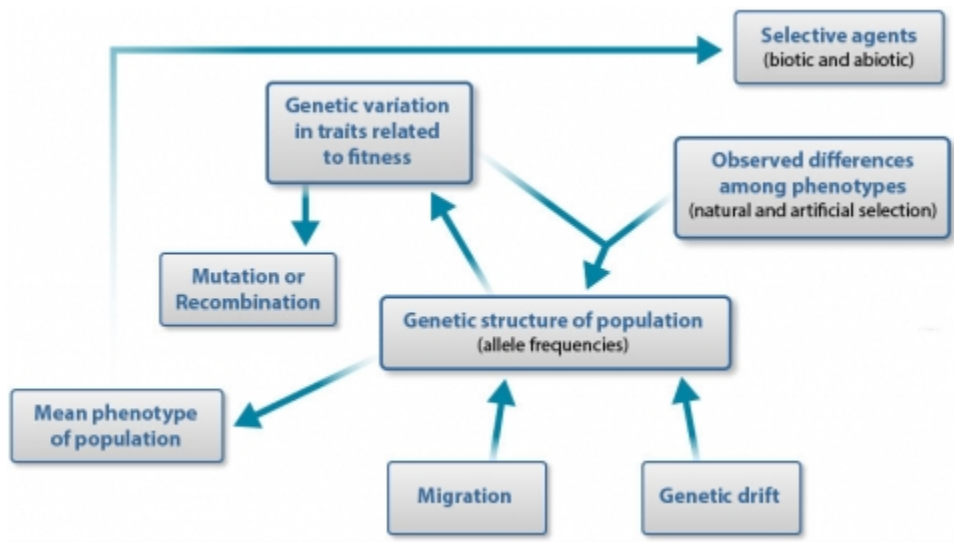
Summary

Fig. 16 Schematic overview of key concepts in population genetics. Source: Conner and Hartl, 2004.

Reflection

The Module Reflection appears as the last "task" in each module. The purpose of the Reflection is to enhance your learning and information retention. The questions are designed to help you reflect on the module and obtain instructor feedback on your learning. Submit your answers to the following questions to your instructor.

1. In your own words, write a short summary (< 150 words) for this module.
2. What is the most valuable concept that you learned from the module? Why is this concept valuable to you?
3. What concepts in the module are still unclear/the least clear to you?

References

Bernardo, R., A. Murigneux, and Z. Karaman. 1996. Marker-based estimates of identity by descent and alikeness in state among maize inbreds. Theoretical and Applied Genetics 93: 262-267.

Conner, J. K., and D.L. Hartl. 2004. A Primer of Ecological Genetics. Sinauer Associates, Sunderland, MA.

Falconer, D.S. and T.F.C. Mackay. 1996. Introduction to Quantitative Genetics. 4th edition. Longman Pub. Group, Essex, England.

Hancock, J.F. 2004. Plant Evolution and the Origin of Crop Species. 2nd edition. CABI Publishing, Cambridge, MA.

National Institutes of Health. National Human Genome Research Institute. "Talking Glossary of Genetic Terms." http://www.genome.gov/glossary/

Pierce, B. A. 2008. Genetics: A Conceptual Approach. 3rd edition. W.H. Freeman, New York.

For Your Information

Frequency of Homozygotes

With continued self-pollination, the heterozygotes will segregate, decreasing the proportion of heterozygotes in the population by half each generation. Notice that the homozygotes can only produce homozygotes.

Table 5 Change in percent heterozygosity in each successive generation.

Generation	Heterozygosity (%)
F_1	100.0
F_2	50.0
F_3	25.0
F_4	12.5
F_5	6.25
F_6	3.12

For each successive generation of offspring resulting from one F1 individual, by the F8 generation, the population is essentially homozygous. When no further segregation for the trait occurs, all progeny derived from that F1 will "breed true" because they are homozygous for the trait. The proportion of plants that are expected to be heterozygous at any gene when starting with a heterozygous F_1 and selfing can be determined by using the formula $(½)^n$, where n = the number of segregating generations, e.g., in F_2 n = 1 and in F_5 n = 4.

Proportion of homozygous plants in any generation is then given by $1 - (½)^n$ which when algebraically converted is equal to: $\dfrac{2^n - 1}{2^n}$

Mating Systems for Crop Species

Table 6

Flower Type	Self-Compatability or Dioecy/Monoecy	Crop Examples
Normally Cross-Pollinated		
Bisexual Flowers	Self-Compatible	Sugarcane, Olive, Amaranth, Avocado, Onion, Carrot, Agave, Sunflower, Kiwi, Pearl millet, Reed canarygrass, Sweet potato
Bisexual Flowers	Self-Incompatible	Radish, Kale, Cabbage, Black mustard, Pineapple, Red Clover, White Clover, Apple, Pear, Cacao, Rye, Alfalfa, Birdsfoot trefoil, Sweet Potato, Buckwheat
Unisexual Flowers	Dioecious	Papaya, Fig, Hops, Hemp, Grape
Unisexual Flowers	Monecious	Mango, Cucumber, Squash, Watermelon, Yam, Rubber, Cassava, Castor bean, Maize, Banana, Coconut, Oil palm
Normally Self-Pollinated		
Bisexual Flowers	Self-Compatible	Barley, Oats, Rice, Triticale, Wheat, Lettuce, Cowpea, Dry bean, Lentil, Chickpea, Peanut, Pea, Soybean, Sesame, Tomato, Tobacco, Coffee, Eggplant, Safflower, Flax, Peach
Predominantly Self-Pollinated, but also Cross-Pollinated to fairly high extent		
Bisexual Flowers	Self-Compatible	Cotton, Sorghum, Rapeseed, Brown mustard

Homozygosity and Heterozygosity

For a given locus, an individual with a genotype of either **AA** or **aa** is homozygous for that gene and is known as a homozygote; the status of the gene is referred to as homozygosity. An individual with the genotype **Aa** is heterozygous for that gene and is called a heterozygote; the status is known as heterozygosity. In the case of polyploid individuals, those with the genotypes **AAAA** (tetraploid) or **aaa** (triploid) would be examples of homozygotes and those with genotypes of **AAaa** (tetraploid) or **AAaaaa** (hexaploid) would be examples of heterozygotes.

The terms homozygous and heterozygous are used to describe the status of single genes or all gene loci within an individual, not within a population. There may be many different alleles of a gene present in a population of individuals, but for each diploid individual there are only two alleles per gene. For each individual, there is one allele from each parent and each allele per gene is present at corresponding loci on homologous chromosomes.

With regard to populations, a homogeneous population would be one in which all individuals in the population would have the same genotype and possess the same alleles for one or more genes. In contrast, a heterogeneous population would be characterized by differing alleles at one or more loci.

Note that a cross between two homozygous parents produces progeny that are homogeneous because all of the individual offspring are genetically identical. However, the offspring would be heterozygous for all loci for which different alleles occurred in the two parents.

Acknowledgements

How to cite this module: Beavis, W., L. Merrick, K. Meade, A. Campbell, D. Muenchrath, and S. Fei. 2016. Population Genetics. *In* Crop Genetics, interactive e-learning courseware. Plant Breeding E-Learning in Africa. Retrieved from **https://pbea.agron.iastate.edu**.