BLOCK 2
Genetic Structure of Human Populations

BLOCK 2 : GENETIC STRUCTURE OF HUMAN POPULATIONS

UNIT 4 HARDY-WEINBERG PRINCIPLE

Contents

Learning Objectives

After reading this unit, you would be able to:

- Discuss the allelic and Genetic frequencies;
- Elucidate Hardy-Weinberg Equilibrium and its assumptions; and
- Explain the applications of Hardy-Weinberg Equilibrium.

4.0 INTRODUCTION

Before we proceed with Hardy Weinberg Law, I would like to give you a brief knowledge about Population Genetics. Population Genetics is a field of Biology that studies the changes in the genetic composition of a biological population over a period of time. The changes result from operation of various micro-evolutionary factors such as natural selection, gene flow, genetic drift and various other processes. These evolutionary processes shape the gene pool of

Contributed by **Dr. Gangaina Kameih**, PhD (Anthropology), Thangmeiband Yumnam Leikai, Imphal West, Manipur

a population. A population evolves with change in the gene pool. Population genetics is the area of research that examines the allele and genotype frequency in a population and attempts to identify various factors that can cause alleles and genotype frequencies to change in a specific group over a period of time. Once human population have been identified, the next step is to find out what evolutionary forces, if any, are operating upon them. To determine whether evolution is taking place at a given locus, population geneticists measure allele frequency for a particular trait and compare this observed frequency with the statistics predicted by mathematical model called as *Hardy-Weinberg equilibrium equation*. In short, Population Genetics is the study of frequency of alleles, genotypes and phenotypes in Mendelian populations, from micro-evolutionary perspectives. The goals of Population Genetics are pursued by formulating mathematical models of gene frequency dynamics and extracting the conclusion to study genetic variation in the actual population and then testing the conclusion against the empirical data.

Box 4.1

Mendelian Population: It is defined as an interbreeding group of organisms which share a common gene pool or a reproductive community of sexual and cross fertility individuals which share a common gene pool.

Gene pool: A total of genes shared by reproductive members of a population or we can also say the collection of all the alleles in a population.

Phenotype: External appearance of an organism resulting from the interaction of its genotype and its environment.

We must understand the concept of genotype and allele frequencies and how to compute them before proceeding to the knowledge of Hardy-Weinberg equilibrium because through genotype and allele frequencies only, we would be able to measure genetic variation in a population and also determine whether the locus studied are in reasonable agreement with their respective Hardy-Weinberg expectations or not.

4.1 GENETIC AND ALLELIC FREQUENCIES

The term frequency is defined as a proportion or percentage used to express as a decimal fraction. The Genetic and allelic frequencies of a sample are used to represent the gene pool of the population. There are always fewer alleles than genotypes; so, the gene pool of a population can be described in fewer terms when the allele frequencies are used. For a locus with two alleles, we will have three genotypes and for a locus with five alleles, there will be fifteen possible genotypes, likewise the possible genotypes vary according to the number of alleles present in a particular locus.

4.1.1 Computing Genotype Frequencies

The genotype frequencies are calculated by simply adding up the number of individuals possessing the genotype and then dividing it by the total number of individuals (N). Thus, for the locus with genotypes *AA*, *Aa* and *aa*, the frequency (F) of each genotype is:

$$\text{F}(AA) = \frac{Number\ of\ AA\ individuals}{N}$$

$$\text{F}(Aa) = \frac{Number\ of\ Aa\ individuals}{N}$$

$$\text{F}(aa) = \frac{Number\ of\ aa\ individuals}{N} \qquad (4.1)$$

The sum of all Genetic frequencies should always be equal to 1.

Example 4.1: Compute genotype frequencies. For better understanding, this example is of two alleles which are codominant. Let us say we visited a tribal village in Manipur and collected 200 samples from 200 respondents and obtained the following genotype numbers *AA* = 72, *Aa* = 96 and *aa* = 32. Now, compute the genotype frequencies from the genotype numbers.

To find out the genotype frequencies, we will simply add up the number of individuals possessing the particular genotype and then divide it by the total number of individuals. It can be seen here that 72 of 200 people have *AA* genotype, therefore, the genotype frequency is $\frac{72}{200} = 0.36$. Likewise, we can do this for all the three genotypes:

$$AA = \frac{72}{200} = 0.36$$

$$Aa = \frac{96}{200} = 0.48$$

$$aa = \frac{32}{200} = 0.16$$

Check Your Progress

1) Compute genotype frequencies from the given genotype numbers of 150 people: *AA* = 54, *Aa* = 72, and *aa* = 24?

.....

.....

.....

.....

Remember, it is very important to check the values after computing, by summing up the obtained genotype frequencies, the values should always add up to 1 which they do now (0.36 + 0.48 + 0.16 = 1.0). In population genetics, though we use genotype frequencies, sometimes we also express in terms of percentages. To express these proportions as percentages, we can perform multiplication

with each proportion: *AA* = 36%, *Aa* = 48%, and *aa* = 16%. When we express in terms of percentage, make sure that these add up to 100%.

Box 4.2

Allele: An alternative form of a gene or DNA.

Codominant: The term describes an arrangement in which alleles in a heterozygote affect the phenotype equally. Neither allele is dominant or recessive.

Genotype: The genetic makeup of an individual at a given locus defined by what has been inherited from both parents.

Locus: A specific location of a gene or other DNA sequence on a chromosome.

4.1.2 Computing Allele Frequencies

The allele frequencies can be calculated by two ways. It can be calculated from the numbers of genotypes or it can be calculated from the frequencies of the genotypes. Computing allele frequencies from the numbers of genotypes is also known as *allele counting method*. Mathematically, for biallelic loci i.e., locus with two alleles (*A* and *a*), the sum of the allele frequencies can be expressed using the following simple formula:

$$p + q = 1 \quad (4.2)$$

This relationship enables us to predict the other allele frequency if we know one, because, using simple algebra, we see that

$$p = 1 - q$$

$$q = 1 - p \quad (4.3)$$

Note that just as the genotype frequencies must add up to 1.0 (p + q = 1), so must the allele frequencies.

4.1.2.1 Allele Counting Method

For biallelic loci, the frequencies (*A* and *a*) of the alleles are usually represented by the symbols *p* and *q*. Based on the number of the genotype, they can be calculated using the following formula:

$$p = \mathrm{f}(A) = \frac{2nAA+nAa}{2N}$$

$$q = \mathrm{f}(a) = \frac{2naa+nAa}{2N} \quad (4.4)$$

Likewise, for three alleles we take the example of ABO locus, the frequencies (*A*, *B* and *O*) of the alleles are usually represented by the symbols *p, q* and *r*.

Based on the number of the genotype, they can be calculated using the following formula:

$$p = \mathrm{f}(A) = \frac{2nAA+nAO+nAB}{2N}$$

$$q = \mathrm{f}(B) = \frac{2nBB + nBO+nAB}{2N}$$

$$r = \mathrm{f}(O) = \frac{2nOO+ nAO+ nBO}{2N} \qquad (4.5)$$

And suppose if we have more than three alleles, we can simply use the letter p with the subscript to refer to different alleles.

4.1.2.1.1 Two Alleles

Example 4.2: Compute allele frequencies from the given genotype numbers. Here, we use the same data from example 4.1 on purpose for you to understand easier. The genotype numbers are

Number of people with genotype AA = 72

Number of people with genotype Aa = 96

Number of people with genotype aa = 32

In the following table, we list the number of A and a alleles and the total number of all alleles for each genotype, and then sum each column:

Genotype	Number of people	Number of A alleles	Number of a alleles	Number of all alleles
AA	72	144	0	144
Aa	96	96	96	192
aa	32	0	64	64
Total	200	240	160	400

Using the equation 4.4, we obtained

$$\text{Relative frequency of } A \text{ allele} = \frac{144+96}{2(200)} = \frac{240}{400} = 0.6,$$

$$\text{Relative frequency of } a \text{ allele} = \frac{64+96}{2(200)} = \frac{160}{400} = 0.4.$$

You must always check the allele frequencies and values should add up to 1.0, which they do (0.6 + 0.4 = 1.0).

Check Your Progress

2) Compute allele frequencies using the allele counting method from the given genotype numbers:

*HPA*1-HPA*1* = 54

*HPA*1-HPA*2* = 72

*HPA*2-HPA*2* = 24

Example 4.3: This example is from the real data collected by Kshatriya et al. (2019) where we studied 5 loci from several villages of Tamenglong district, Manipur. ACE locus was one of the loci, this gene is located on chromosome $17q^23$ and it has two alleles, insertion (*I*) and deletion (*D*). The following table shows the obtained data on ACE locus:

Genotype	Number of people	Number of I alleles	Number of D alleles	Number of all alleles
II	28	56	0	56
ID	45	45	45	90
DD	27	0	54	54
Total	100	101	99	200

Using the equation 4.4, we obtained

$$\text{Relative frequency of } I \text{ allele} = p = \frac{56+45}{2(100)} = \frac{101}{200} = 0.505$$

$$\text{Relative frequency of } D \text{ allele} = q = \frac{45+54}{2(100)} = \frac{99}{200} = 0.495$$

4.1.2.1.2 More than Two Alleles

So far, you have learnt about computing allele frequencies with two alleles. In population genetics, researchers and geneticists also study loci with more than two alleles. In this case, how can we compute allele frequencies which have more than two alleles? The answer is, we compute the allele frequencies in the same manner; as long as the alleles are codominant it is simple to calculate. Example 4.4 shows computations for a locus with three alleles. When we have three alleles, we simply label the allele frequencies as *p*, *q*, and *r*. The table shows the following data:

Example 4.4: How to compute allele frequencies for a locus with three alleles. This example is based on ABO locus, the gene responsible for human ABO blood type and it is found on chromosome number 9. The following data are shown in the table below:

Genotype	Number of people	A alleles	B alleles	O alleles	Number of all alleles
OO	33	0	0	66	66
AA	11	22	0	0	22
AO	32	32	0	32	64
BB	34	0	68	0	68
BO	20	0	20	20	40
AB	70	70	70	0	140
Total	200	124	158	118	400

Using the equation 4.5, we obtained:

$$\text{Relative frequency of } A \text{ allele} = p = \frac{22+32+70}{2(200)} = \frac{124}{400} = 0.310$$

$$\text{Relative frequency of } B \text{ allele} = q = \frac{68+20+70}{2(200)} = \frac{158}{400} = 0.395$$

$$\text{Relative frequency of } O \text{ allele} = r = \frac{66+32+20}{2(200)} = \frac{118}{400} = 0.295$$

As a check, the allele frequencies should add up to 1.0, which they do (0.310 + 0.395 + 0.295 = 1.0).

4.1.2.2 An Alternative Method

As we have mentioned that the allele frequency can be calculated either by the number or the frequency of genotypes. Now we will see how to calculate the allele frequency using the frequency of the genotypes. To compute, we have to add the frequency of homozygote (AA or aa) for each allele to half the frequency of heterozygote's alleles of each type.

$$p = f_{AA} + 1/2\, f_{(Aa)}$$

$$q = f_{aa} + 1/2\, f_{(Aa)} \qquad (4.6)$$

Note that we will obtain the same values of p and q whether we calculate the allelic frequency from the number of genotypes or from the frequency of genotypes.

Example 4.5: Compute allele frequencies using the alternate method i.e. from the genotype frequencies. Here we use the same data from example 4.3 and the data are as follows:

Number of individuals with genotype *II* = 28

Number of individuals with genotype *ID* = 45

Number of individuals with genotype *DD* = 27

In this method, we first need to compute the genotype frequencies from the given genotype numbers. To obtain the genotype frequencies we simply divide the number in each genotype by the total number of genotypes (=100). Now, we obtained:

$$\text{Frequency of genotype } II = \frac{28}{100} = 0.28$$

$$\text{Frequency of genotype } ID = \frac{45}{100} = 0.45$$

$$\text{Frequency of genotype } DD = \frac{27}{100} = 0.27$$

Using the mathematical equation given above (4.6), we can now compute the allele frequencies and obtained:

$$p = I = 0.28 + \frac{0.45}{2} = 0.28 + 0.225 = 0.505$$

$$q = D = 0.27 + \frac{0.45}{2} = 0.27 + 0.225 = 0.495$$

We get the same allele frequencies as with the allele counting method. We will obtain the same values whether we calculate the allelic frequency from the *allele counting method* or from the *frequency of genotypes*. You must check the results carefully because sometimes round off errors occur and therefore, we may not get the exact same results as the allele counting method.

4.2 HARDY-WEINBERG EQUILIBRIUM

The Hardy-Weinberg principle was discovered independently by G.H. Hardy (a British Mathematician) and W. Weinberg (a German physician) in 1908. According to this law, allele and genotype frequency in a population do not change from one generation to another, and remain constant. This principle postulates a set of condition in a population where no evolution occurs. In other words, Hardy Weinberg law presents a hypothetical condition by which a population is assumed to meet conditions such as infinitely large size, occurrence of no mutation, occurrence of no selection, randomly mating and no gene flow occurring in the population. If all these conditions are satisfied, allele frequency will not change from one generation to next generation. This mathematical model is a powerful tool for developing evolutionary models and the law is:

$$p^2 + 2pq + q^2 = 1 \qquad (4.7)$$

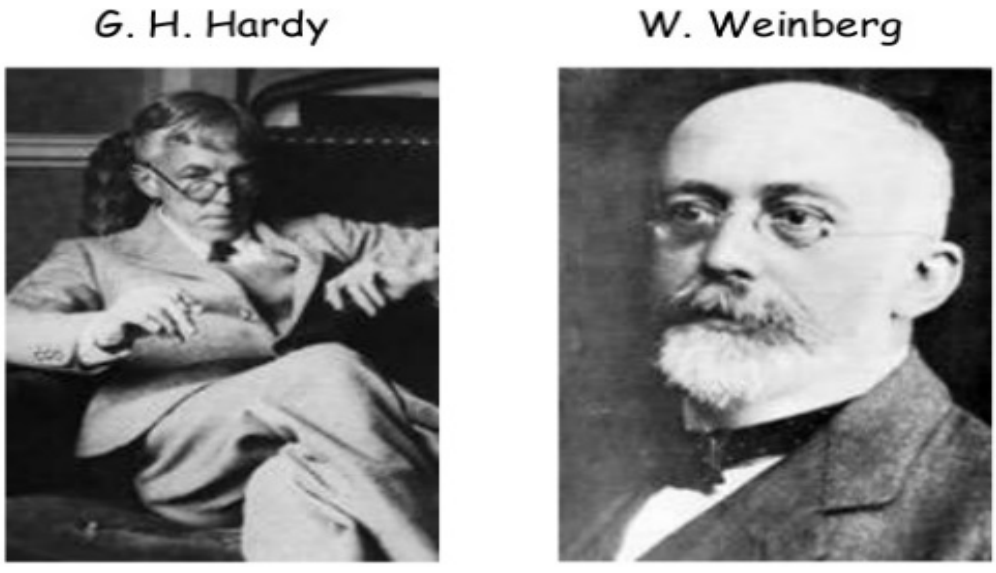


Figure 4.1: G.H. Hardy and W. Weinberg
(***Source:*** https://images.app.goo.gl/qTQZ2U14jRQy2Mzk9)

At the most basic level, Relethford (2012) defined Hardy-Weinberg equilibrium as a mathematical expression describing the expected genotype frequencies in a new generation. Imagine, we have a locus with two alleles, A and a, where p is the frequency of the A allele and q is the frequency of the a allele. Both Hardy and Weinberg showed that the expected genotype frequencies in the next generation are

$$\text{Expected frequency of genotype } AA = p^2$$

$$\text{Expected frequency of genotype } Aa = 2pq$$

$$\text{Expected frequency of genotype } aa = q^2 \qquad (4.8)$$

Using the same data from example 4.2, we have $A = p = 0.6$ and $a = q = 0.4$. From these allele frequencies, we can figure out the expected distribution of genotypes using the Hardy Weinberg equilibrium in the next generation as

$$\text{Expected frequency of genotype } AA = p^2 = (0.6)2 = 0.36$$

$$\text{Expected frequency of genotype } Aa = 2pq = 2(0.6)\,(0.4) = 0.48$$

$$\text{Expected frequency of genotype } aa = q^2 = (0.4)^2 = 0.16$$

Note that the sum of these genotype frequencies adds up to 1, which they do here (0.36 + 0.48 + 0.16 = 1.0).

Table 4.1 summarizes the logic used here to demonstrate Hardy-Weinberg frequencies. You will find other methods of derivation from the text by Hartl and Clark (2007) where he considers the genotypes of all possible offspring from parental genotypes. Figure 4.2 shows the nonlinear relationship between allele frequencies and genotype frequencies. Relethford (2012) has explained the relationship like this "As p increases, the proportion of homozygous genotype AA increases geometrically, and the frequency of homozygous genotype aa decreases geometrically. Note that the frequency of heterozygotes

(*Aa*) increases as *p* increases to reach a maximum at $p = 0.5$, after which the frequency of heterozygotes decreases".

Table 4.1: Using simple probability to derive Hardy-Weinberg equilibrium proportionsa

Genotype of a child	Allele from Father		Allele from Mother	Probability	
AA	*A*		*A*	$p \times p = p^2$	
Aa	*A*	Or	*a*	$p \times q = pq$	$= pq + pq = 2pq$
	a		*A*	$q \times p = qp = pq$	
aa	*a*		*a*	$q \times q = q^2$	

[a]The data presented in this table are based on a single locus with two alleles, A and a, where p is the frequency of the A allele in the parental population and q is the frequency of the a allele in the parental population (Source: Relethford, 2012)

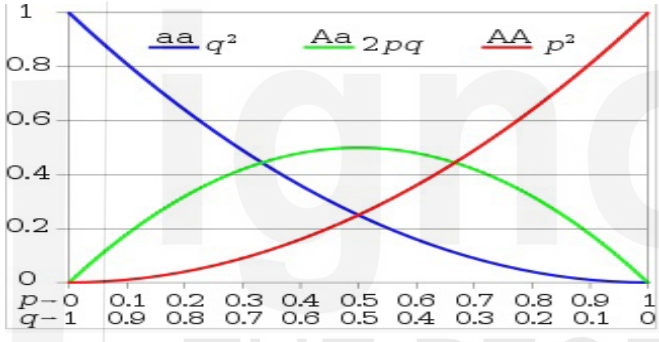


Figure 4.2: Genotype frequencies under Hardy-Weinberg equilibrium for a single locus with two alleles (A and a). The frequency of A allele is p, and the frequency of the a allele is q = 1 - p. The genotype frequencies are $AA = p^2$, $Aa = 2pq$, $aa = q^2$.

With the concept of Hardy-Weinberg equilibrium, we can also predict outcome to some specific questions related to genetic disease which we think will prevail in the next generation. Let us say about Sickle cell anaemia, suppose the frequency of S allele (recessive) is 0.01 and we want to know how many children will be born having sickle cell anaemia (two recessives), In this case, $q^2 = S^2 = (0.01)^2 = 0.0001$, which is 1 in 10,000 offspring. Likewise, if we want to know how many would be carrier we could also compute: $2pq = 2(AS) = 2(0.99)\ (0.01) = 0.0198$, which is 198 in 10,000 offspring.

4.3 ASSUMPTIONS OF HARDY-WEINBERG EQUILIBRIUM

The mathematical concept, called the Hardy-Weinberg principle, is a crucial concept in population genetics. It predicts how gene frequencies will be inherited from generation to generation given a specific set of assumptions. The assumptions of Hardy-Weinberg Equilibrium are discussed in the following paragraphs.

In this case, we assume that we are dealing with a sexually reproducing species (two sexes). We also assume that each generation is discrete and does not overlap with any other generation.

4.3.1 Random Mating

A key assumption of Hardy-Weinberg Equilibrium is random mating. Individuals in populations do not always follow random mating patterns. Non-random mating or assortative mating occurs because of more or less frequent mating among specific individuals than expected in a random mating population. Preferred mating partners can have more similar (positive assortative mating) or dissimilar (negative assortative mating) phenotypes than expected in random mating. Phenotypically similar individuals might have similar genotypes which are observed in consanguineous mating resulting in inbreeding. Assortative mating is however not the same as inbreeding as the former is among phenotypically similar individuals (similar skin color, weight, and height), while inbreeding occurs between individuals with similar genotypes (*AA* x *AA* mating). As with inbreeding, positive assortative mating will increase the proportions of homozygotes. Deviations from random mating will lead to genotype frequencies that are different from those expected under Hardy-Weinberg Equilibrium.

4.3.2 No Mutation

Hardy-Weinberg Equilibrium assumes that there is no mutation. In our simple examples, we have a locus with two alleles, *A* and *a*, and Hardy-Weinberg Equilibrium assumes that these two alleles will be passed on to the next generation, but none of the *A* alleles will mutate into *a*, none of the *a* alleles will mutate into *A*, and neither allele will mutate to a new allele. If this assumption is violated, then there will be a change, albeit small, in the allele frequencies (Relethford 2012). Suppose for example, that a population has 60 *A* alleles and 40 *a* alleles, giving frequencies of 0.6 and 0.4. Imagine that one of the *A* alleles mutates into new form, which we will call *B*. We now have 59 *A* alleles, 40 *a* alleles, and 1 *B* allele. The frequency of the *A* allele is now $\frac{59}{100} = 0.59$. Hardy-Weinberg Equilibrium assumes that mutation does not happen, but in the real world, there will always be a small fraction of alleles that mutate and are passed on to the next generation. Although small, this is a change in allele frequency.

4.3.3 No Genetic Drift

Hardy-Weinberg Equilibrium assumes that there is no genetic drift. Genetic drift is sampling variation, such that allele frequencies will fluctuate randomly over time, sometimes increasing and sometimes decreasing. Large fluctuations in allele frequency are more likely in small populations. Population size refers here to the breeding population size, the number of adults of reproductive age in the population. Hardy-Weinberg Equilibrium assumes that there is absolutely no genetic drift, and in essence assumes that the population is of infinite size. Although this clearly is not possible, in practical terms this translates to the

population being very large, such that there is no perceptible impact of genetic drift (Relethford 2012).

4.3.4 No Natural Selection

Hardy-Weinberg Equilibrium also assumes that there is no Natural selection. Natural selection is Charles Darwin's contribution to evolutionary theory, referring to differential survival and reproduction. It occurs when there is a difference in fitness (the probability of surviving and reproducing) for different genotypes, such that there is a change in allele frequency overtime. One of the classic example of natural selection in human populations is the change in allele frequencies of Hemoglobin S, which is selected for in malarial environments. Here, the higher fitness of the heterozygote (*AS*) relative to the *AA* genotype (susceptibility to malaria) and the *SS* genotype (presence of sickle cell anaemia), is also a classic example of balancing selection.

4.3.5 No Gene Flow or Migration

The last assumption of Hardy-Weinberg Equilibrium discussed here is the assumption that the population is closed and there are no migrants into the population from another population. This assumption means that there is no gene flow entering the population. Gene flow can introduce new alleles into a population, and acts to reduce genetic differences between populations. The "Island Model" proposed by Wright (1931) is the simplest model that shows a meta-population divides into smaller population islands of equal size which exchange genes at the same rate per generation.

4.4 APPLICATION OF HARDY-WEINBERG EQUILIBRIUM

We apply our knowledge of the above assumptions to compare observed reality with theoretical expectation to gain insight into how evolution works. The major use of Hardy-Weinberg Equilibrium is the establishment of a baseline condition- no changes in genotype or allele frequencies- that allows us to model what will happen when one or more of the underlying assumptions of Hardy-Weinberg Equilibrium are not met.

4.4.1 Detecting Deviations from Hardy-Weinberg Equilibrium using the Chi-square Statistic

We can compare observed and expected genotype numbers to explicitly test the hypothesis of equilibrium. This allows us to test the hypothesis that the population is at Hardy-Weinberg Equilibrium (for the specific locus) or not. Imagine that you visit a population of 100 people and find the genotype numbers for the hypothetical locus that we have been using: $AA = 28$, $Aa = 45$, $aa = 27$. These are the observed genotype numbers.

We can now compare observed genotype numbers with the numbers expected under Hardy-Weinberg Equilibrium. We do this by computing the allele

frequencies and then use Hardy-Weinberg Equilibrium to compute the expected genotype numbers. Here we use the allele counting method and find that there are 101 *A* alleles and 99 *a* alle les, for a total of 200 alleles. This gives allele frequencies of $p = \frac{101}{200} = 0.505$ and $q = \frac{99}{200} = 0.495$. We now use Hardy-Weinberg Equilibrium to predict the genotype frequencies: $AA = p^2 = 0.255$, $Aa = 2pq = 0.500$, and $aa = q^2 = 0.245$. Finally we multiply these expected genotype frequencies by the total number of people (=100) to get the expected genotype numbers:

$AA = 0.255(100) = 25.5$

$Aa = 0.5(100) = 50$

$aa = 0.245(100) = 24.5$

We use a chi-square test (a test of significance), a method that is used for comparing observed and expected numbers in statistical research. The chi-square statistic, denoted by the symbol χ^2, is computed using the formula

$$\chi^2 = \sum \frac{(O - E)^2}{E} \qquad (4.9)$$

Here, *O* refers to the observed number for a given genotype and *E* refers to the expected number for a given genotype. The difference between observed and expected (*O* - *E*) is squared and divided by the expected number. The Greek letter sigma (∑) is used in statistics as shorthand for summation; in this case, it means performing the same operations on each genotype and then summing the results.

The computation of the chi-square statistic is presented in the following table with examples. The first three columns of the tables show the genotype, the observed number, and the expected number. The remaining columns walk us through the computation of the chi-square statistic. The allele frequencies were computed from the observed genotype numbers and used to compute the expected number (expected under Hardy-Weinberg equilibrium).

Genotype	**Observed Number (O)**	**Expected Number (E)**	**(O – E)**	**$(O - E)^2$**	**$(O - E)^2/E$**
AA	28	25.5	2.5	6.25	0.2451
Aa	45	50	-5	25	0.5000
aa	27	24.5	2.5	6.25	0.2551

After performing the calculations in this table, we get a chi-square (χ^2) statistic of

$$\chi^2 = 0.2451 + 0.5000 + 0.2551 = 1.0002$$

This value needs to be compared with a standard table of chi-square values for a given probability value, which is the level of statistical significance that we use for testing the hypothesis. Probability value below 0.05 is considered to be statistically significant which means it deviate from Hardy-Weinberg law. The final thing we need to know is a statistic known as degree(s) of freedom, which represents the number of values that are free to vary, symbolized as *df*, is computed as

$$df = n - k - 1 \quad (4.10)$$

Where n is the number of classes and k is the number of independent parameters. We have three genotypes, so $n = 3$. The genotype frequencies are computed from the allele frequencies, p and q, which are the parameters. However, if we know either p or q, we can always compute the other one (because $p + q = 1$), so there is only one independent parameter, and therefore $k = 1$. This means that for the examples discussed now, equation (4.10) gives $df = 3 - 1 - 1 = 1$. A table of chi-square values, available in almost any statistics book or online, shows that for a significance level of 0.05 and $df = 1$, the critical value of the chi-square statistic is 3.841. If the observed chi-square statistic is greater than or equal to the critical value, then we reject the hypothesis of Hardy-Weinberg Equilibrium.

In the example above, the observed chi-square value is 1.0002, which is less than the critical value of 3.841. Therefore, the slight difference that we see between observed and expected genotype numbers is not statistically significant and the population is at Hardy-Weinberg Equilibrium.

Check Your Progress

3) Compare observed and expected genotype numbers to explicitly test the hypothesis of Hardy-Weinberg equilibrium using chi-square test method? The observed genotype numbers are: *AA* = 65, *Aa* = 90, and *aa* = 45.

.....

.....

.....

.....

4.4.2 Hardy-Weinberg Equilibrium and Dominant Alleles

Hardy-Weinberg Equilibrium also has a practical application for computing allele. The allele counting method for computing allele frequencies works only when the alleles are co-dominant. If one allele is dominant and the other is recessive, then the method cannot work. To understand this, consider PTC tasting, a trait in humans that involves the ability to taste the chemical phenylthiocarbamide. Some people can taste this bitter chemical and others cannot. The genetics of PTC tasting is approximated by a model with two alleles, a dominant *tasting* allele (*T*) and a recessive *nontasting* allele (*t*). Using this model, the genotypes and phenotypes are

Genotype	Phenotype
TT	Taster
Tt	Taster
tt	Non taster

There is no way to count the alleles in this case. If someone is a non-taster, we know that person's genotype is *tt* and that we would count two t alleles. However, if someone is a taster, all we know is that this individual could have either genotype *TT* or *Tt*, and we therefore do not know whether to count one or two *T* alleles. We can solve this problem by assuming that this locus is at Hardy-Weinberg Equilibrium, we can write the phenotype frequencies as follows:

$$\text{Frequency of tasters} = \text{frequency of } TT + \text{frequency of } Tt = p^2 + 2pq$$

$$\text{Frequency of non tasters} = \text{frequency of } tt = q^2$$

Because the frequency of non tasters is q^2, we can compute the frequency of q by taking the square root of the frequency of nontasters. Once we have q, we can compute $p = 1 - q$ (from equation 4.3).

Here is an example. Giles et al. (1968) collected data on PTC tasting for 1374 people in the town of Ticul, Yucatan, Mexico, finding 1257 tasters and 117 nontasters. The frequency of non-tasters is $\frac{117}{1374} = 0.085$. We take the square root of this frequency to get the frequency of the *t* allele:

$$q = \sqrt{q^2} = \sqrt{0.085} = 0.292$$

The frequency of the *T* allele is therefore

$$p = 1 - q = 1 - 0.292 = 0.708$$

We can compute allele frequencies when there is dominance and as long as we assume Hardy-Weinberg Equilibrium.

4.5 SUMMARY

When we look at micro-evolution, we are interested in how the relative frequency of alleles can change over time. When the field of genetics was just beginning, some researchers wondered why dominant alleles did not increase to completely replace recessive alleles over time. This question led to a deeper understanding of the mathematical application of Mendel's ideas on a population basis. In 1908, G. Hardy and W. Weinberg independently derived a mathematical relationship between allele and genotype frequencies, and further showed that the allele and genotype frequencies would remain the same generation after generation. If a locus has two alleles *A* and *a* with frequencies *p* and *q* respectively, the relationship with the genotypes of the succeeding generation designated as *AA*, *Aa* and *aa* is given by, $AA = p^2$, $Aa = 2pq$, $aa = q^2$. If the genotype frequencies in the succeeding generation have identical values to

those in the parental generation, the population is said to be at Hardy-Weinberg equilibrium for that locus. In contrast, if Hardy-Weinberg equilibrium is not observed, then evolution is occurring via one of the processes causing change in genotype and allele frequencies, in other words it means assumptions (random mating, infinite size, no mutation, no selection, no genetic drift, and no gene flow) of Hardy-Weinberg Equilibrium are not met or violated.

4.6 REFERENCES

Giles, E., Hansen, A.T., McCullough, J.M., Metzger, D.G., & Wolpoff, M.H. (1968). Hydrogen cyanide and phenylthiocarbamide sensitivity, mid□phalangeal hair and color blindness in Yucatán, Mexico. *American Journal of Physical Anthropology*, *28*(2), pp. 203-212.

Hartl, D.L., & Clark, A.G. (2007). *Principles of Population Genetics*. 4th Edition, Sunderland, Mass: Sinauer Associates.

Kshatriya, G.K., Panmei, T., & Kameih, G. (2019). Alu insertion-deletion polymorphisms in the Tibeto-Burman speaking tribal groups of Manipur, North-East India. *Gene Reports*, *15*, 100372.

Relethford, J.H. (2012). *Human Population Genetics.* (Vol. 7). John Wiley & Sons.

Wright, S. (1931). Evolution in Mendelian populations. *Genetics*, *16*, pp 97.

4.7 ANSWERS TO CHECK YOUR PROGRESS

1) To obtain the genotype frequencies from the given genotype numbers, we simply divide the number in each genotype by the total number of genotypes (=150). Now we obtain:

$$\text{Frequency of genotype } AA = \frac{54}{150} = 0.36$$

$$\text{Frequency of genotype } Aa = \frac{72}{150} = 0.48$$

$$\text{Frequency of genotype aa} = \frac{24}{150} = 0.16$$

2) Using the mathematical equation given in the text (4.4), we can compute the allele frequencies and obtained:

$$p = f(A) = HPA*1 = \frac{2nAA+nAa}{2N}$$

$$HPA*1 = \frac{2\times54+72}{2\times150}$$

$$= \frac{108+72}{300}$$

$$= \frac{180}{300}$$

$$= 0.60$$

$$q = f(a) = HPA*2 = \frac{2naa+nAa}{2N}$$

$$HPA*2 = \frac{2\times24+72}{2\times150}$$

$$= \frac{48+72}{300}$$

$$= \frac{120}{300}$$

$$= 0.40$$

3) As observed genotype numbers is already given, we need to find out the genotype numbers expected under Hardy-Weinberg equilibrium. We do this by computing the allele frequencies and then use Hardy-Weinberg Equilibrium to compute the expected genotype numbers. Using the allele counting method, we find that there are 220 *A* alleles and 180 *a* alleles, for a total of 400 alleles. This gives allele frequencies of $p = = 0.55$ and $q = = 0.45$. We now use Hardy-Weinberg Equilibrium to predict the genotype frequencies: $AA = p^2 = 0.30$, $Aa = 2pq = 0.50$, and $aa = q^2 = 0.20$. Finally we multiply these expected genotype frequencies by the total number of people (=200) to get the expected genotype numbers:

   $AA = 0.30(200) = 60$

   $Aa = 0.50(200) = 100$

   $aa = 0.20(200) = 40$

We use a chi-square test for comparing observed and expected numbers in statistical research.

$$\chi^2 = \sum \frac{(O-E)^2}{E}$$

The computation of the chi-square statistic is presented below. The first three columns of the tables show the genotype, the observed number, and the expected number. The remaining columns are the computation of the chi-square statistic. The allele frequencies were computed from the observed genotype numbers

and used to compute the expected number (expected under Hardy-Weinberg equilibrium).

Genotype	Observed Number (O)	Expected Number (E)	(O – E)	$(O – E)^2$	$(O – E)^2/E$
AA	65	60	5	25	0.417
Aa	90	100	-10	100	1.000
aa	45	40	5	25	0.625

After performing the calculations in this table, we get a chi-square (χ^2) statistic of

$$\chi^2 = 0.417 + 1.000 + 0.625$$

$$= 2.042$$

The observed chi-square value is 2.042, which is less than the critical value of 3.841. Therefore, the slight difference that we see between observed and expected genotype numbers is not statistically significant and the population is at Hardy-Weinberg Equilibrium.

UNIT 5 MECHANISMS OF EVOLUTION

Contents

Learning Objectives

After reading this unit, you will be able to:

- appreciate the mechanisms of evolutionary forces and their importance;
- understand the different mechanisms of evolutionary forces; and
- learn about the example of natural selection, patterns of selection and types of genetic drift.

5.0 INTRODUCTION

The scope of evolution includes investigation of biological history of our planet, genetic diversity of populations, mechanisms of speciation and origin of living beings. Studies on evolution reveal how the different organisms originated during the course of time and what forces are responsible for their adaptation, survival and extinction. Population is the unit of evolution because sum of effect of all forces of evolution affects the entire population and causes

Contributed by **Dr. S.A.A. Latheef**, Department of Genetics and Biotechnology, Osmania University, Hyderabad.

changes from generation to generation. Evolution can be understood as changes taking place over time and include small or large, invisible or visible and non-adaptive and adaptive changes. Evolution is of two types. If changes occur overtime within the organism, it is called micro evolution and if it occurs from one being to another it is macroevolution. Microevolution cause changes in the allele frequency whereas macroevolution leads to speciation (production of new species from earlier species).

5.1 MECHANISMS OF EVOLUTION

There are four known forces of evolution, they are mutation, gene flow, natural selection and genetic drift. Mutation introduces new alleles into the population due to the occurrence of copying errors in DNA replication and transcription. Natural selection (due to the individual differences on the survival and reproductive success) and genetic drift (due to random sampling in small populations) differentially transmit alleles into the next generation. Gene flow, through migration, contributes alleles from one population to other (Vasulu, 2012). The interaction of the evolutionary forces is studied using genetic models to understand the distribution, turnover and diversity of alleles at the particular level. Evolutionary forces disturb the Hardy-Weinberg equilibrium and are responsible for the process of evolution. Interaction of evolutionary forces contributes to variation and spreading alleles within or between populations.

A detailed discussion on various evolutionary forces is given below.

5.1.1 Mutation

It is one of the evolutionary forces. It serves as a raw material of evolution. If evolution has to happen mutation should occur. Mutation occurs in every generation. Mutation is mostly random change in phenotype (colour, size and shape) or genotype forms. Genotype changes involve alteration in the sequence of DNA. Mutation introduces new alleles in the population and changes the allele frequencies. The survival of introduced allele depends on how it can affect the survival fitness of the population. If the introduced new allele is advantageous to the population it is supported by natural selection. Vernon Ingram was the first to identify change in the amino acid of valine in place of glutamic acid in the mutant haemoglobin of patients with sickle cell anaemia. The frequency of mutation in humans is one in 500 million nucleotides (Sudi and Ali– Dunkrah, 2005). Mutation includes all type of heritable changes. The process of introduction of mutation in a gene is called mutagenesis and resulting product is called mutant. The agent used to induce mutation is called a mutagen.

Depending on the frequency of mutant different terms are used. If frequency of genetic mutant is less than 1% it is called variant and if ≥1 it is termed as polymorphism. Mutations are known by different names depending on the number of bases, mechanism and region of localisation of variants as, single nucleotide polymorphisms, indels (insertion and deletion), short tandem and variable number of tandem repeats and copy number variations. Mutation rate is low and varies with the sites of the genome. In some sites its frequency is

higher than others and these are known as hot-spots of mutation. Higher rate of mutations occurs in intronic, repetitive sequences and non-coding regions of mitochondria (hypervariable regions 1 and 2, HV1 and HV2 respectively). Mutations in the HV1 and HV2 are useful to study the maternal history of population using haplo-group or population approaches. The rate of spontaneous mutations ranges from one in 10^4-10^8 gene per generation. Alu transposition element (short interspersed elements or repeats) events occur in every new birth.

Inborn errors of metabolism (IEM) are a group of genetic disorders that are caused by mutations. Their frequency is one in 2500 births across the globe (Jeanmonod et al., 2021) and in India it is one in 2497 births (Lodh and Kerketta, 2013). In the causation of IEM, autosomes, X, Y chromosome and mitochondria are involved. A total of 16000 genes' involvement in IEM has been observed. Information on the IEM is available in the database of "online inheritance in Man". Meta analysis (Waters et al., 2018) of global data showed aminoaciduria (abnormal level of amino acids in urine especially glutamate and aspartate) are highly prevalent IEM (14.7 per 100000 live births). Mutation in the solute carrier family1 member1 (SLC1A1) causes malfunction in the absorption of bicarboxylic amino acid reabsorption in the kidneys. In India, Glucose-6-phosphate dehydrogenase deficiency is one of the highly prevalent IEMs (Lodh and Kerketta, 2013) caused by mutation in G6PD gene. Carriers of G6PD deficiency are asymptomatic till exposed to antimalarial drugs.

Types of mutations, causes and examples are shown in Table 5.1. The effects of mutations include harmful, beneficial and neutral.

5.1.1.1 Harmful Effects

Mutations are responsible for genetic disorders and cancer. The genetic disorders involve single genes both autosomal (Sickle cell anaemia, Thalassemia, Cystic fibrosis, Huntington disease and Marfan syndrome) and sex linked genes (Haemophilia A and Vitamin-D resistant rickets) and also copy number variations or chromosomal abnormalities (Down syndrome, Jacob syndrome, Klinefelter syndrome, Turner syndrome, Super females, Patau syndrome and Edward syndrome). Mutations in BRCA1 and BRCA2 (tumour suppressor genes) cause breast, ovarian, pancreatic and prostate cancers.

5.1.1.2 Beneficial Effects

Carriers of sickle cell trait (HbAS) are known to be protected against *Plasmodium falciparum* malaria. Carriers of sickle cell anaemia with alpha thalassemia showed lower incidence of cardiac complications, jaundice, gall stone, pallor, splenectomy, acute splenic sequestration crises, stroke and avascular than those without alpha thalassemia (Ali Al-Barazanchi et al., 2021).

5.1.1.3 Neutral Effects

Mutations such as silent mutations do not any have effect either positive or negative effect on the organisms and do not change the amino acids they encode. Silent mutations have been shown to interfere in the splicing function of exon splicing enhancers in the desired place in nucleotide sequence (Dickson and Hyman, 2013).

Table 5.1: Type of Mutation, Causes and Example

Type of Mutation	Cause (s)	Example
Dominant	Mutation in single copy of gene.	Achondroplasia
Recessive	Mutation in both copies of gene.	Cystic fibrosis
Morphological	Affect the appearance of individual colour, shape, size	Albinism
Lethal	Cause death of the organism.	Cystic fibrosis, Sickle cell anaemia, achondroplasia
Biochemical	Recognised by the deficiency.	Galactosemia, Phenylketonuria
Resistant	Individual grow even in the presence of infection.	CCR5-delta32 in humans is resistant to HIV
Spontaneous	Errors in DNA replication, transcription, spontaneous lesions and transposition of transposable genetic elements.	Sickle cell anaemia
Induced	Structural changes caused by agents (mustard gas, ethyl urethane, phenol and formaldehyde), radiation, intercalating agents, oxidative agents, deamination, alkylating agents, 5-bromouracil.	Duchenne muscular dystrophy, Congenital myotonic dystrophy
Point (Figure 5.1)	Change of single base pair into another.	Sickle anaemia
(i) Silent or synonymous	Mutated triplet codon code for same amino acid.	In BRCA1, BRAC2 genes cause exon skipping and change protein structure either by creating or inactivating splicing site.
(ii) Missense or non-synonymous	Mutated triplet codon code for different amino acid.	Sickle cell anaemia
(iii) Nonsense	Mutated triplet codon is a premature stop codon,	Cystic fibrosis
Germline	Errors in DNA replication and oxidative damage,	*Familial adenomatous polyposis*
Somatic	Errors in DNA mechanisms or exposure to ultraviolet radiation chemicals and oxidative stress,	Cancer tumours
Copy number variation	Non-homologous end joining, errors in DNA replication and replication of non-contiguous DNA segments,	Huntington's disease

(i) Duplication	Duplication of the gene	Charcot-Marie-Tooth disease type 1 (CMT1
(ii) Deletion	Deletion of the gene	Prader-Willi syndrome
Frame shift	Insertion, deletion of one or more nucleotides leading to the change of reading frame of base sequence.	Cystic fibrosis, Tay-Sachs disease

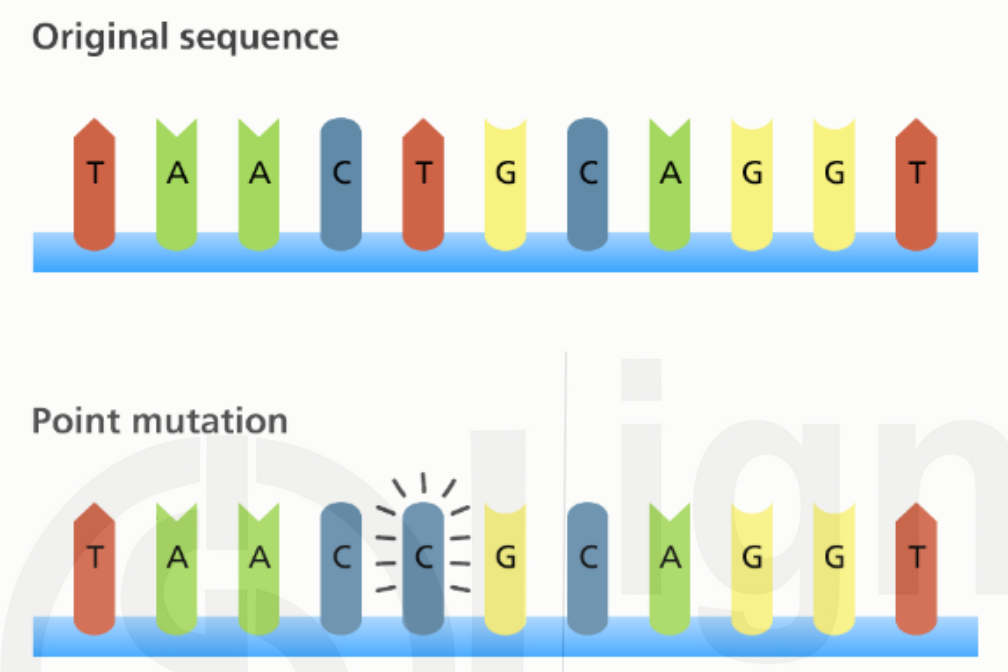


Fig. 5.1 : Mutation
(Source: https://www.yourgenome.org/facts/what-is-a-mutation)

5.1.2 Gene Flow

Diffusion of alleles across populations can be called gene flow. Human beings move from one population and join another population due to various socio-economic reasons, and choose mate and involve in interbreeding. This causes the gene flow or genetic admixture. As a result of gene flow, allele frequency changes in the population left and also the population in which they have joined. Examples of genetic admixture or gene flow (Figure 2) are the Anglo Indians, the American Blacks and the Siddis (Indo-African population). In Latin American countries, admixed populations are seen due to random mating among indigenous populations, Europeans and Africans. Another example is gradient distribution of the B blood group allele. B allele believed to have originated in Asia spread to western countries due to genetic admixture forced by invasions. The barriers to gene flow are geographical, cultural, linguistic, and political factors. For gene flow to occur, not only the large scale migrations and choosing of mates in the settled areas but also the mate selection in particular direction over a long span of time may be responsible. To explain the change in allele frequency contributed by gene flow, Sewall Wright proposed 'island' and 'isolation by distance' models whereas Kimura and Weiss contributed 'stepping stone model'. Stepping stone model assumes that a line or ring of populations exchange migrants with their immediate neighbours only whereas island model

allows exchange of migrants between populations. In contrast, 'isolation by distance model' proposes that gene flow decreases as the geographical distance increases.

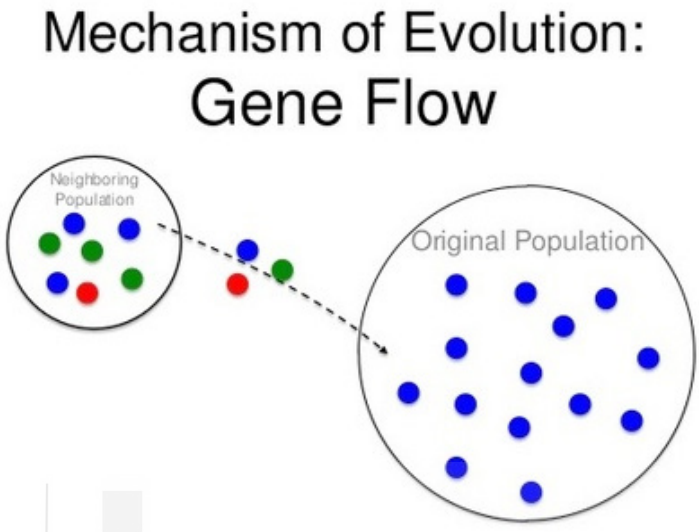


Fig. 5.2 : Gene Flow
(***Source:*** http://anthropologicalconcepts.weebly.com/blog/gene-flow)

5.1.3 Natural Selection

Natural selection is defined as differential survival and reproductive success. Mutation adds new alleles into the population. Natural selection determines the fate of newly added allele in the population. If it is not advantageous or not suitable to the environment, natural selection eliminates it by negative or purifying selection. The survival of allele in particular environment depends on its fitness/ adaptive value/ selective value. Various factors such as mate selection, female fecundity and survival in the reproductive age contribute to the fitness. Those who are fit to survive, have reproductive success and contribute more offspring than less fit people. This is also called adaptation. Natural selection facilitates the adaptation of the organism to the environment and the allelic composition carried by the fit individual is promoted that result in the change in allele frequency of the population due to the replacement of allele composition carried by the less fit individuals. In another words this is called positive selection or diversifying selection. When people of opposite sex carrying different genotypes marry, they pass on unequal alleles to their offspring as a result there may be a difference in the allele frequency of parents or progeny. An individual is the unit of natural selection.

Examples of natural selection are skin colour, physiological and genetic adaptation to high altitude, lactose persistence, sickle cell trait and Duffy blood group.

5.1.3.1 Skin Colour

It formed the basis to classify the humans into different races. Natural selection favoured light skin to those residing far from equator. The light skin allowed ultraviolet rays to enter and produce vitamin D to absorb calcium levels. For

those staying near equator natural selection favoured dark skin to prevent the loss of folic acid (vitamin B9) required for the development of healthy foetuses (Humanorigins.si.edu).

5.1.3.2 Physiological and Genetic Adaptations to High Altitude

Human beings living in plains when enter into the high altitudes, undergo hypoxic stress and develop adaptive responses like increase in the levels of red blood cells, greater lung volume and increased chest dimensions and respiration. These adaptations allow them to stay for short-term. It was observed that long-resident Andeans (altitude of 6961m) of South America have increased cardiac oxygen utilisation, larger lung volumes, less hypoxic pulmonary vasoconstrictor response, narrow alveolar to arterial oxygen gradients, greater uterine artery blood flow during pregnancy and mutations in the genes involved in the regulation of metabolic haemostasis, vascular control and erythropoiesis due to natural selection to support their survival at high altitude.

5.1.3.3 Lactase Persistence

Lactase is an enzyme that hydrolyses lactose, a carbohydrate present in milk, into glucose and galactose. The lactase activity decreases after weaning period in most mammals but some humans continue to produce this enzyme and digest the lactose in the milk. This trait is known as lactase persistence (LP) and has been observed in pastoralist populations. Correlation of estimates of mutation time with dairy practicing activities in Europe and Africa suggested LP trait coevolved with dairying by natural selection due to nutritional advantage associated with the consumption of milk.

5.1.3.4 Duffy Negative Phenotype and Resistance to Malaria

Duffy is one of the blood groups identified in humans. It has six antigens (Fya, Fyb, Fy3-Fy6) on red blood cells (RBC). The FY gene has two alleles FYA and FYB which code for antigens Fya, Fyb which differ by single amino acid. Homozygotes for single nucleotide polymorphism (-33T→C) in the erythroid promoter region of FYB allele have Duffy negative phenotype FY (-a-b) and lack antigens. RBCs lacking Duffy antigens are resistant to *Plasmodium vivax*. The malaria caused by *Plasmodium vivax* has been found to be absent in populations where 95% of them had Duffy negative antigens. Natural selection might have fixed Duffy negative antigens in adapting to *P. Vivax* endemic areas.

Check Your Progress

1) What is the importance of Duffy blood group in Malaria?

.....

.....

.....

.....

5.1.3.5 Sickle Cell Trait and Resistance to *Plasmodium Falciparum* Malaria

Presence of sickling Hb (HbSS) in homozygous state is called sickle cell anaemia and in heterozygote state (HbAS) is termed as sickle cell trait. Sickle

cell anaemia results from point mutation in the beta globin gene at codon 6th position which results in replacing amino acid glutamic acid with valine. Heterozygous carriers of haemoglobin (HbAS) (A=adult and S=Sickling) in malaria endemic areas have shown protection against *Plasmodium falciparum* induced malaria than homozygous carriers of sickling Hb (HbSS). Heterozygotes (HbAS) carrying one sickling allele don't show any symptoms and survive on par with normal individuals. Natural selection favoured heterozygotes carrying genotype (HbAS) in areas endemic to *P. Falciparum* malaria. This is an example of balanced polymorphism. HbAS individuals have been found to have higher fitness than either AA or SS.

Selection is of five types. They are of the following:

1) Selection for the heterozygotes (heterozygotes have higher fitness than homozygote) (Example Sickle cell trait explained under 5.1.3.5).

2) Selection against heterozygotes or under dominance (lower fitness assigned to heterozygotes against two homozygotes). The example cited for this type of selection by Magori and Gould (2006) is described here. Homozygous individuals containing alternate forms of a chromosomal translocation marry, heterozygotes are fit because they have single copy of all genes from both parents and when these heterozygotes marry each other they give birth children among them a fraction of them lack important trans locational segment and not viable. If we observe this fitness phenomenon over generations this can be called as underdominance or selection against heterozygotes.

3) Selection with the co-dominant allele (alleles are co-dominant and one allele (heterozygotes) is favoured). Example for this type of selection is ABO blood group. Co-dominance means no allele can mask the expression of other allele. If an individual inherits B allele from mother and A allele from father both are expressed, he/she has AB blood group. In AB blood group both are co-dominant alleles and in heterozygous state which favoured

4) Selection against dominant allele (expressed as heterozygote). Achondroplasia is an example for this type of selection. This is an abnormality of bone growth occur in 1 in 20-30,000 live births. This abnormality is caused by mutation in fibroblast growth factor receptor 3 (FGFR3) gene. It is inherited as autosomal dominant pattern. The patients of Achondroplasia are characterized by short stature, short arms and legs, fingers during extension appear as trident, unusual large head with a prominent forehead, flat nasal bridge and prominent abdomen and buttocks.

5) Selection against recessive homozygotes (harmful allele is favoured). Example for this type of selection is Tay Sachs disease. Its incidence is 1 in 250-350 people. Most commonly observed in Ashkenazi Jews, South eastern Quebec and among Cajun of Louisiana. It is an autosomal recessive disorder (inheritance of two copies of abnormal gene from both

parents). Homozygotes are characterized by deficiency of hexosaminidase, An enzyme which leads to failure of breakdown of GM2-ganglioside (glycosphingolipids), accumulation of GM2-ganglioside and progressive deterioration of central nervous system. The disease is due to mutation in hexosaminidase subunit alpha gene (HEXA) and 80 mutations are observed in this gene. Cherry red spot in the macula of eye is characteristic symptom of this disease. Eye, ear, brain and cognitive abnormalities are observed in these patients and life threatening complications develop by the age of 15 years. Individuals receiving abnormal gene from one parent is called carrier who don't develop the disease.

5.1.3.6 Modes or Patterns of Selection

Patterns or modes of natural selection are of four types. 1. Stabilising selection 2. Directional selection 3. Disruptive/ diversifying selection. 4. Balanced Selection.

5.1.3.6.1 Stabilising Selection

This type of selection takes place when both the environment and the genetic composition of population are stable. Intermediate or heterozygote phenotype or allele is favoured and two extreme phenotypes or alleles are eliminated. Examples of stabilising selection are 1. height distribution in a population. 2. Based on birth weight new born are categorised into low birth weight (<2500g), high birth weight (>4500g) and average birth weight (2500-4500g). Two extreme birth weight newborns have less fitness, high morbidity and mortality and therefore natural selection favours average birth weight babies.

5.1.3.6.2 Directional Selection

This type of selection occurs when there is change in the environment in a particular direction and shift takes place in mean of the character of a distribution. Extreme phenotype or allele that has higher fitness and ability to produce more offspring is favoured and fixed while eliminating the less fit phenotype or allele. Examples are increased body size in response to cold environment, antibiotic resistance in microbes due to the development of new mutant and change of Biston betularia (peppered moth) from light colour in pre-industrial era to black colour in post industrial era to protect from predators.

5.1.3.6.3 Disruptive/ Diversifying Selection

Disruptive selection happens when homogenous populations are separated into different adaptive groups. Extreme phenotypes or alleles are favoured and average phenotype or allele is eliminated. Fifteen species of Passeriforms (order of birds) or Darwin's finches (Figure 5.3) (song birds) found on the Galapagos archipelago and Cocos Island of Pacific Ocean under administrative control of Ecuador are closely related but differ in the size or shapes of their beaks due to adaptation to consuming different food sources.

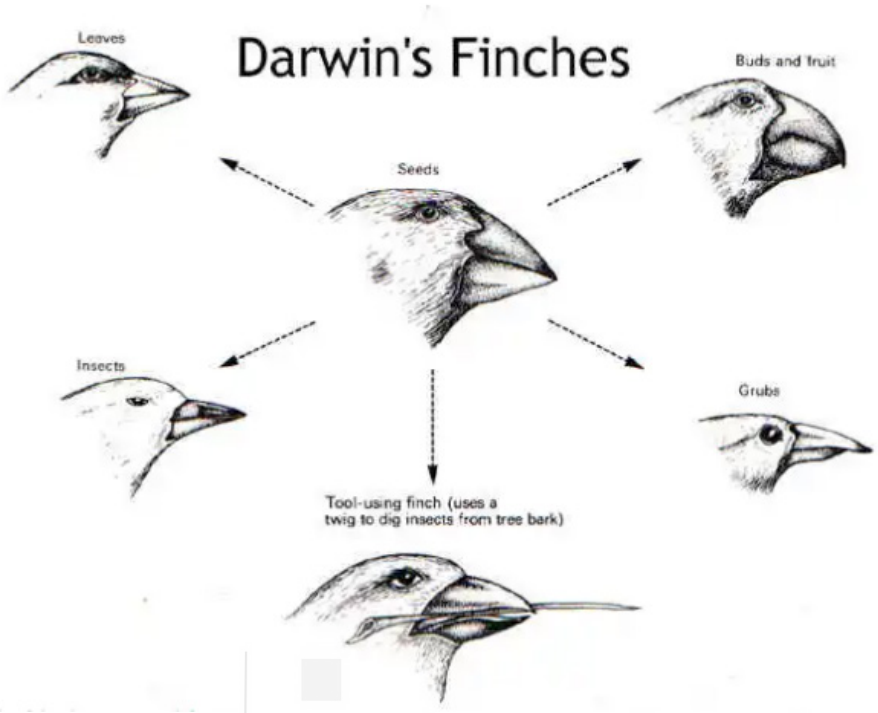


Fig. 5.3 : Darwin's Finches

(***Source:*** https://shrabontibhowmik.wordpress.com/2016/10/02/darwins-theory-of-natural-selection-and-how-did-it-start/)

5.1.3.6.4 Balanced Selection

Heterozygotes than homozygotes of either spectrum have higher fitness, are favoured and contribute more offspring to the next generation. This type of selection is also called heterozygote advantage or over-dominance. The best example for this type of selection is advantage of sickle cell trait in areas endemic to *Plasmodium falciparum* malaria. The details of sickle cell trait have been discussed under the example of natural selection in this unit.

To investigate the selection in human populations, Crow formulated an index. This index has mortality and fertility components. These are influenced by several factors such as variation in fertility, age at marriage, menarche, age of death and survival to reach to fertility age. To account for fertility and mortality during conception, Johnston and Kensinger (1971) have proposed modified Crow index to measure selection in the population. Studies conducted on selection intensity among tribes of India showed higher index of mortality compared to fertility and in urban populations due to public health facilities and higher socio-economic conditions decreased index of mortality and fertility was observed.

5.1.4 Genetic Drift

The National Human Genome Research Institute, Maryland, United States of America, defined genetic drift as random fluctuations in the frequencies of alleles from generation to generation due to chance events. Genetic drift is of two types and they are Bottle neck effect and Founder effect.

2) What is Genetic Drift?

.....

.....

.....

.....

5.1.4.1 Bottleneck Effect

Pandemic like flu, severe acute respiratory syndromes, human immunodeficiency virus, COVID-19, natural calamities like tsunami, earthquakes and droughts and man- made events like exploding of nuclear and hydrogen bombs reduce to the effective size of the population and change genetic diversity. Effective size of a population refers to those interbreeding individuals who by mating contribute offspring to next generation. As a result of reduction in effective size of population the allele frequency will be different before the action of genetic drift which resembles the neck of the bottle restricting the flow of genes. Therefore, this is called bottle neck effect (Figure 5.4). Bottle neck effect reduces the genetic diversity and change the allele frequency. The example for bottle neck effect is Greenlandic Inuit. Analysing exome data of 18 individuals it was proposed that population reduction has been affecting the Inuit over the 15,000 years. Alleles that are rare (variant in TBC1D4 associated with type 2 diabetes and indel in SEMA4C linked to sucrose isomaltose) in East Asian or European population were found to be common in Inuit (Prohaska et al., 2019).

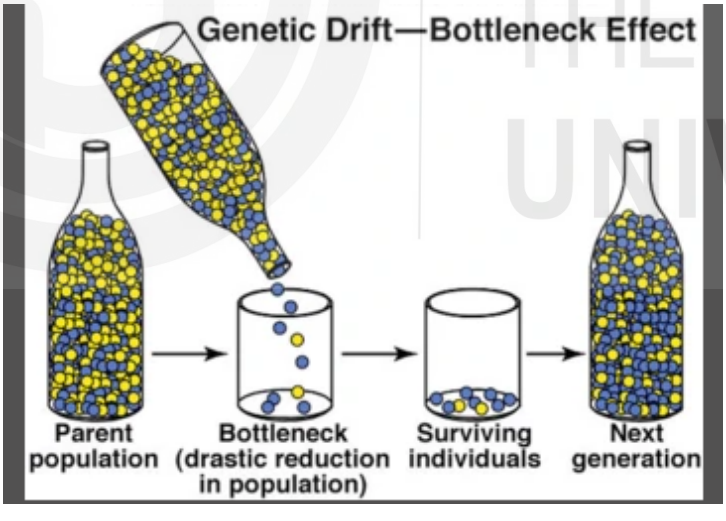


Fig. 5.4 : Bottleneck Effect

Source: https://dragonflyissuesinevolution13.wikia.org/wiki/Bottleneck_Effect)

5.1.4.2 Founder Effect

Founders are early settlers or ancestors who established new colony or population in foreign lands. The founding of new colony or subpopulation may also take place due to exploration of island or new area, warfare, survival from ship wreck, repeated migration over time or waves of migration from their lands to new landscapes. The founders of subpopulation carry random sample of genes. The new subpopulation after a few generations may be characterised

by the absence of alleles or higher prevalence of rare alleles against the larger population (Figure 5.5). Repeated migrations of humans during different time periods establish new subpopulation whose genetic composition will be different from the original population. This is called serial founder effect. Examples of founder effect are continental distribution of Y chromosome and mitochondrial haplo groups and tracing of genetic signature of African people in the population of America, Europe and South Asia, support evidence on out of Africa origin of human beings. Unusual frequency of morphological and genetic traits in north Indian population such as higher frequency of lactose malabsorption, lack of isoenzyme ALDH-1, AK2, pc, K, cde and A2 genes; high frequency of G6PD deficiency in Naga and Gd variant in Bodos and absence of Gd- variant in Adi and Hmar populations are some of the examples of founder effect in India (Vasulu, 2012).

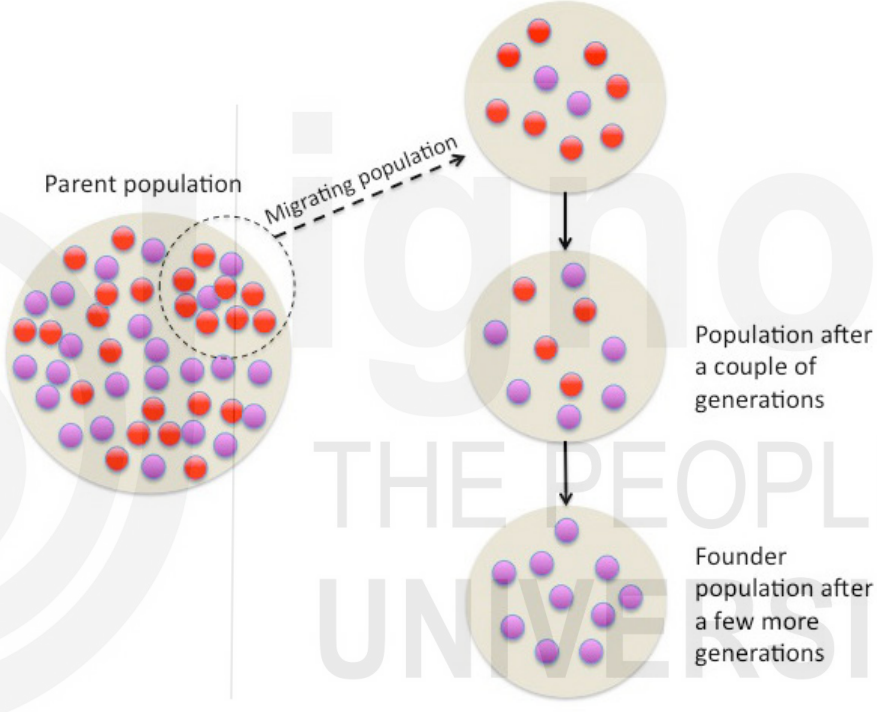


Fig. 5.5 : Founder Effect

(***Source:*** https://commons.wikimedia.org/w/index.php?curid=47951519)

5.2 SUMMARY

Evolution can be understood as changes taking place over time and include small or large, invisible or visible and non-adaptive and adaptive changes. The four forces of evolution are known and they are mutation, gene flow, natural selection and genetic drift. Mutation introduces new alleles into the population due to the occurrence of copying errors in DNA replication and transcription. Natural selection (due to the individual differences on the survival and reproductive success) and genetic drift (due to random sampling in small populations) differentially transmit alleles into the next generation. Mutations may cause harmful, beneficial and neutral effects. To explain the change in allele frequency contributed by gene flow, Sewall Wright proposed 'island' and 'isolation by distance' models whereas Kimura and Weiss contributed

'stepping stone model'. Natural selection is defined as differential survival and reproductive success. Natural selection determines the fate of newly added allele in the population. If it is not advantageous or not suitable to the environment, natural selection eliminates it by negative or purifying selection. Selection is of five types and they are (1) Selection for the heterozygotes (heterozygotes have higher fitness than homozygote) (2) Selection for against heterozygotes (lower fitness assigned to heterozygotes against two homozygotes) (3) Selection against codominant alleles (alleles should be codominant and one allele (heterozygotes) is favoured) (4) Selection against dominant alleles (expressed as heterozygote) and (5) Selection against recessive homozygotes (harmful allele is favoured). Patterns or modes of natural selection are (1) Stabilising selection (2) Directional selection (3) Disruptive/ diversifying selection. (4) Balanced Selection. Genetic drift is a random fluctuation in the frequencies of alleles from generation to generation due to chance events. Genetic drift is of two types and they are (1) Bottle neck effect and (2) Founder effect.

5.3 REFERENCES

Achondroplasia. Available at https://rarediseases.org/rare-diseases/achondroplasiaAbzhanov,

A. (2010). Darwin's Galapagos finches in modern biology. Philos Trans R Soc Lond B Biol Sci., 365(1543), 1001-1007

Ali Al-Barazanchi, Z.A., Abdulateef, S.S., & Hassan, M.K. (2021). Co-Inheritance of α-thalassemia gene mutation in patients with sickle cell disease: Impact on clinical and hematological variables. Niger J Clin Pract, 24, 874-82

Dean, L. (2005). Blood groups and red cell antigens [Internet]. Chapter 9, The Duffy blood group. Bethesda (MD): National Center for Biotechnology Information (US). Retrieved from: https://www.ncbi.nlm.nih.gov/books/NBK2271.

Defining evolution. National Centre for Science Education. Retrieved from https://ncse.ngo/defining-evolution.

Dey, S., & Kapoor, A.K. Paper no.14 Human origin and evolution. Module no.10 Forces of evolution pp 1-12. ePG Pathashala , Anthropology.

Dickson, E.T., & Hyman. P. (2013). Brenner's Encyclopedia of Genetics (2nd ed.). Retrieved from https://www.sciencedirect.com/topics/medicine-and-dentistry/silent-mutation.

Franceschini, G., Sirtori, C.R., Capurso, A., Weisgraber, K.H., & Mahley, R.W. (1980). A-IMilano apoprotein, decreased high density lipoprotein cholesterol levels with significant lipoprotein modifications and without clinical atherosclerosis in an Italian family. J. Clin. Invest 66 (5), 892–900

Galton, D.J., Mattu, R., Needham, E.W., & Cavanna, J. (1996 June). Identification of putative beneficial mutations for lipid transport. Z Gastroenterol, 34 Suppl 3, 56-8

Genetic drift. Retrieved from https://www.genome.gov/genetics-glossary/Genetic-Drift.

Gerbault et al. (2011 March 27). Evolution of lactase persistence: An example of human niche construction. Philos Trans R Soc Lond B Biol Sci., 366(1566), 863-77

Jeanmonod, R, Asuka, E., & Jeanmonod, D. (Updated 2021 July 20). Inborn errors of metabolism. StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing. Retrieved from: https://www.ncbi.nlm.nih.gov/books/NBK459183/

Julian, C.G., & Moore, L.G. (2019). Human genetic adaptation to high altitude: Evidence from the Andes. Genes (Basel), 10(2), 150

Kameih, G. Paper no. 8 Human population genetics. Module no.1. Definition and scope of population genetics. pp1-13. ePG Pathashala, Anthropology.

Langhi, D.M. Jr., & Bordin, J.O. (2006 October). Duffy blood group and malaria. Hematology, 11(5), 389-98

Libert et al. (1998 March). The deltaccr5 mutation conferring protection against HIV-1 in Caucasian populations has a single and recent origin in North-eastern Europe. Hum Mol Genet, 7(3), 399-406

Lodh, M., & Kerketta, A. (2013). Inborn errors of metabolism in a tertiary care hospital of Eastern India. Indian Pediatr 50, 1155–1156.

Magori K & Gould F. (2006). Genetically engineered underdominance for manipulationofpestpopulations:adeterministicmodel.Genetics.172:2613-2620. doi:10.1534/genetics.105.051789.

Mikkola et al. (1997 February). Molecular mechanism of a mild phenotype in coagulation factor XIII (FXIII) deficiency: A splicing mutation permitting partial correct splicing of FXIII A-subunit mRNA. Blood, 15; 89(4), 1279-87

Modern human diversity- Skin color. Retrieved from https://humanorigins.si.edu/evidence/genetics/human-skin-color-variation/modern-human-diversity-skin-color.

Prohaska et al. (2019 March 21). Human disease variation in the light of population genomics. Cell, 177(1), 115-131

Singh, S., & Kshatriya, G.K. Paper no. 8 Human population genetics. Module no.5. Mutation and genetic load, pp.1-20. ePG Pathashala, Anthropology. Paper no. 8 Human population genetics. Module no.12. Models of Natural Selection pp.1-26. ePG Pathashala, Anthropology.

Spencer, H.G. Population genetics. Retrieved from https://www.eolss.net/sample-chapters/c02/E4-31-05-04.pdf

Sudi, I.Y., & Ali-Dunkrah, U. (2005). Mutation and its role in biotechnology. Nig J Bioetch, 16 (1), 1-29

Taysachs disease. Available at https://rarediseases.org/rare-diseases/tay-sachs-disease/.

Vasulu, T. (2012). Hardy-Weingberg equilibrium. MANE-001. Block 2. Human population genetics, Unit 2. (2012). Meaning and scope of population genetics. MANE-001. Block 2. Human population genetics, Unit 1.

Waters et al. (2018 December). Global birth prevalence and mortality from inborn errors of metabolism: A systematic analysis of the evidence. J Glob Health, 8(2), 021-102

5.4 ANSWERS TO CHECK YOUR PROGRESS

1) Duffy blood group is one of the 45 blood groups identified in human beings. Duffy blood group individuals lacking Duffy antigens are resistant to *Plasmodium vivax* induced malaria. Refer to sub-section 5.1.3.4.

2) It is a random fluctuation in the frequencies of alleles from generation to generation due to chance events. Refer to sub-section 5.1.4.

UNIT 6 GENETIC POLYMORPHISMS

Contents

Learning Objectives

After reading the unit, you will be able to:

- Define the concept of genetic polymorphism;
- Explain genetic polymorphism with respect to serological, biochemical and molecular markers;
- Discuss the use of studying polymorphic markers;
- Understand the concept of balanced and transient polymorphism; and
- Explain models contributing to maintenance of genetic polymorphism.

Contributed by **Dr. Ketaki Chandiok**, PhD (Anthropology), Delhi

6.0 INTRODUCTION

Before we begin with the discussion on genetic polymorphism, it is essential that we understand what a gene is? A gene can be defined as a segment of the DNA that specifies the sequence of amino acids in a particular protein. Throughout the life cycle of a cell, it is the DNA that directs the cellular functions and exists in an uncoiled granular form. However, during the life cycle of the cell, the normal activities of the cell might cease and the cell divides. Cell division results in the production of new cells. At this stage, the DNA is highly coiled and visible under a microscope as a discrete structure called chromosomes. During the early stage of cell division, when the chromosomes become visible, they are made up of two strands or two DNA molecules that are joined together at a constricted area called the centromere. Chromosomes are present in identical sets of two (or in pairs). Humans have 46 chromosomes or 23 sets of chromosomes. Of these, 22 pairs are autosomes and other two are sex chromosome that is X and Y. The sex chromosomes determine the sex that is either male or female. The autosomes are responsible for all physical characteristics of an individual except primary sex determination.

Human cells contain around 28000-30000 genes (Deloukas et al., 1998). These genes code for important and necessary information that determine molecular traits that are passed on from parents to their offspring. Genes encode various traits like hair color, eye color, skin color, hair texture, etc. Any change in the DNA sequence brings about change in the genetic information, which brings about change in the phenotypic expression and also the associated biological function. The changes in the DNA sequence is known as a mutation. Physical anthropologists are concerned in understanding visible human variations as they are interested not only in identifying the factors that produce visible physical variation but also the underlying genetic determinants that dictate it. Genetic variations arise due to the differences in the DNA sequence among populations from the wild type form. Each and every individual has two sets of genomes, one maternal and one paternal. Therefore, at each genetic location (locus), the alleles from the maternal and the paternal side, can either have identical DNA sequence or slightly differing DNA sequence. The wild type form in a population refers to individuals with normal phenotype. The wild type form is usually possessed by the majority of the individuals in the population. In contrast to this, mutant type refers to individuals with a phenotype that varies from the normal population. These variations can also be referred to as homozygous if the alleles on both the chromosomes are identical or heterozygous if they differ on any one of the chromosomes.

6.1 GENETIC POLYMORPHISM

The term "polymorphism" is a combination of two Greek words "poly" meaning multiple and "morph" meaning form, can be defined as a mendelian trait, which exists in a population, in at least two different forms. Ford, 1940 defines genetic polymorphism refers to the occurrence together in the same habitat of two or more discontinuous forms or phases of a species in such proportions that the

rarest of them cannot be maintained by recurrent mutations. In simpler words, genetic polymorphism refers to the occurrence in the same population of two or more than two alleles at the same locus in the same population, such that the frequency of the rarer allele is always greater than one percent and the rarer allele is maintained in the population, not merely be recurrent mutations (Cavalli-Sforza and Bodmer, 1971). In the nutshell, polymorphisms can be defined as the variations in the DNA sequence that are present in the population, with the frequency of the variation being greater than 1 percent. In other words, it can be said that mutation frequency is more than 1 percent in a population, it is a polymorphism. Insertions-deletions polymorphisms, single nucleotide polymorphisms, restriction site polymorphisms or restricted fragment length polymorphism etc. are some of the examples of genetic polymorphisms.

Check Your Progress

1) What is Genetic Polymorphism?

.....

.....

.....

.....

Most of the polymorphic traits studied are known to be genetic in nature. The polymorphic traits are inherited in a Mendelian fashion. In case a population has more than one version of the same trait, then the population is said to be polymorphic for that trait, such as blood group type A, type B and type O. Genetic polymorphisms confer diversity within a population and hence are used to study population diversity, migration patterns and relationship between and within various populations and also the potential affinities of populations.

A few genetic markers which show polymorphism are presented below:

6.2 SEROLOGICAL MARKERS

Blood groups are the best examples of serological markers. Blood groups (that is ABO and Rh) are one of the most important serological markers that play an important role in blood transfusion and organ transplantation. The ABO blood group system discovered by Landsteiner is one of the examples of tri-allelic inheritance. These markers follow Mendelian inheritance. The genetic system for the ABO blood group is located in chromosome number 9 and has three major alleles, A, B and O. In this, the A and B allele behaves as co dominant and O as recessive. The phenotypes A, B, AB and O are present in all populations but vary in proportions in different parts of the world. The ABO blood group is one of the most studied genetic polymorphism systems in the Indian subcontinent with B allele frequencies being highest for most of the populations in the Indian subcontinent.

The Rh system is one of the most complex polymorphic systems in humans. The antibody was discovered by Levine and Stetson in 1939. Later, it was reported

to be located on chromosome number 1. The red blood cells are designated as Rh+ and Rh- depending upon the presence or absence of major D antigen. The antigens for the Rh system are defined as products of three loci:C,D and E and the antigens C,c,D,d,E and e have been studied extensively. For the Rh system, high frequency of D allele has been reported, with the CDe haplotype being the most common.

The blood group systems are important for understanding population diversity and human genetic ethnic ancestry.

6.3 BIOCHEMICAL MARKERS

Individuals in a population are known to differ on the basis of biochemical markers for example human red cell enzymes like adenylate kinase, phosphoglucomutase, G6PD and serum proteins like haptoglobin, transferrin, immunoglobulins etc., hemoglobins and thalassaemia etc.

Glucose 6 phosphate deficiency is one of the most well studied polymorphisms. The G6PD gene is located on the long arm of the X chromosome (Xq^28) region chromosome. It is a 18 kb long gene consisting of around 13 exons, transcribed to a 2.269 Kb messenger RNA with 1.545 Kb coding region. The G6PD deficiency in the male subjects can be detected using the simplest fluorescent spot test by Beutler and Mitchell, 1968. The method relies on fluorescence of NADPH, generated by an amount of G6PD enzyme. In Hong Kong, a routine screening for G6PD deficiency is done for newborns. The commonest variant of G6PD, the G6PD Canton, has been reported from South China, where it has been sequenced and was reported to be due to a mutation in the nucleotide position 1376 of the cDNA, G to T, resulting in a missense mutation at amino acid position 459, for Arg to Leu (Cavalli-Sforza and Bodmer, 1971).

6.4 MOLECULAR MARKERS (DNA POLYMORPHISM)

A DNA polymorphism is a sequence difference compared to a reference standard that is present in at least 1-2% of a population. The genetic polymorphism occurs in the DNA sequences in numerous ways. There can be different types of genetic polymorphisms, depending upon the number of nucleotides involved in the polymorphism and also the reason about how those variations have occurred.

DNA polymorphism can be studied in the following forms:

6.4.1 Single Nucleotide Polymorphism (SNP)

These are the most common type of all the polymorphisms. Human genome, that is the complete genetic material in a cell consists of about 3 billion base pairs. SNP occurs in every 300 nucleotides, meaning that there are around 10 million SNPs that are present in the human genome. The nucleotide base pair substitutions that occur can be between the purines (that is A, G) or pyrimidine

(C, T) are known as transitions. Transitions make up two thirds of all the SNP. Another type of base pair substitution occurs between a purine and a pyrimidine. This is known as a transversion. The single nucleotide polymorphisms occur either in the coding region of the genes or its non-coding region. The SNP that occurs within the coding region can be either synonymous or non-synonymous. The synonymous changes do not affect the amino acid sequence whereas the non-synonymous changes the sequence of the amino acid. Further, the non-synonymous SNPs can be either missense or nonsense. Missense mutation refers to the change in the single base pair that causes substitution of some different amino acids. In contrast to this, non- sense mutation refers to the change in the DNA sequence, leading to a premature stop codon or producing a non-functional, incomplete and truncated protein.

6.4.2 Restriction Fragment Length Polymorphisms (RFLP)

Restriction site polymorphisms (RSPs) are a subset of the SNPs that cause either a loss or gain of a restriction site. A restriction site can be defined as a location where a particular restriction enzyme cuts the DNA sequence, at a particular DNA sequence location, resulting in pieces of DNA. The change in the nucleotide sequence at a particular site location from the original one, enables the enzyme to cut the DNA into pieces. This results in creation of a different length of DNA piece in an individual and another DNA piece in some another individual. This is restriction site polymorphisms. The RSPs can also be described as restriction fragment length polymorphism (RFLP).

Expression SNPs refers to the SNPs present in the splicing regions of the gene that can affect the gene splicing, lead to m RNA degradation, or the non-coding RNA sequences. They can occur either upstream or downstream of the gene.

6.4.3 Variable Number of Tandem Repeats (VNTR)

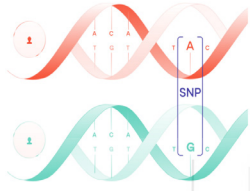
These are arrays of 2 or more base pair core units, located in the non- coding region of the genome, adjacent to each other. Variable number of tandem repeats refer to a condition where the number of nucleotides in the core unit is variable or is not known. On the basis of the size of the core unit, they can be categorized as either

a) mini-satellites (10-60 bp) is a collection of moderately sized arrays, usually 10-60 base pairs of tandemly repeated DNA sequences that are dispersed over considerable portions of the nuclear genome. For example: GATACCCCAAAG GATACCCCAAAG GATACCCCAAAG is an array of 12 nucleotide repeats from 3-20 kbp.

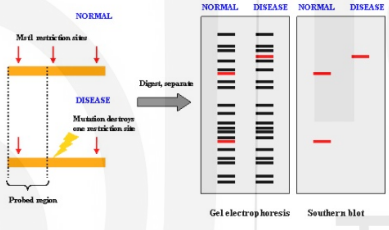
b) microsatellites (<10 bp) or short tandem repeats is a small array of tandem repeats of a simple nucleotide sequence which is usually less than 10 base pairs. For example, GATA GATA GATA GATA GATA GATA has 4 bases repeated 6 times. TA TA TA TA TA TA is a dinucleotide repeat and TAT TAT TAT TAT TAT is a trinucleotide repeat. The STRs were first used in the Persian Gulf War in 1991 for identification of human remains.

6.4.4 Insertion Deletion Polymorphisms

These refer to genetic variations in which a sequence of DNA is either inserted or deleted from the gene. The frequency of the occurrence of the insertion deletion polymorphism is only about one tenth of the frequency of the SNPs. Studies have shown that around 90 percent of the insertion deletions are of 1-10 nucleotides only. Around 9 percent have sequences involving 11 to 100 nucleotides, whereas only 1 percent have sequences greater than 100 nucleotides (Mullaney et al., 2010). Alu Indels is one of the most widely studied insertion deletion polymorphism for human evolutionary and diversity studies, because of their known ancestral state, identification by descent, a wide and common occurrence and stability.

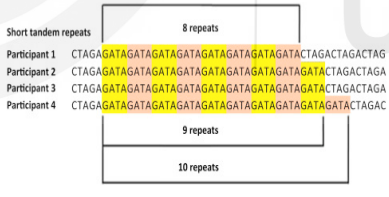


SNP (https://medium.com/@sanogenetics/snp-of-the-week-58e23927c188)



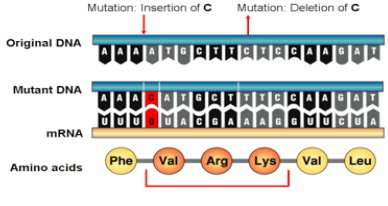
Restriction Fragment Length Polymorphisms (RFLP) or Restriction site polymorphisms (RSPs)

(https://www.ncbi.nlm.nih.gov/probe/docs/techrflp/)



Variable Number of Tandem Repeats

(https://www.sciencedirect.com/topics/immunology-and-microbiology/short-tandem-repeat)



Insertion Deletion (https://www.pathwayz.org/Tree/Plain/BASE+DELETIONS+%26+INSERTION)

Fig. 6.1 Various forms of Genetic Polymorphism

6.4.5 Lineage Markers

6.4.5.1 Mitochondrial Markers

DNA in humans is not present only in the nuclear regions of the cells, but also the mitochondria of the human cells and is known as the mitochondrial DNA (mtDNA) consisting of 16,569 nucleotides. The mtDNA is transmitted from the mothers only and hence both males and females receive it through the mothers only. Mitochondrial DNA can be obtained from hair shaft, skeletal remains etc, which are poor sources of nuclear DNA Mitochondrial markers present on the mitochondrial haploid genome are used for tracing the maternal ancestral lineage. This is because of their uni-parental mode of inheritance. The mitochondrial genome consists of multiple copies of mt DNA.

6.4.5.2 Y Chromosomal Markers

They can be used for tracing the paternal ancestral lineages. The Y chromosome is transmitted unchanged from one generation to the next. Few changes occur, but they do not produce any effect as 98% of the DNA is non-coding.

6.5 BIOCHEMICAL POLYMORPHISMS VERSUS DNA POLYMORPHISMS

The traditionally studied polymorphisms are the serum protein polymorphisms, blood group polymorphisms and red cell enzymes etc have been studied for long. These polymorphisms are in the functional region of the genome and are under the influence of natural selection and therefore, there is restricted variation in it. In contrast to this, the DNA polymorphisms are found in the nonfunctional regions of the genome and have much higher levels of variation, making DNA polymorphisms extremely useful for population genetic studies. There are numerous polymorphisms that are known at the level of DNA. Therefore, DNA polymorphism helps in study of haplotype variation, which is more informative in human population genetic studies.

6.6 APPLICATIONS OF GENETIC POLYMORPHISM

The applications of genetic polymorphisms in anthropology are numerous. A few of them are listed below:

1) Understanding the population history and its affinities with other populations. For understanding the affinities, the DNA sequence of the individuals from different populations is studied.

2) Understanding and inferring the demographic histories of the populations.

3) Understanding the mutation patterns and its reconstruction.

4) Dating occurrences of the mutations in the populations.

5) Mapping of the disease genes and risk prediction.

6) Tracing the trials of the disease and other genes.

7) Identification of genetic matches for organ and marrow donation.

8) Understanding the influence of genetic factors on drug response and metabolism like explaining the underlying unexpected drug effects, clarifying the understanding of the molecular mechanisms, evaluating the clinical relevance etc.

9) Identification of an individual and paternity testing.

6.7 SIGNIFICANCE OF GENETIC POLYMORPHISM

Genetic polymorphisms are known to affect the protein functions and play an important role in human health. Genetic polymorphism plays an important role in the understanding of the biochemical pathways. It is important to note that the human polymorphisms are present in almost all the racial groups known, although the frequency may vary between different populations. An understanding of the gene frequency helps one to know about the evolutionary past and also helps in the prediction of the biological future.

The polymorphic markers can also be used in personal identification. In addition to this, the polymorphic markers also help in paternity testing of the children or fetus, identification of forensic samples in murder cases, rape or assault cases. The polymorphic markers also find utility in population studies as well. With technology advancement numerous polymorphic markers have been reported and the gene frequency data are used for studying the evolution of the human race. Previously, classical serological and biochemical markers were studied. The practical utility of these markers was limited due to the limited number of possible genotypes available at each loci. However, the discovery of the hyper variable DNA loci, tries to ameliorate this problem. Here, it is important to note that not the study of one or two loci, but understanding the distribution of a large number of genetic loci can be used to understand the genetic relationship among populations. While understanding the genetic relationship between populations, it is necessary that proper markers are selected and appropriate statistical tools are used. Recording population variation is necessary as it helps in understanding the mechanism causing variation and also helps in understanding the molecular basis of diseases.

6.8 BALANCED AND TRANSIENT POLYMORPHISM

6.8.1 Balanced Polymorphism

Balanced polymorphism is characterized by the presence of both the alleles for the gene, rather than having two copies of a single allele alone. Here both the versions of the alleles are maintained. In nature, balanced polymorphism occurs

when there is a stable frequency of the alleles overtime due to selection. This type of polymorphism is characterized by heterozygous condition and brings about heterozygote advantage. Individuals that carry both the versions of the alleles are more fit and are able to survive and reproduce, as compared to those having two copies of the either form. Balanced polymorphism results when an allele that is harmful in the homozygous state may produce a heterozygote, the reproductive fitness of which exceeds the homozygote. Hence, balanced polymorphism is when natural selection favors heterozygote over both the homozygous recessive and dominant. The superiority of the heterozygous over the homozygotes is responsible for maintenance of balanced polymorphism in a population and is referred to as heterozygote superiority or overdominance or single locus heterosis. E.B. Ford coined the term balanced polymorphism to the evolutionary process that maintains the heterozygote over a long period.

Balanced polymorphism neither allows elimination of alleles nor its fixation. In balanced polymorphism, the heterozygote is fitter than the mutant or the wild type homozygous and is able to survive and hence is selected. Therefore, the mutation is maintained in an intermediate frequency and it is not fixed. Balanced polymorphism is an important factor that is known to steadily maintain the polymorphism in a population for a long period of time and hence is important in conferring genetic variability to a population. This allows the individuals in a population to adapt to the rapidly changing environmental conditions.

A few examples of balanced polymorphism are presented below:

Sickle Cell Disease

Sickle cell disease is an example of balanced polymorphism. Sickle cell disease is caused through a single nucleotide mutation in the beta globin gene. This causes the production of abnormal hemoglobin characterized by sickle shaped red blood cells. These sickle shaped red blood cells are fragile and clog the blood vessels resulting in organ damage, severe anaemia and heart failure. This autosomal recessive disorder is characterized by joint pain, swollen spleen, and severe infections.

Malaria is known to be quite common in the tropical regions of Africa. The heterozygous carriers of sickle cell condition are resistant to malaria (a condition caused by *Plasmodium falciparum* parasite; characterized by chills and fever). Those who are heterozygous for the sickle cell trait, have a mixture of both the normal erythrocytes as well as the sickle shaped cells. In contrast to these, the homozygous carriers of the sickle cell gene have a higher frequency of the sickle cell shaped red blood cells in the population. The work by A.C. Allison in 1954 reported that in spite of the deleterious characteristics of the homozygote, the sickle cell mutation is maintained in the population as the heterozygote possesses a selective advantage, which is of being malaria resistant. Therefore, the heterozygote advantage of the sickle cell red blood cells, provides a protective effect as the malarial parasite cannot grow in these cells. Therefore, in areas where malaria is prevalent, those with sickle cell heterozygote condition, tend to survive and increase in frequency over the normal homozygotes as they are susceptible to malaria.

In this context, sickle cell disease provides a classic example of balanced polymorphism, where heterozygotes have a state of co-existence with the disadvantageous homozygotes in the population. Here, balanced selection or polymorphism results through a balance of the selection against the homozygous sickle cell sufferers and selection against the normal homozygous due to malaria. Therefore, there is an increased fitness as the heterozygote survives at a rate much higher than the mutation rate.

G6PD Deficiency

G6PD deficiency is another example of balanced polymorphism, where natural selection acts in two opposite directions. G6PD deficiency is a sex linked enzyme deficiency. The deficiency is associated with hemolytic anemia, which is a life threatening condition. The condition is associated with certain environmental conditions like inhaling pollen, eating fava beans or certain drugs, or contracting some infections. Studies from Africa have reported that hemi zygous males and heterozygous females for the G6PD enzyme deficiency are at a less risk for having malaria. Since both the heterozygous and hemi zygotes are selected for, the mutant allele should predominate in the malaria exposed environment. However, this does not happen. This is because of natural selection. Those with enzyme deficiency, that is hemizygous males and homozygous females are selected out of the population due to anaemia. In this context, natural selection acts in two directions with the hemizygous males selecting for the mutant alleles as it protects against malarial infection and selecting against due to enzyme deficiency. This is balanced polymorphism.

Other examples of balanced polymorphism include beta thalassemia, ovalocytosis, Tay-Sachs Disease, cystic fibrosis, PKU etc.

6.8.2 Transient Polymorphism

Transient polymorphism is when there is a progressive replacement of one allele by another allele of the gene. The rare allele starts gaining an overall advantage, thereby spreading in the population. This will reduce the frequency of the normal allele. Here, one of the alternate forms of the allele is progressively replaced by another due to inheritance. This happens under the influence of strong environmental selective pressure, which eliminates one of the alleles from the gene pool. Transient polymorphism is when there is changing gene frequency over time. Herein a form is gradually replaced by another one. It can be thought of as a temporary situation that can be characterized as a by product of directional selection. In case of transient polymorphism, the mutation gradually fixes itself due to selection, and there is an elimination of the alternative alleles, which can lead to an increase in its frequency overtime.

A classic example of transient polymorphism is the phenomenon of industrial melanism occurring in the moth species of Europe and the United States. E.B. Ford, a British ecological geneticist, was the first one to demonstrate the phenomenon of industrial melanism as an effect of natural selection. The moth *Biston betularia*, undergoes a mutation at a single locus to produce dark and melanic form. The mutant allele in this case is dominant and any gamete

containing this mutant allele produces a melanic form. In 1848, the first melanic form was found in a collection from Manchester. However, by 1895, around 95% of the collected forms were dark forms, also referred to as *carbonaria*.

During the 1950s, H.B.D Kettlewell, through a series of 12 observations and mark recapture, demonstrated that the light and dark forms of the moth species, was preyed upon differently by the birds. Kettlewell findings reported that the birds selectively caught and ate more moths of the form that did not match the background. In the industrialized areas of England, the walls and the tree trunks on which the moths used to rest were darkened by the pollutants from the factories. Due to this, the *carbonaria* possessed a selective advantage. In contrast to this, in the rural areas, the light form enjoyed a selective advantage. The carbonaria form produced by recurrent mutations, protected it from predation by birds. The light form still exists in rare numbers in nature, highlighting that the selection takes a great amount of time to eliminate the recessive allele.

Fig. 6.2 : Transient Polymorphism (https://ib.bioninja.com.au/higher-level/topic-10-genetics-and-evolu/103-gene-pools-and-speciati/polymorphisms.h)

Check Your Progress

2) What is Balanced and Transient Polymorphism?

.....

.....

.....

.....

6.9 MODELS EXPLAINING MAINTENANCE OF GENETIC POLYMORPHISM

6.9.1 Relationship between-sickle Cell Anemia and Malaria

Malaria along with the sickle cell anaemia condition is high in tropical regions of Africa. The sickle cell disease is less common in Caucasians due to less occurrence of malaria in the region. Carriers with the heterozygous sickle cell trait are more resistant to malaria, than the normal homozygotes. It is a well known fact that malaria is quite common in the tropical regions of Africa.

Individuals with the sickle shaped red blood cells have a selective advantage as the malarial parasite *Plasmodium falciparum* cannot grow in these cells. The parasite spends the first part of its life cycle in the mosquito *Anopheles gambiae* salivary glands. When the infected mosquito bites an individual, the parasite enters the red blood cells and is transported to the liver. The blood cells burst releasing the parasite throughout the body. Cultural development along with the emergence of agriculture lead to an increase of *Anopheles* mosquitoes owing to the breeding ground provided due to cultivation of crops. The adaptation to agricultural mode of subsistence is one of the key determinants explaining the occurrence of the malarial parasites in the red blood cells of the humans. Those with sickle cell diseases can change their intracellular environment, so that the parasite never develops in them. In this context, the normal individuals have higher mortality and lower fertility rates in malarial environments as compared to those with sickle cell disease, thereby, contributing more people to the succeeding generations. Therefore, as the malaria rose, the selective pressures gives rise to changes in the shape of RBC to sickle cell shape from elliptical. Regions with settlements of those with a large number of sickle cell carriers have escaped the devastating malaria incidence. As of today, as well, sickle cell disease is more prevalent in agricultural societies as compared to hunter and gatherer societies.

6.9.2 X linked Polymorphism

There are 4 types of genetic polymorphisms that are X linked, these are: colour blindness, G6PD deficiency, Xg blood group and Xm. Here we will discuss the first two polymorphisms as the last two are not associated with any selective effects, so the forces that maintain them cannot be explained fully.

G6PD deficiency is a X linked recessive trait and occurs largely in populations residing in regions where malaria is prevalent and is associated with resistance to malaria. G6PD is a sex linked enzyme deficiency and is known to affect around 400 million people worldwide. The deficiency is associated with hemolytic anaemia, which is a life threatening condition and is also under the influence of conditions like eating fava beans, inhaling pollen, taking drugs, and catching certain infections. Studies from Africa have reported that hemizygous males and heterozygous females for the enzyme deficiency are less likely to develop malaria, revealing a selective advantage. In areas with malaria attack cells carrying the normal allele are more prone to rapid destruction by the parasite. The G6PD enzyme deficiency is responsible for the lower concentration of a substance called reduced glutathione, which is necessary for the growth of the malarial parasites.

Colour blindness/ colour vision deficiency is a X-linked recessive trait wherein the majority of the colour blind people can see colours barring red, green or blue. Individuals with normal colour vision can distinguish between colours by the addition of three colours i.e., red, green and blue. The red green colour blindness can be of the following main types:

1) Protan (red blind) type with the subtypes of absolute/ strong colour blindness (protanopia) and partial/ mild (protanomalia).

2) Deutran (green blind) type with the subtypes of absolute/strong colour blindness (deuteranopia) and partial/ mild (deuteranomalia).

In most populations there are about three times as many deutan males as protan males. Studies have shown that there are differences between populations for the total frequency of colour blindness. The more primitive populations have a lower frequency, wherein the hunting gathering communities had a prevalence of 1-2 percent, the industrial society had a prevalence of about 7 percent. The colour blindness condition was a disadvantage to the primitive people as it impaired their inability to distinguish between flora and fauna that is essential for livelihood. The negative selection was more so relaxed in modern societies, which explains its higher prevalence in such societies.

Check Your Progress

3) Explain the relationship between Sickle Cell anaemia and Malaria

.....

.....

.....

.....

6.10 SUMMARY

Genetic polymorphism, refers to the occurrence of multiple forms of single allele/gene in an individual of a population, such that the rarest one cannot be maintained by recurrent mutations and the frequency of neither of which is less than 1 percent. The polymorphism can be either serological, biochemical and molecular. The polymorphism that results when the allele is in the process of replacement is transient. As the favoured allele is fixed in the population it becomes monomorphic. Balanced polymorphism is the presence of both the alleles for the gene, rather than having two copies of a single allele alone.

6.11 REFERENCES

Crow, J.F. & Kimura M. (1970). *An Introduction to Population Genetics Theory*. New Jersey: Blackburn Press.

Deloukas, P., Schuler, G. D., Gyapay, G., Beasley, E. M., Soderlund, C., Rodriguez-Tome, P., & Bentley, D. R. (1998). A physical map of 30,000 human genes. *Science*, 282 (5389), 744-746.

Li, C.C. (1976). *First Course in Population Genetics*. USA: Boxwood Press.

Cavalli Sforza, L.L. & Bodmer, W.F. 1971. *The Genetics of Human Populations*. San Francisco. USA: W.H. Freeman,

Ford, E. B. (1940). *Polymorphism and Taxonomy. In The New Systematics*, ed. Huxley, J., pp 493-513. Oxford: Clarendon Press.

Dobzhansky, T. (1951). *Genetics and the Origin of Species*. New York: Columbia University Press.

Gardner, E.D, Simmons, M.J & Snustad, D.P. (2003). *Principles of Genetics,* 8th Edition, New York: John Wiley and Sons.

Simmons, S & Simmons, M.J. (2003). *Principles of Genetics,* 3rd Edition, New York: John Wiley and Sons.

6.12 ANSWERS TO CHECK YOUR PROGRESS

1) Ford (1940) defines genetic polymorphism refers to the occurrence together in the same habitat of two or more discontinuous forms or phases of a species in such proportions that the rarest of them cannot be maintained by recurrent mutations. In simpler words, genetic polymorphism refers to the occurrence in the same population of two or more than two alleles at the same locus in the same population, such that the frequency of the rarer allele is always greater than one percent and the rarer allele is maintained in the population, not merely be recurrent mutations (Cavalli-Sforza and Bodmer, 1971)

2) Balanced polymorphism is characterized by the presence of both the alleles for the gene, rather than having two copies of a single allele alone. Here both the versions of the alleles are maintained. In nature, balanced polymorphism occurs when there is a stable frequency of the alleles overtime due to selection. This type of polymorphism is characterized by heterozygous condition and brings about heterozygote advantage. Individuals that carry both the versions of the alleles are more fit and are able to survive and reproduce, as compared to those having two copies of the either form.

   Transient polymorphism is when there is a progressive replacement of one allele by another allele of the gene. The rare allele starts gaining an overall advantage, thereby spreading in the population. This will reduce the frequency of the normal allele. Here, one of the alternate forms of the allele is progressively replaced by another due to inheritance. This happens under the influence of strong environmental selective pressure, which eliminates one of the alleles from the gene pool.

3) Malaria is known to be quite common in the tropical regions of Africa. The heterozygous carriers of sickle cell condition are resistant to malaria (a condition caused by *Plasmodium falciparum* parasite; characterized by chills and fever). Those who are heterozygous for the sickle cell trait, have

a mixture of both the normal erythrocytes as well as the sickle shaped cells. In contrast to these, the homozygous carriers of the sickle cell gene have a higher frequency of the sickle cell shaped red blood cells in the population. The work by A.C. Allison in 1954 reported that in spite of the deleterious characteristics of the homozygote, the sickle cell mutation is maintained in the population as the heterozygote possesses a selective advantage that is of being malaria resistant. Therefore, the heterozygote advantage of the sickle cell red blood cells, provides a protective effect as the malarial parasite cannot grow in these cells. Therefore, in areas where malaria is prevalent, those with sickle cell heterozygote condition, tend to survive and increase in frequency over the normal homozygotes as they are susceptible to malaria.