COURSE GUIDE

BIO 403
POPULATION GENETICS

Course Team Dr. Idowu, A. Taiwo (Developer/Writer) - UNILAG
Abiodun E. Adams (Coordinator) - NOUN
Dr. Yisa Abraham Gana (Course Reviewer) - NOUN
Dr. Maureen N. Chukwu (HOD) - NOUN

NATIONAL OPEN UNIVERSITY OF NIGERIA

National Open University of Nigeria
Headquarters
University Village
Plot 91, Cadastral Zone
Nnamdi Azikiwe Expressway
Jabi, Abuja

Lagos Office
14/16 Ahmadu Bello Way
Victoria Island, Lagos

e-mail: centralinfo@nou.edu.ng
URL: www.nou.edu.ng

Reviewed 2021

Printed 2023

ISBN: 978-058-443-9

CONTENTS **PAGE**

Introduction

BIO 403: Population Genetics is a 400- level, 2 credit-unit course in Biology. It is a second semester undergraduate course offered to students admitted in the School of Science and Technology, and School of Education who are offering Biology or related programmes.

This course guide tells you briefly what the course is all about, what course materials you will be using and how you can work your way through these materials. It gives you some guidance on your Tutor-Marked Assignments.

There is/are Self-Assessment Exercise(s) within the body of a unit. The exercise(s) is/are an overview of the unit to help you assess yourself at the end of every unit.

What You Will Learn in This Course

This course contains 14 units, which cover the genetic structure of a population. The genetic structure of a population is assessed in terms of genotypic and allelic frequencies. You will learn that two populations may be different because they have different genetic structures. In an ideal population, in which no forces of change are acting, a randomly interbreeding population would show constant genotypic and allelic frequencies for a given locus. However, the frequency of a given allele in a population can be changed by recurrent mutation, selection, or migration or by random sampling effect.

You will also learn that population genetics is the experimental and theoretical study of the pattern of inherited variation in populations and its modulation in time and space.

Course Aims

This course aims to enable you understand the genetic composition of a population and the forces that change that composition.

Course Objectives

There are some specific objectives set out in other to achieve the overall aim of this course. Every unit has specific objectives in addition. The unit objectives are stated in behavioural/achievable terms at the beginning of each unit. They are meant for you to read before you start working through the unit. You can also refer to them as you work through the course unit to check how far you are progressing. At the end of the unit, endeavour to refer to the objectives again to ensure you have achieved them. In this

way, you would be sure that you have done what you are required to do. The comprehensive objectives are given below.

On successful completion of this course, you should be able to:

- appreciate that there is extensive genetic variation in most natural populations
- discuss how Hardy-Weinberg principle shows the relationship between allelic and genotypic frequencies
- state how inbreeding increases the frequency of homozygotes
- explain that mutation is the original source of genetic variation
- enumerate that genetic drift may result in loss of genetic variation
- explain that gene flow can introduce new alleles into a population.

Working through This Course

In this course, you are advised to devote your time in reading through the material. You are required to do all that has been stipulated in the course: study the course units, read the recommended reference textbooks and do all the unit(s) self-assessment exercise(s) and at some points, you are required to submit your assignment (TMAs) for assessment purpose. You should therefore avail yourself the opportunity of being present during the tutorial sessions so that you would be able to compare knowledge with that of your mates.

Course Materials

You are to be provided with the two major course materials. These are:

1. Course Guide
2. Study Units

The course comes with a list of recommended textbooks. These textbooks are supplement to the course materials so that you can avail yourself of reading further. Therefore, it is advisable you acquire some of these textbooks and read them to broaden your scope of understanding.

Study Units

This course is divided into 3 modules with a total of 14 units as follows:

Module 1

Unit 1 Basic Concepts
Unit 2 Determination of Genotypic Frequency from Counts
Unit 3 Determination of Allelic Frequency from Counts

Unit 4 Determination of AllelicFrequency Given Various Conditions
Unit 5 Sex-Linked Traits

Module 2

Unit 1 Hardy-Weinberg Populations
Unit 2 Application of Hardy-Weinberg Equations to Real Populations
Unit 3 Application of Hardy-Weinberg Principle to One Locus Multiple Allelic System
Unit 4 Sex-Linked Traits

Module 3

Unit 1 Testing Populations for Hardy-Weinberg Equilibrium
Unit 2 Evolutionary Changes in Genetic Structure of Populations
Unit 3 Genetic Drift
Unit 4 Effect of Migration or Gene Flow on Evolutionary Changes
Unit 5 Effect of Selective Forces on Evolutionary Changes

Assessment

There are two components of the assessment for this course:

1. The Self-Assessment Exercise
2. The Tutor-Marked Assignment (TMAs)

Self-Assessment Exercise (SAQ)

The exercise(s) within each unit is/are meant to test your understanding of the concepts in the unit. It is non-grading and as such does not add up to your grade in the course.

Tutor-Marked Assignment (TMA)

The TMA is the continuous assessment component of your course. It accounts for 30% of the total score you will obtain in this course.

Final Examination and Grading

In this course, the final examination will be of two hours duration and have a value of 70% of the total course grade. Questions for the examination will reflect the types of exercises, examples and tutor-marked assignments. All areas of the course will be assessed.

Use the time between finishing the last unit and sitting for the examination to revise the entire course. You might find it useful to review your tutor-marked assignments and comment on them before the examination.

Summary

Population genetics is the study of genetics at the population level. Also, it should be appreciated that extensive genetic variation occurs in most populations. By counting the numbers of different genotypes in a population, we can calculate genotypic and allelic frequencies. We can also assess changes in population by measuring the effects of mutation, migration, genetic drift and the breeding systems

MAIN COURSE

CONTENTS **PAGE**

MODULE 1

UNIT 1 BASIC CONCEPTS

CONTENTS

1.0 INTRODUCTION

In **BIO 201: Genetics 1**, you learnt that individuals may differ genetically. This implies that differences may exist at individual level. At the population level too, differences may also exist. One may therefore say that population A is genetically different from population B with respect to certain attributes. Such attributes include frequency of genotypes (genotypic frequency) and frequency of alleles (allelic frequency) in the populations being compared.

2.0 OBJECTIVES

At the end of this unit, you should be able to:

- discuss population genetics
- highlight the differences in scope between Population Genetics and Basic Genetics taken at 200 Level.

3.0 MAIN CONTENT

3.1 Differences between Transmission and Population Genetics

Bio-Genetics deals with transmission genetics or classical genetics. You could have observed that it is primarily concerned with genetic processes that occur between individuals, and how genes are passed from one individual to another. Thus, the unit for transmission genetics is the individual. Population genetics on the other hand, studies heredity in groups of individuals for traits that are determined by one or only a few genes. Therefore, in population genetics the unit of study is the population. A population can be defined as a group of individuals that exist in time and space and can interbreed. Such a population is referred to as Mendelian population.

IN TEXT QUESTIONS

What is Bio-Genetics?

Answer: Bio-Genetics deals with transmission genetics or classical genetics.

3.2 Historical Development of Population Genetics

The basic laws of heredity are the law of segregation and the law of independent assortment as formulated by Mendel in his experiment on garden peas. However, it was later realised that there are questions about the population that cannot be addressed by mere application of Mendelian laws of heredity especially at the population level. It was also realised that mathematical models are required to describe structure of populations and to answer many questions that cannot be answered in transmission genetics. Population genetics therefore evolved as a branch of genetics that describes in mathematical terms, the consequences of Mendelian inheritance on the population level. Therefore, a student of population genetics requires basic skill in Mathematics. Almost all the mathematical foundations of genetic changes in populations were developed in a short period of time during the 1920s and 1930s by three men: R. A. Fisher, J. B. S. Haldane and S. Wright. Some measure of disagreement emerged among these men, but they disagree on which evolutionary processes were more important, not on how the processes worked. Since the 1960s, excitement has arisen in the field of population genetics, primarily on three fronts. First, the high-speed computer has made it possible to do a large amount of arithmetic in a very short period of time; thus, complex simulations of real populations

can be added to the repertoire of the experimental geneticist. Secondly, electrophoresis has provided a means of gathering the large amount of empirical data necessary to check some of the assumptions used in mathematical models. The information and interpretation of the electrophoretic data have generated some controversy about the role of "neutral" evolutionary changes in natural populations. Last, newer techniques of molecular genetics are being used to analyse the relationships among species and the rate of evolutionary processes. We shall consider these issues later in the course.

3.3 Frequently Asked Questions (FAQs) in Population Genetics

Questions frequently asked and studied by population geneticists include the following:

1. How much genetic variation is found in natural populations, and what processes control the amount of variation observed?
2. What evolutionary processes shape the genetic structures of populations?
3. What processes are responsible for producing genetic divergence among populations?
4. How do biological characteristics of a population, such as breeding system, fecundity, and age structure, influence the gene pool of the population?

3.4 Answers to the Frequently Asked Questions

To answer these questions, population geneticists frequently develop mathematical models and equations to describe what happens to the gene pool of a population under various conditions. As a result of regularity of the meiotic process in gamete formation and the ensuing mechanisms of gamete fusion during sexual reproduction, a great deal of predictability concerning the genetic structure of a population can be modelled. An example is the set of equations that describes the influence of random mating on the allele and genotypic frequencies of an infinitely large population, a model called the **Hardy-Weinberg law**, which we will discuss later in this course. At first, students often have difficulty grasping the significance of models such as these, because the models are frequently simple and require numerous assumptions that are unlikely to be met by organisms in the real world. Beginning with simple models is useful, however, because with such models we can examine what happens to the genetic structure of a population when we deliberately violate one assumption after another and then in combination. Once we understand the results of the simple models, we can incorporate more realistic conditions into the equations. In the case

of the Hardy-Weinberg law, the assumptions of infinitely large size and random mating may seems unrealistic, but these assumptions are necessary at first for simplifying the mathematical analysis. We cannot begin to understand the impact of nonrandom mating and limited population size on allelic frequencies until we first know what happens under the simpler conditions of random mating and large population size.

3.5 The Gene Pool: A Central Concept in Population Genetics

To study the genetic structure of a Mendelian population, population geneticists must first describe the gene pool of the population quantitatively because members of a Mendelian population share from a common gene pool. A gene pool can defined as the alleles among the reproductive members of a population from which gametes can be drawn. We will find Mendelian genetic principles essential in order to understand the genetic composition of populations. For example, the principle of segregation (that two alleles in a heterozygous individual are equally represented in the gametes produced by that individual), when this is applied to all individuals in a population, it allows us to understand the overall genetic variation in that population. Let us first describe some examples of genetic variation before discussing the principles of population genetics that explain the basis of this variation and how it may change.

3.6 Types of Genetic Variation in Populations

Genetic variants can directly affect the morphology or colour of an organism in a natural or wild population. A classic example of genetic variation that has fascinated evolutionary biologists in England and France for decades is the colour and banding pattern of *Cepaea nemoralis,* a European land snail common in gardens and uncultivated fields. The morphological variation is determined by the alleles at a group of closely linked genes, so that their nails vary strikingly in colour and pattern from yellow, unbanded types to yellow, banded types to brown, unbanded ones. The frequencies of the alleles that determine yellow colour and bandedness vary on a large scale from one part of Europe to another and on a small scale from one field to another.

In one detailed study, the frequency of yellow, unbanded types over a 200 meter stretch on an earthen dam in the Netherlands varied from 60% at one end to 40% at the other. Furthermore, these populations remained at the same frequencies over a 12-year period, suggesting that evolutionary forces were maintaining the differences. For example, it appears from other studies that the different forms confer protective

camouflage (or crypsis) from the predation of birds, depending upon the habitat in which they are found. When the snails are resting on grass blades, the bands are vertical, so the banded snails blend in with the grass. But when the snails are in the ground litter, the unbanded brown snails are the most difficult to detect.

In a number of species, variation in chromosome structure has been observed among different individuals. Recently, new techniques developed to examine morphological details of chromosomes have uncovered many new chromosomal variants, and it is likely that numerous other small variations in chromosome structure, such as deletions, replications, and inversions, are still to be discovered. One of the most thoroughly studied examples of chromosomal variation was initiated by Theodosius Dobzhansky, a Russian geneticist who moved to the United States in the 1920s. He and his students analysed the inversion variants on the third chromosome of the fruit fly *Drosophila pseudoobscura*, a widespread species in the mountains of western North America.

In the 1930s, Dobzhansky began a long-term study of variation in this inversion. Some of these data show the frequencies of four inversion types over thirty years of sampling at a site in New Mexico. Initially, the inversion type AR (Arrowhead) was most common, but in 1965 type PP (Pikes Peak) increased in frequency, only to decline in later years. Two other inversion types, CH (Chiricahua) and ST (Standard), were always present, but at low frequencies in the samples from this population.

These inversions, although inherited as single units like alleles (see unit 4), generally contain hundreds of genes. Because different inversions may have different alleles at these loci, they can substantially affect the survival of individual flies. In fact, some of the changes observed over time may have been caused by differential survival of various inversion to DDT, an insecticide that was first applied in t he 1940s.

Most alleles that cause genetic disease in humans occur in very low frequencies. However, some disease alleles have higher frequencies in certain human populations than in others. For example, the frequency of the sickle-cell allele at the β-globin locus is quite high in many African populations and in populations with African ancestry, such as African-Americans. The sickle-cell allele also occurs in some Mediterranean and Asian populations.

The presence of the sickle-cell allele can be ascertained by biochemical examination of the β-globin molecule. Variant forms of a protein can be detected by **gel electrophoresis,** a technique that allows the separation of different proteins extracted from blood, tissues, or whole organisms.

The process is carried out by imposing an electric field across a gelatinous supporting medium (such as a polyacrylamide or starch gel) into which the protein has been placed. The proteins are allowed to migrate for a specific amount of time and are then stained with various protein-specific chemicals, resulting in bands on the gel that permit the relative mobility of a specific protein to be determined. The relative mobility is generally a function of the charge, size and shape of the molecule. Two proteins that are the products of different alleles can have different mobilities because the differences in sequence result in a change in charge size, or shape of the molecule.

The three β-globin genotypes (homozygotes for the normal and sickle- cell allele plus a heterozygote) have different banding patterns when they are run on an electrophoretic gel. The normal A_1A_1 genotype has only band that migrates relatively fast, as in the first column, while individuals with sickle-cell disease, genotype A_2A_2 , also have only one band, but it migrates more slowly. The heterozygous individuals, A_1A_2, who have the sickle-cell trait (they are carriers) have two bands, each band representing the protein synthesised from the respective allele. Sometimes these genotypes at the β-globin locus are represented as AA, AS, and SS, instead of A_1A_1, A_1A_2, A_2A_2.

The reason for the high frequency of this disease allele is that heterozygotes (those who carry this allele and the normal allele) have a relatively high resistance to malaria. The relative immunity of sickle- cell carriers to malaria is thought to occur because the red blood cells of heterozygotes are sickle-shaped, a state that inhibits infection by the malaria parasite. When we compare the frequency of the sickle-cell allele and the past incidence of malaria on a geographic scale (before the extensive eradication of the mosquito that carries the malaria parasite), we find an extremely high concordance.

IN TEXT QUESTION

Various forms of protein can be detected by _____

Answer: gel electrophoresis

4.0 CONCLUSION

Population genetics evolved as a branch of basic genetics you offered at your 200 Level **(BIO 201: Genetics 1).** In population genetics, consequences of Mendelian inheritance are described at the population level.

5.0 SUMMARY

Population genetics is different from transmission genetics. The course BIO 201-Genetics I offered in your 200 level is basic genetics that focuses on transmission genetics. In population genetics, genetic structure of populations is studies with reference to different traits found in the population.

6.0 TUTOR-MARKED ASSIGNMENT

1. Outline the differences between population and transmission genetics.
2. What is meant by saying two human populations do not share from the same gene pool?

7.0 REFERENCES/FURTHER READING

Klug, W. S.; Cummings, M.R.; Spencer, C.A. & Palladino, M.A. (2009). *Concepts of Genetics*. Pearson Int. Ed. (9th ed.). Pearson Benjamin Cummings.

Peter, J. Russel. (nd). *Genetics.* New York: HarperCollins College Publishers.

Russel, P. J. (1996). *Genetics* (7th ed.). New York: HapperCollins College Publishers.

Weaver, R. F & Hedrick, P. H. (1997). *Genetics* (3rd ed.). WCB Publishers.

UNIT 2 DETERMINATION OF GENOTYPIC FREQUENCY FROM COUNTS

CONTENTS

1.0 INTRODUCTION

Recall that an allele is one of two or more alternative forms of a gene. Many traits in living organisms are controlled by a pair of alleles. We shall first shift our attention to a situation whereby a trait is controlled by only two alleles i.e. a pair of alleles. The term "genotypic" is an adjective derived from the term genotype. Genotype can be defined as the genetic (or allelic) constitution of an organism. Thus, the phrase genotypic frequency refers to the frequency of the different genotypes present at a particular locus in an individual.

2.0 OBJECTIVE

At the end of this unit, you should be able to:

- determine genotypic frequencies from actual counts.

3.0 MAIN CONTENT

3.1 Genetic Structure of Populations

To study the genetic structure of a population, one must first describe the gene pool of the population quantitatively. This is done by determining, first, the genotypic frequency and then, the allelic frequency in the population. In this unit, our attention will be focused on calculation of genotypic frequency. Determination of allelic frequency will be considered in the subsequent unit.

3.2 Determining the Three Different Genotypes

To be able to determine the frequency or proportion of the three genotypes in a two allelic system, the genotypes must be distinguishable. This is why determination of genotypic frequency from actual counts is usually done with a two allelic system whereby all the genotypes are distinguishable. Note that ability to distinguish the different genotypes depend on the dominance relationship between the alleles. There are three main different types of dominance relationship: complete dominance, incomplete dominance, and co-dominance. It is only when the dominance relationship is incomplete dominance or co- dominance that it is possible to distinguish the different genotypes possible at that particular locus.

3.2.1 Complete Dominance (Not Suitable for Determination of Genotypic Frequency by Actual Counts)

In complete dominance, it is not possible to differentiate between the homozygous dominant and the heterozygous class, because both have the same phenotypes. In most situations, the normal or wide wild-type, allele is completely dominant over mutant alleles. The basis of this will be more clear when we later examine the biochemical products of genes and find that one functioning copy of a gene is generally enough to produce sufficient gene product for a wild-type phenotype. For example, the albino allele in humans does not produce functioning gene product, and when an individual is homozygous for this allele, the biochemical pathway for the pigment melanin is blocked. However, one copy of the wild-type allele, as is present in heterozygous, allows sufficient production of the pigment melanin so that normal pigmentation results.

In some instances, the mutant allele is dominant over the wild-type; that is, the wild type is recessive. There are 4,458 entries representing dominant loci in McKusick's 1994 catalog of human traits, many of them are rare diseases. For example, the most common type of dwarfism, achondroplasia, is dominant, so that heterozygous exhibits the mutant phenotype. In subsequent unit of this course, we will discuss some aspects of Huntington's disease, a fatal neurological disorder that is also dominant.

3.2.2 Incomplete Dominance (Suitable for Determination of Genotypic Frequency by Actual Counts)

In incomplete dominance, it is possible to distinguish between the homozygous dominant and the heterozygous class. Therefore, determination of genotypic frequency from population counts is

possible. When Carl Correns, one of Mendel's disciples, experimentally crossed red-flowered four-o'clocks with white-flowered ones, the F_1 plants were unlike either parent, being an intermediate pink instead. When he allowed pink F_1 plants to self-fertilise, the F_2 ratio of phenotypes was 1 red: 2 pink: 1 white, or ¼, ½, ¼. Although at first glance, these results appear to deviate from Mendel's, what Correns actually observed in the F_2 was a 1:2:1 phenotypic ratio corresponding to the 1:2:1 genotypic ratio expected based on Mendelian principles. Because neither parental phenotype was fully expressed in the F_1, the trait is said to exhibit **incomplete dominance;** that is, all genotypes have different phenotypes, with the heterozygote's phenotype intermediate between the two homozygotes.

However, the type of dominance for a given genetic variant is not always easy to designate. For example, in what appears to be complete dominance, the observable phenotype of the heterozygote is the same as that of the dominant homozygote, but may show incomplete dominance on a biochemical level. For example, RR homozygotes and Rr heterozygotes at the gene affecting seed shape in garden peas, both have the same round seed phenotype. However, these genotypes differ biochemically: the heterozygote has starch grains that are intermediate between the two homozygotes in type and amount. For this reason, we must be specific about the level of the trait – for example, gross phenotypic or biochemical – we are describing.

3.2.3 Co-dominance (Suitable for Determination of Genotypic Frequency by Actual Counts)

As in incomplete dominance, it is possible to distinguish between the homozygous dominant and the heterozygous class in co-dominance allelic relationship. Therefore, determination of genotypic frequency from population counts is possible. In some cases, the traits associated with both alleles are observable in the heterozygote. Such alleles are said to exhibit co-dominance. For example, individuals heterozygous for red blood cell antigens, genetically determined chemicals on the membranes of red blood cells, often exhibit properties of both alleles. A well-known case is type AB in the ABO blood group system, which has the antigens from two different alleles. Another blood group system, called MN, has two alleles, generally denoted as L^M and L^N, and three genotypes, L^ML^M, L^ML^N, and L^NL^N. (Sometimes these genotypes are called MM, MN, and NN, respectively). The expected proportions of genotypes in the offspring are also shown, and it can be seen that they follow the Mendelian segregation pattern.

Although the example of flower colour in four-o'clocks illustrates incomplete dominance on a gross phenotypic scale, close examination of

the pink flowers of heterozygote show that both red plastids and white plastids are present. (Plastids are a type of cell organelle). In other words, for this trait we find incomplete dominance on the gross phenotypic level and co-dominance on a sub-cellular level.

3.3 Calculation of Genotypic Frequency

A frequency is a percentage or a proportion, and always ranges between 0 and 1. If 43% of the people in a group have red hair, the frequency of red hair in the group is 0.43. To calculate the genotypic frequencies at a specific locus, we count the number of individuals with one particular genotype and divide this number by the total number of individuals in the population. We do this for each of the genotypes at the locus. The sum of the genotypic frequencies should be 1. Consider a locus that determines the pattern of spots in the scarlet tiger moth, *Panaxia dominula*. Three genotypes are present in most populations, and each genotype produces a different phenotype i.e. the three genotypes are distinguishable. E.B. Ford collected moths at one locality in England and found the following numbers of genotypes: 452 BB, 43 Bb, and 2bb, out of a total of 497 moths. The genotypic frequencies (where f= frequency of) are therefore:

f(BB) = 452/497 = 0.909
f(Bb) = 43/497 = 0.087
f(bb) = 2/497 = 0.004
Total = 1.000

IN TEXT QUESTION
Genotypic frequency range from _____ to _____ Answer: 0 to 1

4.0 CONCLUSION

Population genetics relates the process of individual heredity and development to the genetic composition of populations and to changes in that composition in time.

5.0 SUMMARY

Populations of living organism have different genetic structures because different populations do not share from the same gene pool. This issue is central to population genetics.

6.0 TUTOR-MARKED ASSIGNMENT

1. Traits controlled by alleles with complete dominance relationship are not suitable for population genetic analysis. Why?
2. The number of individuals living in a village is 200. A study showed that the numbers of individuals in the village with different M-N blood group phenotypes are as follow:

Phenotype	**No. of individuals**
M	60
MN	100
N	40

3. Calculate (a) the genotypic frequency and (b) the allelic frequency.

7.0 REFERENCES/FURTHER READING

Klug, W. S.; Cummings, M.R.; Spencer, C.A. & Palladino, M.A. (2009). *Concepts of Genetics*. Pearson Int. Ed. (9th ed.). Pearson Benjamin Cummings.

Peter, J. Russel (nd). *Genetics.* New York: Harper Collins College Publishers.

Russel, P. J. (1996). *Genetics* (7th ed.). New York: Harper Collins College Publishers.

Weaver, R. F. & Hedrick, P. H. (1997). *Genetics* (3rd ed.). WCB Publishers.

UNIT 3 DETERMINATION OF ALLELIC FREQUENCY FROM COUNTS

CONTENTS

1.0 INTRODUCTION

Although genotypic frequencies at a single locus are useful for examining the genetic structure of a population, in most cases allelic frequencies are more useful in describing the gene pool. The use of allelic frequencies offers several advantages over the genotypic frequencies. First, in sexually reproducing organisms, genotypes break down to alleles when games are formed, and alleles, not genotypes, are passed from one generation to the next. Consequently, only alleles have continuity over time, and the gene pool evolves through changes in the frequencies of alleles. Furthermore, there are always fewer alleles than genotypes, and so the gene pool can be described with fewer parameters when allelic frequencies are used. For example, if there are three alleles segregating at a particular locus, six genotypes will form and frequencies must be calculated for each to describe the gene pool.

IN TEXT QUESTION
Genotypic frequencies at a single locus are useful for examining the _____ Answer: genetic structure of a population,

2.0 OBJECTIVE

At the end of this unit, you should be able to:

- determine allelic frequencies from actual counts.

3.0 MAIN CONTENT

3.1 Calculation of Allelic Frequency from Observed Number of Genotypes

Allelic frequency may be calculated in two ways, either from:

- the observed numbers of different genotypes at a particular locus
- or the genotypic frequencies.

First, we can calculate the allelic frequencies directly from the numbers of genotypes. In this method, we count the number of alleles of one type at a particular locus and divide it by the total number of alleles at that locus in the population:

$$\text{Allelic frequency} = \frac{\text{Number of copies of a given allele in the population}}{\text{Sum of all alleles in the population}}$$

As an example, imagine a population of 1,000 diploid individuals with 353 *AA*, 494 *Aa*, and 153 *aa* individuals. Each *AA* individual has two *A* alleles, while each *Aa* heterozygote possesses only a single *A* allele.

Therefore, the number of *A* alleles in the population is (2 x the number of *AA* homozygotes) + (the number of *Aa* heterozygotes) or (2 x 353) + 494, = 1,200.

Because every diploid individual has two alleles, the total number of alleles in the population will be twice the number of individuals, or 2 x 1,000.

Using the equation given above, the allelic frequency is 1,200/2,000 = 0.60.

When two alleles are present at a locus, we can use the following formula for calculating allelic frequencies:

$$P = f(A) = \frac{(2 \times \text{number of } AA \text{ homozygotes}) + (\text{number of } Aa \text{ heterogotes})}{(2 \times \text{total number of individuals})}$$

$$q = f(a) = \frac{(2 \times \text{number of } aa \text{ homozygotes}) + (\text{number of } Aa \text{ heterogotes})}{(2 \times \text{total number of individuals})}$$

3.2 Calculation of Allelic Frequency from Genotypic Frequency

The second method of calculating allelic frequencies goes through the step of first calculating genotypic frequencies as demonstrated previously.

In this example f(*AA*) = 0.353; f(*Aa* = 0.494; and f(*aa*) = 0.153. From these genotypic frequencies, we calculate the allele frequencies as follows:

- *p* = f(*A*) = (frequency of the *AA* homozygote) + (1/2 x frequency of the *Aa* heterozygote)
- *q* = f(*A*) = (frequency of the *aa* homozygote) + (1/2 x frequency of the *Aa* heterozygote)

The frequencies of two alleles, f(*A*) and f(*a*), are commonly symbolised as *p* and *q*. The allelic frequencies for a locus, like the genotypic frequencies, should always add up to 1. This is because in a one gene locus model that has only two alleles, 100% (i.e., the frequency = 1) of the alleles are accounted for by the sum of the percentages of the two alleles. Therefore, once the frequency of one of the alleles (e.g. p) is known, and since p + q=1, *q* can easily be obtained by q = 1- *p.*

IN TEXT QUESTION

The frequencies of two allele are symbolized by _____ and _____ *p* and *q*

4.0 CONCLUSION

Only alleles have continuity over time, and the gene pool evolves through changes in the frequencies of alleles. In most cases allelic frequencies are more useful in describing the gene pool. Thus, determination of allelic frequencies given various circumstances is very crucial in population genetics.

5.0 SUMMARY

Allelic frequencies could be determined directly by counting the actual numbers of individuals in different genotypic classes in the population. Also, if the genotypic frequencies are known, allelic frequencies could equally be determined. There are formulae respectively for calculating allelic frequencies from actual counts given the two conditions respectively:

Given the number of individuals in different genotypic classes (i.e. from number of genotypes):

P= f(A)= (2 x number of *AA* homozygotes) + (number of *Aa* heterogotes)
(2 x total number of individuals)

q= f(a)= (2 x number of *aa* homozygotes) + (number of *Aa* heterogotes)
(2 x total number of individuals)

From Genotypic Frequencies:

- *p* = f(*A*) = (frequency of the *AA* homozygote) + (1/2 x frequency of the *Aa* heterozygote)
- *q*= f(*A*) = (frequency of the *aa* homozygote) + (1/2 x frequency of the *Aa* heterozygote)

However, once the frequency of one of the alleles (e.g. p) is known, and since p+q=1, *q* can easily be obtained by:

q = 1- *p*

6.0 TUTOR-MARKED ASSIGNMENT

Calculate the allelic frequencies from the genotypic frequencies obtained in the Tutor Marked Assessment question in Unit 2. Comment whether determination of allelic frequencies from actual counts and from genotypic frequency gives the same answer.

7.0 REFERENCES/FURTHER READING

Klug, W. S.; Cummings, M.R.; Spencer, C.A. & Palladino, M.A. (2009). *Concepts of Genetics*. Pearson Int. Ed. (9th ed.). Pearson Benjamin Cummings.

Peter, J. Russel (nd). *Genetics.* New York: Harper Collins College Publishers.

Russel, P. J. (1996). *Genetics* (7th ed.). New York: Harper Collins College Publishers.

Weaver, R. F. & Hedrick, P. H. (1997). *Genetics* (3rd ed.). WCB Publishers.

UNIT 4 DETERMINATION OF ALLELIC FREQUENCY GIVEN VARIOUS CONDITIONS

CONTENTS

1.0 INTRODUCTION

It is possible to obtain allelic frequencies given a variety of situations. It is necessary to master the various methods very well.

2.0 OBJECTIVE

At the end of this unit, you should be able to:

- determine allelic frequencies from observation made on phenotypic and genotypic counts in the population.

3.0 MAIN CONTENT

3.1 Method of Calculation given actual Counts in Multiple Allelic Systems

Multiple Allelic System

Suppose we have three alleles – A^1, A^2, and A^3 – at a locus, and we want to determine the allelic frequencies.

Here, we employ the same rule that we used with two alleles: we add up the number of alleles of each type and divide by the total number of alleles in the population:

$$p = f(A^1) = \frac{(2 \times A^1A^1) + (A^1A^2) + (A^1A^3)}{(2 \times \text{total number of individuals})}$$

$$q = f(A^2) = \frac{(2 \times A^2A^2) + (A^1A^2) + (A^2A^3)}{(2 \times \text{total number of individuals})}$$

$$r = f(A^3) = \frac{(2 \times 3^2A^3) + (A^1A^3) + (A^2A^3)}{(2 \times \text{total number of individuals})}$$

To illustrate the calculation of allelic frequencies when more than two alleles are present, we will use data from a study on genetic variation in "milkweed beetles". Walter Eanes and his coworkers examined allelic frequencies at a locus that codes for the enzyme phosphoglucomutase (PGM). Three alleles were found at this locus; each allele codes for a different molecular variant of the enzyme. In one population, the following numbers of genotypes were collected:

AA	=	4
AB	=	41
BB	=	84
AC	=	25
BC	=	88
CC	=	32
Total	=	274

The frequencies of the alleles are calculated as follows:

$$f(A) = p = \frac{(2 \times 4) + 41 + 25}{(2 \times 274)} = 0.135$$

$$f(B) = q = \frac{(2 \times 84) + 41 + 88}{(2 \times 274)} = 0.542$$

$$f(C) = r = \frac{(2 \times 32) + 88 + 25}{(2 \times 274)} = 0.323$$

As seen in these calculations, we add twice the number of homozygotes that possess the allele and one times each of the heterozygotes that have the allele. We then divide by twice the number of individuals in the population, which represents the total number of alleles present. Note that in the top part of the equation for each allelic frequency, we do not add all the heterozygotes, because some of the heterozygotes do not have the allele; for example, in calculating the allelic frequency of A, we do not add the number of *BC* heterozygotes in the top part of the equation. BC individuals do not have an A allele. We can use the same procedure for calculating allelic frequencies when four or more alleles are present.

3.2 Method of Calculation given Genotypic Frequencies in Multiple Allelic Systems

The second method for calculating allelic frequencies from genotypic frequencies can also be used here. This calculation may be quicker if we have already determined the frequencies of the genotypes. The frequency of the homozygote is added to half of the heterozygote frequency because half of the heterozygote's alleles are *A* and half are *a*. If three alleles (A^1, A^2 and A^3) are present in the population, the allelic frequencies are:

$$p = f(A^1) = f(A^1A^1) + 1/2\, f(A^1A^2) + 1/2f(A^1A^3)$$
$$q = f(A^2) = f(A^2A^2) + 1/2\, f(A^1A^2) + 1/2f(A^2A^3)\ r$$
$$= f(A^3) = f(A^3A^3) + 1/2\, f(A^1A^3) + 1/2f(A^2A^3)$$

Although calculating allelic frequencies from genotypic frequencies may be quicker than calculating them directly from the numbers of genotypes, more rounding error occur. As a result, calculations from direct counts are usually preferred. Calculating genotype frequencies and allelic frequencies are illustrated for a one-gene locus, three allelic cases as illustrated below:

Sample calculating of genotypic and allelic frequencies for hemoglobin variants among Nigerians where multiple alleles are present

Hemoglobin Genotypes

AA	AS	SS	AC	SC	CC	Total
2,017	783	4	173	14	11	3,002

Calculation of Genotypic Frequencies

$$\text{Genotypic frequency} = \frac{\text{Number of individuals with the genotype}}{\text{Total number of individuals}}$$

$$f(SS) = \frac{4}{3{,}002} = 0.0013 \qquad f(AA) = \frac{2017}{3{,}002} = 0.672 \qquad f(AC) = \frac{173}{3{,}002} = 0.058$$

$$f(AS) = \frac{783}{3{,}002} = 0.261 \qquad f(SC) = \frac{14}{3{,}002} = 0.0047 \qquad f(CC) = \frac{11}{3{,}002} = 0.0037$$

Calculation of allelic frequencies from the number of individuals with a particular genotype

$$\text{Allelic frequency} = \frac{\text{Number of copies of a given allele in the population}}{\text{Sum of all alleles in the population}}$$

(2xnumber of *SS* individuals) + (Number of *AS* individuals) + (number of *SC* individuals)

f(S) = (2 x total number of individuals)

(2 x 4) + 783 +14) = 805

$$f(S) = \frac{(2 \times 3{,}002)}{6{,}004} = 0.134$$

(2x number of *AA* individuals) + (Number of *AS* individuals) + (number of *AC* individuals)

f(A) = (2 x total number of individuals)

(2 x 2,017) + 783 +173) = 4,990

$$f(A) = \frac{(2 \times 3{,}002)}{6{,}004} = 0.831$$

(2x number of *CC* individuals) + (Number of *AC* individuals) + (number of *SC* individuals)

f(C) = (2 x total number of individuals)

(2 x 11) + 173 +14) = 209

$$f(C) = \frac{(2 \times 3{,}002)}{6{,}004} = 0.035$$

Calculation of allelic frequencies from the frequency of particular genotypes

$f(S) = f(SS) + 1/2\ f(AS) + 1/2\ f(SC)$
$f(S) = 0.0013 + 1/2 \times (0.261) + 1/2 \times 0.0047 = 0.134$
$f(A) = f(AA) + 1/2\ f(AS) + 1/2\ f(AC)$
$f(A) = 0.672 + 1/2 \times 0.261 + 1/2 \times 0.0587 = 0.831$
$f(C) = f(CC) + 1/2\ f(SC) + 1/2\ f(AC)$
$f(C) = 0.01037 + 1/2 \times 0.0047 + 1/2 \times 0.0058 = 0.0035$

Data are from Livingstone, F.B. 1973. Data on the Abnormal Hemoglobins and Glucose-6-Phosphate Dehydrogenase Deficiency in

Human Populations of 1967-1973. Contributions in Human Biology No. 1. Museum of Anthropology, University of Michigan.

IN TEXT QUESTION

How is allelic frequency calculated?

Answer

$$\text{Allelic frequency} = \frac{\text{Number of copies of a given allele in the population}}{\text{Sum of all alleles in the population}}$$

4.0 CONCLUSION

Allelic frequencies can be calculated from actual observations made in the population in a multiple allelic system. The technique is an extension of the two allelic cases.

5.0 SUMMARY

Determination of allelic frequencies from population counts requires the use of appropriate formulae.

6.0 TUTOR-MARKED ASSIGNMENT

See section 3.2 for a sample question and model calculation.

7.0 REFERENCES/FURTHER READING

Klug, W. S.; Cummings, M.R.; Spencer, C.A. & Palladino, M.A. (2009). *Concepts of Genetics*. Pearson Int. Ed. (9th ed.). Pearson Benjamin Cummings.

Peter, J. Russel (nd). *Genetics.* New York: Harper Collins College Publishers.

Russel, P. J. (1996). *Genetics* (7th ed.). New York: Harper Collins College Publishers.

Weaver, R. F. & Hedrick, P. H. (1997). *Genetics* (3rd ed.). WCB Publishers.

UNIT 5 SEX-LINKED TRAITS

CONTENTS

1.0 INTRODUCTION

Calculation of allelic frequencies at an X-linked locus is slightly more complicated, because males have only a single X-linked allele. However, we can use the same rules we used for autosomal loci. Remember that each homozygous female carries two X-linked alleles; heterozygous females have only one of that particular allele, and all males may have only a single X-linked allele.

2.0 OBJECTIVE

At the end of this unit, you should be able to:

- calculate allelic frequency for sex-linked traits.

3.0 MAIN CONTENT

3.1 Method of Calculation

To determine the number of alleles at an X-linked locus, we multiply the number of homozygous females and the number of hemizygous males. We next divide by the total number of alleles in the population. When determining the total number of alleles, we add twice the number of females (because each female has two X-lined alleles) to the number of males (who have a single allele at X-linked loci). Using this reasoning, the frequencies of two alleles at an X-linked locus (SA and Xa) are determined with the following equations:

$$p = f(X^A) = \frac{(2 \times X^AX^A \text{ females}) + (X^AX^a \text{ females}) + (X^AY \text{ males})}{(2 \times \text{number of females}) + (\text{number of males}}$$

$$q = f(X^a) = \frac{(2 \times X^aX^a \text{ females}) + (X^AX^a \text{ females}) + (X^aY \text{ males})}{(2 \times \text{number of females}) + (\text{number of males})}$$

Allelic frequencies at an X-linked locus can be determined from the genotypic frequencies by:

$$p = f(X^A) = f(X^AX^A) + 1/2\, f(X^AX^a) + f(X^AY)$$

$$q = f(X^a) = f(X^aX^a) + 1/2\, f(X^AY) + f(X^aY)$$

You should strive to understand the logic behind these calculations, not just to memorise the formulas. If you fully understand the basis of the calculations, you will not need to remember the exact equations and will be able to determine allelic frequencies for any situation.

IN TEXT QUESTION

Why is twice the number of female added when calculating total number of allele? Answer: because each female has two X-lined alleles

4.0 CONCLUSION

Allelic frequencies can be determined from actual counts and genotypic frequencies for sex-linked traits.

5.0 SUMMARY

Appropriate formulae can be used to determine genotypic and allelic frequencies in sex-linked cases. It should be taken note that males are hemizygous for sex-linked traits since they have only one x chromosome.

6.0 TUTOR-MARKED ASSIGNMENT

A small community is made up of 400 people (220 females: 180 males). Ten females are colour blind while 40 males are colour blind. The rest have normal vision. Although normal vision is controlled by a completely dominant allele, assume that it is possible to distinguish between homozygous normal vision and heterozygous individuals. Determine (a) genotypic frequencies and (b) allelic frequencies.

7.0 REFERENCES/FURTHER READING

Klug, W. S.; Cummings, M.R.; Spencer, C.A. & Palladino, M.A. (2009). *Concepts of Genetics*. Pearson Int. Ed. (9th ed.). Pearson Benjamin Cummings.

Peter, J. Russel (nd). *Genetics.* New York: Harper Collins College Publishers.

Russel, P. J. (1996). *Genetics* (7th ed.). New York: Harper Collins College Publishers.

Weaver, R. F. & Hedrick, P. H. (1997). *Genetics* (3rd ed.). WCB Publishers.

MODULE 2

UNIT 1 HARDY-WEINBERG POPULATIONS

CONTENTS

1.0 INTRODUCTION

In the last Module, we learnt how to determine genotypic and allelic frequencies in a population. It was also stressed that it is crucial to know all the genotypic classes to be able to analyse the genetic structure of a population in terms of its genotypic and allelic frequencies. Autosomal traits controlled by co-dominant and intermediate dominant alleles offered an opportunity to do this because all the genotypes can be distinguished.

However, many traits especially many inherited diseases result from disease-causing alleles that are recessive to wild type alleles, an important question is how common are the heterozygous carriers of the disease causing allele in a population? It is possible to know the phenotypic frequencies of healthy individuals and those affected by the disease. Can this be used to predict the frequency of heterozygous carriers? Certainly no because healthy individuals consist of two genotypic classes: homozygous dominant and heterozygous classes. In analysing the genetic structure of a population, Hardy-Weinberg principle becomes very important.

2.0 OBJECTIVES

At the end of this unit, you should be able to:

- state the Hardy-Weinberg Law
- highlight the 5 assumptions of Hardy-Weinberg equilibrium.

3.0 MAIN CONTENT

3.1 Importance of Hardy-Weinberg Principle

The key to understand the importance of Hardy-Wendy principle lies in establishing a quantitative relationship between phenotype, genotype, and allele frequencies within and between generations. The Hardy- Weinberg law**,** named for the two men – G.H. Hardy and W. Weinberg – who independently developed it in 1908, clarifies the relationships between genotype and allele frequencies within a generation and from one generation to another. To derive a general law describing the inheritance of alleles in a population, it is necessary to make certain simplifying assumptions about the nature of the population, the individuals within it, and the genes these individuals carry.

3.2 Hardy-Weinberg Assumptions

Derivation of the Hardy-Weinberg law depends on five such assumptions:

1. The population includes a very large, virtually infinite, number of individuals.
2. The individuals mate at random in the sense that each individual's genotype at the locus of interest does not influence his or her choice of a mate.
3. No new mutations appear in the gene pool.
4. There is no migration into or out of the population.

5. There are no genotype-dependent differences in the ability to survive to reproductive age and transmit genes to the next generation.

Populations that satisfy all these five assumptions are said to be at Hardy-Weinberg equilibrium.

3.3 Explanation of the Key Points in Hardy-Weinberg Assumptions

3.3.1 Random Mating

The first assumption is random mating, which means that the probability that two genotypes will mate is the product of the frequencies (or probabilities) of the genotypes in the population. If the *MM* genotype makes up 90% of a population, then any individual has a 90% chance (probability = 0.9) of mating with a person with an *MM* genotype. The probability of an *MM* by *MM* mating is (0.9)(0.9), or 0.81.

Deviations from random mating come about for two reasons: choice or circumstance. If members of a population choose individuals of a particular phenotype as mates more or less often than at random, the population is engaged in **assortative mating**. If individuals with similar phenotypes are mating more often than at random, positive assortative mating is in force; if mating occurs between individuals with dissimilar phenotypes more often than at random, negative assortative mating, or **disassortative mating**, is at work.

Deviations from random mating also arise when mating individuals are either more closely related genetically or more distantly related than individuals chosen at random from the population. **Inbreeding** is the mating of related individuals, and **outbreeding** is the mating of genetically unrelated individuals. Inbreeding is a consequence of pedigree relatedness (e.g. cousins) and small population size.

One of the first counter intuitive observations of population genetics is that deviations from random mating alter genotypic frequencies but not allelic frequencies. Envision a population in which every individual is the parent of two children. On the average, each individual will pass on one copy of each of his or her alleles. Assortative mating and inbreeding will change the zygotic (genotypic) combinations from one generation to the next, but will not change which alleles are passed into the next generation. Thus genotypic, but not allelic, frequencies change under nonrandom mating.

IN TEXT QUESTION: Compare and contrast between Inbreeding and outbreeding: Answer: **Inbreeding** is the mating of related individuals, and **outbreeding** is the mating of genetically unrelated individuals.

3.3.2 Large Population Size

Even when an extremely large number of gametes are produced in each generation, each successive generation is the result of a sampling of a relatively small portion of the gametes of the previous generation. A sample may not be an accurate representation of a population, especially if the sample is small. Thus, the second assumption of the Hardy- Weinberg equilibrium is that the population is infinitely large. A large population produces a large sample of successful gametes. The larger the sample, the greater the probability that the allelic frequencies of the offspring will accurately represent the allelic frequencies in the parental population. When populations are small or when alleles are rare, changes in allelic frequencies of the offspring will accurately represent the allelic frequencies in the parental population. When populations are small or when alleles are rare, changes in allelic frequencies take place due to chance alone. These changes are referred to as **random genetic drift**, or **genetic drift**.

3.3.3 No Mutation or Migration

Allelic and genotypic frequencies may change through the loss or addition of alleles through mutation or migration (immigration or emigration) of individuals from or into a population. The third and fourth assumptions of the Hardy-Weinberg equilibrium are that neither mutation nor migration causes such allelic loss or addition in the population.

3.3.4 No Natural Selection

The final assumption necessary to the Hardy-Weinberg equilibrium is that no individual will have a reproductive advantage over another individual because of its genotype. In other words, no natural selection is occurring (Artificial selection, as practiced by animal and plant breeders, will also perturb the Hardy-Weinberg equilibrium of captive populations).

In summary, the Hardy-Weinberg equilibrium holds (is exactly true) for an infinitely large, randomly mating population in which mutation, migration, and natural selection do not occur. In view of these assumptions, it seems that such equilibrium would never be characteristic of natural populations. However, this is not the case. Hardy-Weinberg equilibrium is approximated in natural populations for two major reasons. First, the consequences of violating some of the assumptions, such as no mutation or infinitely large population size, are small. Mutation rates, for example, are on the order of one change per locus per generation per 10^6 gametes. Thus, there is virtually no

measurable effect of mutation in a single generation. In addition, populations do not have to be infinitely large to act as if they were. As we will see later in this course, a relatively small population can still closely approximate Hardy-Weinberg equilibrium. In other words, minor deviations from the other assumptions can still result in a good fit to the equilibrium; only major deviations can be detected statistically. Second, the Hardy-Weinberg equilibrium is extremely resilient to change because, regardless of the perturbation, the equilibrium is usually reestablished after only one generation of random mating. The new equilibrium will be, however, at the new allelic frequencies – the Hardy- Weinberg equilibrium does not "return" to previous allelic values.

3.4 Hardy-Weinberg Law

3.4.1 Statement of the Law

The five key points explained above are the highlights of Hardy- Weinberg Law. The law states that: In a large random mating population with no selection, mutation, or migration, the gene frequencies and the genotype frequencies are constant from generation to generation.

Note the five key points: large, random mating, no selection, mutation and migration.

3.4.2 Hardy-Weinberg Expression

The relationship between gene frequencies and genotype frequencies is of the greatest importance because many of the many deductions about population genetics and quantitative genetics rest on it. The relationship is as follow:

Given: gene frequencies of dominant allele = p and q.

Gene frequency of recessive allele = q

Then Hardy-Weinberg genotypic frequencies = p^2, 2pq, and q^2

This is illustrated in the Table below:

	Genes in Parents		Genotypes in Progeny
	A_1 A_2		A_1A_1 A_1A_2 A_2A_2
Frequency	p q		P^2 2pq q^2

Sum alleles must add up to one, $p + q = 1$.

Thus, $(p + q)^2 = 1^2$

The binomial expansion gives: $p^2 + 2pq + q^2 = 1$

IN TEXT QUESTION
State the Hardy-Weinberg genotypic frequency
Answer: Then Hardy-Weinberg genotypic frequencies = p^2, 2pq, and q^2

4.0 CONCLUSION

In situations of complete dominance, when the heterozygotes cannot be distinguished from the homozygotes (dominant), it is impossible to determine allelic and genotypic frequencies from actual counts. In such circumstance (provided the population is in equilibrium, Hardy- Weinberg equation given by $p^2 + 2pq + q^2 = 1$ should be used.

5.0 SUMMARY

The Hardy-Weinberg Law describes what happens to allelic and genotypic frequencies of a large population, when there is random mating, no mutation, migration, or natural selection. If these conditions are met, allelic frequencies do not change from generation to generation, and the genotypic frequencies stabilize after one generation in the proportion $p^2 + 2pq + q^2$, where p and q represent the frequencies of alleles in the population.

6.0 TUTOR-MARKED ASSIGNMENT

1. List the 5 Hardy-Weinberg assumptions.
2. Are there Hardy-Weinberg populations in the real world?

7.0 REFERENCES/FURTHER READING

Klug, W.S.; Cummings, M.R.; Spencer, C.A. & Palladino, M.A. (2009). *Concepts of Genetics.* Pearson Int. Ed. (9th ed.). Pearson: Benjamin Cummings.

Peter, J. Russel (n.d). *Genetics.* Harper Collins College Publishers. Russel,

P. J. (1996). *Genetics* (7th ed.). New York: Harper Collins College Publishers.

Weaver, R. F. & Hedrick, P. H. (1997). *Genetics* (3rd ed.). WCB Publishers.

UNIT 2 APPLICATION OF HARDY-WEINBERG EQUATIONS TO REAL POPULATIONS

CONTENTS

1.0 INTRODUCTION

The conditions to attain Hardy-Weinberg equilibrium are impossible or at least very difficult to meet in real populations. All populations have finite sizes, mutations do occur, migration into and out of a population is common; and many genotypes of interest such as those that cause lethal childhood diseases, affect ability to survive and reproduce. Nevertheless even when many of the assumptions of Hardy-Weinberg do not apply, the equation derived on the basis of these assumptions is remarkably robust at providing estimates of genotype and phenotype frequencies in real populations.

2.0 OBJECTIVE

At the end of this unit, you should be able to:

- apply Hardy-Weinberg law to one locus two allelic system.

3.0 MAIN CONTENT

3.1 Application of Hardy-Weinberg Law to One Locus, Two-Allelic System

The binomial equation enables us to use information on genotype and allele frequencies to predict the genotype frequencies of the next generation. Suppose, for example, that in a population of 100,000 people carrying the recessive allele a for albinism, there are 100 *aa* albinos and 1800 *Aa* heterozygous carriers. To find what the frequency of heterozygous carries will be in the next generation, you compute the allele frequencies in the parent population:

98,100 *AA* individuals; 1800 *Aa* individuals and 100 *aa* individuals - 196,200 *A* alleles + 1800 *A* alleles; 1800 *a* alleles + 200 *a* alleles

One of 200,000 total alleles the frequency of the A allele is 198,000/200,000 =

99/100 = 0.99; thus $p = 0.99$, and the frequency of the allele is

2,000/200,000 = 1/100 = 0.01

Thus, q = 0.1

The Hardy-Weinberg equation for the albino gene in this population is:

$p^2 + 2pq + q^2 = 1$

(0.99)2 + 2 (0.99 x 0.01) + (0.01)2 = 1

0.9801 + 0.0198 0.0001 = 1

It thus predicts that in the next generation of 100,000 individuals, there will be:

100,000 x 0.9801 = 98,010 *AA* individuals,

100,000 x 0.0198 = 1980 *Aa* individuals,

100,000 x 0.00001 = 10 *aa* individuals,

This example shows that in one generation, the genotype frequencies have changed slightly. A natural question is: have the allele frequencies also changed?

Recall that the initial frequencies of the *R* and *r* alleles are *p* and *q* respectively, and that $p + q = 1$. You can use the rules for computing allele frequencies from genotype frequencies to compute the alleles' frequencies of the next generation.

From the Hardy-Weinberg equation, you know that p^2 of the individuals are *RR*, whose alleles are all *R*, and 2*pq* of the individuals are *Rr*, 1/2 of whose alleles are *R*. Similarly, q^2 of the individuals are *rr*, whose alleles are all *r*, and 2*pq* of the individuals are *Rr*, 1/2 of whose alleles are *r*. If $p + q = 1$, then $q = 1-p$ and the frequency of the *R* allele in the next generation population is:

$p^2 + 1/2\ [2p(1 – p)] = p^2 + p(1 – p)$

$= p^2 + p – p^2 = p$

Using these equations to calculate the allele frequencies of A and a in the second generation of 100,000 individuals, some of whom are albinos, we find:

- for *p*: 0.98 + 0.99 – 0.98 = 0.99 = the frequency of the *A* allele
- for *q*: 0.0001 + 0.01 – 0.0001 = the frequency of the *a* allele

These frequencies are the same as those in the previous generation. Thus, though the genotype frequencies have changed slightly from the first generation to the next, the allele frequencies have not changed. Note that this is true of both the dominant and the recessive alleles.

4.0 CONCLUSION

Mendelian segregation, when applied to population genetics has the property that random mating results in an equilibrium distribution of genotypes after only one generation.

5.0 SUMMARY

The simplest case of one locus, two allelic systems illustrates the important use of Hardy-Weinberg Principle. If a population is not in equilibrium, the principle can be used to predict the genotypic proportions in the next generation because equilibrium is attained after one generation of random breeding assuming that other forces that can disturb population equilibrium are not at play.

6.0 TUTOR-MARKED ASSIGNMENT

About 70% of all white North Americans can taste the chemical phenylthiocarbamide, and the remainder cannot. The ability to taste is determined by the dominant allele T, and the inability to taste is determined by the recessive allele t. If the population is assumed to be in Hardy-Weinberg equilibrium, what are the genotypic and allelic frequencies in this population?

7.0 REFERENCES/FURTHER READING

Klug, W.S.; Cummings, M.R.; Spencer, C.A. & Palladino, M.A. (2009). *Concepts of Genetics.* Pearson Int. Ed. (9th ed.). Pearson: Benjamin Cummings.

Peter, J. Russel (n.d). *Genetics.* Harper Collins College Publishers. Russel,

P. J. (1996). *Genetics* (7th ed.). New York: Harper Collins College Publishers.

Weaver, R. F. & Hedrick, P. H. (1997). *Genetics* (3rd ed.). WCB Publishers.

UNIT 3 APPLICATION OF HARDY-WEINBERG PRINCIPLE TO ONE LOCUS MULTIPLE ALLELIC SYSTEM

CONTENTS

1.0 INTRODUCTION

It is helpful to review application of Hardy-Weinberg principle to the simple two-allelic system before attempting to know the application of the principle to multiple allelic systems. This is because application to multiple allelic systems is an extension of the knowledge of the simple two-allelic system.

2.0 OBJECTIVE

At the end of this unit, you should be able to:

- determine allelic and genotypic frequencies in a multiple allelic system.

3.0 MAIN CONTENT

3.1 Multiple Allelic System

The Hardy-Weinberg principle can also be extended to more than two alleles. Recall that multiple allelic system is when there are more than two alleles that can occupy one locus. Let us illustrate this with human ABO blood group system, which has three alleles: I^A, I^B, I^O. Table 1 gives the frequencies of the resulting genotypes and phenotypes. As for two alleles, the genotypic frequencies can be calculated as the product of particular male and female gametic frequencies. For example, the frequency of genotype I^AI^A is $(p)(p) = p^2$. Overall, the genotypic frequencies are equal to the square of the following trinomial (because there are three alleles):

$$(p + q + r)^2 = p^2 + 2pq + q^2 + 2pr + 2qr + r^2$$

Check table 1 to see that you understand the genotypes associated with these frequencies.
Note that because of the recessitivity of allele I^O, two genotypes have the A phenotype (I^AI^A and I^AI^O) and two have the B phenotype (I^BI^B and I^BI^O). The frequency of the O phenotype, using the Hardy-Weinberg principle, is the square of the frequency of the recessive allele I^O, that is r^2; the frequencies of the A and B phenotypes are $p^2 + 2pr$ and $q^2 + 2qr$, respectively; and the frequency of the AB phenotype is 2pq. In the United States Caucasian population, the frequencies of alleles I^A, I^B and I^O are approximately 0.28, 0.06 and 0.66, resulting in phenotypic frequencies of 0.45, 0.08, 0.03, and 0.44 for ABO blood group types A, B, AB, and O, respectively. Luckily, the rarest phenotype, type AB, is also the universal recipient, so people with this phenotype can receive transfusions from individuals of any ABO blood group type.

Table 1: The Hardy-Weinberg Principle Applied to the ABO Blood Group System

Female Gametes (Frequencies)	**Male**	**Gametes**	**(Frequencies)**
	$I^A(p)$	$I^B(q)$	$I^O(r)$
$I^A(p)$	$I^A\ I^A\ (p)^2$ A	$I^A\ I^B(pq)$ AB	$I^A\ I^O(pr)$ A
$I^B(q)$	$I^A\ I^B\ (pq)$ AB	$I^B\ I^B(q^2)$ B	$I^B\ I^O\ (qr)$ B
$I^O(r)$	$I^A\ I^O(pr)$ A	$I^B\ I^O(qr)$ B	$I^O\ I^O(pr)$ O

An important consequence of having many alleles at a locus is that, in general, most individuals in a population are heterozygotes. For example, at the *HLA-A* and *HLA-B* loci, over 90% of the individuals are heterozygotes and less than 10% are homozygotes. To illustrate this calculation, the Hardy-Weinberg heterozygosity for the ABO system is:

$$H = 2pq + 2pr + 2qr$$

$$= 2(0.28)(0.06) + 2(0.28)(0.66) + 2(0.06)(0.66)$$

$$= 0.482.$$

IN TEXT QUESTION:
An important consequence of having many alleles at a locus is that, in general, most individuals in a population are _____.
Answer: Heterozygous

4.0 CONCLUSION

A good understanding of the two allelic systems is a prerequisite to understanding multiple allelic systems. It is helpful to understand and practice problem solving in two allelic systems very well because multiple allelic cases can simply be described as extensions of two allelic systems.

5.0 SUMMARY

When there are many alleles (which means many more heterozygous than homozygotes), the simplest approach is to calculate the Hardy- Weinberg homozygosity and subtract it from unity. The remainder is the Hardy-Weinberg heterozygosity.

6.0 TUTOR-MARKED ASSIGNMENT

1. In a large natural population a plant species, there are three alleles at one locus. The alleles are A^W, A^I and A^F. The following results were obtained in a population study.

   A^SA^S = 0.04
   A^IA^I = 0.09
   A^FA^F = 0.25
   A^SA^I = 0.12
   A^SA^F = 0.20
   A^IA^F = 0.30

2. What are the allelic frequencies in the population?

7.0 REFERENCES/FURTHER READING

Klug, W.S.; Cummings, M.R.; Spencer, C.A. & Palladino, M.A. (2009). *Concepts of Genetics.* Pearson Int. Ed. (9th ed.). Pearson: Benjamin Cummings.

Peter, J. Russel (n.d). *Genetics.* Harper Collins College Publishers.

Russel, P. J. (1996). *Genetics* (7th ed.). New York: Harper Collins College Publishers.

Weaver, R. F. & Hedrick, P. H. (1997). *Genetics* (3rd ed.). WCB Publishers.

UNIT 4 SEX-LINKED TRAITS

CONTENTS

1.0 INTRODUCTION

Recall that Sex-linked or X-linked traits are traits controlled by genes located on the X chromosomes. Well known examples in human are colourblindness and haemophilia. Thus, males are hemizygous for sex-linked traits because they inherit only one X chromosome from their mothers unlike females that inherit 2 X chromosomes: one from each parent. Thus, population genetic analysis of X-linked traits is slightly different as compared to autosomal traits.

IN TEXT QUESTION

Sex-linked or X-linked traits are traits controlled by genes located on the _____ Answer: X chromosomes

2.0 OBJECTIVE

At the end of this unit, you should be able to:

- determine genotypic and allelic frequencies for sex-linked traits using Hardy-Weinberg principle.

3.0 MAIN CONTENT

3.1 Application of Hardy-Weinberg Principle to X-linked Traits

For X-linked alleles in females, the Hardy-Weinberg frequencies are the same as those for autosomal loci:

p^2 (X^AX^A), $2pq$ (X^AX^B), and q^2 (X^BX^B).

In males however, the frequencies of the genotypes will be p(X^AY) and q (X^BY), the same as the frequencies of the alleles in the population. For this reason, recessive X-linked traits are more frequent among males than females. To illustrate this concept, consider red-green colour blindness, which is an X-linked recessive trait. The frequency of the

colourblind allele varies among human ethnic groups; the frequency among African Americans is 0.039. At equilibrium, the expected frequency of colourblind males in this group is q = 0.039, but the frequency of colourblind females is only q^2

$= (0.039)^2 = 0.0015.$

When random mating occurs within a population, the equilibrium genotypic frequencies are reached in one generation. However, if the alleles are X-linked and the sexes differ in allelic frequency, the equilibrium frequencies are approached over several generations. This is because males receive their X-chromosome from their mother only, while females receive an X chromosome from both the mother and the father.

Consequently, the frequency of an X-linked allele in males will be the same as the frequency of that allele in their mothers, whereas the frequency in females will be the average of that in mothers and fathers. With random mating, the allelic frequencies in the two sexes oscillate back and forth in each generation, and the difference in allelic frequency between the sexes is reduced by half each generation. Once the allelic frequencies of the males and females are equal, the frequencies of the genotypes will be in Hardy-Weinberg proportions after one more generation of random mating.

4.0 CONCLUSION

Recessive X-linked traits are more frequent among males than among females because the frequency of an X-linked recessive trait in male is the same as the allelic frequency while it is equal to the Weinberg frequency in the female.

5.0 SUMMARY

Analysis of X-linked traits in population genetics is different for males and females because of the different chromosomal constitutions of the two sexes. In females, the Hardy-Weinberg frequencies are the same as those for autosomal loci: p^2 (X^AX^A), $2pq$ (X^AX^B), and q^2 (X^BX^B). In males however, the frequencies of the genotypes will be $p(X^AY)$ and q (X^BY), the same as the frequencies of the alleles in the population.

6.0 TUTOR-MARKED ASSIGNMENT

About 7% of men are colour-blind in consequence of a sex-linked recessive gene. Assuming Hardy-Weinberg equilibrium, what proportions of women are expected to be (i) carriers, and (ii)

colourblind? (iii) In what proportion of marriages are both husband and wife expected to be colourblind?

7.0 REFERENCES/FURTHER READING

Klug, W. S.; Cummings, M.R.; Spencer, C.A. & Palladino, M.A. (2009). *Concepts of Genetics*. Pearson Int. Ed. (9th ed.). Pearson Benjamin Cummings.

Russel, P. J. (1996). *Genetics* (7th ed.). New York: Harper Collins College Publishers.

Weaver, R. F. & Hedrick, P. H. (1997). *Genetics* (3rd ed.). WCB Publishers.

MODULE 3

UNIT 1 TESTING POPULATIONS FOR HARDY-WEINBERG EQUILIBRIUM

CONTENTS

1.0 INTRODUCTION

In the last Module, Unit 1 Sections 4.0 and 4.1, the conditions for conformity and deviation from Hardy-Weinberg Equilibrium were explained. Due to sampling errors, it is not likely that the observed numbers of different genotypes will be exactly the same as what is expected even if the population is in Hardy-Weinberg Equilibrium. The question to ask is how much deviation from the expected numbers should be considered significant to reject the hypothesis (null hypothesis) that the population is not likely to be in Hardy-Weinberg Equilibrium?

2.0 OBJECTIVES

At the end of this unit, you should be able to:

- distinguished the differences between the observed population count is statistically difference from Hardy-Weinberg expectation
- use chi-square test for analysis.

3.0 MAIN CONTENT

3.1 Testing Whether a Population is in Hardy-Weinberg Equilibrium

Testing for Hardy-Weinberg Proportions

The Hardy-Weinberg law can be used as a null hypothesis to which the genetic structure of any particular population can be compared. If the observed genetic structure does not match the expected structure based on the law, one can begin to ask about which of the assumptions are being violated. To determine whether the genotypes of a population are in Hardy-Weinberg proportions, we first compute p and q from the observed frequencies of the genotypes. Note: it is important not to take the square roots of the homozygote frequencies to obtain allele frequencies because in doing that, you already assume that the population is in equilibrium. Thus, allele frequencies should be calculated from the sum of homozygote frequencies plus one half the heterozygote frequencies – for a more complete discussion. Once we have obtained these allelic frequencies, we can calculate the expected genotypic frequencies (p^2, $2pq$, and q^2) and compare these frequencies with the actual observed frequencies of the genotypes using a **chi- square test**. The chi-square test gives us the probability that the difference between what we observed and what we expect under the Hardy-Weinberg law is due to chance.

To illustrate this procedure, consider a locus that codes for transferring (a blood protein) in the red backed vole, *Clethrionomys gapperi.* Three genotypes are found at the transferring locus: *MM, MJ,* and *JJ.* In a population of voles trapped in the Northwest Territories of Canada in 1976, 12 *MM* individuals, 53 *MJ* individuals, and 12 *JJ* individuals were found. To determine if the genotypes are in Hardy-Weinberg proportions, we first calculate the allelic frequencies for the population using our familiar formular:

$$P = \frac{(2 \times \text{number of homozygotes} + (\text{number of heterozygotes})}{(2 \times \text{total number of individuals})}$$

Therefore:

$$P = f(M) = \frac{(2 \times 12) + (53)}{(2 \times 77)} = 0.50$$

$$q = 1 - p = 0.50$$

Using p and q calculated from the observed genotypes, we can now compute the expected Hardy-Weinberg proportions for the genotypes:

f(MM) = p^2=$(0.50)^2$ = 0.25; *f(M)*= 2*pq* = 2(0.50(0.50) = 0.50; and *f(JJ*) = q^2 = $(0.50)^2$ =0.25. However, for the chi-square test, actual numbers of individuals are needed, not the proportions. To obtain the expected numbers, we simply multiply each expected proportion times the total number of individuals counted (*N*), as show below:

		EXPECTED	**OBSERVED**
f(MM)	= p^2 x N		
	= 0.25 x 77 =	19.3	12
f(M)	= 2*pq* x N		
	= 0.50 x 77 =	38.5	53
f(JJ)	=q^2 x N		
	= 0.25 x 77 =	19.3	12

With observed and expected numbers, we can compute a chi-square value to determine the probability that the differences between observed and expected numbers could be the result of chance. The chi-square is computed using the same formula that we employed for analysing genetic crosses; that is, *d*, the deviation, is calculated for each class as (observed – expected); d^2, the deviation squared, is divided by the expected number e for each class; and chi-square (X^2) is computed as the sum of all d^2/e values.

For the above example, X^2 = 10.98, we need to find this value in the chi-square table under the appropriate degrees of freedom. This step is not as straightforward as in our previous X^2 analysis. In those examples, the number of degrees of freedom was the number of classes in the sample minus 1. Here, however, while there are three classes, there is only one degree of freedom, because the frequencies of alleles in a population have no theoretically expected values. Thus, *p* must be estimated from the observations themselves. So one degree of freedom is lost for every parameter (p in this case) that must be calculated from the data. Another degree of freedom is lost because, for the fixed number of individuals, once all but one of the classes has been determined, the last class has no degree of freedom and is set automatically. Therefore with three classes (*MM*, *MJ*, and *JJ*), two degrees of freedom are lost, leaving one degree of freedom.

In the chi-square table under the column for one degree of freedom, the chi-square value of 10.98 indicates a *p* value less than 0.05. Thus, the probability that the differences between the observed and expected values are due to chance is very low. That is, the observed numbers of genotypes do not fit the expected numbers under Hardy-Weinberg law.

4.0 CONCLUSION

A careful consideration of Hardy-Weinberg conditions will reveal that it is impossible for real populations to satisfy these conditions. However to a large extent, some real populations approach Hardy-Weinberg proportions. Testing of populations for Hardy-Weinberg proportions is important in population genetics.

5.0 SUMMARY

Chi-square test is a statistical tool used to test a population for Hardy-Weinberg equilibrium. To perform the test, actual numbers of individuals are needed, not the proportions. To obtain the expected numbers, we use the formula below:

$$X^2 = \sum(o-e)^2/e$$

6.0 TUTOR-MARKED ASSIGNMENT

Consider the problem in Unit 3 of Module 2. Test whether the population is in Hardy-Weinberg equilibrium.

7.0 REFERENCES/FURTHER READING

Klug, W.S.; Cummings, M.R.; Spencer, C.A. & Palladino, M.A. (2009). *Concepts of Genetics.* Pearson Int. Ed. (9th ed.). Pearson: Benjamin Cummings.

Russel, P. J. (1996). *Genetics* (7th ed.). New York: Harper Collins College Publishers.

Weaver, R. F. & Hedrick, P. H. (1997). *Genetics* (3rd ed.). WCB Publishers.

UNIT 2 EVOLUTIONARY CHANGES IN GENETIC STRUCTURE OF POPULATIONS

CONTENTS

1.0 INTRODUCTION

When the population is large, randomly mating, and free from mutation, migration, and natural selection, no evolution occurs. The genetic structure of the population is in equilibrium. For many populations, however, the conditions required by the Hardy-Weinberg law do not hold. Populations are frequently small, mating may be nonrandom, and mutation, migration, and natural selection may be occurring. In these circumstances, allelic frequencies do change, and the gene pool of the population evolves in response to the interplay of these processes. In the following sections we discuss the role of four evolutionary processes – mutation, genetic drift, migration, and natural selection – in changing the allele frequencies of a population. We also discuss the effects of nonrandom mating on genotype frequencies. We first consider violations of the Hardy-Weinberg equilibrium assumptions one by one. We then consider several cases where two assumptions are violated simultaneously. Be aware, however, that in real populations the entire assumption one by one.

2.0 OBJECTIVE

At the end of this unit, you should be able to:

- discuss the term mutation.

3.0 MAIN CONTENT

3.1 Mutation

One process that can potentially alter the frequencies of alleles within a population is mutation. Usually, a mutation converts one allelic form of

a gene to another. The rate at which mutations arise is generally low, but varies among loci and among species.

Ultimately, mutation is the source of all new genetic variation; new combinations of alleles may arise through recombination, but new alleles only occur as a result of mutation. Thus, mutation provides the raw genetic material upon which evolution acts. Most mutations will be detrimental and will be eliminated from the population. A few mutations, however, will convey some advantages to the individuals that possess them and will spread through the population. Whether a mutation is detrimental or advantageous depends upon the specific environment, and if the environment changes, previously harmful or neutral mutations may become beneficial.

Mutations also have the potential to affect evolution by contributing to the second step in the evolutionary process – by changing the frequency of alleles within a population.

Consider a population that consists of 50 individuals, all with the genotype *AA*. The frequency of *A(p)* is 1.00 [(2 x 50)/100]. If one *A* allele mutates to *a*, the population now consists of 49 *AA* individuals and 1 *Aa* individual. The frequency of *p* is now [(2 x 49) + 1] = 0.99. When another mutation occurs, the frequency of *A* drops to 0.98. If these mutations continue to occur at low but steady rate over long periods of time, the frequency of *A* will eventually decline to zero, and the frequency of a will reach a value of 1.00.

The mutation of *A* to *a* is referred to as a forward mutation. For most genes, mutations also occur in the reverse direction; a may mutate to *A*. These mutations are called reverse mutations, and they typically occur at a lower rate than forward mutations. The forward mutation rate – the rate at which A mutates to a (*A* → *a*) – symbolised with *u*; the reverse mutation rate – the rate at which a mutates to A (a → A) – is symbolised with *v*. Consider a hypothetical population in which the frequency of *A* is *p* and the frequency of *a* is *q*. We assume that the population is large and that no selection occurs on the alleles. In each generation, *a* is proportion *u*, the actual number mutating depends upon both u and the frequency of *A* alleles.

Given that we have a forward mutation rate increasing the frequency of *a* and a reverse mutation rate decreasing the frequency of *a*, it is intuitively easy to see that eventually, the population achieves equilibrium, in which the number of alleles undergoing forward mutation is exactly equal to the number of alleles undergoing reverse mutation.

The equilibrium frequency for *a* is:

$$q = \frac{u}{u + v}$$

and equilibrium value for p is

$$p = u + v \frac{v}{}$$

Consider a population in which the initial allelic frequencies are $p = 0.9$ and $q = 0.1$, and the forward and reverse mutation rates are $u = 5 \times 10^{-5}$ and $v = 2 \times 10^{-5}$ respectively, (these values are similar to forward and reverse mutation rates observed for many genes). In the first generation, the change in allelic frequency is:

$$\Delta p = vp - up$$
$$= (2 \text{ x } 10\text{-}5 \text{ x } 0.1) - (5 \text{ x } 10\text{-}5 \text{ x } 0.9)$$
$$Ap = -0.000043$$

The frequency of *A* decreases by only four-thousandths of 1%. A equilibrium, the frequency of the *a* allele, *q*, equals:

$$q = u + v \frac{u}{}$$

$$q = (5 \text{ x } 10\text{-}5) + (2 \text{ x } 10\text{-}5) \frac{5 \times 10^{-5}}{} = 0.714$$

If no other processes act on a population after many generations, the alleles will reach equilibrium. Therefore, mutation rates determine the allelic frequencies of the population in the absence of other evolutionary processes. However, because mutation rates are so low, the change in allelic frequency due to mutation pressure is exceedingly slow.

4.0 CONCLUSION

Mutation is the source of genetic variations in populations. However, it can cause only a small change in allelic frequency per generation.

5.0 SUMMARY

Mutation can proceed in the forward or backward direction. The rate of forward mutation is u while that of backward mutation is v. At equilibrium, the frequencies of 'A' and 'a' alleles are given by p = v/(u+v) and q = u/(u+v) respectively.

6.0 TUTOR-MARKED ASSIGNMENT

If an allele, A, mutates to a with a frequency of 1 in 10,000 and back-mutates with a frequency of 1 in 100,000, and if the three genotypes have equal fitness, what will be the genotypic frequencies at equilibrium in a random-mating populations?

7.0 REFERENCES/FURTHER READING

Klug, W.S.; Cummings, M.R.; Spencer, C.A. & Palladino, M.A. (2009). *Concepts of Genetics.* Pearson Int. Ed. (9th ed.). Pearson: Benjamin Cummings.

Russel, P. J. (1996). *Genetics* (7th ed.). New York: Harper Collins College Publishers.

Weaver, R. F. & Hedrick, P. H. (1997). *Genetics* (3rd ed.). WCB Publishers.

UNIT 3 GENETIC DRIFT

CONTENTS

1.0 INTRODUCTION

Another major assumption of the Hardy-Weinberg law is that the population is infinitely large. Real populations are not infinite in size, but frequently they are large enough that expected ratios are realised and chance factors have small effects on allelic frequencies. Some populations are small, however, and in these groups, chance factors may produce large changes in allelic frequencies. Random change in allelic frequency due to chance is called **genetic drift**, or simply ***drift*** for short. Sewall Wright which laid much of the theoretical foundation of the discipline, championed the importance of genetic drift in the 1930s, so sometimes genetic drift is called the *Sewall Wright effect* in his honour.

To see how chance can play a big role in altering the genetic structure of a population, imagine a small group of humans inhabiting a South Pacific Island. Suppose that the population consists of only ten individuals, five of whom have green eyes and five have brown eyes. For this example, we assume that eye colour is determined by a single locus (actually a number of genes controls eye colour) and the allele for green eyes is recessive to brown (*BB* and *Bb* coders for brown eyes and *bb* codes for green). The frequency of the allele for green eyes is 0.6 in the Island population. A typhoon strikes the island, killing 50% of the population; five of the inhabitants perish in the storm. Those five individuals who die all have brown eyes. Eye colour in no way affects the probability of surviving the fact that only those with green eyes survive is strictly the result of chance. After the typhoon, the allelic frequency for green eyes is 1.0. Evolution has occurred in this

population – the frequency of the green-eye allele has changed from 0.6 to 1.0, simply as a result of chance.

Now, imagine the same scenario, but this time with a population of 1,000 individuals. As before, 50% of the population has green eyes and 50% has brown eyes. A typhoon strikes the island and kills half the population. How likely is it that, just by chance, all 500 people who perish will have brown eyes? In a population of 1,000 individuals, the probability of this occurring by chance is extremely remote. This example illustrates an important characteristic of genetic drift – chance factors are likely to produce significant changes in allelic frequencies only in small populations.

Chance deviations from expected proportions arise from a general phenomenon called 'sampling error. Just by chance or by "**error**" sample may deviate from the larger pool; the smaller the sample, the larger the potential deviation.

Flipping a coin is analogous to the situation in which sampling error occurs. When two flip a coin, we expect 50% heads and 50% tails. If we flip the coin 1,000 times, we will get very close to that expected fifty-fifty ratio. But, if we flip the coin only four times, we would not be surprised if by chance we obtain 3 heads and 1 tail, or even all tails. When the sample – in this case the number of flips – is small, the sampling error can be large. All genetic drift arises from such sampling error.

2.0 OBJECTIVES

At the end of this unit, you should be able to:

- define genetic drift
- measure genetic drift in a population
- discuss variation due to genetic drift
- state the forms of genetic drift.

3.0 MAIN CONTENT

3.1 Measuring Genetic Drift

Genetic drift is random, and thus we cannot predict what the allelic frequencies will be after drift has occurred. However, since sampling error is related to the size of the population, we can make predictions about the magnitude of genetic drift. To determine the magnitude of genetic drift, we must know the effective population size, which equals

the equivalent number of adults contributing gametes to the next generation, the effective population size is:

$$Ne = (4 \times Nf \times Nm)/Nf + Nm$$

Where N_f equals the number of breeding females and N_m, equals the number of breeding males.

Students often have difficulty understanding why this equation must be used – why the effective population size is not simply the number of breeding adults. The reason is that males, as a group, contribute half of all genes to the next generation and females, as a group, contribute the other half. Therefore, in a population of 70 females and 2 males, the two males are not genetically equivalent to two females; each male contributes 1/2 x 1/2 = 0.25 of the genes to the next generation, whereas each female contributes 1/2 x 1/70 = 0.007 of all genes. The small number of males disproportionately influences what alleles are present in the next generation. Using the above equation, the effective population size is *Ne* = (4 x 70 x 2) (70 + 2) = 7.8, or approximately 8 breeding adults. What this means is that in a population of 70 females and 2 males, genetic drift will occur as if the population had only four breeding males and four breeding females. Therefore, genetic drift will have a much greater effect in this population than in one with 72 breeding adults equally divided between males and females.

Other factors, such as differential production of offspring, fluctuating population size, and overlapping generations can further reduce the effective population size.

3.1.1 The Dunker's in Eastern Pennsylvania

Small breeding units that lack gene flow are genetically isolated from other groups and often experience considerable genetic drift, even though surrounded by much larger populations. A good example is a religious sect, known as the **Dunkers,** found in eastern Pennsylvania. Between 1719 and 1729, fifty Dunker families emigrated from Germany and settled in the United States. Since that time, the Dunkers have remained an isolated group, rarely marrying outside of the sect, and the number of individuals in their communities has always been relatively small.

The Pennsylvania frequencies were different from the frequencies of the German population from which the Dunkers descended. The most likely explanation for the unique Dunker allelic frequencies observed is that genetic drift has produced random change in the gene pool.

3.2 Variation Due to Genetic Drift

The amount of variation among populations resulting from genetic drift is measured by the **variance of allelic frequency,** which equals:

$$Sp_2 = \frac{pq}{2N_e}$$

Where Ne equals the effective population size, p and q equal the allelic frequencies. A more useful measure is the **standard error** of allelic frequency, which is the square root of the variance of allelic frequency:

$$S_p = \sqrt{\frac{pq}{2N_e}}$$

The standard error can be used to calculate the 95% confidence limits of allelic frequency, which indicate the expected range of p in 95% of such populations. The 95% confidence limits equal approximately $p + 2s_p$. Suppose, for example, that p = 0.8 in a population with N_e equal to 50. The standard error in allelic frequency is sp = (p_q/N_e) = 0.04. The 95% confidence limits for p are therefore $p + 2s_p = 0.72 \leq p \leq 0.88$.

To interpret the 95% confidence limits, imagine that 100 populations with N_e of 50 have p initially equal to 0.8. Genetic drift may cause allelic frequencies in some populations to change; the 95% confidence limits tell us that in the next generation, 95 of the original 100 populations should have *p* within the range of 0.72 to 0.88. Therefore, if we observe a change in p greater than this, say from 0.8 to 0. 86, we know that the probability that this change will occur by genetic drift is less than 0.05. Most likely we would conclude that some process other than genetic drift contributed to the observed change in allelic frequency.

3.3 Forms of Genetic Drift

3.3.1 Small Populations

Genetic drift arises when population size remains continuously small over many generations. Undoubtedly, this situation is frequent, particularly where populations occupy marginal habitats, or when competition for resources limits population growth. In such populations, genetic drift plays an important role in the evolution of allelic frequencies. Many species are spread out over a large geographic range. This can result in a species consisting of numerous populations of small size, each undergoing drift independently. In addition, human intervention such as the clear cutting of forests can result in the

fragmentation of previously large continuous populations into small subdivided ones with each showing genetic drift.

3.3.2 Founder Effect

Another way in which genetic drift arises is through **founder effect.** Founder effect occurs when a population is initially established by a small number of breeding individuals. Although the population may subsequently grow in size and later consist of a large number of individuals, the gene pool of the population is derived from the genes present in the original founders. Chance may play a significant role in determining which genes were present among the founders, and this has a profound effect upon the gene pool of subsequent generations. Founder effects have frequently been used to explain the subsequent evolution of new species, but their importance to the process of species formation is currently under intense study and debate.

3.3.3 Bottleneck Effect

A third form of sampling error, called bottleneck effect, also played an important role in the population of Tristan da Cunha. Bottleneck effect is a form of genetic drift that occurs when a population is drastically reduced in size. During such a population reduction, some genes may be lost from the gene pool as a result of chance.

You should note that bottleneck effect does not involve migration, but can occur in form of disaster, war or other forms of accident.

The gene pool of Tristan da Cunha has been influenced by genetic drift in the form of founder effect, small population size, and bottleneck effect.

Consider the inhabitants of Tristan da Cunha, a small, isolated island in the South Atlantic. This island was first permanently settled by William Glass, a Scotsman, and his family in 1817. They were joined by a few additional settlers, some shipwrecked sailors, and a few women from the distant island of St. Helena, but for the most part the island remained a genetic isolate. In 1961, a volcano on Tristan da Cunha erupted, and the population of almost 300 inhabitants was evacuated to England.

4.0 CONCLUSION

Genetic drift in small populations can result in large chance changes in allelic frequency. Thus, it may be the cause of the low heterozygosity observed in some endangered species.

5.0 SUMMARY

As a result of small population size, chance effects called genetic drift can change allelic frequency in a population and therefore cause evolution. It operates in populations by virtue of their small sizes that cause sampling errors or/and through bottleneck and founder effects.

6.0 TUTOR-MARKED ASSIGNMENT

It is known that genetic drift that changes allelic frequencies affects small populations than large populations. Which of the populations A and B below is likely to be more affected by genetic drift?

	A		B	
Population	Males	Females	Males	Females
	80	120	50	150
Total	**200**		**300**	

7.0 REFERENCES/FURTHER READING

Klug, W.S.; Cummings, M.R.; Spencer, C.A. & Palladino, M.A. (2009). *Concepts of Genetics.* Pearson Int. Ed. (9th ed.). Pearson: Benjamin Cummings.

Russel, P. J. (1996). *Genetics* (7th ed.). New York: Harper Collins College Publishers.

Weaver, R. F. & Hedrick, P. H. (1997). *Genetics* (3rd ed.). WCB Publishers.

UNIT 4 EFFECT OF MIGRATION OR GENE FLOW ON EVOLUTIONARY CHANGES

CONTENTS

1.0 INTRODUCTION

One of the assumptions of the Hardy-Weinberg law is that the population is closed and not influenced by other populations. Many populations are not completely isolated, however, and exchange genes with other populations of the same species. Individuals migrating into a population may introduce new alleles to the gene pool and alter the frequencies of existing alleles. Thus migration has the potential to disrupt Hardy-Weinberg equilibrium and may influence the evolution of allelic frequencies within populations. The term migration usually implies movement of organisms. In population genetics, however, we are interested in the movement of genes. Movement of genes takes place only when organisms or gametes migrate, and contribute their genes to the gene pool of the recipient population. This process is also referred to as **gene flow**.

In text Question.

Movement of genes takes place when organisms _____Answer. Migrate

Gene flow has two major effects on a population. First, it introduces new alleles to the population. Second, when the allelic frequencies of migrants and the recipient population differ, gene flow changes the allelic frequencies within the recipient population. Through exchange of genes, different populations remain similar, and thus, migration is a homogenising force that tends to prevent populations from accumulating genetic differences among them.

2.0 OBJECTIVES

At the end of this unit, you should be able to:

- define gene flow
- discuss how gene flow between populations can introduce new alleles and, in some cases, can rapidly change allelic frequencies.

3.0 MAIN CONTENT

3.1 Evolutionary Effect of Migration or Gene Flow

To illustrate the effect of migration on allelic frequencies, we will consider a simple model in which gene flow occurs in only one direction, from population I to population II. Suppose that the frequency of allele *A* in population **I** (p_I) is 0.8 and the frequency of *A* in population **II** (p_{II}) is 0.5. In every generation, some individuals migrate from population 1 to population II, and these migrants are a random sample of the genotypes in population I. After migration, population II actually consists of two groups of individuals: the migrants with pi = 0.8, and the residents with *pii* = 0.5. The migrants now make up a proportion of population II, which we designate m. The frequency of *A* in population II after migration (p'_{II}) is:

$$p'_{II} = mp_I + (1-m)p_{II}$$

We see that the frequency of *A* after migration is determined by the proportion of A alleles in the two groups that now comprise population II. The first component, mp_I, represents the *A* alleles in the migrants – we multiply the proportion of the population that consists of migrants (m) by the allelic frequency of the migrants (p). The second component represents the *A* alleles in the residents, and equals the proportion of the population consisting of residents (*1-m*) multiplied by the allelic frequency in the residents (p_{II}). Adding these two components together gives us the allelic frequency of *A* in population II after migration.

The change in allelic frequency in population II as a result of migration:

$$(\Delta p) = p'_{II} - P_{II}$$

In the previous equation, we found that p'_{II} equaled $mp_1 + (1-m)p_{II}$, so the change in allelic frequency can be written as:

$$Ap = mp_I + (1-m)p_{II} - p_{II}$$

Multiplying (1 – m) by p_{II} in the above equation, we obtain:

$$(\Delta p) = mp_I + p_{II} - mp_{II} - p_{II}$$

$(\Delta p) = mp_I - p_{II}$, therefore by factorisation, the final equation is:

$$(\Delta p) = m(p_I - p_{II})$$

This final equation indicates that the change in allelic frequency from migration depends upon two factors; the proportion of the migrants in the final population and the difference in allelic frequency between the migrants and the residents. If no differences exist in the allelic frequency of migrants and residents ($p_I - p_{II} = 0$), then we can see that the change in allelic frequency is zero. Populations must differ in their allelic frequencies in order for migration to affect the makeup of the gene pool. With continued migration, p_I and p_{II} become increasingly similar, and, as a result, the change in allelic frequency due to migration decreases. Eventually, allelic frequencies in the two populations will be equal, and no further change will occur.

An additional point to be noted is that migration among populations tends to increase the effective size of the populations. Genetic drift causes populations to diverge. Migration, on the other hand, reduces divergence among populations, effectively increasing the size of the individual populations.

4.0 CONCLUSION

Gene flow between populations can introduce new alleles and can cause changes in allelic structure leading to evolutionary changes in populations.

5.0 SUMMARY

Migration between populations is known as gene flow. Its effect on allelic frequency is given by $(\Delta p) = m(p_I + p_{II})$.

6.0 TUTOR-MARKED ASSIGNMENT

1. The gene frequency for gene ‘a’ causing albinism (an autosomal recessive trait) in the two villages (A & B) is 0.1 and 0.4 respectively. The news of an impending famine caused a representative sample of 50 migrants out of the total 1000 inhabitants of village A to leave Village A for B.
2. Determine the new gene frequency in the recipient village after one generation of migration (Note: ‘m’ in the formula represents the proportion and not the number of migrants).
3. What is the change in allelic frequency due to migration?
4. Hence, explain whether the recipient population has undergone evolutionary change as a result of one generation of migration.

7.0 REFERENCES/FURTHER READING

Klug, W. S.; Cummings, M.R.; Spencer, C.A. & Palladino, M.A. (2009). *Concepts of Genetics*. Pearson Int. Ed. (9th ed.). Pearson Benjamin Cummings.

Russel, P. J. (1996). *Genetics* (7th ed.). New York: Harper Collins College Publishers.

Weaver, R. F. & Hedrick, P. H. (1997). *Genetics* (3rd ed.). WCB Publishers.

UNIT 5 EFFECT OF SELECTIVE FORCES ON EVOLUTIONARY CHANGES

CONTENTS

1.0 INTRODUCTION

In the previous units in this module, we have examined three major evolutionary processes capable of changing allelic frequencies and producing evolution – mutation, genetic drift, and migration. These processes alter the gene pool of a population, and they certainly influence the evolution of a species. However, mutation, migration and genetic drift do not result in adaptation.

2.0 OBJECTIVES

At the end of this unit, you should be able to:

- highlight the effect of natural selection on allelic frequencies
- enumerate the effect of natural selection effect on evolutionary changes.

3.0 MAIN CONTENT

3.1 Adaptation

Adaptation is the process by which traits evolve that make organisms more suited to their immediate environment; these traits increase the organism's chances of surviving and reproducing. Adaptation is responsible for the many extraordinary traits seen in nature – wings that enable a hummingbird to fly backward, leaves of the pitcher plant that capture and devour insects, brains that allow humans to speak, read, and

love. These biological features and countless other exquisite traits are the product of adaptation. Genetic drift, mutation, and migration all influence the pattern and process of adaptation, but adaptation arises chiefly from natural selection.

3.2 Natural Selection

In Text Question

Natural selection is a _____ force Answer: Dominant

Natural selection is the dominant force in the evolution of many traits and has shaped much of the phenotypic variation observed in nature. Natural selection can be defined as differential reproduction of genotypes. It simply means that individuals with certain genes produce more offspring than others; therefore those genes increase in frequency in the next generation. Through natural selection, traits that contribute to survival and reproduction increase over time. In this way organisms adapt to their environment.

3.3 Fitness and Coefficient of Selection

Darwinian fitness is defined as the relative reproductive ability of a genotype. An adaptive value of 1 is assigned to a genotype that produces the most offspring. The fitness of the other genotypes is assigned relative to this. For example, suppose that the genotype G^1G^1 on the average produces 8 offspring, G^1G^2 produces an average of 4 offspring, and G^2G^2 produces an average of 2 offspring. The G^1G^1 genotype has the highest reproductive output, so its fitness is 1 ($W_{II} = 1.0$). Genotype G^1G^2 produces on the average 4 offspring for the 8 produced by the most fit genotype, so the fitness of G^1G^2 (W_{12}) is 4/8 =
0.5. Similarly, G^2G^2 produces 2 offspring for the 8 produced by G^1G^1, so the fitness of G^2G^2 (W_{22}) is 2/8 = 0.25.

A related measure is the selection coefficient, which is a measure of the relative intensity of selection against a genotype. The selection coefficient is symbolized by s and equals 1-W. In our example, the selection coefficients for G^1G^1 are $s = 0$; for G^1G^2, $s = 0.5$; for G^2G^2, $s = 0.75$.

The change in allelic frequency that results from natural selection can be calculated using the "table method". Begin by listing the genotypes (A^1A1, A^1A^2, and A^2A^2) and their initial frequencies. If random mating has just taken place, the genotypes are in Hardy-Weinberg proportions and the initial frequencies are p^2, $2pq$, and q^2. We then list the fitness for each of the genotypes, W_{11}, W_{12}, and W_{22}. Now, suppose that selection occurs and only some of the genotypes survive. The contribution of each genotype to the next generation will be equal to the initial frequency of the genotype multiplied by its fitness. For A^1A^1 this will be $p^2 \times W_{11}$.

3.4 Selection against a Recessive Trait

When a trait is completely recessive, both the heterozygote and the dominant homozygote have a fitness of 1, while the recessive homozygote has reduced fitness, as shown below:

Genotype	Fitness
AA	1
Aa	1
aa	1-s

If the genotypes are initially in Hardy-Weinberg proportions, the contribution of each genotype to the next generation will be the frequency times the fitness.

AA	$p^2 \times 1 = p^2$
Aa	$2pq \times 1 = 2pq$
aa	$q^2 \times (1-s) = q^2 - sq^2$

The mean fitness of the population is $p^2 + 2pq + q^2 - sq^2$. Since $p^2 + 2pq + q^2 = 1$, the mean fitness becomes $1 - sq^2$, and the normalized genotypic frequencies after selection are:

A	$\frac{p^2}{1-sq^2}$
Aa	$\frac{2pq}{1-sq^2}$
aa	$\frac{q^2 - sq^2}{1-sq^2}$

Genotypic frequencies after selection.

This gives:

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

$$\Delta p = q' - q$$

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

Allelic frequency after selection

and finally,

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

3.5 Selective Mating

Selective mating implies non-random mating between members of a population as has been pointed out before in Unit 1 of Module 2 in this course. Recall that deviations from random mating come about for two reasons: choice or circumstance. If members of a population choose individuals of a particular phenotype as mates more or less often than at random, the population is engaged in **assortative mating**. If individuals with similar phenotypes are mating more often than at random, *positive assortative mating* is in force; if matings occur between individuals with dissimilar phenotypes more often than at random, *negative assortative mating,* or **dissassortative mating**, is at work.

3.5.1 Coefficient of Inbreeding (f)

This is a measure used to determine the effect of inbreeding on genotypic frequencies in a population. The largest value f can take is 1while the least value is 0. In large, random mating populations, it is generally assumed that f is close to zero, because any inbreeding that may occur is between very distant relatives and thus will have very little effect on the inbreeding coefficient. The formula for determining inbreeding coefficient in a population is given by:

$f = (2pq – H)/2pq$

Many species of plants have mating systems that include both self-fertilization and random mating. If the proportion of self-fertilisation is high, nearly all individuals in the population should be homozygotes. Thus, either at the individual or the population level, inbreeding increases homozygosity. An example is the plant species *Avena fatua*, a highly self-fertilising species of wild oats, which is widespread in many areas around the Mediterranean Sea.

4.0 CONCLUSION

Natural selection can result in the elimination of detrimental alleles by selection against homozygotes or in the maintenance of genetic variation by selecting for heterozygotes.

5.0 SUMMARY

Natural selection is the dominant force in the evolution of many traits and has shaped much of the phenotypic variation observed in nature. Darwinian fitness is defined as the relative reproductive ability of a genotype.

6.0 TUTOR-MARKED ASSIGNMENT

The data below shows the result of analysis of Population Y for incidence of ABO blood group. The ABO blood group is determined by 3 multiple autosomal alleles (I^A, I^B, and I^O) at a particular locus.

Table 1a: Incidence of ABO blood group in Populations Y

Population	$I^A I^A$	$I^A I^B$	$I^B I^B$
Y	480	150	120

i. Determine the inbreeding coefficient (F) of population Y.
ii. Comment briefly on the implication of your answer.

Ans: (i) The relevant formula is f = (2pq – H)/2pq

To determine 2pq: q2 = 120/750 = 0.16 therefore q = 0.4. This implies that
p = 0.6. Therefore 2pq = 2 x 0.6 x 0.4 = 0.48.
To calculate H, the observed frequency of heterozygous individuals, we have:

H = 150/750 = 0.2. Therefore, f = (0.48 – 0.2)/0.48 = 0.58

iii. The coefficient of inbreeding is quite high. This implies that mating between relatives (consanguineous mating) is common in the population.

7.0 REFERENCES/FURTHER READING

Klug, W. S.; Cummings, M.R.; Spencer, C.A. & Palladino, M.A. (2009). *Concepts of Genetics*. Pearson Int. Ed. (9th ed.). Pearson Benjamin Cummings.

Russel, P. J. (1996). *Genetics* (7th ed.). New York: Harper Collins College Publishers.

Weaver, R. F. & Hedrick, P. H. (1997). *Genetics* (3rd ed.). WCB Publishers.