

Indira Gandhi
National Open University
School of Social Sciences

Block

2

HUMAN POPULATION GENETICS

UNIT 1

Meaning and Scope	5
--------------------------	----------

UNIT 2

Hardy-Weinberg Equilibrium	24
-----------------------------------	-----------

UNIT 3

Genetic Polymorphism	55
-----------------------------	-----------

Expert Committee

Prof. P. Dash Sharma (Retd.)
Dept. of Anthropology
Ranchi University, Ranchi

Prof. P. Veerraju (Retd.)
Dept. of Human Genetics
Andhra University
Visakhapatnam

Prof. S.M.S. Chahal
Dept. of Human Biology
Punjabi University, Patiala

Prof. A. Papa Rao
Dept. of Anthropology
Sri Venkateswara University
Tirupati

Dr. Roli Mathur
Scientist 'C'
Division of Basic Medical Sciences
Indian Council of Medical Research
New Delhi

Dr. Seema Kalra
Assistant Professor
Dept. of Biochemistry
IGNOU, New Delhi

Faculty of Anthropology SOSS, IGNOU

Dr. Rashmi Sinha, Reader
Discipline of Anthropology
IGNOU, New Delhi

Dr. P. Venkatramana
Assistant Professor
Discipline of Anthropology
IGNOU, New Delhi

Dr. Rukshana Zaman
Assistant Professor
Discipline of Anthropology
IGNOU, New Delhi

Dr. Mitoo Das
Assistant Professor
Discipline of Anthropology
IGNOU, New Delhi

Dr. K. Anil Kumar
Assistant Professor
Discipline of Anthropology
IGNOU, New Delhi

Programme Coordinator: Dr. Rashmi Sinha, SOSS, IGNOU, New Delhi

Course Coordinator: Dr. P. Venkatramana, SOSS, IGNOU, New Delhi

Content Editor

Dr. Gautam K. Kshatriya
Associate Professor, Dept. of Anthropology
University of Delhi, Delhi

Block Preparation Team

Unit Writers

Dr. T.S. Vasulu (Unit 1 and 2)
Associate Professor
Biological Anthropology Unit
Indian Statistical Institute
203, B.T. Road, Kolkata

Prof. Suraksha Agrawal (Unit 3)
Department of Medical Genetics
Sanjay Gandhi Post Graduate
Institute of Medical Sciences
Lucknow

Cover Design

Dr. Mitoo Das
Asstt. Professor
Discipline of Anthropology
SOSS, IGNOU
New Delhi

Authors are responsible for the academic content of this course as far as the copyright issues are concerned.

Print Production

Mr. Manjit Singh
Section Officer (Pub.), SOSS, IGNOU, New Delhi

July, 2012

© Indira Gandhi National Open University, 2012

ISBN-978-81-266-6134-3

All rights reserved. No part of this work may be reproduced in any form, by mimeograph or any other means, without permission in writing from the Indira Gandhi National Open University.

Further information on Indira Gandhi National Open University courses may be obtained from the University's office at Maidan Garhi. New Delhi-110 068.

Printed and published on behalf of the Indira Gandhi National Open University, New Delhi by the Director, School of Social Sciences.

Laser Typeset by : Tessa Media & Computers, C-206, A.F.E.-II, Okhla, New Delhi

Printed at :

BLOCK 2 HUMAN POPULATION GENETICS

The science of population genetics is an integral component of biological anthropology in understanding human evolution and in pursuing the goal of human origins vis-à-vis the divergence of human groups. Population genetics provides underpinning for all evolutionary biology. The word evolution includes all changes, large or small, visible and invisible, adaptive and non-adaptive. In this context, the existence of genetic polymorphisms in human populations becomes immensely important in seeking answers to our understanding of the evolutionary significance of these variations. Stated simply, population genetics is the study of genetic variation through gene and genotype frequencies in a population, and predicting the way they would change or would be maintained over a period under the differential effects of various micro-evolutionary forces. There are four basic evolutionary forces: mutation, natural selection, genetic drift and gene flow. Mutations are copying errors during DNA replication and transcription, which introduce new alleles into the population. Natural selection is the differential transmission of alleles into the next generation due to the consequences of functional differences on an individual's survival and reproductive success. Genetic drift is the differential transmission of alleles into the next generation as a result of random sampling, and has the greatest potential impact in small populations. Gene flow spreads alleles from one population into another via migration, making them more genetically similar to each other, and countering genetic differentiation by drift. These four evolutionary forces are reflected in patterns of diversity, measured by the numbers of different alleles at a gene locus, the frequencies of each allele, and the interrelatedness of each allele to the others present at the same time. To understand why genetic diversity accumulates to a particular level, why it has its observed distribution, and how its turnover occurs, we investigate the interaction of these four evolutionary forces using population genetic models (Relethford and Harding 2001). Besides, population structure, inbreeding and the Mendel's principles of inheritance and associated factors that violate those principles, also produce differential genetic variation. It is also worth mentioning here that the concept of neutral or nearly neutral mutations along with genetic drift has been consistent with many observations of genetic variations. It is seen for example approximately 0.08% of the nucleotide base pairs (bp) in human DNA vary among individuals. Why these and not others? One explanation is that selection favors functionally different DNA alleles in different circumstances. Another is that DNA variation is tolerated when the alleles of a gene are functionally equivalent. The former explanation clearly applies to some variation, but the latter explanation, formalized as neutral theory, is invoked most often. Either way, the aim of population genetics is to model the dynamics of evolutionary change within and between populations. Moreover, anthropological genetics also deals with change in genetic diversity, population size in the context of evolutionary changes pertaining to *Homo sapiens*. Thus, assuming a simple population genetic model evolutionary population size of humans is estimated to be about 10 000. Population genetic models have also been used to estimate the time depth of typical diversity in the human nuclear genome, and it is suggested to be approximately 800 000 years. Studies on genetic diversity have revealed that population differences are mainly due to the presence of low-frequency alleles that have not diffused far from their geographic place of origin.

Further, by studying formal models of gene frequency change, population geneticists therefore hope to shed light on the evolutionary process, and to permit the consequences of different evolutionary hypotheses to be explored in a quantitatively precise way.

The field of Population Genetics was firmly established with the pioneering work of R.A. Fisher, J.B.S. Haldane and Sewall Wright in 1920's and 1930's. Their achievement was to integrate the principles of Mendelian genetics, which were rediscovered at the turn of century, with Darwinian natural selection. Though the compatibility of Darwinism with Mendelian genetics is today taken for granted, in the early years of the twentieth century it was not. Many of the early Mendelians did not accept Darwin's 'gradualist' account of evolution, believing instead that novel adaptations must arise in a single mutational step; conversely, many of the early Darwinians did not believe in Mendelian inheritance, often because of the erroneous belief that it was incompatible with the process of evolutionary modification as described by Darwin. By working out mathematically the consequences of selection acting on a population obeying the Mendelian rules of inheritance, Fisher, Haldane and Wright showed that Darwinism and Mendelism were highly compatible; which played a key part in the formation of the 'neo-Darwinian synthesis', and thus explains why population genetics came to occupy such a pivotal role in evolutionary theory.

With the advent of different laboratory techniques and various polymorphic serological and biochemical markers, the impetus in studies was gained to a large extent. Later, the power of molecular methods also gave a new dimension to the field of population genetics. Now, with the help of different techniques, and with completion of the various genome projects, the methods of "functional genomics" are scaling up to identify the roles of novel genes. Inevitably increasing attention is being paid to the significance of genetic variations in populations. Prompted by the high incidence of multifactorial diseases as a group, the medical community has become acutely aware of the need to understand the basic structure of genetic variations in populations in order to determine the aspects of the variation that can cause disease. Although, multifactorial diseases have received great attention in the recent past, but the scope of population genetics actually is much broader because the anthropological genetics not only explains the causes of human diversity and the evolutionary history but also focus on understanding adaptation to local environments.

Thus, Population genetics is the key to our understanding of human variation, and by linking medical and evolutionary themes; it enables us to understand the origins and impacts of our genomic differences. Despite current limitations in our knowledge of the locations, sizes and mutational origins of structural variants, the overall growth in this field has brought new insights into recent human adaptation, genome biology and disease association studies. Population genetics provides models for investigating the balance of evolutionary forces acting on genetic diversity. Studies that use these models have found that the evolution of contemporary human genetic diversity has occurred over the past several hundred thousand years or longer. Our species is geographically widespread, but shows low levels of differences among population groups suggesting persistent levels of gene flow as well as dispersal. It is difficult to classify humans into groups by their DNA profiles, and impossible to successfully apply a biological concept of race of diversity within living human populations.

The origin of modern human genetic diversity is still widely debated. Genetic data indicates the importance of Africa in modern human evolution, in line with the observations from the fossil record of the first appearance of modern anatomical form in Africa. Whether Africa is the only region that we can trace our ancestors to, or whether it is the primary region remains to be seen. Some genetic evidence does suggest ancient contributions in southern Asia, a region where the fossil evidence for replacement is equivocal. It may be the case that our origins are best described as 'mostly (but not exclusively) out of Africa' (Relethford and Harding 2001).

UNIT 1 MEANING AND SCOPE

Contents

- 1.1 Introduction
 - 1.1.1 Inheritance-Historical Development
- 1.2 Unit of Study
 - 1.2.1 Gene Pool
 - 1.2.2 Breeding Isolation
- 1.3 Mating Patterns
 - 1.3.1 Random Mating
 - 1.3.2 Assortative Mating
- 1.4 Inbreeding
 - 1.4.1 Consanguineous Marriages
 - 1.4.2 Types of Inbreeding/Consanguineous Marriages
 - 1.4.3 Consequences of Inbreeding
 - 1.4.4 Measures of Inbreeding
 - 1.4.5 Inbreeding Coefficient
 - 1.4.5.1 Inbreeding in Families (Pedigree)
 - 1.4.5.2 Inbreeding in Populations (Fp)
 - 1.4.6 Extent of Consanguineous Marriages in India
- 1.5 Summary
 - Suggested Reading
 - Sample Questions

Learning Objectives



After reading this unit, you will be able to:

- define biological inheritance and describe its historical development;
- discuss breeding isolation and its implications in human population genetics;
- define and outline various mating patterns;
- explain inbreeding and types of consanguineous marriages; and
- measure inbreeding in families and in populations, its consequences of inbreeding/consanguineous marriage.

1.1 INTRODUCTION

Population Genetics can be defined as the study of genetic transmission in populations of interbreeding organisms. It is the study of the nature and source of the inherited differences and involves predicting the changes that may take place in relative frequencies of different genes that may be found in a population, and determining a condition under which equilibrium between forces affecting their frequencies may be obtained.

The science of population genetics deals with Mendel's laws and other genetic principles as they affect entire populations of organisms. The organisms may be

human beings, animals, plants or microbes. The populations may be natural, agricultural or experimental. The environment may be city, farm, field or forest. The habitat may be soil, water or air. Because of its wide ranging purview, population genetics cuts across many fields of modern biology. A working knowledge has become essential in genetics, evolutionary biology, systematic, plant breeding, animal breeding, ecology, natural history, forestry, horticulture, conservation and wildlife management. A basic understanding of population genetics is also useful in medicine, law, biotechnology, molecular biology, cell biology, sociology and anthropology.

Population genetics also includes the study of the various forces that result in evolutionary changes in species through time. By defining the framework within which evolution takes place, the principles of population genetics are basic to broad evolutionary perspectives on biology. Many oddities in biology become comprehensible in the light of evolution: they result from shared ancestry among organisms, and they attest to the unity of life on earth. One of the purposes of population genetics is to study the mechanism of origin and maintenance of genetic variability. The genetic variability is studied in terms of polymorphism of various genetic markers, as polymorphism usually refers to alternative hereditary forms that can easily be distinguished from each other and whose inheritance is clearly understood. Human populations are polymorphic for a large number of genetic loci. These polymorphic loci are useful for constructing linkage maps, genetic counselling, paternity testing and studying the evolutionary relationships of human populations. The association between gene frequencies and environmental factors has been made for number of polymorphic loci. It seems that this type of association study is an effective way to make an inference about selective mechanisms, which could be done by having the data on the distribution of gene frequencies among various populations, provided that the population is endogamous. If there are gene frequencies for a number of loci in a population, the heterozygosity for individual locus can be calculated and average heterozygosity per locus of a population is obtained. The average heterozygosity indicates the magnitude of genetic variation that exists within a population.

1.1.1 Inheritance – Historical Development

Probably no area of biology arouses more interest or creates more controversy than evolution, which deals with the origin of living organisms, the genetic diversity of populations, the mechanisms of speciation, and the biological history of our planet. Evolution is the branch of biology that attempts to explain how to tens or hundreds of millions of different species of plants, animals and micro organisms arouse in the course of several billion years of earth's history. Evolutionary science analyzes the various forces that cause species to adapt, to change and to eventually become extinct.

Evolution and genetics are inextricably connected because the biological changes that occur in organisms through time are due to changes in their hereditary information changes in their genes. To a geneticist, evolution refers to changes in gene and genotype frequencies that arise and accumulate through time in populations and organisms (it is populations, not individual organisms, which evolve). These cumulative changes in gene and genotype frequencies are subject to natural selection, a process originally proposed by Charles Darwin. Charles Darwin, in his landmark work 'On the Origin of Species by Means of Natural

Selection' altered the course of intellectual history in the Western World. Like Newton, Copernicus, Galileo and others, Darwin proposed a mechanistic explanation for natural phenomena. Darwin argued that the diversity of life on earth could be accounted for by the operation of simple and observable processes that are part of our everyday experience. Darwin made the assumptions that phenotypic variation is partly determined by hereditary differences between individuals. It is remarkable that the mechanisms of heredity transmission were not then known because as Darwin himself realised, the theory of evolution by natural selection depended critically on the way in which hereditary information is passed from generation to generation in interbreeding populations, which he himself was unable to account for. Darwin realised that for his theory of natural selection to be plausible and for the adaptive modifications, observed as phenotypes, to be passed on from one generation to the next, a mechanism of hereditary transmission was required otherwise the offspring will not resemble their parents and fitness-enhancing traits will not spread through the population but he was not able to shed light on this. The fact that he could not explain the mechanism of inheritance was the main objection to his theory of evolution. He accepted a weak form of Lamarck's use and disuse theory to explain origin of variation and attempted to work out a way for explaining mechanism of inheritance by assuming that in response to environmental stimuli, somatic cells of body would release entities called 'gemmules'. According to him, these 'gemmules' carried information on the traits of the organism and would accumulate in germ cells and get passed on to the next generation. But he could not prove this. He knew that for selection to operate in a population and gradually alter it, continuous supply of variation was required. There were scholars who believed in importance of selection for creating new adaptive changes. Others doubted that selection acting on continuous variations was sufficient to transform one species into another. Still others advocated Lamarckism. It took the discovery of genes and mutations in the 20th century to render natural selection feasible and unavoidable as an explanation for evolution.

Historically the first and still the most conclusive evidence for the existence of genes come from the phenomenon of segregation of traits observed in the offspring of hybrids between individuals or strains that differ in some recognizable aspect. The principle behind the transmission was for the first time demonstrated by Austrian monk Gregor J. Mendel (July 20, 1822 – January 6, 1884). His publications on the results obtained from the *very carefully selected* breeding experiments on '*hybridisation*' conducted on *Pisum sativum* in monastery gardens of Austria for eight years (1856-1863) indicated that the transmission of these (seven) characters (height, size and shape of seed, colour of flowers etc. among pea plants) are particulate in nature. The hereditary transmission of the above traits follows specific pattern what has termed as law of independent assortment and law of segregation. According to Mendel's laws, given the parental phenotype/genotype combination, it is possible to predict the likelihood genotype/phenotypes of the offspring. Another important aspect of the discovery of Mendel's laws is that the transmission of hereditary traits from parents to offspring follows simple binomial expectation with stochastic perturbation, which is more likely to be true in a large sample size.

The Mendelian's concept of hereditary unit is contradictory to the erstwhile popular Greek school of thought '*blending theory of inheritance*' that involves mixing nature of inheritance of characters from parents to offspring. In this regard,

it is important to mention that Darwin has proposed ‘*pangenesis*’ to explain the mechanism of inheritance and in doing so; he postulated ‘*gemmules*’ particle nature of inheritance. According to Darwin, The ‘*gemmules*’ is produced by the cells and could diffuse to other parts of the organism and produce new cells etc. However, this did not stand the scrutiny of the developments of empirical experiments of hereditary by others like Francis Galton etc. As against these backdrops, independently and at the same time (1860s), the Mendel’s hereditary experiments have demonstrated the existence of particular nature of hereditary units, and rightly, he located its existence in germ cells.

The principle of segregation and dominance and principle of independent assortment was formulated by Gregor Mendel in 1866 under such peculiar circumstances that the scientific world failed to recognize or appreciate it until after a lapse of 34 years. Mendel was not primarily a biologist but a monk in the Augustinian Monastery at Brunn, Austria (now Brno, Czechoslovakia). After seven years of experimental work in the monastery gardens, he presented the results of his experiments, together with the generalizations we now know as ‘Mendel’s Law’, at two meetings of the Natural History Society of Brunn in 1865. The results and the theory were printed in the annual proceedings of the society, which appeared and distributed to libraries in Europe and America in 1866.

Possibly, people expected that the results obtained by Mendel is more related to hybridisation and are of importance to breeding animals and the implications that Mendel’s discovery of hereditary units, the secret behind the heredity and its implications to overall biology, was possibly realised by a few, until after three decades later by the rediscovery of his laws in 1900 simultaneously by Hugo De Vries in Holland, Carl Correns in Germany and Erich Von Tschermak in Austria, who found Mendel’s forgotten paper and proclaimed its importance. Immediately his conclusions began to be confirmed and extended by experiments carried on in various parts of the world on many kinds of plants and animals.

It was from the Mendel’s rediscovery of laws of hereditary that describes how the characters that he selected – e.g., height, seed colour – among plants follows simple principles of transmission, it became clear that the *cause for the hereditary transmission of a particular (Mendelian) character* is governed by what Mendel hypothesised as ‘*hereditary unit*’. This hereditary unit was later described as ‘gene’ by William Bateson and the subject that deals with the heredity and its transmission rules and regulations is the discipline of ‘Genetics’. The hereditary transmission of traits, its variation or extent of diversity among regional populations, how its changes over time, what are the factors and causes that influence these changes all are important for us and the study that deals with it in brief is the ‘human population genetics’. It also deals with theoretical and empirical studies to understand the how different traits change over time and factors governed its mechanism at the population level, at amino acid, enzyme, molecular level etc.

1.2 UNIT OF STUDY

One of the major intellectual challenges of early twentieth century genetics was the integration of Mendelian genetics and the theory of evolution by natural selection. The resulting synthesis combined the postulates of evolution by natural

selection presented. Darwin with the statistical rules of genetic transmission in the population to construct mathematical theory of evolution. This synthetic theory of evolution, based on population genetics, is sufficient to account for the evolution of the great variety of life forms on earth. Thus, Population Genetics, as we know it today, arose from the need to reconcile Mendel with Darwin.

It may not be less worth mentioning here that genotype frequencies are determined in part by pattern of mating. To deduce the genotype frequencies from the estimated allele frequencies, British mathematician, G.H.Hardy and German Physician, W.Weinberg independently formulated a model in 1908 describing the relationship between the gene and genotype frequencies, and this relationship is now popularly known as Hardy-Weinberg principle or Hardy-Weinberg equilibrium (H.W.E), which eventually laid down the foundation of genetic analysis of population.

Population genetics deals with the changes in gene frequency over generations in a population. In the context of evolution, according to Darwin, it is the species living in a given eco-niche evolves over time and this is brought by what he called as 'natural selection', the driving force for the evolutionary change among species. The unit of study in the biological realm of Darwinian evolution is the species. Apart from the phenotypic characters that is/are characteristic (identity) of given species — what we can understand from the common English usage that is quite apt to the topic, for example, "birds of the same feather flock together", — interbreeding within a species is prime distinguishing character that defines a species, the study unit. This is in a way identified by characteristic species specific phenotypic traits. In general, in nature, mating is restricted to within the species; as such the phenotypic and genetic characters are transmitted through generations and become restricted to a given species. Therefore, mating within a species is compatible and matings between species are unlikely and are supposed to be incompatible. However, Man was successful in bringing out interspecies matings between related species e.g. Lion and Tiger, Horse and Donkey, etc. in artificial circumstances (cross-breeding) creating some unusual types such as 'tigon' and 'mule' (was supposed to be infertile). The mating within a species or a group restricts the transmission of characters through generations within the group or a species. This is similar to water pond or pool/lake where the water is contained and restricted to its geographical space.

In general, in Man, human populations live in groups over a neighbourhood area in a wide geographical or eco-regions. In general, the marriages are expected to be within a particular community, group or clan or caste etc. Marriages within a community or group or clan, is said to be endogamous. Similar to water tight compartments, the endogamy within a community restricts transmission of genetic characters within its community, over generations. In population genetics, the endogamous populations, where marriages are restricted within a community are construed as 'gene pools' where genes/alleles or consisting of a variety of traits with different frequencies confined to the population or the gene pool.

1.2.1 Gene Pool

Gene pool refers to the endogamous population where the marriages are restricted within the group, restricting the genes within its group or community. This is also called as Mendelian population (A reproductive community of individuals

which share in a common gene pool). How do we investigate whether a population is endogamous? or how do we investigate the extent of a Mendelian population or a gene pool among human populations?

The boundaries of a gene pool in human populations can be investigated by the extent of endogamy that is practiced in a local/regional population and by studying the marital alliances or the pattern of mating. It is not always easy to estimate the extent of endogamy, except in small isolated populations with known or defined barriers that restricts marriages within the population. These barriers could be geographical and/or cultural. Examples where geographical factors play major barriers to marriages within in the community include island populations or a few tribal populations who live in hunting and gathering sustainable livelihood in interior jungles or high mountain ranges of difficult terrain or inhospitable environmental conditions. Even when the population is large, they might subdivide to form several endogamous populations. In this case, the cultural factors that regulate marriages lead to endogamy within in the community. Identification of these cultural isolates helps us to investigate or define the 'gene pool. At empirical level one gets information about marriages within the community and estimates "endogamy index". It is defined as per cent occurrence of number of marriages within the community to the total number of marriages. The index ranges from 0 to 100 and a value of above 95% can serve as indication to suggest the population is an endogamous and can be considered as a gene pool. The index can be investigated for each generation to get an idea to track changes in endogamy levels in the recent past. In case one would like to investigate the status of endogamy beyond 5 or 6 generations, one can choose other informative data. One such data is marriages among surnames or clans. This can help us to get to know the relative levels of endogamy in the recent past beyond several generations.

1.2.2 Breeding Isolation

In human populations, breeding isolation can give an estimate the spatial distribution of gene pool. The breeding isolation can be investigated by marital distance between the birth places of the spouses in a population. In general there is a tendency to marry within the communities over a neighbourhood region. The distance between the birth places of the spouses gives an extent of spatial distribution of gene pool. In general, in traditional rural, tribal societies and in among some urban societies a positively skewed marital distribution is expected. A low mean marital distance (MMD) is supposed to be characteristic of traditional societies. In urban and metropolitan societies is expected to show higher mean and higher variance of MMD. In a study conducted among the regional populations of *Yanadi* (a tribal population from Andhra Pradesh) who differ in their subsistence pattern in islands, coastal and plateau and hill forest areas shows the typical positively skewed MMD. The mean also varies from hunting gathering to agricultural societies, lower among hunting gathering and higher the mean among agricultural societies of the same tribe. It also shows the spatial distribution of extent of gene pool among the Yanadi regional populations (Figure. 1.1).

The urban population tend to have large and higher values of MMD and their gene pool extend over a large spatial distance. Among these populations they are characteristic of lower endogamy rates, which become difficult to define them as breeding isolates or gene pools. Another difficulty is in urban communities

are tend to show higher interbreeding with other communities, which poses problems for considering them as breeding isolates or gene pools for population genetic studies. Though there are very few studies about the interbreeding or admixture in various populations, but in general such rates are about 1-2 per cent among the rural endogamous groups and among urban communities it can vary as high as 10%. Due to socio-economic reasons such admixture or interbreeding is increasing among urban societies.

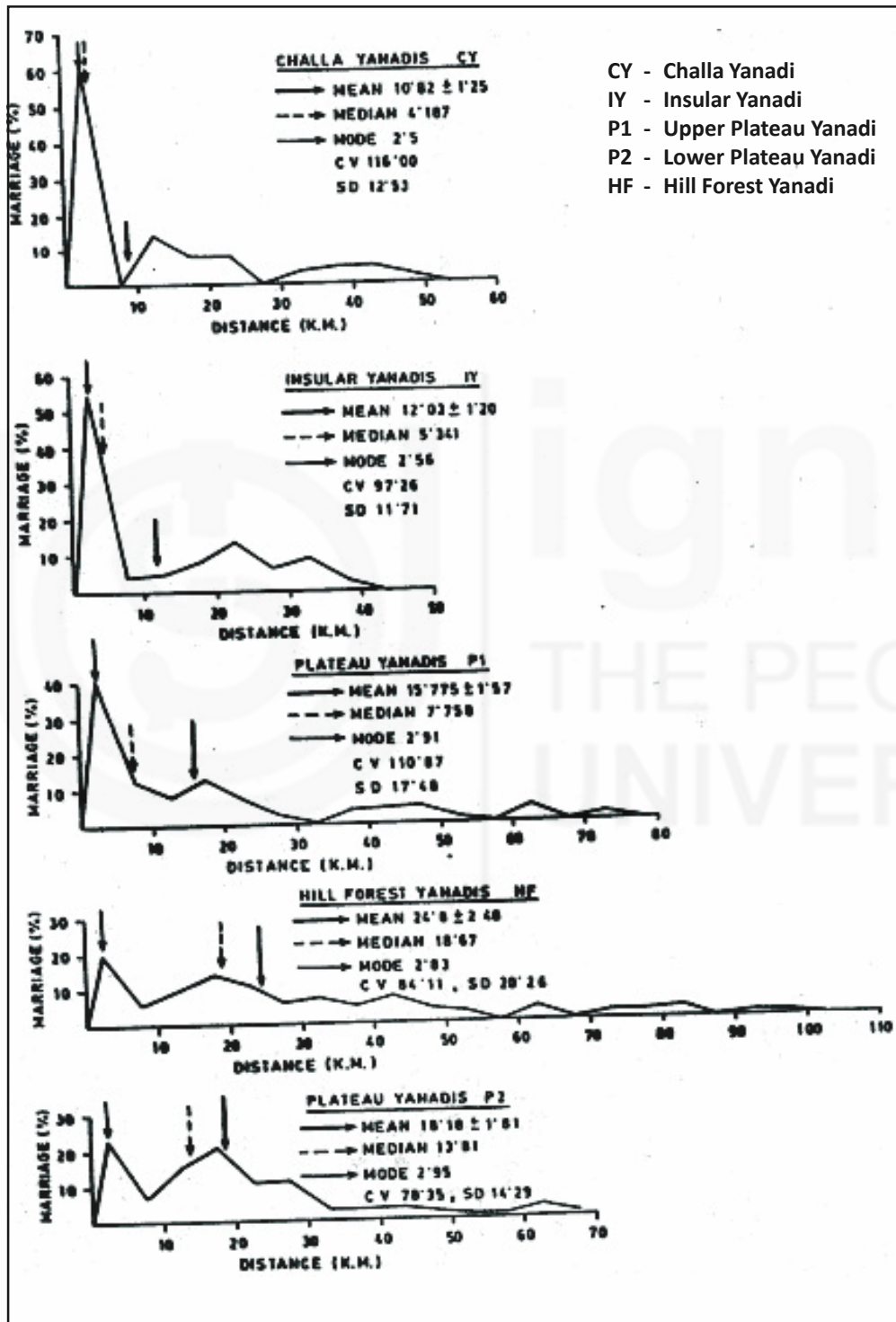


Fig.1.1: Distribution of matrimonial distance in regional breeding populations of the Yanadi tribe

1.3 MATING PATTERNS

Hereditary transmission of characters from generation to generation depends on type and selection of males and females that they participate in sexual selection leading to reproduction. Both in animals and human populations mating behaviour is very complex and it is non-random. The selection of male and female is characterised by a variety of patterns and certain pattern is characteristic of some groups. Charles Darwin while elucidating his hypothesis of origin of species by natural selection, considered sexual selection is an important criterion that plays a role in the successive reproduction that might lead to proliferation of desired qualities in the offspring. Those who are not so successful in selecting a mate will have least possibility of successful reproduction of their abilities being transmitted in the next generation. These abilities or characters stands eliminated in successive generations. In human populations, there is a wide variety of mating patterns and marriage norms, rules and regulations. These vary in different societies and communities.

Human populations are organized conglomerate entities. It generally consists of demographic, cultural, linguistic, political sub structuring; each with its own defined functional domain. Of these several entities, culture is one important component of human populations, which can have a profound biological significance. For example, populations have varying rules and regulations in choosing a mate or marriage partner. Charles Darwin in 1871 in his book 'Descent of Man and selection in relation to sex', recognized sexual or mate selection as important causative factor as that of natural selection in Evolution.

The non-random nature of choice of mate or mate selection can lead to significant genetic changes. There are different types of mating. The genetic significance of each type of mating can be investigated theoretical and its theoretical expectation of the possible genetic changes over generations can be worked out or marriages in diverse societies. These are described below.

1.3.1 Random Mating

Random mating is also referred as 'panmixia'. Panmixia or Random mating refers to the situation in a population where there is equal opportunity for both the sexes to mate and reproduce irrespective of any choice or selection. Random mating refers to a situation where absence of cultural rules and regulations that involve preferences or avoidances of choice or selection of marriage partners. If an individual mate or marry without any criteria of selection or choice what so ever, it is random mating. Random mating can also refer to those mating with respect to individual genotype or phenotype, though they may be marrying individuals within the community. Therefore, the random mating also refers to the characters which are not involved, either directly or indirectly, in the choice or selection by the respective mates or marriage partners.

Random mating is one important criterion in Hardy-Weinberg equilibrium which ensues, in a large population, stability of gene frequency over generations. Otherwise deviations from random mating, a common phenomenon among human populations, is an important factor which changes the gene frequency.

1.3.2 Assortative Mating

In all human populations, from simple societies of tribes and to complex societies of urban in metropolis, people usually like to select their marriage partners, mates or spouse, i.e., non-randomly, with desirable characters of their liking. This is referred as 'assortative mating', an important factor of non-random selection of mates that exists in all human societies. Assortative mating is one of the deviations from random mating assumption of Hardy-Weinberg equilibrium. According to H W Law the gene frequency in a population is expected change as a result of non-random mating due to assortative mating.

Choice of mate selection is more concerned directly with respect to some specific phenotypic characters (rarely genotypes are involved) and it might involve, indirectly, other characters which are associated or related to selected traits or characters among spouses or mates. For example, choice of marriage partner could be based on such desired or preferred traits e.g., skin colour, age preferences, intelligence, social status etc. In case one selects intelligence as the criterion it might be associated with high social status, which may not be desired trait.

In general, the institution of marriage, in human populations, prescribes rules and regulations that allow members of the community choice to select his/her spouse or marriage partner. Assortative mating refers to the preferential marriages (or mating) between individuals that involve either similar traits such as intelligence or phenotypic characters such as height, skin colour etc. or it might involve different traits of phenotypic characters. Assortative mating is one type of non-random mating that commonly occurs among human populations.

Assortative mating could be of two types: positive and negative mating. In positive assortative mating (homogamy) type both the spouses choose or select their mates on the basis of common phenotype or of similar characteristics. Conversely, the negative assortative (heterogamy) mating types are those where the partners select their mates based on different or dissimilar phenotypic traits.

For example, if tall and fair people choose to marry those with similar traits of equally tall and fair peoples, then it is assortative mating involving phenotypic characters of tall and fair skin colour. Similarly the opposite can also happen where people choose their mates unlike themselves or with dissimilar or opposite traits. In general, in some societies, in mate selection, there is preference (or selective advantage value) for fair skin colour people to get married. Conversely, people with dark skin colour are avoided or least preferred as ones' spouse. In a patrilineal society, it is observed that males with dark skin colour prefer to marry fair skin brides. In these communities, there is negative assortative mating involving selection of dissimilar characters (dark and fair skin colour) between the males and females. In simple positive assortative mating people marry their spouse who are similar to them and in negative assortative mating spouse avoid similar characters or marry their spouse with dissimilar or diverse phenotypic traits.

1.4 INBREEDING

In many societies marriages takes place between relatives. In general, breeding among close individuals within a group is referred as 'inbreeding', and it is

applicable in case of plants and animals. The extent of inbreeding depends on the amount of common ancestry shared by the parents of the 'inbred' child (the progeny of a consanguineous marriage or inbreeding) or in genetic terms it is the proportion of genes that the parents have in common. It is opposite to outbreeding referring to breeding with the individuals from outside the group or unrelated individuals.

In human societies the marriages contracted among relatives and are referred as 'consanguineous marriages' or 'inbreeding'. Consanguineous is more with reference to cultural context and inbreeding is more with reference to general breeding that include Man and other organisms.

1.4.1 Consanguineous Marriages

In general relatives are referred as blood relatives (see introduction above for details) as per the belief that blood is responsible for hereditary transmission. In Greek the word 'Consanguin' refers to blood and relatives are referred as 'consanguines'. The other word for relatives is kin. Hence, marriages contracted between relatives (consanguines) or among kin are referred as consanguineous marriages.

In several human societies there is a preference and practice of marriages among close relatives. Possibly owing reasons due to social, cultural and economic reasons, several societies practice consanguineous marriages.

1.4.2 Types of Inbreeding/Consanguineous Marriages

There are different types of inbreeding and/or consanguineous marriages that are practiced that differ among communities or societies. Self fertilization is one of the closest form of inbreeding, which is observed among some flowering-plants, some hermaphroditic animals. The types of close inbreeding that are disapproved in most of the human societies but rarely observed in Man include sib-mating, parent-offspring. In early history of mankind, e.g., in ancient Egypt, the royal families had followed brother-sister marriages for some generations. This was based on the belief that royal blood is pure and to maintain the purity of blood they followed sib-sib mating. The Egyptian queen Cleopatra is one such example of generations of sib-sib mating in the royal family.

The cousin mating or marriages among near-relatives are practiced in several human populations. These different types of close marriages can be investigated by drawing a pedigree with all the information of relationship between husband and wife in a family that can extend to several generations. Aunt-nephew and Uncle-niece marriages are the marriages that occur across two different generations. In Indian populations Uncle-niece marriages are more common to Aunt-nephew marriages. Depending on the type of consanguineous relationship between husband and wife, they are categorized as first degree, second degree (and etc.) consanguineous marriages. The first cousin marriages are categorized as first degree consanguinity. This involves individuals who belong to same generation but come from different families whose parents share a common ancestor (grand parental level). These are referred as first cousin marriages, marriage among cousins who share a common grand parent. An individual if he

or she marries his/her paternal or maternal daughter or son (uncle's daughter or son respectively) the cousin marriage is the first cousin marriage. Similarly in case of second cousin marriages, the second cousins share a common ancestor at great grand parent level. In between there could be marriages between individuals that belong to two generations e.g., one-and-half cousin marriages etc.

The cousin marriages could be of different types: for example, parallel, cross and double cousin marriages etc. The marriage between the cousins who were descendents of two brothers or sisters (or whose (either) parents are either brothers or sisters) is referred as parallel cousin marriages. An individual marries his or her paternal uncle's daughter or son is referred as parallel cousin marriage. The word 'parallel' refers to same sex (either brothers or sisters) at the parental level. In case of cross cousin marriages, it is the marriage between the offspring (cousins) who's (either) parents are brother and sisters. In cross cousin marriage an individual marries his or her maternal uncle's daughter or son is the cross cousin marriage type. Apart from the above two there could be one more type of cousin marriages, it is called as double cross cousin marriages. In double cross cousin marriage, the marriages between a brother and a sister of a family marry their cousins a sister and a brother of their uncle or aunts respectively. In cross cousin marriage either a boy or a girl marries his or her uncle's daughter or son (cousin) respectively. In double cross cousin marriage type both a boy and a girl marry their uncle's or aunts' girl or son (cousins) respectively.

The prevalence of consanguineous marriages varies across different populations across regions and through time (generations). At least in the early history of Man such close inbreeding should have been very common as a result of population structure that is conducive of small population size, isolation and subsistence pattern of hunting and gathering and early agricultural life. It should have been a common practice in the history of mankind till 19-20th century or two centuries ago before the era of breeding and principles of inheritance and its consequences were investigated. Charles Darwin, who married his cousin Emma Wedgwood, had of the opinion: it is likely that we are all descendents of cousin marriages.

1.4.3 Consequences of Inbreeding

One of the important aspects of inbreeding is that it increases the chance of some of the recessive alleles with deleterious effects to get expressed in the inbred (the offspring of consanguineous marriages) which get transmitted from one or more common ancestors carrying (harbour) the rare recessive alleles.

Theoretically, inbreeding in a population, in general, increases the number of homozygotes at any autosomal locus. In case of recessive Mendelian diseases this can lead to increased risk of genetic diseases among the inbred children. In the same logic, in case the alleles are not deleterious in nature, but are advantageous, increase of such genes through inbreeding in a population is expected to be beneficial. The consequence of inbreeding whether it is deleterious or advantageous depends on the status of the allele that is involved in the expression among the inbred and transmission from the common ancestors.

Such cases pose health and survival problems of deleterious or debilitating diseases. This is health and socio-economic burden to the society and to the country. Increase of deleterious characters as a result of inbreeding is referred as 'inbreeding load'.

Inbreeding in a small population lead to decrease in the genetic diversity, and it may lead to decreased Darwinian fitness of an organism. In case if these results to less capability to survive in changing environment, or eco-niche, with inbreeding such characters will eventually allow the species to extinction. Such characters are supposed to have less Darwinian fitness. Such decline in reproductive performance and fitness with inbreeding in a population is referred as 'inbreeding depression'.

In case the population is large enough, some of the undesirable characters which are deleterious, over long inbreeding in a population, suppose to get eliminated by selection and thus attributes to better fitness and increased survival advantage to the population.

The above theoretical expectations differ from the empirical studies conducted on the consequences of inbreeding and its effects in human populations. These results are not uniform across populations and there has been a debate on the deleterious or harmful effects of inbreeding in human populations. This could be partly due to problems of study design, the methodology used and the history of inbreeding in the population. Some studies show significant differences with respect to some characters considered among the inbreeding and non-inbred population, while others have not found such differences in other populations studied. These studies on inbreeding effects are concerned with characters such as fertility, offspring mortality, morbidity, diseases, and rare genetic disorders, etc.

1.4.4 Measures of Inbreeding

The consanguineous marriages or mating in a population increases the chance of inheriting two identical copies of a homologous gene in the offspring from one or more common ancestors. The possibility that an inbred-individual gets two identical copies (homozygote) of such homologous genes due to inbreeding is described as 'identical by descent' (or 'ibd'). And the two alleles are called 'autozygous'. Apart from the case of close kin marriages that can lead to increased likelihood of both the identical copies to get expressed in the offspring by descent, it can also happen by chance alone. In cases of chance occurrence of homozygosity of two identical copies being expressed in an individual independent of inheritance from a common ancestry or absence of inbreeding, then such homozygous alleles are said to be described as 'identical by state' (or 'ibs') and the two alleles are called as 'allozygous'. The extent of autozygosity depends on the type of consanguinity followed in a family or in a population. The degree or extent of consanguinity or inbreeding in a family or in a population can be quantified by suitable measures.

There are at least two types of inbreeding measures. These measures can be calculated or estimated at different levels or type of data: pedigree data, population data and based on variety of traits: phenotypic, genetic and genomic information. The two measures are:

- 1) A genetic similarity measure to find out how or to what extent two related individuals are genetically similar; and
 - 2) The effect of inbreeding on the fitness of inbred offspring, described as 'inbreeding depression'
- a) **Inbreeding depression:** Inbreeding is expected to reduce the fitness of an individual when compared to non-inbred population. With inbreeding within variation will reduce and between variations is expected to increase. The extent of Inbreeding depression can be estimated in a pedigree among the inbred and outbred individuals and in a population. This can be investigated in case of single trait or multiple loci and phenotypic characters. The measures to estimate the extent of inbreeding depression include deviation from the mean of phenotypic characters, or deviation or differences between frequencies of a trait between the outbred and inbred individuals in a population and taking care of the variance of the trait concerned.

1.4.5 Inbreeding Coefficient 'F'

It is the measure of extent of genetic similarity among inbred individuals as a result of inbreeding. Sewall Wright (1922,1923) defined inbreeding coefficient as the 'correlation between uniting gametes' however such correlation can also occur because of other types of mating e.g., assortative mating.

Malecot (1948) defined inbreeding coefficient 'F' based on the probability that the homologous genes of uniting gametes are 'identical by descent (ibd)'.

Therefore, F is the probability of autozygosity

Inbreeding coefficient == coefficient of consanguinity
 == coefficient of ancestry

An inbred individual has genes that are likely to be same with the ancestor.

The inbreeding coefficient can be calculated (estimated) by different methods for pedigrees and populations and for a variety of data as well.

1.4.5.1 Inbreeding in Families (Pedigree)

In case of individual inbreeding coefficient (F_x) from a detailed pedigree where individual x is an inbred offspring as a result of consanguineous marriage of the parents and/or grand parents etc. The inbreeding coefficient of the individuals can be calculated as follows:

The inbreeding coefficient of an individual, according to Sewall Wright (1922) is

$$F_x = \sum [(1/2)^{n_1 + n_2 + 1} (1 + F_A)]$$

Where x is the inbred individual

n_1 is the number of generations from the inbred (x) to the male common ancestor

n_2 are the number of generations from the inbred (x) to the females common ancestor

F_A is the inbreeding coefficient of the ancestor

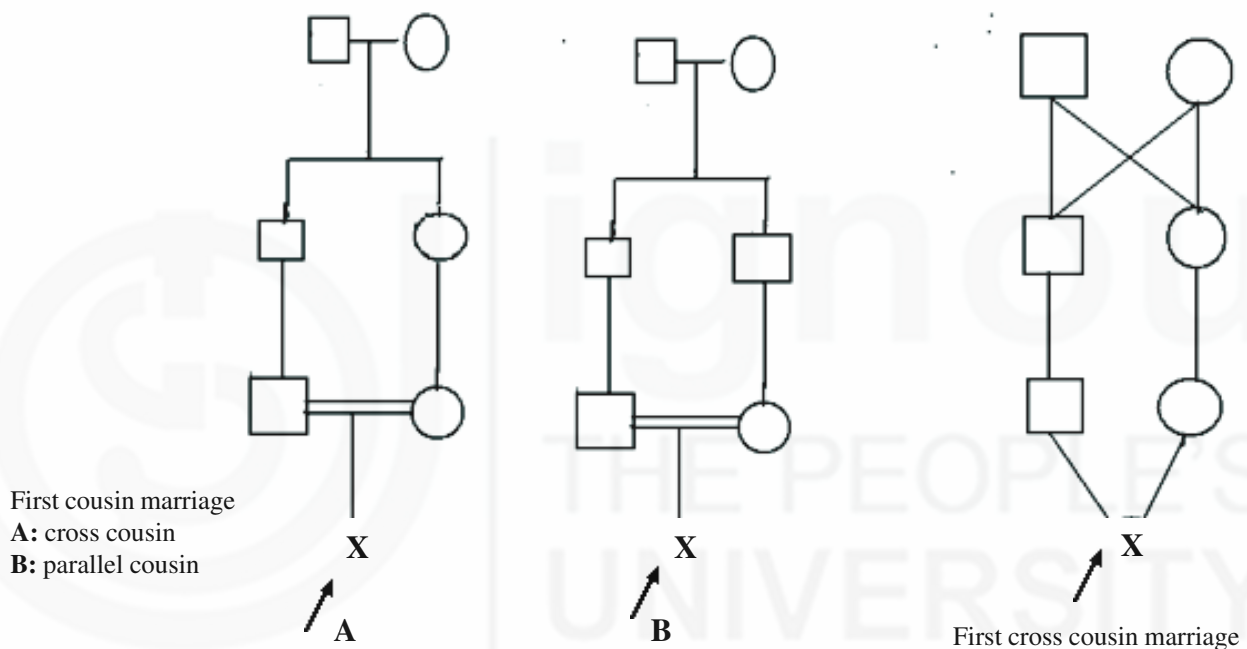
The inbreeding coefficient (F) can also be calculated from path analysis:

$$F_X = \sum [(1/2)^{n_p - 1} (1 + F_A)] \text{ and}$$

$$F_X = \sum [(1/2)^{n_a} (1 + F_A)]$$

Where, n_p is the number of paths that connects x to the common ancestor, and n_a is the number of ancestors from the inbred to the ancestor in the earliest generation. F_A is the inbreeding coefficient of the common ancestor. If the common ancestor is non-inbred, then F_A is zero (in both the methods).

The value of F ranges from 0 to 1. If F is zero then there is no inbreeding in the population and higher the value higher the inbreeding coefficient. This states proportion of common genome shared by the inbred as a result of inbreeding. The offspring of first cousin marriage is 0.0625 or 6.25% of the genome is supposed to be same with that of the common ancestor.



In the above diagram first cross cousin and parallel cousin marriages are shown. X is the inbred (↗), his or her parents are cousins and an example of consanguineous marriage (=).

The right hand side of the diagram is short form of representing the inbreeding in the pedigrees. Here the ancestors are not themselves are inbred, so that the term $(1 + F_A)$ is zero for the above pedigree.

The inbreeding coefficient of the x in the above can be calculated as per the above formulae: The pedigree shows two loops. Here F_A is zero as the ancestors are unrelated and not inbred.

There are 5 ancestors (a) in the first loop starting from x.

Another 5 ancestor (a) in the other loop starting from x.

If one wants to count number of paths, then there are six paths (lines) (p) that connect x through one common ancestor and another six paths (lines) (p) through the other common ancestor.

By using any one of the formulae it is easy to calculate F_x for example:

$$F_x = \sum [(1/2)^{n_a} (1 + F_A)] = (1/2)^5 + (1/2)^5 (1) = (1/16)$$

$$F_x = \sum [(1/2)^{n_p-1} (1 + F_A)] = (1/2)^{6-1} + (1/2)^{6-1}(1) = (1/16)$$

The inbreeding coefficient of a first cross/parallel cousin marriage offspring is 0.0625. Similarly it is possible to calculate the inbreeding coefficient of different type of related marriages practiced in the population through pedigree method. The same can be extended to complex pedigree structures with multiple loops extending to several generations as well.

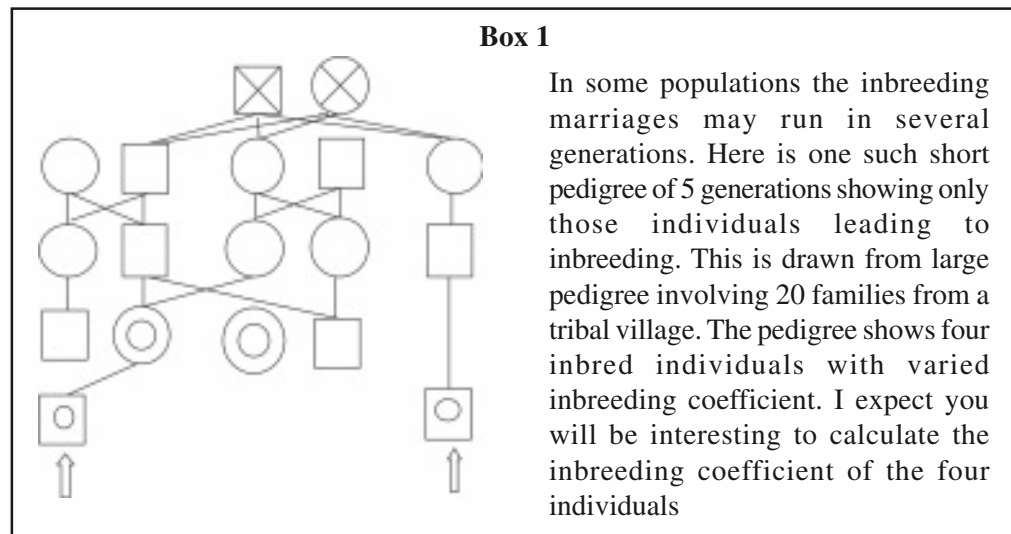
The above inbreeding coefficient refers to autosomal loci. But the same can be extended to estimate the inbreeding coefficient for sex-linked chromosomes (F_s) as well. In that case there is no inbreeding coefficient for a male offspring and the path if there are two consecutive males F_s will be zero.

Inbreeding coefficient for different type of consanguineous marriages in human populations is shown below. More close the relationship the higher the inbreeding coefficient. In general the above is very simple example of individual cases of consanguineous or related marriages to illustrate the inbreeding by pedigree method. The different inbreeding coefficients of known related marriages are shown in Table 1.1. However, in empirical study, the pedigree information is a tedious exercise to draw large family data and from them to identify the complex inbreeding that might run several generations.

Table 1.1: Type of consanguineous marriage and inbreeding coefficient

Type of consanguineous or related marriage	Inbreeding coefficient	
Uncle – Niece or Aunt – Nephew (UN/AN)	1/8	0.025
First cousin marriage (cross, parallel) IC	1/16	0.0625
First cousin once removed IC1	1/32	0.03125
Second cousin 2C	1/64	0.015625
Double cross first cousin marriage D1C	1/16	0.0625
Double cross second cousin marriage D2C	1/32	0.03125

Here for example, we have a one such complex inbreeding drawn from field work studies among tribal villages which shows one such large complex inbreeding network involving several families (see box 1). Indeed to calculate the inbreeding coefficient for such complex pedigree will be difficult. But now a days there are several methods of drawing the related marriages and luckily there are softwares on the internet services they can provide quick way of drawing the pedigree of complex nature and to calculate the inbreeding coefficient as well.



1.4.5.2 Inbreeding in Populations (F_p)

By knowing different types of inbreeding coefficient in a population it is possible to calculate the inbreeding coefficient for a particular population where inbreeding is preferred by tradition. This will be the average inbreeding coefficient practiced in the population.

Suppose in a survey one has observed different types of consanguineous types of marriages with each occurring at some frequency or proportion.

If ' c_i ' is the proportion of 'ith' types of marriages ('ith' e.g., = Uncle-Niece marriages, First Cross Cousin, Second Cross Cousin once removed, Double cross cousin etc. marriages observed in a population)

The inbreeding coefficient in the population is:

$$F_p = \sum (c_i \cdot F_i)$$

Where, c_i is the proportion of type (ith) of consanguineous marriage and F_i is the inbreeding coefficient of **ith** type of consanguineous marriage.

1.4.6 Extent of Consanguineous Marriages in India

In India several communities practice consanguineous marriages and the extent of consanguineous marriages vary. For example a review work indicates that the extent of average inbreeding coefficient among Muslim population in different states of the country vary from 0.007 in West Bengal to 0.26 in Madhya Pradesh (Table 1.2). In case of castes and tribes, such marriages in different regions show an average of 0.0325. The extent of consanguineous marriages and average inbreeding levels practiced in the country is shown in Fig. 1.2. (Malhotra and Vasulu, 1999).

Table 1.2: Extent of consanguineous marriages and inbreeding among muslims in different states of India

State/ Region	No. of Marriages	Type of Marriage			Inbreeding Coefficient
		UN	I.C	Total	
Andhra Pradesh	356	3.37	32.87	36.2	0.025
Tamil Nadu	6116	0.88	16.29	19.5	0.012
Kerala	215	—	16.74	22.3	0.012
Maharashtra	2014	—	14.00	20.71	0.010
Madhya Pradesh	351	7.41	19.09	59.3	0.026
Rajasthan	412	—	31.55	41.3	0.022
Uttar Pradesh and Delhi	1000	—	27.7	49.4	0.020
West Bengal	1483	—	11.4	27.6	0.010
	835	—	20.36	22.2	0.013
	471	—	8.28	19.3	0.007

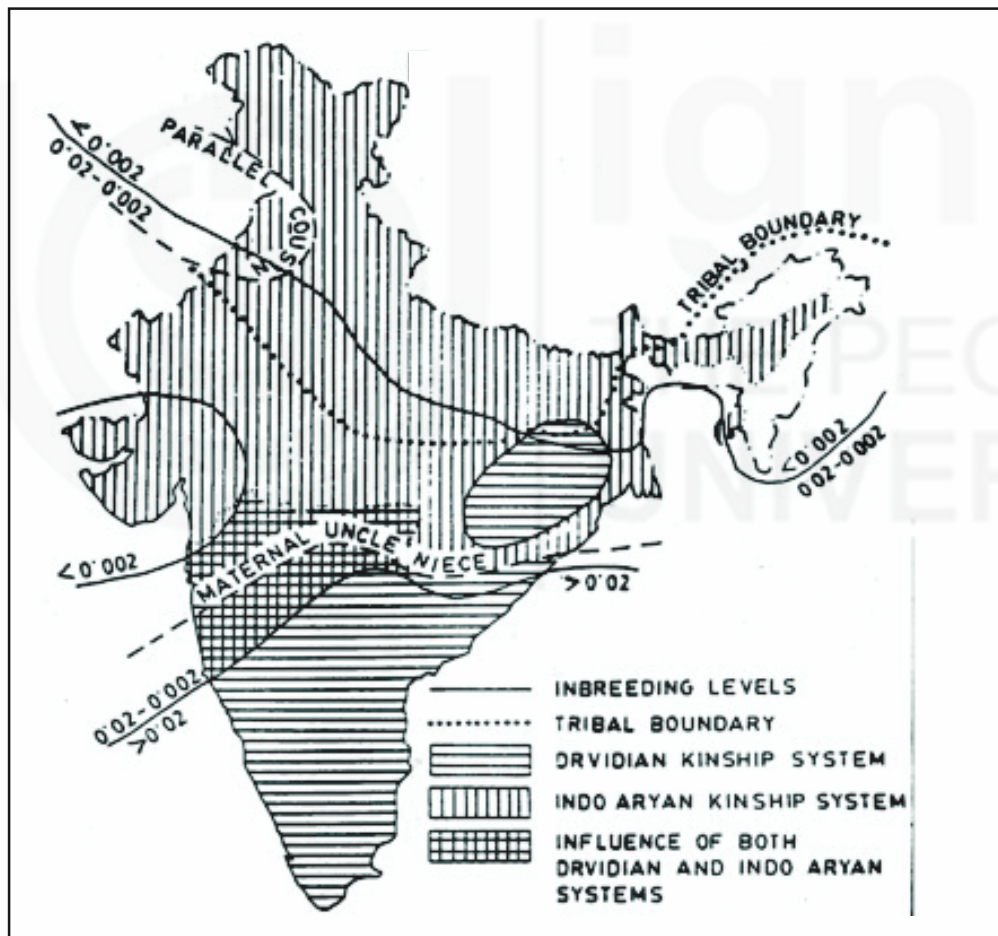


Fig.1.2: Extent of consanguineous marriages and inbreeding levels in India

1.5 SUMMARY

- 1) An understanding of the human population genetics starts with an idea of the concept of heredity, historical development of the concept and to the present knowledge of the concept of Mandelian hereditary unit and genetics and its developments.

- 2) In human populations the study of population genetics is the gene pool, the study unit. To identify and investigate the extent of gene pool requires understanding of the breeding isolation, especially who marries whom and where. This demographic information helps us to understand the extent of endogamy.
- 3) Sexual selection is one of the important mechanism that can significantly influence the pattern of gene transmission in a population. In case of random mating, where everybody can marry anybody as per HW equilibrium the genetic pattern is expected to be influenced by selection, drift and mutational pressures.
- 4) In human populations mating is non-random. Each population follows a variety of rules and regulations in selecting their marriage or mating partners. There are preferences or criterion of mate selection, one such criterion is based on e.g., morphological characters, intelligence, socio-economic status and is referred as assortative mating. One can choose to marry a partner with desirable similar characters as ones mate, where it is called as positive assortative mating or dissimilar or opposite characters where it is called as negative assortative mating.
- 5) Several human populations prefer to marry their own relatives. Such related mating or marriages are called inbreeding. Consanguineous marriages in Man are one of the common type of marriage. There are a variety of types of related marriages and all result in inbreeding. Inbreeding leads to a greater chance of inheriting two copies of the homologous genes from one or more common ancestors. This is referred as identity by descent (ibd). In case of inbreeding there is decrease in variability of a trait. There is decrease in heterozygosity of several characters. This decrease in survival and fertility due to inbreeding is called inbreeding depression.
- 6) Theoretically it is possible to investigate the extent of genetic similarity between related individuals. One such measure is the inbreeding coefficient. By knowing types of consanguineous type or inbreeding practiced, one can calculate the inbreeding coefficient for different types of related marriages. One can also calculate the inbreeding coefficient in a population and examine the trends of levels of inbreeding practiced over time and region among diverse populations.

Overall this unit gives an idea of developments of the concept of inheritance how it is developed from the earlier time to the current understanding of genetics. This gives an idea of what is the unit of study in human population genetics, and how we investigate the breeding isolation in Man. It gives an idea of what is random mating, non-random mating, deviations from random mating in Man. Inbreeding and consanguineous marriages and how to measure them, the pattern and levels of inbreeding in Indian populations.

Suggested Reading

Charles Darwin 1859. *Origin of Species by Natural Selection*.

Charles Darwin 1871. *Descent of Man and Selection in Relation to Sex*.

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

Crawford M.H and P.L Workman, 1973. *Methods and Theories of Anthropological Genetics*. Albuquerque: University of New Mexico Press.

Meaning and Scope

Harrison, G.A and A.J Boyce. 1972. *The Structure of Human Populations*. Oxford: Clarendon Press.

Malhotra, K.C and T.S Vasulu. 1993. *Structure of Human Populations in India*. In. "Human Population Genetics: A centennial tribute to J.B.S.Haldane".(ed.). P.P. Majumder.207-233. New York: Plenum Press.

Curt Stern. 1973. *The Principles of Human Genetics*. (3rd). San Francisco: W.H. Freeman.

Sample Questions

- 1) What is gene pool? How do we investigate the extent of gene pool in human populations.
- 2) What is Assortative mating and how it is different from random mating and inbreeding.
- 3) What are different measures of inbreeding?



UNIT 2 HARDY-WEINBERG EQUILIBRIUM

Contents

- 2.1 Introduction
 - 2.2 Hardy Weinberg Equilibrium (HWE)
 - 2.2.1 Importance and Implications of Hardy Weinberg Equilibrium
 - 2.3 Applications in Human Population Genetics
 - 2.4 Departure from HWE
 - 2.4.1 Factors Affecting Change in Gene Frequency
 - 2.4.1.1 Mutation
 - 2.4.1.2 Genetic Drift
 - 2.4.1.3 Natural Selection
 - 2.4.1.4 Gene Flow
 - 2.4.1.5 Genetic Equilibrium
 - 2.5 Summary
- Suggested Reading
- Sample Questions

Learning Objectives



After reading this unit, you will be able to:

- define what is Hardy-Weinberg Equilibrium or Law;
- depict the importance of HW Equilibrium and the field of population genetics;
- explain the method how to estimate the genotype and phenotype frequencies from HW theorem and to calculate in empirical situation; and
- evaluate the theory behind the deviations from H-WE, especially the gene frequency changes with respect to Mutation, Genetic drift, Selection, Gene flow and how to investigate them in empirical situations in human populations.

2.1 INTRODUCTION

Living organisms are endowed with unique abilities, traits that allow them to survive in a given environment. These traits or abilities may show or exhibit enormous variations within species and across species. Some of these traits are unique to that species; some traits are common within and across species with little variation, these are adaptive characters and gives survival advantage.

These traits are the '*phenotypic*' forms that can be observed as a quantitative trait (or measurable) or classified as types or categories. These traits are hereditary and transmitted across generations: either in the same form or in slight variable form. At times some new traits or variations of the trait appear among the offspring. Some of these traits are governed by '*genes*' or located in the '*genome*' of an organism. The nature of heredity of some of these traits could be complex and/or it could follow some simple principles of transmission.

Human population genetics deals with how these traits or variation change in a population over space and time (generations)? What are the factors that influence the variation of these traits in the population? To what extent these traits are hereditary and are influenced by environment? Can we understand them by simple theoretical models? Can we study how different forces operate differentially in different populations to give a characteristics distribution of gene and genotype frequencies?

Population genetics is the study of gene and genotype frequencies in populations of interbreeding organisms (small or large, natural or artificial) and predicting the way these frequencies are maintained or changed under the combined influence of various factors. It is concerned with applying models of gene frequency change involving different factors in the context of Mendelian genetics to examine evolution in a quantitative manner. In order to understand the pattern of allele frequencies we need to have a defined population, in this case a '*Mendelian population*'. Dobzhansky (1951) defined it as the reproductive community of individuals which share a common gene pool. Evolutionary studies involve reconstructing past demographic events that have led to the present day diversity patterns. Use of various models allows one to examine interplay of various factors and make inferences about the past based on present day data. But one has to be careful about interpreting results obtained from any model considering that all models have some assumptions inherent to them.

2.2 HARDY-WEINBERG EQUILIBRIUM

During the early 1900s people were interested to validate the Mendelian laws of genetics to other organisms, including Man. Are there Mendelian traits in Man?

Mendelian Genetics in Man

BOX 2.1

Brachydactyly in a population: All the mating types are:

The mating types include individuals Normal and Brachydactyly (a Dominant Mendelian trait)

N – Normal and B –Brachydactyly

- Both parents are normal — All the offspring are N
- One parent is N and the other B (heterozygous)
- One parent is N the other B
- Both parents are brachydactyly B (heterozygous)
- Both parents are brachydactyly B (homozygous)

- NN – NN = All the offspring N
- NB – NN = $\frac{1}{2} : \frac{1}{2} = N : B$
- BB – NN = all are B
- NB – NB = 1 N : 1 B: 2 B
- BB – BB = All the offspring B

Of the 5 possible combinations of parental mating types 4 types of matings results brachydactyly offspring.

Therefore, B are more frequent than N in a population as per Mendelian expectation. However Normal individuals are more frequent than Brachydactyly in a population.

there is apparent contradiction between what is observed and what is expected (Mendelian)!

At Cambridge one research scholar was studying 'brachydactyly' – a trait characteristic of small or short digital length ('*brachy*' and '*dactyl*' in Greek

means 'short' and 'digit' respectively) than the normal type. The trait runs in some families. Does 'brachydactyly' follow Mendelian principles? The results of the study showed that '*brachydactyly*' is dominant Mendelian trait and the pedigrees showed 3:1 ratio of brachydactyly to normal offspring. This has invoked an important and interesting question? If it is a dominant trait, there will be more and more brachydactyly individuals in the population, but normal individuals are more frequent than brachydactyly individuals (See Box 2.1)

GH Hardy has solved the puzzle theoretically and published the theorem in Science (Hardy, 1908). GH Hardy's proof illustrates that the gene (or allele) frequency, — here in this case, frequency of brachydactyly individuals in a population, — will not increase over generations, but remain the same, under equilibrium conditions or in the absence of confounding variables. In 1908, Dr. W. Weinberg independently also published similar results (Weinberg, 1908) and is called as HWEquilibrium. (See Box 2.2)

BOX 2.2: Historical anecdotes: HWEquilibrium/Law

In 1908, a German physicist Dr. Weinberg published similar results on Mendelian genetics in a German journal. It was discovered by Dr. Curt Stern (publication in Science 1943), and the Hardy theorem was rightfully referred as HW Law or HWEquilibrium. However, in 1903, at least five years earlier, two scientists have considered similar such possibility of change in gene frequency. They are: WE Castle 1903 in America and Karl Pearson 1903 in England. These two papers considers the question of equilibrium state of gene frequency and change in gene frequency partially with respect to some factors.

a) What is H-W EQUILIBRIUM/LAW?

HWE states that in a randomly mating population of sufficiently large size, and in the absence of the influencing factors such as; mutation, migration, selection, genetic drift and inbreeding, the gene and genotype frequencies will remain constant from generation to generation.

The mathematical proof of invariance of gene frequency under given assumptions, require:

- a) simple knowledge of school algebra and
- b) basic concepts of Mendelian genetics (See Box 2.3).

The proof in case of autosomal '*biallelic*' trait is given in Box 2.4. (for further reading see references)

BOX 2.3: Basic concepts

Phenotype : A trait or a character that is observed as types or measurable and is transmitted from parents to offspring. Some phenotypes are complex with unknown genotypes, and some are directly governed by hereditary units (genes).

Gene : The causative factor of hereditary transmission of traits (phenotypes) and are located in the chromosomes (the hereditary materials in cell nucleus and in mitochondria).

Allele : Genes, the causative factor of hereditary transmission can exist or express in different forms and are referred as ‘alleles’.

Codominant: Where both the alleles are equally expressive in the offspring.

Recessive : The alleles whose expression is suppressed at phenotypic level. The heterozygote offspring of a recessive allele will express the phenotype of the dominant allele.

Haploid : Organisms which carry one set of chromosomes.

Haplotype : It is short form of ‘Haploid genotype’. Refers to genetic markers located on one chromosome. A haplotype can be identified by SNP (single nucleotide polymorphism).

Diploid : Organisms which carry two sets of chromosomes, each set derived from either of the parent. Man is diploid and carries two sets of chromosome (2N).

A diploid individual can carry two copies (alleles) of the gene in each of the chromosome that he or she gets from his or her parents. The two copies could be of the same type (form/status) or of different type (form/status).

Homozygous: The two alleles that an individual carries are of the same or identical types.

Heterozygous: The two alleles that an individual carries are of different type.

Genotype : Is the combination of alleles that a diploid individual can carry in each of the chromosomes.

For example, in case of a ‘biallelic’ gene say A, B two forms (alleles) of the gene that occur in each of the two sets of the chromosomes. There could be three different genotypes: AA, AB, BB.

AA and BB : two different homozygotes (genotype).

AB = BA : heterozygote (genotype).

The box shows the “Punnet’s square” – method of scoring different combination of genotypes based on the male and female gametes or mating types. This can be extended to multiple alleles.

Polymorphism: If a gene exists in more than one form or morph (alleles) and that occurs in stable frequency in a population.

Punnet’s square		
		Male Gamete
		A B
Female gamete	A	AA BA
	B	AB BB
		Genotype

BOX 2.4: Hardy-Weinberg theorem or principle: Proof

In case of genetic trait that has positive family history in a population, let us assume that the gene is biallelic and therefore the two alleles are: B1 and B2 and let

‘p’ is the frequency of ‘B1 allele’ and

‘q’ is the frequency of B2 allele,

N is the total individuals and

So that $(p + q = 1)$ or $p = (1 - q)$ or $q = (1 - p)$

An individual in the population can have three types of genotypes: B1B1, B1B2, and B2B2. And let the frequency of the above three genotypes in the parental population are: P, H and Q respectively.

	Gene (alleles)		Genotypes		
	B1	B2	B1B1	B1B2	B2B2
Frequencies	p	q	P	H	Q

If there are total N individuals in the sample, there will be

P individuals with genotype B1B1 type,

H individuals with genotype B1B2 type and

Q individuals with genotype B2B2 type

So that sum of $(P + H + Q) = N$,

Assuming all the individuals of the three genotypes are equally fertile, then given the genotypes, one can calculate the frequencies ‘p’ and ‘q’ in the population, by gene counting method:

The gene (allele) frequency ‘p’ = $[P + \frac{1}{2}(H)]/N = (B1B1)/N + \frac{1}{2}(B1B2)/N$, and

The gene (allele) frequency ‘q’ = $[Q + \frac{1}{2}(H)]/N = (B2B2)/N + \frac{1}{2}(B1B2)/N$

This is the gene (allele) frequencies of ‘p’ and ‘q’, which are also the gametes produced in the population.

Only some of the gametes form zygotes that will eventually become individuals in the next generation. The allele (gene) frequency in the zygote is unchanged provided there is no reproductive advantage of either of the allele and the zygotes formed represent a large sample of the parental gametes.

Random mating between individuals is equivalent to random union among their gametes. Therefore, in the next generation, the genotype frequencies among the zygotes (fertilized eggs) are the result of random union of two types of gametes. The genotype frequencies among the progeny are therefore can be worked out by Punnet’s square. Or it is the multiplication of the frequencies of the gametic types produced by the parents. Viz.,

Genotype			Allele		
B1B1	B1B2	B2B2	B1	B2	
Genotype	B1B1	2B1B2	B2B2		
Frequency	p^2	$2pq$	q^2		
Absolute freq.	P	H	Q		

BOX 2.4 (Contd.)

Hardy-Weinberg theorem or principle: Proof

In a population there will be three different types of genotypes among males and females, who will mate randomly and they will give rise to their offspring who will represent the same genotypes in the next generation. We will have to work out the frequencies of offspring genotypes given the three genotypes among the male and female parents. This is worked out easily by Punnet's square: the frequencies of different mating types among the male and female genotypes in the population and different possible genotypes among the offspring are.

Frequency of different mating types and the offspring genotypes

		Male Parent — Genotype		
Female Parent		B1B1	B1B2	B2B2
Freq.		p^2	$2pq$	q^2
B1B1	p^2	p^4	$2p^3q$	p^2q^2
B1B2	$2pq$	$2p^3q$	$2p^2q^2$	$2pq^3$
B2B2	q^2	p^2q^2	$2pq^3$	q^4

Once we know the possible offspring genotypes as a result of random mating among the three parental genotypes we can calculate the expected frequencies among the offspring genotypes for different combination of mating types in the population. Given the three genotypes six possible mating types are possible in the population and each mating type will give rise to offspring of different possible combination of genotypes. These are worked out in the following table and this gives the allele frequencies in the offspring population in the next generation:

Frequency of different mating types and the offspring genotypes

Parent		Offspring — Genotype		
Female X Male		B1B1	B1B2	B2B2
(Genotype)	Freq.			
B1B1 X B1B1	p^2	p^4	—	—
B1B1 X B1B2	$4p^3q$	$2p^3q$	$2p^3q$	—
B1B2 X B1B1				
B1B1 X B2B2	$2p^2q^2$	—	$2p^2q^2$	—
B2B2 X B1B1				
B1B2 X B1B2	$4p^2q^2$	p^2q^2	$2p^2q^2$	p^2q^2
B1B2 X B2B2	$4pq^3$	—	$2pq^3$	$2pq^3$
B2B2 X B1B2				
B2B2 X B2B2	q^4	—	—	q^4
1		$p^2 (p^2 + 2pq + q^2)$	$2pq (p^2 + 2pq + q^2)$	$q^2 (p^2 + 2pq + q^2)$
		p^2	$2pq$	q^2

Thus the genotypic frequencies in the offspring remain the same in two successive generations, assuming that allele frequencies are not influenced by selection, mutation and mating is random and there is no differential fertility and mortality and the population is large.

The above is true for *autosomal loci* and can be extended for multiple loci. It is also true for sex-linked trait. Here the gene frequencies will oscillate (by $\frac{1}{2}$) between two sexes in successive generation and will soon reach to equilibrium.

2.2.1 Importance and Implications of HWE

What are the implications and why it is so important? In brief, it is the fundamental theorem of population genetics.

- **Methodology:** Tells us how to calculate (or estimate) the allele frequency or genotype frequency from observed phenotypes in an empirical situation. It can help us to investigate how many alleles are governed by a phenotypic trait.
- **Evolution:** It is quantitative way of understanding the mechanism of evolutionary factors and its influences. Evolution is a dynamic and complex phenomenon and it is hardly possible to study evolution in the laboratory conditions. It gives insights into the inter-relationship between the forces and how to study the effects of each of these forces and the gene frequency. (See the box 2.5 for the relationship between gene frequencies and genotype frequencies).
- It is the benchmark criterion to test whether a new trait is in equilibrium or if not how to test the reasons for the deviations.
- It helps us in genetic counselling to expect the likelihood of a child being homozygous for a recessive deleterious trait given the parental genotype. It helps in forensic science in cases like identification of suspects, parent-offspring disputes etc.
- **Quantitative Genetics:** HWE helps us to investigate complex genetic traits, to estimate the role of environment and genetic components, spatial distribution of gene frequency etc.

Further implications of this principle are as under:

- In case in a population a particular trait or character is in HWE, (the converse) it does not mean that the assumptions are satisfied. (The theoretical proof is complicated and it is available).
- The allele frequencies remain constant from generation to generation. This means that hereditary mechanism itself does not change allele frequencies. It is possible for one or more assumptions of the equilibrium to be violated and still not produce deviations from the expected frequencies that are large enough to be detected by the goodness of fit test.
- When an allele is rare, there are many more heterozygotes than homozygotes for it. Thus, rare alleles will be impossible to eliminate even if there is selection against homozygosity for them.
- For populations in HWE, the proportion of heterozygotes is maximal when allele frequencies are equal ($p = q = 0.50$), and when this happens the heterozygote frequency will be $0.50 (2 \times 0.50 \times 0.50)$. Unless HWE is violated (as in selective loss of homozygotes), heterozygosity can never be more than 0.50 at any biallelic locus. The relationship between gene frequency and genotype frequency is illustrated in Box 2.5.

BOX 2.5**The relationship between gene frequency and genotype frequency**

It is interesting to know the relation between the gene and genotype frequency for a biallelic loci which is under H-W equilibrium. The graph shows the changes in the three genotype frequencies as against the change in allele frequencies A_1 and A_2 from 0 to 1 in Cartesian coordinates (drawn on x and y axis).

A_1A_1/A_2A_2 – homozygotes, A_1A_2 – heterozygotes

Relationship between genotype frequency and gene frequency
- biallelic GENE

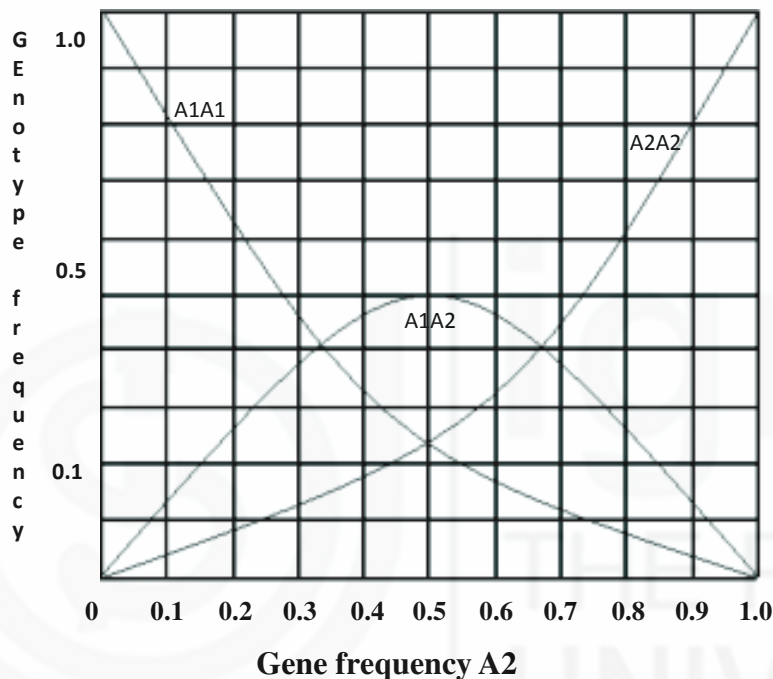


Fig: Left top curve – A_1A_1 , Right top curve – A_2A_2 , Lower curve – A_1A_2

The graph shows two interesting properties of the HWE:

- ✓ The frequency of the heterozygotes can reach to a maximum of 50%
- ✓ And this can occur when the gene frequency of ' p ' = ' q ' = 50%
- ✓ When one of the gene frequencies of an allele is low, the rare allele predominantly occurs as heterozygotes and there will be few heterozygotes.

- An application of HWE is that when the frequency of an autosomal recessive disease (e.g., sickle cell disease, hereditary hemochromatosis, congenital adrenal hyperplasia) is known in a population and unless there is reason to believe HWE does not hold in that population, the gene frequency of the disease gene can be calculated. Likewise, the carrier rate may be calculated for autosomal recessive disorders if the disease gene frequency is known. For example, phenylketonuria (PKU) occurs in $1/11,000$ (q^2), which gives a heterozygote carrier frequency of approximately $1/50$ [$2pq(1-q)$]. If the diseased individuals (q^2) are deducted from the whole population, the carrier rate in normal individuals approximates to [$2q/1+q$].

- It has to be remembered that when HWE is tested, mathematical thinking is necessary. When the population is found in equilibrium, it does not necessarily mean that all assumptions are valid since there may be counterbalancing forces. Similarly, a significant deviance may be due to sampling errors (including Wahlund effect), misclassification of genotypes, measuring two or more systems as a single system, population substructure, failure to detect rare alleles and the inclusion of non-existent alleles. The Hardy-Weinberg laws rarely holds true in nature (otherwise evolution would not occur). Organisms are subject to mutations, selective forces and they move about, or the allele frequencies may be different in males and females. The gene frequencies are constantly changing in a population, but the effects of these processes can be assessed by using the Hardy-Weinberg law as the starting point.
- The direction of departure of observed from expected frequency cannot be used to infer the type of selection acting on the locus even if it is known that selection is acting. If selection is operating, the frequency of each genotype in the next generation will be determined by its relative fitness (W). Relative fitness is a measure of the relative contribution that a genotype makes to the next generation. It can be measured in terms of the intensity of selection (s), where $W = 1 - s$ [$0 < s < 1$]. The frequencies of each genotype after selection will be $p^2 W_{AA}$, $2pq W_{Aa}$, and $q^2 W_{aa}$. The highest fitness is always 1 and the others are estimated proportional to this. For example, in the case of heterozygote advantage (or overdominance), the fitness of the heterozygous genotype (Aa) is 1, and the fitnesses of the homozygous genotypes negatively selected are $W_{AA} = 1 - s_{AA}$ and $W_{aa} = 1 - s_{aa}$. It can be shown mathematically that only in this case a stable polymorphism is possible. Other selection forms, underdominance and directional selection, result in unstable polymorphisms. The weighted average of the fitnesses of all genotypes is the mean fitness. It is important that genetic fitness is determined by both fertility and viability. This means that diseases that are fatal to the bearer but do not reduce the number of progeny are not genetic lethal and do not have reduced fitness (like the adult onset genetic diseases: Huntington's chorea, hereditary hemochromatosis). The detection of selection is not easy because the impact on changes in allele frequency occurs very slowly and selective forces are not static (may even vary in one generation as in antagonistic pleiotropy).
- All discussions presented so far concerns a simple biallelic locus. In real life, however, there are many loci which are multiallelic, and interacting with each other as well as with the environmental factors. The Hardy-Weinberg principle is equally applicable to multiallelic loci but the mathematics is slightly more complicated. For multigenic and multifactorial traits, which are mathematically continuous as opposed to discrete, more complex techniques of quantitative genetics are required.

2.3 APPLICATION IN HUMAN POPULATION GENETICS

The behaviour of HW principle under different assumptions is the discipline of 'population genetics', which describes, primarily, the changes in gene frequency that are influenced by demographic factors, population structure variables,

historical, random events, sampling fluctuations and evolutionary factors of selection and mutation. In simple the four main factors that influence the gene frequency in a population are: mutation and genetic drift (non-systematic factors), migration and selection (systematic factors). The genetic drift is effective, more specifically, in populations whose size is small or limited e.g., an isolate or an island population or a small endogamous population. These are described in detail below:

- For example, HWE has helped us to find out to investigate the number of alleles of ABO locus and how to calculate the gene frequency of ABO locus (e.g., Bernstein has given the method of correction) (see Box 2.6)
- We were able to understand how HbS despite its deleterious effect it maintains its equilibrium in the population.
- HWE helps to understand the some of the health problems in some isolated populations, whose propagation is the result of genetic drift, and selection or inbreeding etc.
- HWE has forensic applications in solving problems related to disputed paternity, to provide evidence in case of crime to detect the culprit, property or biological inheritance cases.
- It helps in understanding the complex genetic disorders, to be able to estimate the contribution of genetic versus environmental effects.
- HWE helps to understand to investigate the human origins, the role of selection versus demographic effects on the genetic diversity in a population.

BOX 2.6

HWE – Gene frequency estimation: Gene counting method

Given the information about the genotypes, HWE helps us to estimate the allele frequency by 'gene counting method' (how many alleles a genotype contains). For example,

- As each homozygote carries two alleles and each
- Heterozygote carries one allele, therefore, estimate of an allele frequency in a population of size N individuals (or 2N alleles) will be

$$= (2 \text{ homozygotes} + \text{heterozygote}) * (1/2N)$$
- In a population there will be three genotypes and their absolute frequency will be say N₁, N₂ and N₃ (where N₁ + N₂ + N₃ = N). If there are two alleles say 'A' and 'a' with a frequency 'p' and 'q' respectively (where p + q = 1).
- By gene counting method assuming HW law the gene (allele) frequencies of p = (1/2N) * (2N₁ + N₂), q = (1/2N) * (2N₃ + N₂) and p = (1 - q)

2.4 DEPARTURE FROM HWE

In general, the factors that are assumed to be non-operative under HWE are hardly realised in the living systems. The living system (populations or organisms) are structured (non-random entities) and are influenced by multiple and interactive factors that operate through space and time. With the help of HW equilibrium it is possible to investigate and estimate the effect of these individual forces that change gene frequency in human populations.

2.4.1 Factors Affecting Change in Gene Frequency

The four aspects of the H-W assumptions are:

- i) Demographic:
 - a) Size, mating, fertility and mortality, and migration
- ii) Evolutionary:
 - a) Mutation, selection, gene flow
- iii) Population structure:
 - a) Social and Cultural factors
 - i) Matings and Marriage specifications that regulate the marriage or mating type in a population.
 - ii) Non-random mating – Sexual selection of mates
- iv) Ecological:
 - a) Population bottle-neck events:
 - i) Pandemic: disease, earthquake etc.
 - ii) Historical: wars etc.,

Of the above factors, for the present academic purpose, we will be dealing a few factors and examine how these factors change or influence the gene frequency in a population and how to estimate them in empirical situation.

2.4.1.1 Mutation

Mutation is a random change in phenotypic or genotypic forms that occur once a while in a population. The probability or likelihood of occurrence, in a population, in general, is of the order of one over several lakhs or tens of thousand of individuals.

For example, several of the Mendelian syndromes and disorders that have been discovered in human populations are the result of mutation. In general, it is observed to be a single mutation or point mutation. At molecular level, mutation primarily refers to changes in the DNA sequences (or SNPs — Single Nucleotide Polymorphism) in the genome of an individual (population) with phenotypic manifestations resulting to non-normal cases, some of them are clinically or medically identified as diseases or a syndrome. If one can search *web* resources, there is a data base created by Hopkins institute and or by NIH (America) on a list of Mendelian syndromes, which can be found by a search criterion OMIM (Online Mendelian Inheritance in Man). One can also find such data bases from a variety of sources.

Some examples will help us to get an idea of mutation and its effects. Sickle cell anaemia (or HbS condition), is a disease related to Haemoglobinopathies. Its inability to synthesize Oxygen (O_2) to its full capacity by an individual who is suffering from the disease or the trait, which results a risk to survival liability. This is identified as due to a point mutation or single mutation at the 6th position of the α -globin chain of the haemoglobin gene. The single aminoacid substitution (α_6 Glu to Val) changes the haemoglobin structure, which is phenotypically identified as sickle cell shaped form (or half moon shaped form) of the RBC.

Mutation is an important factor or ingredient leading to the appearance of new characters in the population. The fate of the mutation as a new character in a population depends on its advantage or disadvantage that it can impinge to the

survival fitness of the population. For example, mutation is a significant evolutionary force which can change allele frequency variation in a population under a favourable environment. How we can know the relation between the mutation and allele frequency change from HWE.

a) **Change in gene frequency due to mutation (μ):**

Random changes that happen at the DNA sequence, especially at the coding region of the gene can create an allele which can alter the gene frequency in the population in successive generations. This can be investigated theoretically given the mutation rate per generation in the population. This has been shown in Box 2.7 for bi allelic loci.

The theoretical results suggest that the change in gene frequency of a mutant allele, after 't' generations, depends on the initial allele (gene) frequency before mutation and the mutation rate of the allele per generation in the population.

This is an important result and can help to calculate change in gene frequency after 't' generations given the mutation rate (μ) per generation and the initial gene frequency in the population.

b) **Rate of mutation (μ):**

Though the mutation is random, but the rate of mutation varies. It is site specific – there are 'hot-spots' where mutation rate is more frequent than in other parts of the genome. In general, the coding part of the gene does not support mutation to occur, as a result of proof reading process and functional importance of the codons. However, mutations occur at higher rate in the intronic region, and in the repeat sequences than in exons or codons. Also the mitochondrial non-coding parts, viz., hyper variable regions HV1, HV2 in the D-loop has higher rate of mutation than in nuclear genome.

BOX 2.7

Change in gene frequency due to mutation (μ)

If there are two alleles 'A' and 'a' with its frequencies ' p_0 ' and ' q_0 ' at the initial stage (say at time ' t_0 ') in a population and ' μ ' is the mutation rate that changes allele 'A' to 'a' per generation, then gene frequency (g.f.) of 'A' will decrease by an amount ' μp_0 ' in the first generation. Therefore the g.f. of 'A' allele in the first generation after mutation will be:

$$P_1 = p_0 - \mu p_0 = (1 - \mu) P_0$$

In the (next) second generation the gene frequency is expected to be:

$$\begin{aligned} P_2 &= P_1 - \mu P_1 = (1 - \mu) P_0 - \mu(1 - \mu) P_0 \\ &= (1 - \mu) P_0 \quad [(1 - \mu)] \\ &= (1 - \mu)^2 P_0 \end{aligned}$$

After 't' generations the g.f. of 'A' is expected to be

$$P_t = (1 - \mu) P_{t-1} = (1 - \mu)^2 P_{t-2} \dots = (1 - \mu)^t P_0$$

When μ is very small $(1 - \mu)^t$ can be approximately equated to $e^{-\mu t}$, (where e is natural logarithm to base e), therefore, gene frequency after 't' generations will be

$$P_t \sim P_0 e^{-\mu t}$$

Therefore, the mutations that occur at HV1 and HV2 regions of mitochondrial genome help us to investigate the short-term evolution or micro-evolutionary

trends in sub-populations. This has helped us to address some of the questions of human origins or to verify the Darwin's hypothesis that the Africa is the origin of Man. This also helps to enquire the antiquity and past genetic history of diverse populations and their diversity and relationship with other human populations.

2.4.1.2 Genetic Drift

Genetic drift is an important non-systematic evolutionary force. To understand the concept of genetic drift, let us know what the word 'drift' conveys, in general. One of the descriptions for the word 'drift' in the English Dictionary is: "move aimlessly from one place or activity to another" – this is more with reference to things or events that we experience with practical world e.g., drifting by air, wind and water or ocean. Similar phenomena can also happen with respect to gene frequency in a small population. In small populations, as a result of population-events such as pandemic diseases, earthquakes etc., the population size is drastically reduced which can have significant effect on the genetic diversity and gene frequency: for example, the gene frequency can drift from one generation to generation randomly leading to either loss or fixation of alleles over generations (in the absence of other interfering factors). In small populations or due to demographic and ecological effects the population size drastically reduced to a fraction (or a random sample) of the original population with allelic representation different from the original population. In these cases, there will be random changes in gene frequency, which appear to drift at varying frequencies in successive generations in an erratic manner. For example, the studies on the origins of Man, suggest that decreasing heterozygosity and linkage disequilibrium levels away from Africa are supportive of the role of genetic drift among human populations.

To understand how the genetic drift can happen or possible, one can investigate and/or understand by attempting some simple examples or simulation exercises. These are available on the online resources. One such example is illustrated in Box 2.8.

a) **Bottle-neck effect**

Genetic Drift can happen in a variety of ways due to different events that populations experiences in empirical situation. These have been referred as part of ecological factors that disturb the population size (see 2.4.1). Historically the world has experienced several pandemic diseases in the past: e.g. Syphilis, Plague, leprosy, malaria, HIV infection, etc. which has killed or wiped out bulk of the population. The natural geographical events like earthquakes, tsunami etc. had killed vast majority of the populations. Even the political and man interfering events like explosion of atomic bombs, world wars etc. have affected the demographic size of the populations. Each such event is followed by a drastic reduction in population sizes. In genetic terms it means reduction in genetic diversity (at the time of the event), and those survived will have different allelic profile or gene frequency and the stability of a particular allele over generations depends on the demographic structure of the population.

'Breeding individuals' part of the demographic structure of a population is of particular genetic importance. They are capable of mating and producing children. They will be a fraction (of the total population) who contribute to the next generation or gene pool and is referred as 'effective size' (N_e).

BOX 2.8**Simple exercises to understand the genetic drift**

There are different ways to replicate to illustrate the random drift phenomena. One such simple example could be the following:

- Start with a jar that contain with N number of blue, red, yellow balls.
- At the first step blindly or randomly take out (say e.g., by hand) some balls and put them in the second bottle.
- Then from the second bottle, take some balls (e.g., by hand) and put in the third bottle.

If you have started with large sample (N) of mixed coloured balls you can repeat the same. Otherwise, at the third/fourth bottle you can count how many of red, blue and yellow balls. Compare the outcome with the original number of red, blue and yellow balls at the start. They will differ from the original number at the start. You may also find the absence of a particular colour at the fourth (or nth) bottle.

In case of Genetic Drift, similar such random sampling of gene frequency changes happen over successive generations in a small population. One can search several such simple examples on the online resources on genetic drift – bottleneck effect, founder effect etc.

Genetic drift can alter the 'effective size' of a population and change the genetic diversity. After successive generations, the gene frequency in the population will be significantly different from the gene frequency before genetic drift. This is similar to the bottle neck, where the narrow neck of the bottle restricts the flow and this event is referred as 'bottle neck-effect' in population genetics. Such bottle neck effect resulting to sudden population size reduction had been experienced by several human populations in the past historical times affecting the genetic structure: genetic diversity, gene frequency changes.

b) Founder effect

The word 'founders' refers to the ancestors or the earliest settlers who colonised or founded the new population in alien territories. It could be an historical adventure of war fare, or exploration to a new island or new area or it could also be due to chance factors like surviving from a sudden calamities like ship wreck, etc. or it could be serial migration of people at different timing to other places: in all the cases, a few founders start living and establishing a new subpopulation.

In genetic scenario, the few founders represent a random sample of the genes from the original population or gene pool from which they got separated. It is possible that, some of the rare alleles that are in the large population, by chance, may not be present in the founder individuals. It could be that, among the founders, especially if the founders are related, by chance, some of alleles may be of a higher frequency than the original population. Therefore, in the new colony after generations the gene pool will have either absence of the allele or higher frequency of the rare allele than when compared to the original population.

c) **Serial founder effect**

It is possible that people or organisms migrate repeatedly over time or waves of migration from a region to found new colonies. Such repeated waves of migration at different time periods produce successive subpopulations or gene pools whose genetic profile will be different. There appears to be waves of out of Africa migration to other continents that had happened at different time periods in the past, whose genetic signature can now be traced among the extant populations in South Asia, Europe, and America etc. The mitochondrial, X and Y chromosomal haplogroup distribution of continental populations can be explained as a result of founder effect of out of Africa hypothesis of human origins.

d) **Empirical studies of founder effect in Man**

The importance of 'Founder effect' as significant evolutionary factor has been outlined by German evolutionary biologist Ernst Mayr (1942). Founder effect is the "*The establishment of a new population by a few original founders (in an extreme case, by a single fertilized female) which carry only a small fraction of the total genetic variation of the parental population.*» This is sampling effect especially the genetic composition and evolution of the successive generations entirely depends upon the few founders. A few examples illustrating the role of genetic drift in the gene frequency changes are shown in Box 2.9

BOX 2.9**Studies on Genetic Drift**

- **Tristan da Cunha is an island;** the few hundred individuals (<300) living on the island are mostly the (15) descendants (8 males and 7 females) who had founded the island in 1816-1908. Three of the founders were Asthma sufferers and there is high incidence of Asthma in the population. In a study of the 9 Y-chromosome haplotypes of the island, seven of them are traced to its 7 male founders.
- **Amish population, USA:** All most all the Amish population (~249K) descended from about 200 founders from German during 18th century. The population is endogamous, they show high frequency of genetic disorders as a result of founder effect that include dwarfism, metabolic disorders, unusual distribution of blood types, metabolic disorders etc.
- **'Blue Fugates' of Appalachian, Kentucky, USA:** In 1800, Martin Fugate and his wife settled in trouble some creek in Kentucky. They carried recessive gene methemoglobinemia (met-H). Due to deficiency of an enzyme diaphorase (NADH methemoglobin reductase) met-H levels rise and this gives rise to reduced oxygen-carrying capacity. This gives a tinge of blue skin of the homozygous condition. Isolation and inbreeding has caused to increase of blue people which are traced to the founders Fugates.
- **India:** In the northeast populations, some of them live in geographical isolation, practice endogamy show unusual frequency of a few genetic

traits which are expected to be due to genetic drift and founder effect. Some of them include:

- Complete lack of A2, cde, K, pc, and AK2 genes, lack of isozyme ALDH-1 (Roychoudhury and Nei 1997), a high prevalence (about 50%) of lactase malabsorption (Flatz 1987).
- Low frequency of AIBG*2 allele (Juneja et al. 1989), high frequency of G6PD deficiency in Naga (Seth and Seth 1971), absence of ‘Gd_’ variant in Adi and Hmar and high frequency of this variant in Bodos (Saha et al. 1990).
- Continuing from classical genetic observations, unique and rare allele frequency of microsatellite loci among the Adi subpopulations (Krithika et al. 2005). High frequency of susceptibility of tuberculosis in some clans of tribes, stomach cancer, high incidence of cardio deaths etc.
- Absence of attached ear lobe among the Nandiwalas in Maharashtra.
- Population size reduction and allele frequency changes among Ahmedias of Kashmir population.

2.4.1.3 Natural Selection

Charles Darwin (and Wallace) has described natural selection as one of the important factor (key mechanism) of evolution. Natural selection happens where there is differential rate of reproductive success among different genotypes (underlying the phenotype, or trait or observed character). How selection operates at the molecular (genome) level for example, especially change in gene frequency considered, theoretically, in population genetics.

Due to differential reproductive success involving these variant of the trait, there will be more offspring with the variant than those individuals with other variant of the trait. In Darwinian sense ‘*fitness*’ (‘*Darwinian Fitness*’) refers to ability to contribute successfully to the next generation. This is also referred as ‘adaptive value’ or ‘selective value’. Therefore, if the differences of fitness are in a way associated with the presence or absence of a particular allele (or gene) in the individual’s genotype then selection operates at the genetic level.

When a gene is subjected to selection (or under selective pressure), its frequency in the offspring is not the same as in the parents (or in the previous generation) as parents with different genotypes pass on their genes *unequally* to the next generation. This leads to change in gene frequency and consequently also of genotype frequency, as a result of selection (of a particular gene). The theoretical investigation of change in gene frequency of an allele under selection pressure is more complex, than factors like mutation, migration. There could be different situations under which selection can operate in a population and different situations need to be incorporated in theoretical models. Here we will consider a few of those situations (types of selection) in a more descriptive way, rather than theoretically, which is beyond the scope of the present purpose.

Theoretically, selection is measured by 'fitness' (' W ') or by selection coefficient (' s '). *Fitness* refers to 'relative rate of survival'. The selection coefficient (' s ') is defined as ($1 - W$) and the value varies between 0 and 1. Once the fitness is quantified and defined the different types of dominance can be taken as degrees of dominance with respect to fitness (this is different from the dominance effect of the gene). In general, most mutant genes are completely recessive compared to the wild type as can be observed from phenotypic form of the trait. This does not imply that the heterozygotes are equally fit when compared to homozygote.

Before we get to know the effect of selection on gene frequency, it is required to know different types of selection and its fitness values. Some of the known selection types are: no dominance, partial dominance, complete dominance, over dominance. The fitness values for the four types are shown below (See Box 2.10). The change in gene frequency with respect the four types of selection (with fitness values) are given in Box 2.11.

BOX 2.10			
Types of dominance or degree of dominance and fitness			
	A_2A_2	A_1A_2	A_1A_1
a. No dominance:	I ————— I	I ————— I	I
	$1 - s$	$1 - 1/2s$	1
	A_2A_2	A_1A_2	A_1A_1
b. Partial dominance:	I ————— I	I ————— I	I
	$1 - s$	$1 - hs$	1
	A_2A_2		A_1A_2, A_1A_1
c. Complete dominance:	I ————— I	I ————— I	I
	$1 - s$		1
	A_2A_2	A_1A_1	A_1A_2
d. Over dominance:	I ————— I	I ————— I	I
	$1 - s_2$	$1 - s_1$	1

a) Types of selection

Selection is a systematic force and operates in different ways. Selection takes place when there is differential fitness of a heritable trait. Based on the effect on the allele frequencies, the selection can be seen operating into three types.

Directional selection: occurs one extreme value or allele is selected. In case if one of the allele of a variety of the trait has greater fitness and producing more offspring of that allele or a variety, then the selection is said to be directional. The effect of directional selection is fixation of allele with greater fitness and the loss of the allele with least fitness. For example: well known cases come from the parasitic world, especially resistance to antibiotics in case of some of the vector-borne diseases. Initially as a result of antibiotic the parasite growth comes down to zero, but the parasites develops some mutant or new variant which gets resistance against the antibiotics or better fitness in the presence of antibiotics, in due course, the less fit variant is replaced by new variant which can survive against antibiotics. This can be illustrated as a shift in the mean of the character of a distribution (See box 2.12).

BOX 2.11: Change in gene frequency under selection

First we will consider the basic formulae for the change in gene frequency that is achieved in one generation of selection. Under the similar notation that has been used above for other factors (p = gene freq. of A_1 and q = gene freq. of A_2), the below table shows the genotype frequencies under HWE before selection to the allele for the three genotypes (first line).

	Genotypes			
	A_1A_1	A_1A_2	A_2A_2	Total
Initial frequency	p^2	$2pq$	q^2	1
Coefficient of selection	0	0		s
Fitness	1	1	$1 - s$	
Genetic contribution	p^2	$2pq$	$q^2 (1-s)$	$(1 - sq^2)$

Here we consider selection acting on the recessive genotype A_2A_2 with a selection coefficient: 's' acting against it. This will have a differential fitness to the genotypes that will be as given in the second line. By multiplying the initial frequency by the fitness values gives the frequency of each genotype after selection. This is the third line – the genetic contribution to allow to selection to operate over life cycle. Therefore, after selection, there will be a loss of fitness that is proportional to an amount $(1 - sq^2)$. From this we can calculate the frequency of A_2 gametes produced (frequency of A_2 genes in the progeny). The new g.f. is (where $p = (1 - q)$)

$$q_1 = [q^2 (1 - s) + pq] / (1 - sq^2)$$

$$= [q - sq^2] / (1 - sq^2)$$

The change in gene frequency Δq , resulting from one generation of selection is

$$\Delta q = q_1 - q = sq^2 (1 - q) / (1 - sq^2)$$

This tells us that the effect of selection on gene frequency depends not only on the intensity of selection s, but also on the initial gene frequency (of the recessive allele).

Different type of selection

What we have considered above is in general selection with respect to recessive allele q under selective pressure. But there are variety (or types) of selection that can act on the allele frequency. Depending upon the type of selection the change in gene frequency will consequently change. These are shown in the following table.

Initial freq. & fitness of Genotypes			New gene frequency due to selection at q	Change in gene frequency
A_1A_1	A_1A_2	A_2A_2	q_1	$\Delta q = q_1 - q$
p^2	$2pq$	q^2		
1.	1	$1 - \frac{1}{2}s$	$1 - s (q - \frac{1}{2}sq - \frac{1}{2}sq^2) / (1 - sq)$	$-\frac{1}{2}sq (1-q) / (1 - sq)$
2.	1	$1 - hs$	$1 - s (q - hspq - sq^2) / (1 - 2hspq - sq^2)$	$-spq[q + h(p-q)] / (1 - 2hspq - sq^2)$
3.	1	1	$1 - s (q - sq^2) / (1 - sq^2)$	$-[sq^2 (1-sq)] / (1 - sq^2)$
4.	$1-s$	$1-s$	$1(1-sq + sq^2) / (1 - s(1-q^2))$	$+ [sq^2 (1 - sq)] / [1 - s(1 - q^2)]$
5.	$1-s_1$	1	$1 - s_2 (q - s_2q^2) / (1 - s_1p^2 - s_2q^2)$	$+ [pq(s_1p - s_2q)] / (1 - s_1p^2 - s_2q^2)$

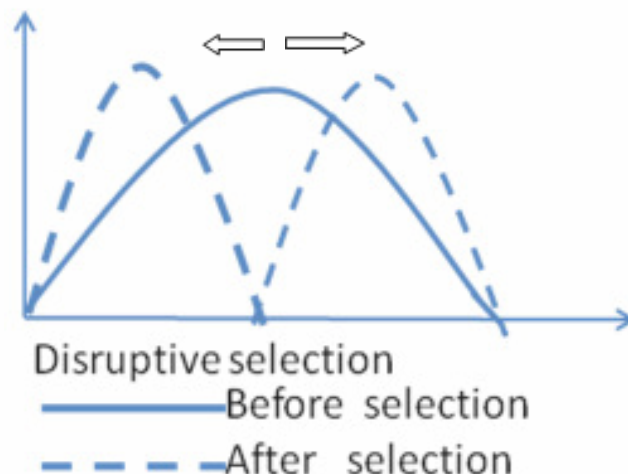
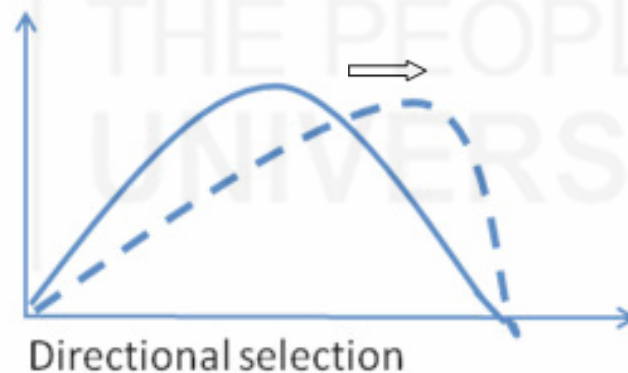
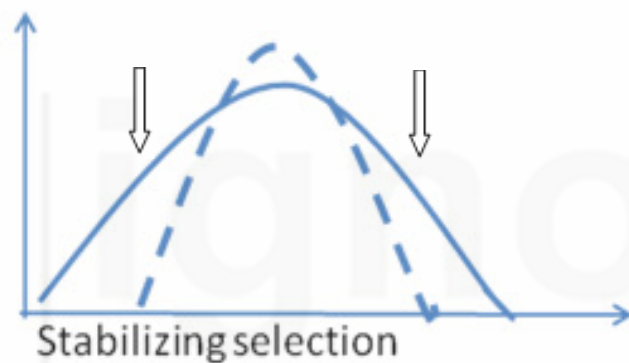
Above the different types of dominance are:

1. No dominance, selection against A_2
2. Partial dominance of A_1 : selection against A_2
3. Complete dominance of A_1 : selection against A_2
4. Complete dominance of A_1 : selection against A_1
5. Over dominance: Selection against A_1A_1 and A_2A_2 (Applicable to any degree of dominance with fitnesses expressed relative to A_1A_2)

Stabilizing selection: The extremes are selected in favour of the middle. In case of stabilizing selection, the two extreme values of a trait or alleles will have lower fitness than the intermediate value or the heterozygote alleles of a trait. One of the well known examples includes birth weight. The average birth weight of offspring ranges between 2500g (5.5pounds) to 4500g (10 pounds). Offspring with weight less than 2500g are low birth weight and greater than 4500g are the heavy babies and both have less chance of survival. As a result the selection favours the offspring with the average birth weight. Stabilizing selection is also the reason in case of height distribution in a population. This can be illustrated as a change in mean values of the distribution (See Box 2.12)

BOX 2.12

Different types of selection. Change in gene frequency in case of Stabilizing, Directional and disruptive (balancing) selection



Balanced Selection: In case of balanced selection, the heterozygotes have higher fitness than either of the homozygotes. This is also called heterozygous advantage or over-dominance. The best example is the sickle cell anaemia. In non-malarial environment the homozygote state of the sickle cell anaemia will have low fitness and as a result the allele gets lost in the population in due course of time. However, in malarial environment, Homozygote sickle cell anaemic individuals have the better fitness as equal to the normal homozygote individuals; as such both the alleles will be maintained in the population. (See Box 2.12).

Disruptive selection: both the extreme value (alleles) of a trait gets selected. It is one form of balanced selection. In case of disruptive selection, the extreme values or the alleles (low and high) of a trait will have a higher fitness when compared to the average value. As a result of disruptive selection the extreme values will increase as against the average values of the trait. This can be explained as leading to bimodal distribution (See Box 2.12).

b) Opportunity for natural selection

In general, to investigate natural selection in human populations is complex. Since natural selection operates on fertility and mortality, it can help us to get an overall idea of operation of natural selection. Indeed, Crow (1958) has formulated an index (Crow's Index) to examine the maximum intensity of (natural) selection that is more applicable for human populations; the index is based on the demographic components of fertility and mortality rates. According to Crow, *“there can be selection only if, through differential survival and fertility, individuals of one generation are differentially represented by progeny in succeeding generations. The extent to which this occurs is a measure of ‘total selection intensity’. It sets an upper limit on the amount of genetically effective selection.”*

The total selection intensity (as defined by Crow) has two components: A fertility component (I_f) and mortality (I_m). The fertility and mortality patterns depend on several factors that vary across populations such as age at marriage, menarche, and survival to reach to fertility age, variation in fertility and age of death etc. Likelihood of these occurring needs to be calculated based on age-sex structure.

The fertility and mortality also include embryonic development and birth; these have been incorporated to make it more rigorous and efficient estimate by Johnston and Kensinger (1971). More details of the Crow Index and the relationship are given in Box 2.13.

The estimates of ‘total intensity of selection’ have been studied in wide diverse populations. In Indian scenario, tribal populations show larger ‘Index of mortality than fertility components. There is also an overall declines in I_m and I_f among urban communities as a result of socio-economic and public health facilities. More details of the trends of the Crow's Index in Indian populations are described by Gautam (2009).

BOX 2.13**Index of opportunity for selection****Crow (1958) and Johnston & Kensinger (1971)**

The total selection intensity (I_t), is computed based on

I_m = index of opportunity for natural selection due to pre-reproductive mortality (mortality from birth to reproductive age, i.e. below 15 years).

I_f = index of opportunity for natural selection due to fertility.

X = average number of live births per women who have completed their reproductive life span (aged 45 years and above).

V_f = variance (average deviation from mean) of number of live births.

P_d = proportion of pre-reproductive deaths.

P_s = proportion of survivors from birth to reproductive ages.

The proportion of pre-reproductive deaths (P_d) is calculated from children ever-born to mothers aged 45 years and above (who have completed their fertility) and pre-reproductive deaths.

The proportions of survivors were calculated by subtracting P_d from 1:

$$I_t = I_m + I_f/P_s$$

$$I_m = P_d/P_s \quad \& \quad P_s = 1/P_d$$

$$I_f = V_f/X^2$$

The crow's Index of opportunity for selection was modified by Johnston and Kensinger (1971) to account for the survival and mortality component during conception, before the birth of an infant.

This include I_{me} = the selection due to prenatal mortality, P_{ed} = the probability to die before birth,

P_b = the probability to survive till birth, I_{mc} the index of total selection due to postnatal mortality,

P_d = the probability to die before reaching reproductive age,

P_s = the likelihood to survive till reproductive age, I_f = selection due to fertility,

V = variance due to fertility among women who had completed their fertility,

X is the mean number of births, P_d and P_s are proportion of deaths and survivors.

The modified total intensity index I_t is:

$$I_t = I_{me} + (I_{mc}/P_b) + (I_f/P_b) P_s$$

$$I_{me} = P_{ed}/P_b, \quad P_b = 1 - P_{ed}$$

$$I_{mc} = P_d/P_s$$

$$P_s = (1 - P_d)$$

$$I_f = V/X^2$$

2.4.1.4 Gene Flow

a) Migration

Migration or gene flow is an important factor that can change the gene frequency. Emigration or immigration of individuals between populations can alter or change in the gene frequency. In genetic terms it is either loss of genetic diversity due to emigration or increase of genetic diversity due to immigration of individuals. There is loss of gene flow from a gene pool or gain of gene flow into a subpopulation from other gene pool. The quantitative estimate of the effect of migration in case of an allele at a single locus has been estimated by Bernstein and it has been shown (Box 2.14).

BOX 2.14

Change in gene frequency due to migration (m) / gene flow or genetic admixture

Suppose if migration is unidirectional from mainland to a nearby island and is random, then suppose

‘m’ is the rate of migration per generation from mainland to island

a) p_i be the frequency of gene A in immigrating individuals

‘ p_0 ’ is the frequency of gene A in the island

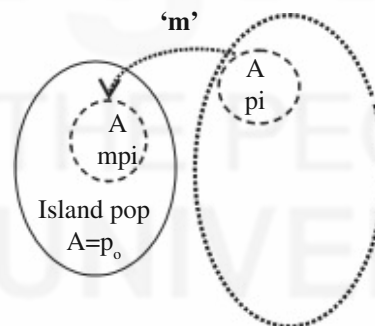
b) The gene freq of A in the *island after migration* is

$$p_{am} = m p_i + (1 - m) p_0$$

The change in gene freq in one generation is

$$\begin{aligned} p_{am} - p_0 &= [m p_i + (1 - m) p_0] - p_0 \\ &= m (p_i - p_0) + p_0 - p_0 \end{aligned}$$

$$m = (p_{am} - p_0) / (p_i - p_0)$$



NB: This is based on Bernstein's formula for an allele at a single locus

The effect of migration rate (m) on allele frequency in a population is the proportion of differences of allele frequency in the island population (p_i) before and after migration (p_{am}) to the difference between allele frequency in the migrant population (p_0) and the island population (p_i). The above formula can be extended for a multiple loci by using least square or maximum likelihood estimate procedures. It can also be worked out based on gene identity method.

b) Genetic admixture

Gene flow can happen between two subpopulations through random mating or admixture or marriages. The American Blacks, Anglo Indians, are examples of genetic admixed populations. The Latin American countries are populated by admixed populations contributed by native tribes, African, European and other settlers. The estimates of admixture proportion can be estimated for a gene located at a specific locus of interest or for a set of genetic markers located at different loci. The above formula (Box 2.14) can be used to estimate the 'm' the admixture

in a hybrid population. It is also possible to estimate the admixture proportions based on genetic distances and from principal component analysis (for multilocus allele frequencies).

c) Barriers to gene flow

Human populations live over wide geographical regions forming local subpopulations; these subpopulations are formed as a result of endogamy which is promoted by geographical, cultural, linguistic, political and other factors. The same factors form barriers for gene flow and restrict the admixture, intermarriages etc. between the local populations. In India caste, geographical isolation, cultural, linguistic, political factors play a major role in restricting the gene flow or admixture or intermarriages between groups.

d) Theoretical Models of gene flow

These factors are important to consider estimating or modelling the gene flow between populations. In population genetic point of view, there is a decrease in genetic diversity with the increasing distance or geographical location of the populations. This gives spatial pattern of gene frequency clines, which help us to understand the geographic variation of genetic markers across populations and regions.

Since gene flow can occur in different scenarios, there are a variety of theoretical models to account for different situations of spatial gene exchange or flow. For example, Sewall Wright has proposed 'island', 'neighbourhood' and 'isolation by distance' models and 'steppingstone' model by Kimura and Weiss.

Island model: It is the simple situation similar to island population. Suppose the population is distributed among a few close (equi distance) islands, each of population size N . The people tend to marry within each of the islands and gene flow is restricted, in the sense that there is equal immigration between islands, hence 'island model'. Suppose the mating takes place at random in each of such island or insular populations. The gene frequency in each of the island will differ with respect to total population (of all the islands). The theoretical results show that the deviation in such island model is exactly the variance in allele frequency among the islands. The number of homozygotes in the total population is always larger than expected from HW proportions in that population. The result is known as '*Wahlund principle*'. For a two-allele polymorphism, the genotypic proportions in the total population are:

$$AA : p_0^2 + V, \quad Aa : 2p_0q_0 - 2V \quad \text{and} \quad aa : q_0^2 + V.$$

These proportions are similar to those population practicing inbreeding with inbreeding coefficient 'F' ($F = V/p_0q_0$). Where P_0 is the gene frequency of allele A and V is the variance of the gene frequency among the islands. "*The change in heterozygote frequency is twice the covariance among populations in the frequency of the allele in the heterozygote, and this may be positive or negative.*" (Christiasen and Feldman, 1986). One other model proposed by Sewall Wright is Neighbourhood model.

Steppingstone model: The island model is too realistic to realise, therefore other models have been proposed which is more close to geographically structured populations. Kimura and Weiss (1964) proposed the '*stepping stone model*'. In

'one-step-linear (one dimensional) stepping stone model, the populations are arranged, rather in a linear fashion, on a long chain. The migration occurs between the neighbouring populations. This situation allows the distant populations with least migration between them are expected to behave differently than the neighbouring populations that are expected to change the gene frequency of the extreme populations as against the neighbouring populations. Kimura and Weiss (1964) have shown that the correlation in gene frequencies (r) between demes decreases approximately exponentially as a function of the number of steps (x) between demes.

This is expected to lead to clines in the gene frequency or geographical clines of the allele frequency.

Isolation by distance model: This was proposed by Sewall Wright, which is in a similar to the stepping stone model in a continuously distributed population.

2.4.1.5 Genetic Equilibrium

The evolutionary forces of mutation, selection, and drift may oppose each other to create a dynamic equilibrium in which allele frequencies no longer change.

In a randomly mating population without selection or drift to change allele frequencies, and without migration or mutation to introduce new alleles, the Hardy-Weinberg genotype frequencies persist indefinitely. Such an idealized population is in a state of genetic equilibrium. In reality, the situation is much more complicated; selection and drift, migration and mutation are almost at work changing the population's genetic composition. However, these evolutionary forces may act in contrary ways to create a dynamic equilibrium in which there is no net change in allele frequencies. This type of equilibrium differs fundamentally from the equilibrium of the ideal Hardy-Weinberg population. In a dynamic equilibrium, the population simultaneously tends to change in opposite directions, but these opposing tendencies cancel each other and bring the population to a point of balance. In the ideal Hardy-Weinberg equilibrium, the population does not change because there are no evolutionary forces at work. However, opposing evolutionary forces can create a dynamic equilibrium within a population.

Box 2.15

Calculating Equilibrium Allele Frequencies with Balancing Selection

Genotypes:	AA	Aa	aa
Relative fitnesses:	$1 - s$	1	$1 - t$
Frequencies:	P^2	$2pq$	q^2
Average-relative fitness:	$\bar{w} = P^2 \times (1-s) + 2pq \times 1 + q^2 \times (1-t)$		
Frequency of A in the next generation after selection:	$P' = [P^2 (1 - s) + (1/2) 2pq] / \bar{w} = p(1-sp) / \bar{w}$		
Change in frequency of A due to selection:	$\Delta p = P' - p = pq(tq-sp) / \bar{w}$		
At equilibrium, $p\Delta = 0$	$P = t / (s + t)$ and $q = s / (s + t)$		

Balancing Selection

One type of dynamic equilibrium arises when selection favors the heterozygotes at the expense of each type of homozygote in the population. In this situation, called *balancing selection* or *heterozygote advantage*, one can assign the relative fitness of the heterozygotes to be 1 and the relative fitness of the two types of homozygotes to be less than 1:

Genotype:	AA	Aa	Aa
Relative fitness	$1-s$	1	$1-t$

In this formulation, the terms $1-s$ and $1-t$ contain selection coefficients that are assumed to lie between 0 and 1. Thus, each of the homozygotes has a lower fitness than the heterozygotes. The superiority of the heterozygotes is sometimes referred to as ‘*overdominance*’.

In cases of heterozygote advantage, selection tends to eliminate both the A and ‘a’ alleles through its effects on the homozygotes, but it also preserves these alleles through its effects on the heterozygotes. At some point these opposing tendencies balance each other, and a dynamic equilibrium is established. To determine the frequencies of the two alleles at the point of equilibrium, one must derive an equation that describes the process of selection, and then solve this equation for the allele frequencies when the opposing selective forces are in balance that is, when the allele frequencies are no longer changing (Box 2.15).

At the balance point, the frequency of A is $p = t/(s + t)$,

and the frequency of a is $q = s/(s + t)$

As an example, let’s suppose that the AA homozygotes are lethal ($s = 1$) and that the aa homozygotes are 50 percent as fit as the heterozygotes ($t = 0.5$). Under these assumptions, the population will establish a dynamic equilibrium when $p = 0.5/(0.5 + 1) = 1/3$ and $q = 1/(0.5 + 1) = 2/3$.

Both alleles will be maintained at appreciable frequencies by selection in favour of the heterozygotes – a condition known as a balanced polymorphism.

In humans, the disease sickle-cell anaemia is associated with a balanced polymorphism. Individuals with this disease are homozygous for a mutant allele of the α -globin gene, denoted Hb^s , and they suffer from a severe form of anaemia in which the haemoglobin molecules crystallize in the blood. This crystallization causes the red blood cells to assume a characteristic sickle shape. Because sickle-cell anaemia is usually fatal without medical treatment, the fitness of $Hb^s Hb^s$ homozygotes has historically been 0. However, in some parts of the world, particularly in tropical Africa, the frequency of the Hb^s allele is as high as 0.2. With such harmful effects, why does the Hb^s allele remain in the population at all?

The answer is that there is moderate selection against homozygotes that carry the wild-type allele Hb^A . These homozygotes that carry the wild-type allele Hb^A . These homozygotes are less fit than the $Hb^s Hb^A$ heterozygotes because they are more susceptible to infection by the parasites that cause malaria, a fitness-reducing disease that is widespread in regions where the frequency of the Hb^s allele is high.

We can schematize this situation by assigning relative fitness to each of the genotype of the $\hat{\alpha}$ -globin gene:

Genotype:	$Hb^S Hb^S$	$Hb^S Hb^A$	$Hb^A Hb^A$
Relative fitness:	$1 - s$	1	$1 - t$

If one assumes that the equilibrium frequency of Hb^S is $p = 0.1$ – a typical value in West Africa – and if one notes that $s = 1$ because the $Hb^S Hb^S$ homozygotes die, one can estimate the intensity of selection against the $Hb^A Hb^A$ homozygotes because of their greater susceptibility to malaria:

$$P = t / (s + t)$$

$$0.1 = t / (1+t)$$

$$t = (0.1)/(0.9) = 0.11$$

This result tells us that the $Hb^A Hb^A$ homozygotes are about 11 percent less fit than the $Hb^S Hb^A$ heterozygotes. Thus, the selective inferiority of the $Hb^S Hb^S$ and $Hb^A Hb^A$ homozygotes compared to the heterozygotes creates a balanced polymorphism in which both alleles of the $\hat{\alpha}$ -globin gene are maintained in the population.

Various other mutant Hb alleles are found at appreciable frequencies in tropical and subtropical regions of the world in which falciparum malaria is – or was – endemic. It is plausible that these alleles have also been maintained in human populations by balancing selection.

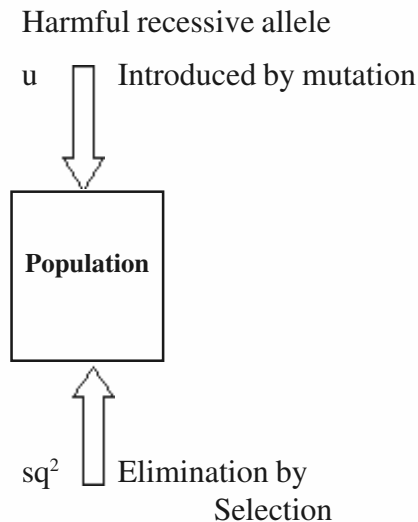
Mutation-Selection Balance

Another type of dynamic equilibrium is created when selection eliminates deleterious alleles that are produced by recurrent mutation. For example, let's consider the case of a deleterious recessive allele a that is produced by mutation of the wild-type allele A at rate u . A typical value for u is 3×10^{-6} mutations per generation. Even though this rate is very low, over time, the mutant allele will accumulate in the population, and, because it is recessive, it can be carried in heterozygous condition without having any harmful effects. At some point, however, the mutant allele will become frequent enough for aa homozygotes to appear in the population, and these will be subject to the force of selection in proportion to their frequency and the value of the selection coefficient s . Selection against these homozygotes will counteract the force of mutation, which introduces the mutant allele into the population.

If one assumes that the population mates randomly, and if one denotes the frequency of A as p and that of a as q , then one can summarize the situation as follows:

Mutation:		Selection:		
Produces a		eliminates a		
$A \rightarrow a$	Genotype:	AA	Aa	aa
rate = u	Relative fitness:	1	1	$1-s$
	Frequency:	P^2	$2pq$	q^2

Mutation introduces mutant alleles into the population at rate u , and selection eliminates them at rate sq^2



Mutation-selection balance for a deleterious recessive allele with frequency q . Genetic equilibrium is reached when the introduction of the allele into the population by mutation at rate u is balanced by the elimination of the allele by selection with intensity s against the recessive homozygotes.

When these two processes are in balance, a dynamic equilibrium will be established. We can calculate the frequency of the mutant allele at the equilibrium created by mutation – selection balance by equating the rate of mutation to the rate of elimination by selection:

$$u = sq^2$$

Thus, after solving for q , we obtain $q = \sqrt{u/s}$

For a mutant allele that is lethal in homozygous condition, $s = 1$, and the equilibrium frequency of the mutant allele is simply the square root of the mutation rate. If one uses the value for u that was given above, then for a recessive lethal allele the equilibrium frequency is $q = 0.0017$. If the mutant allele is not completely lethal in homozygous condition, then the equilibrium frequency will be higher than 0.0017 by a factor that depends on $1/\sqrt{s}$. For example, if s is 0.1, then at equilibrium the frequency of this slightly deleterious allele will be $q = 0.0055$, or 3.2 times greater than the equilibrium frequency of a recessive lethal allele.

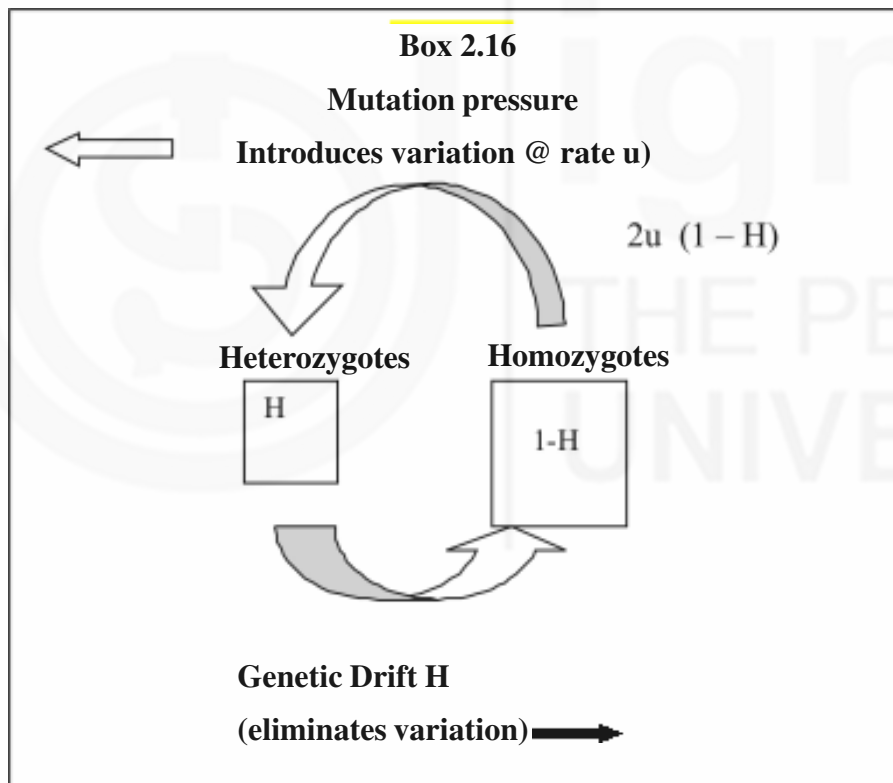
Studies with natural population of *Drosophila* have indicated that lethal alleles are less frequent than the preceding calculations predict. The discrepancy between the observed and predicted frequencies has been attributed to partial dominance of the mutant alleles—that is, these alleles are not completely recessive. Natural selection appears to act against deleterious alleles in heterozygous condition as well as in homozygous condition. Thus, the equilibrium frequencies of these alleles are lower than one would otherwise predict. Selection that acts against mutant alleles in homozygous or heterozygous condition are sometimes called purifying selection.

Mutation-Drift Balance

The random genetic drift eliminates variability from a population. Without any counteracting force, this process would eventually make all populations

completely homogeneous. However, mutation replenishes the variability that is lost by drift. At some point, the opposing forces of mutation and genetic drift come into balance and a dynamic equilibrium is established. The genetic variability can be quantified by calculating the frequency of heterozygotes in a population- a statistic called the heterozygosity, which is symbolized by the letter H . The frequency of homozygotes in a population- often called the *homozygosity*- is equal to $1-H$. Over time, genetic drift decreases H and increases $1-H$, and mutation does just the opposite as shown in the figure below (Box 2.16).

Let's assume that each new mutation is selectively neutral. In a randomly mating population of size N , the rate at which drift decreases H is $\frac{1}{2N}H$. The rate at which mutation increases H is proportional to the frequency of the homozygotes in the population ($1-H$) and the probability that one of the two alleles in a particular homozygote mutates to a different allele, thereby converting that homozygote into a heterozygote. This probability is simply the mutation rate μ for each of the two alleles in the homozygote; thus, the total probability of mutation converting a particular homozygote into a heterozygote is 2μ . The rate at which mutation increases H in a population is therefore equal to $2\mu(1-H)$.



When the opposing forces of mutation and drift come into balance, the population will achieve an equilibrium level of variability denoted by H . This equilibrium value of H can be estimated, by equating the rate at which mutation increases H to the rate at which drift decreases it:

$$2\mu(1-H) = \frac{1}{2N}H$$

By solving for H , the equilibrium heterozygosity at the point of mutation-drift balance is obtained as :

$$\hat{H} = \frac{4N\mu}{4N\mu + 1}$$

Thus, the equilibrium level of variability (as measured by the heterozygosity) is a function of the population size and the mutation rate.

If one assumes that the mutation rate is $\mu = 1 \times 10^{-6}$, one can plot $\hat{H}$ for different values of N . For $N < 10,000$, the equilibrium frequency of heterozygotes in the population will be quite low; thus, drift dominates over mutation in small populations. For N equal to $1/\mu$, the reciprocal of the mutation rate, the equilibrium frequency of heterozygotes would be 0.8, and for even greater values of N , the frequency of heterozygotes increases asymptotically towards 1. Thus, in large populations, mutations dominate over drift; every mutational event creates a new allele, and each new allele contributes to the heterozygosity because the large size of the population protects the allele from being lost by random genetic drift.

Values of $\hat{H}$ in natural populations vary among species. In the African cheetah, for example, $\hat{H}$ is 1 percent or less among a sample of loci, suggesting that over evolutionary time, population size in this species has been small. In humans, $\hat{H}$ is estimated to be about 12 percent, suggesting that evolutionary time population size has averaged about 30,000 to 40,000 individuals. Estimates of population size that are derived from heterozygosity data are typically much smaller than estimates obtained from census data. The reason for this discrepancy is that the estimates based on heterozygosity data are *genetically effective* population sizes—sizes that take into account restrictions on mating and reproduction, as well as temporal fluctuations in the number of mating individuals. The genetically effective size of a population is almost less than the census size of a population.

(**Source:** *Principles of Genetics* (2006) by D. Peter Snustad and Michael J. Simmons. John Wiley & Sons (Asia Edition) PP. 750-754.

2.5 SUMMARY

- 1) Understanding of Population genetics principles, requires the basic concepts of Mendelian genetics: the result of segregation, the concept 'gene', 'phenotype', 'genotype', 'dominant', 'recessive' traits, 'allele' etc. Parental mating types and expected distribution of genotypes among the offspring.
- 2) Hardy-Weinberg equilibrium is the solution to an intriguing question: what happens to gene frequency of a dominant character over generations in a population. With three times more frequent than normal does this will increase over generations?
- 3) HWE law states that under the absence of intervening factors, especially in a large population, given random mating, no selection of any sort, no mutation and absence of demographic factors like migration, differential fertility and mortality etc., the allele frequency remain constant over generations. This can be proved theoretically, easily, for a 'biallelic locus and it can be extended to multilocus as well.
- 4) The importance of HWE: it gives a methodology to estimate the allele frequency in a population based on phenotypic/genotypic information of the parental mating types. It helps us to investigate the relationship between change in gene frequency with respect to mutation, migration, selection,

genetic drift etc. The entire investigation is the kernel of a branch of biomathematics or the new field: 'population genetics' and 'quantitative genetics'.

- 7) HWE is the bench mark of qualitative test to check whether a trait, an allele, SNP, is in equilibrium. It tells how to distinguish between the effects of evolutionary forces from the demographic factors.
- 8) Mutation is a non-systematic and random, but rate of mutation is site specific. Mutations are more frequent at hot-spots and are rare at the 'conserved region'. The mitochondrial non-coding genome has a higher frequency of mutations than the nuclear genome.
- 9) Genetic drift is a non-systematic force which can lead to significant changes in gene frequency in a small population. If an allele is rare in a small population, it can get lost or get fixed in the population over generations.
- 10) Founder effect is one form of genetic drift. The founders are a sample (represent a fraction of the genetic diversity) of original populations. The descendents of a few founders have the gene frequency that is dependent on the genetic composition and genetic structure of the founders. It can also happen as bottleneck effect, especially as a result of sudden population size reduction in a population, due to reasons such as natural causes or man-made causes or socio-cultural regulations. There could be serial founder effect as a result of waves of migration at different times. The mitochondrial investigation of human origins suggests that the human origins and migration to other continents appears as a result of serial founder effect from Africa.
- 11) Natural selection is one of the complex systematic forces that can influence significant changes in gene frequency. Selection can operate in multitude ways and it is a slow process than to the effect of migration or admixture etc.
- 12) Selection basically operates at differential fertility and mortality levels. It is measured as 'fitness' the ability to leave offspring and refers to 'relative rate of survival'. It is measured by 'selection coefficient' ('s') which is a function of fitness (W). The fitness or selection coefficient differs with respect to the type of dominance: complete, partial, over etc.
- 13) The effect of 'directional selection' to shift the mean allele frequency towards its extremes. Or it could be stabilizing selection that shifts the allele frequency of extreme alleles as a result the heterozygote frequency will increase. Or it could be disruptive selection where the extreme allele frequency increases as against the heterozygote frequency.
- 14) Selection can also be measured based on demographic factors of fertility and mortality trends. Crow's Index of opportunity for selection measures total selection intensity that a population can experience which depend on two components, fertility and mortality.
- 15) Gene flow (migration/admixture) is a systematic factor which can bring rapid changes in gene frequency within a short period. In general, human populations follow a variety of restrictions or regulations that restrict gene

flow between and within populations. The barriers for gene flow could be because of culture or due to geographical, political, religious and linguistic etc.

- 16) There are theoretical models to investigate the effect of spatial gene flow or population structure between populations. Island model, stepping stone model, neighbourhood model help us to investigate the spatial gene flow in different situations of population structure.

Suggested Reading

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

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

Cavalli Sforza L.L. and Bodmer W.F. 1971. *The Genetics of Human Populations*. Sanfrancisco. USA: W.H. Freeman,

Falconer, D.S, 1980. *Introduction to Quantitative Genetics*. Second edition. London and New York: Longman.

Christiansen, B, Freddy, and Feldman W Marcus. 1986. *Population Genetics*. Victoria: Blackwell Scientific Publications (Australia) Pvt. Ltd.

Gautam, R.K (2009). Opportunity for natural selection among the Indian population: secular trend, covariates and implications. J. Biosoc. Sci. 41:705-745.

Kimura, M. and G.H, Weiss 1964. The steppingstone model of population structure and the correlation with distance. Genetics. 49:561-576.

Majumder, PP. 1993. Human Population Genetics. A centennial tribute to JBS Haldane. Ed. PP Majumder. New York: Plenum Press.

Malhotra, K.C. 1988. Statistical Methods in Human Population Genetics. Ed. KC Malhotra. Calcutta: Eka Press.

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

Sample Questions

- 1) A total of 120 individuals were tested for M, N blood group and the observed genotype frequencies of MM, MN and NN are 34, 62 and 24 respectively. Calculate the gene (allele) frequencies?
- 2) If 'i' is the mutation rate ($i = 10^{-5}$) per generation for a gene frequency of A then how many generations are required to reduce the gene frequency by a factor of $\frac{1}{2}$.
- 3) What is Hardy-Weinberg equilibrium? Explain why HWE is important in genetic of populations?
- 4) In case in a population the observed gene frequencies of a particular bi-allelic locus are in HW equilibrium for the locus, does this imply the population satisfies the assumptions of the HW equilibrium? Explain?
- 5) What is genetic drift and how it operates in populations? Explain with Examples.

UNIT 3 GENETIC POLYMORPHISM

Contents

- 3.1 Introduction
- 3.2 Balanced Polymorphism
- 3.3 Transient and Balanced Polymorphisms
- 3.4 Serological Markers
- 3.5 Biochemical Polymorphisms
- 3.6 Molecular Markers
 - 3.6.1 Repetitive DNA Sequence Variants
 - 3.6.2 Non- Repetitive DNA Sequence Variants
 - 3.6.3 Lineage Markers
- 3.7 Tools for Studying Polymorphisms
- 3.8 Genetic Markers and Disease
- 3.9 Genetic Mapping of Disease Gene on Human Chromosome Using Polymorphic Markers
- 3.10 Use of Polymorphic Markers in Forensic Testing
- 3.11 Use of Polymorphic Markers in Population Studies
- 3.12 Summary
 - Suggested Reading
 - Sample Questions

Learning Objectives



After reading this unit, you will be able to:

- define the concept of genetic polymorphism;
- explain genetic polymorphism with respect to serological, biochemical and molecular markers;
- explain the genetic markers in disease association; and
- discuss the use of polymorphic markers in population and forensic studies.

3.1 INTRODUCTION

Genetic polymorphism can be defined as the occurrence together in the same population two or more than two alleles such that the frequency of rare allele is always $>1\%$, and is maintained in the population not merely by the recurrent mutation. Polymorphism can be in a coding region (coding region means the portion of DNA which code for a gene, it may be synonymous or non-synonymous) or more commonly, in the noncoding regions (which does not code for functional region), often vary by ethnicity. Basic information about the types, frequencies and distribution of common polymorphisms are essential not only for the understanding of pathological entities, but also to know our evolutionary past and provide guidance about our biological future. The most common polymorphism in our genome are single base pair sequence variation i.e. SNP

but other types like copy number changes, insertions, deletions, duplications and rearrangements also occur. The methods to assess this diversity is variable. Few examples of polymorphic markers are listed in table 3.1.

Table 3.1: Example of Genetic polymorphisms

Type of marker	Year	No. of loci	Properties
Blood groups	1910-1960	~20	May need fresh blood, rare antisera. Genotype cannot always be inferred from phenotype because of dominance. No easy physical localization.
Electrophoretic mobility variants of serum proteins	1960-1975	~30	May need fresh serum, specialized assays, no easy physical localization often limited polymorphisms
Human Leucocyte Antigens (HLA)	1970	1 (multi locus haplotype)	One linked set highly informative. Can only test for linkage to 6p21.3
DNA RFLPs	1975	>105 (potentially)	Two allele markers, maximum heterozygosity 0.5, initially required Southern blotting, now PCR. Easy physical localization
DNA VNTRs (minisatellites)	1985-	>104 (potentially)	Many alleles, highly informative can be typed by southern blotting easy physical localization. Tend to cluster near ends of chromosomes.
DNA VNTRs (microsatellites) (di-, tri-, and tetranucleotide repeats)	1989-	105 (potentially)	Many alleles, highly informative Can be typed by automated multiplex PCR, easy physical localization. Distributed throughout genome
DNA SNPs	1998-	106 (potentially)	Less informative than microsatellites. Can be typed on a very large scale by automated equipment, without gel electrophoresis, etc.

3.2 BALANCED POLYMORPHISM

When natural selection favors heterozygotes over both homozygotes, the result is balanced polymorphism. It accounts for the persistence of an allele even though it is deleterious when homozygous. Some of the examples are given below:

Sickle Cell Disease

It is an autosomal recessive disorder that causes anemia, joint pain, a swollen spleen, and frequent, severe infections. It illustrates balanced polymorphism because carriers are resistant to malaria, an infection by the parasite *Plasmodium falciparum* that causes cycles of chills and fever. The parasite spends the first stage of its life cycle in the salivary glands of the mosquito *Anopheles gambiae*. When an infected mosquito bites a human, the malaria parasite enters the red

blood cells, which transport it to the liver. The red blood cells burst, releasing the parasite throughout the body.

It is known since long that malaria is a quite common in the tropical regions of Africa. Sickle shape red blood cells provide selective advantage as malarial parasite cannot grow in these cells. Therefore, along with malaria the sickle cell anemia also increased in these parts of Africa. The sickle cell disease is less common in Caucasians due to the less frequency of malaria. This shows the heterozygous advantage of sickle cell as it provides protective effect.

The rise of sickle cell disease goes hand in hand with the cultural development with the advent of cultivation of crops gave a breeding ground to *Anopheles* mosquitoes as the malaria rose the selective pressures gave rise to the change in the shape of the RBCs from elliptical to sickle shaped and when it occurred in homozygous condition the disease was caused otherwise it had selective advantage. The spread of sickle cell disease is associated to the migratory events. Africa by people migrating from Southern Arabia and India, or it may have arisen by mutation directly in East Africa.

Settlements with large numbers of sickle cell carriers escaped devastating malaria. They were therefore strong enough to clear even more land to grow food-and support the disease-bearing mosquitoes. Even today, sickle cell disease is more prevalent in agricultural societies than among people who hunt and gather their food.

G6PD Deficiency

It is a sex-linked enzyme deficiency. It affects 400 million people throughout the world. It results into hemolytic anemia which is life-threatening. It is under the influence of certain environmental conditions like eating fava beans, inhaling certain types of pollen, taking certain drugs, or catching certain infections. It has been seen in Africa that hemizygous males and heterozygous males for this enzyme deficiency are at less risk for malaria again revealing a selective advantage for heterozygotes. Therefore, natural selection acts in two directions hence it could be one of the example of balanced polymorphism.

3.3 TRANSIENT AND BALANCED POLYMORPHISMS

Polymorphism occurs when two or more clearly different phenotypes exist in the same population of a species-in other words, the occurrence of *more than one form or morph*.

A transient polymorphism is one that is changing in frequency over time. In transient polymorphism, one form is gradually being replaced by another. As the name implies, it represents a temporary situation as a by-product of directional natural selection.

The phenomenon of industrial melanism occurs in a number of moth species in Europe and the United States. The British ecological geneticist, E. B. Ford, first called attention to this phenomenon as a way of demonstrating the effect of natural selection in nature (as opposed to artificial selection experiments which have long enjoyed success in the lab). Ford noted that a light colored moth species,

Biston betularia, occasionally undergoes mutation at a single locus to produce a dark or melanic individual. Since the mutant allele is dominant, any gamete containing this mutant will produce a melanic individual upon syngamy. The first melanic specimen in this species was found in a collection from Manchester, England dated 1848, but by 1895 about 95% of all collected specimens were dark morphs, referred to as the form *carbonaria*. In a series of 12 observations and mark-recapture experiments during the 1950s, H. B. D. Kettlewell demonstrated that the two forms (light and dark) were differentially preyed upon by birds. He found that the birds selectively caught and ate more individuals of the form that did not match its background as compared to the one that was masked. In industrialized areas of England where the substrate (walls and tree trunks) upon which the moths rested were darkened by pollutants in the smoke poured out by factories, the *carbonaria* form possessed a selective advantage. Rural areas, unaffected by pollutants, afforded the light form an adaptive advantage. The environmental change brought on by the industrial revolution did not produce the *carbonaria* form (which presumably appeared from time to time due to recurrent mutation); it only protected the dark moths from bird predation (the agent of natural selection). The fact that the light form still exists in rare numbers in industrialized areas testifies to the amount of time selection requires to eliminate a recessive allele.

Mendelian Population

A population is a group of individuals who share a common gene pool where the characters are transmitted in a Mendelian fashion from one generation to the next generation. A group of individuals within which marriages are performed is called a Mendelian population. In a given Mendelian population, which is under Hardy-Weinberg equilibrium, the resultant genotype and phenotype frequencies are more or less permanently established.

3.4 SEROLOGICAL MARKERS

Blood groups are the best cited examples of serological markers. Both ABO and Rh are quite important serological markers as they can be used to study population diversity. These blood groups cause newborn hemolytic diseases. They also have a role in blood transfusion and also solid organ transplantation. They follow mendelian inheritance. ABO blood groups were discovered by Landsteiner in 1900 and are cited as a best example of triallelic inheritance. Blood groups can be tested by using antisera and red blood cells using simple agglutination techniques. Presently more advanced molecular techniques are also used.

The ranges of phenotypes in humans are a direct result of genetic variations which act together with environmental and behavioral factors to produce diversity. The identification of gene polymorphisms, which control the blood group antigen expression, contributes to the understanding of the biological significance of blood group systems. In addition to assisting in the characterization of allelic variations, the identification of gene polymorphisms allows us to estimate the processes involved in the formation of different populations (the founder effect, genetic drift, migration, etc.). Thus, blood group gene polymorphisms are valuable predictors of genomic ethnic ancestry.

3.5 BIOCHEMICAL POLYMORPHISMS

There is marked difference between individuals on the basis of biochemical markers like G6PD, human enzymes and proteins etc. This has been explained in the above section. However, here we would like to throw some light on the molecular basis of G6PD variants.

Molecular Basis of G6PD Variants

The G6PD gene, located on chromosome Xq28 region, is 18 Kb long consisting of 13 exons transcribed to a 2.269 Kb messenger RNA with 1.545 Kb of coding regions. The commonest variant in South China, G6PD Canton, has been sequenced and was found to be due to a mutation at nucleotide (nt) position 1376 of cDNA, G to T, resulting in a missense mutation in amino acid position 459, Arg to Leu. With improved DNA technology, the whole cDNA sequence can be amplified and screened for mutation directly. PCR technique and restriction analysis has been used.

World Incidence and Distribution of G6PD Deficiency

G6PD deficiency in male subjects can be detected easily by a number of screening tests. The simplest one is the fluorescent spot test developed by Beutler and Mitchell which relied on the fluorescence of NADPH, generated by an adequate amount of G6PD enzyme. This test can also be done on blood sample dried on filter paper similar to the Guthrie cards. In Hong Kong, the routine screening of newborns have included test for G6PD deficiency.

3.6 MOLECULAR MARKERS

Although ~99% DNA is known to be similar between individuals but still sequence differences exist between individuals in non-coding regions of the genome and such polymorphic regions are useful for various kinds of analyses in population genetic studies. A genetic marker can be a nucleotide sequence of variable length, varying from a single base pair to several hundred base pairs. Selection of markers for any study is dictated by the nature and purpose of the study. The more commonly used markers in population genetics studies can broadly be grouped as follows:

3.6.1 Repetitive DNA Sequence Variants

Tandem Repeats

Besides the interspersed repeats (SINEs and LINES), Tandem repeats are the other kind of repeated elements found in the genome. These are highly variable tandemly repeated arrays of 2 or more base pair core units in the non-coding regions of the genome and are located adjacent to each other. On the basis of size of the core unit, they are categorised into minisatellites (10-60 bp¹), Short Tandem Repeats (STRs) or microsatellites (<10 bp). When the number of nucleotides in the core unit is not known or is variable then it is called Variable Number Tandem Repeats (VNTRs).

Insertion/ Deletion Polymorphisms

An InDel or Insertion-Deletion polymorphism refers to insertion or deletion of a DNA sequence of variable length in the genome. The concerned DNA sequence

may vary in length from a single nucleotide to several hundred nucleotides. They are widely spread across the genome and constitute around 1.5 million of more than 10 million polymorphisms known in humans.

Alu InDels – *Alu* Insertion/ Deletion polymorphisms (*Alu* InDels) involve *Alu* sequences that are characterized by the cleavage action of *AluI* restriction endonuclease.

Properties of *Alu* sequences such as their known ancestral state, identity by descent, wide occurrence and stability make them ideal markers for human evolutionary and diversity studies.

3.6.2 Non- Repetitive DNA Sequence Variants

Single Nucleotide Polymorphisms (SNPs or *Snips*)

SNP or Single Nucleotide Polymorphism is a single nucleotide (base pair) change in a DNA sequence. As with all polymorphisms, for an alteration to be considered a snip it must be present in $\geq 1\%$ of the population being considered. They make up about 90% of all the human gene sequence variation. SNPs may be present in coding regions (exons) or non-coding regions (introns) or intergenic regions.

Restriction Fragment Length Polymorphisms (RFLPs) are the characteristic pattern of fragments of DNA produced when a DNA sequence is cleaved by specific enzymes belonging to endonuclease class of enzymes. The property of these enzymes that enables them to cleave DNA segment only at specific locations known as restriction sites have led to their use in detecting genetic differences on the basis of absence or presence of restriction sites.

3.6.3 Lineage Markers

Mitochondrial Markers

Maternally inherited mitochondrial genome consists of multiple copies of circular mitochondrial DNA or mtDNA. Markers present on this haploid genome are primarily used for tracing maternal ancestral lineage(s) in populations because of their uniparental inheritance.

Y-chromosomal Markers

Like mtDNA, Y- chromosome has a uniparental inheritance but in the male line and can thus be used for tracing paternal ancestral lineages. In absence of recombination, Y-chromosome is more or less transmitted unchanged from one generation to next and the few changes that may occur usually do not have any effect as around 98% of the DNA is in non-coding region.

Hardy-Weinberg equilibrium

The Hardy-Weinberg equilibrium predicts that in the absence of evolutionary forces, both allelic and genotypic frequencies remain constant in a population and that if the equilibrium is disturbed a new equilibrium will be reached within one generation based on the allelic frequencies of the remaining population. The conditions that must be met for the predictions of the Hardy-Weinberg equilibrium to be valid are described below:

- 1) Random mating: Mating patterns must randomly reflect the entire breeding population, with no dependence on genotype or closeness of relationship (either positive or negative).

- 2) No sex bias in allelic frequencies: The distribution of alleles must be the same in both sexes.
- 3) All genotypes equally viable and fertile: There must not be any selective advantages or disadvantages. This is seldom true in a real population, and often must be taken into account in terms of evolutionary pressures.
- 4) Mutation rate too low to alter ratios: The basic assumption is that alleles are stable through many generations and are not altered or degraded significantly by mutation. In practice this is generally not a serious problem.
- 5) Closed population (no in or out migration): The “population” that is being considered must be a constant one. Introduction of new genes into the breeding pool or loss of genes from the breeding pool by migration between “populations” can distort trends.
- 6) Population must be large: The population must be large enough so that there are no confounding effects due to genetic drift (random events altering allelic frequencies by pure chance) or due to “founder” effects, where a recessive gene becomes fixed in a population because too many of its members are descendants of a single individual.

The Hardy-Weinberg law can also be applied to multiple alleles and X-linked alleles. The genotypic frequencies expected under Hardy-Weinberg equilibrium will differ according to the situation.

3.7 TOOLS FOR STUDYING POLYMORPHISMS

Both conventional and advanced techniques are used to study polymorphisms. Conventional techniques are blood groups by carrying out simple agglutination techniques or protein electrophoresis for studying the protein polymorphisms etc. Under advanced techniques are the tools for studying molecular markers, and the foremost requirement for carrying out molecular analysis of any kind is the availability of the genetic material. As mentioned earlier, DNA is the focal point of human diversity and disease-association studies by virtue of the fact that it is the blueprint of our existence. There are several techniques for isolating DNA such as manual methods (like Phenol Chloroform, Salting-out) and kits. The technique of DNA isolation or extraction varies depending on the starting material, but, it is the technique of PCR which is the most useful for DNA analysis.

Polymerase Chain Reaction (PCR): It involves cycling of DNA sample through a series of heating and cooling cycles with the required raw materials and enzymes to achieve its exponential amplification. The technique has come a long way since its invention. Instead of having to manually maintain the heating and cooling cycles, automated thermal cyclers are now available; and instead of having to add fresh polymerase (earlier derived from *E. coli*) after every cycle because of its denaturation due to heating, thermally stable DNA polymerases such as *Taq* DNA Polymerase are now made use of.

Amplification of DNA by PCR has found applications in a variety of fields ranging from forensics to archaeology; study of variation and evolution to mutation detection; gene mapping and cloning and DNA sequencing to epidemiology among several others.

Restriction Digestion: It is the method of cutting DNA sequences into fragments using restriction endonucleases or enzymes that cut at specific recognition sites. This generates DNA fragments of varying lengths producing a variation pattern known as Restriction Fragment Length Polymorphisms (RFLPs). The variation may be produced in response to absence or presence of particular SNP(s) or an insertion or deletion event in that region and is recognised in the form of banding pattern. Resulting fragments are separated according to molecular size using gel electrophoresis. There are several classes of endonucleases- Type I, Type II, Type III and Type IV but the most commonly used restriction enzymes are of type II and they cleave DNA fragment at specific sites within or close to the recognition sequence. Most of these enzymes cut palindromic sequences.

The technique is useful in detection of mutations/ SNPs. It is also used to detect VNTRs. The technique has been widely used for constructing physical maps of the genome, genetic linkage maps; in forensic testing; and in epidemiological and evolutionary studies.

Electrophoresis: It is one of the few techniques that has been in use since the beginning of study of classical genetic markers and is still in use for molecular markers. It is the method of separating macromolecules (both proteins and nucleic acids) on the basis of size, electric charge or other physical properties under the influence of electric field.

Sequencing: DNA sequencing refers to establishing the exact sequential arrangement of bases in a stretch of DNA. Knowledge of exact sequence of bases in a gene is crucial especially in ascertaining the function of genes. This is also important as the disease-causing alterations in the genes can then be identified.

The selection of technique and markers depends upon the purpose of study. In the following section we have discussed the uses of polymorphic markers.

Uses of polymorphisms: All the markers listed in table 3.1 can be used for population diversity studies. Now a days most extensively studied markers are Single nucleotide polymorphisms. Genomics and specially SNP research can be used to improve health care through gene therapy, to yield new targets for drug discovery, to renew the process of drug development and to discover new diagnostics.

3.8 GENETIC MARKERS AND DISEASE

Understanding the genetic basis of complex human diseases (like hypertension, cardiovascular disease, diabetes etc.) has been increasingly emphasized as a means of achieving insight into disease pathogenesis, with the ultimate goal of improving preventive strategies, diagnostic tools, and therapies. Genetic approaches to complex disorders thus offer great potential to improve our understanding of their pathophysiology, but they also offer significant challenges. These can be studied either using linkage analysis. In linkage analysis we use families and try to find out which polymorphic marker is near to the disease gene and then try to map the gene on the human genome. The other approach is where we study populations of both types of individuals. One would be those suffering with a disease and the other would be who are not suffering with the disease. We take different polymorphic markers and study in these two sets of samples. Then we

compare both the groups and if both the groups differ significantly at these markers we propose that these markers may be associated with the disease.

Association studies can be a very powerful approach for finding genetic determinants of a complex disorder. It has been suggested that if hundreds of thousands of single nucleotide polymorphisms (SNPs) were identified across the genome, then it would be possible to perform genome-wide association studies to identify the regions of linkage disequilibrium around disease susceptibility genes. In addition, they noted that much smaller sample sizes would be required to detect association than to detect linkage. The SNP Consortium is rapidly identifying single nucleotide polymorphisms, and within next several years, genome-wide association studies may become a reality.

These association studies can result into positive association or negative association. Some time they result into false positive or false negative results. The following general guidelines, summarized in Table 3.2, may be useful for genetic association studies. First, are the candidate gene(s) under study should be biologically reasonable. Several factors can determine the appropriateness of a candidate gene. If human genetic linkage studies have identified a chromosomal region linked to a disease, or if an animal model for a disease is influenced by a particular gene or syntenic chromosomal region, positional candidate genes in such genomic regions warrant strong consideration. In addition, the biologic plausibility of a candidate gene for involvement in disease pathogenesis is important. However, obvious limitations of this candidate approach are the large number of potential candidate genes for complex diseases and the reality that only known genes can be investigated. Although candidate genes can be selected for study on this basis, they should not be ruled out on the basis of our current understanding of disease pathophysiology- important new insights may be missed if potential candidate genes must fit into current pathophysiologic models.

Table 3.2: Evaluation of candidate gene case-control association studies

Issue	Key Questions	Possible Solutions
Selection of candidate	Is candidate gene biologically reasonable?	Demonstration of biologically functional effect
Gene polymorphism	Is the candidate gene a positional candidate?	Within linked region in man or systemic from animal model
Population stratification	Are cases and controls matched?	Matching on ethnicity Family-based association designs Negative results with multiple unlinked markers
Hardy-Weinberg (H-W) equilibrium	Is control group in H-W equilibrium?	Calculation of H-W equilibrium with goodness-of-fit test (2 alleles) or simulation (multiple alleles)
Multiple comparisons	How many alleles were tested?	Bonferroni correction
	How many genetic loci were tested?	Estimation of empirical P values

A second criterion in evaluation of case-control association studies is the careful selection of cases and control subjects. Do the case subjects meet appropriate criteria for disease affection? Are control subjects free from symptoms of disease, associated intermediate phenotypes, and potential confounders? Have control subjects been exposed to relevant environmental influences involved in disease pathogenesis while remaining clearly unaffected? Were the cases and controls matched on demographic and environmental factors? Was consideration of population stratification included, either by attempting to match ethnicity or by typing unlinked markers.

A third criterion in the evaluation of case-control studies is assessment of Hardy-Weinberg equilibrium in the markers studied within the control group. Hardy-Weinberg equilibrium indicates that the genotype frequencies can be determined directly from the allele frequencies; failure to demonstrate Hardy-Weinberg equilibrium could result from genotyping errors, inbreeding, genetic drift, mutation, or population substructure. Hardy-Weinberg equilibrium can be readily assessed with a goodness-of-fit chi square test for biallelic markers; for markers with multiple alleles (such as short-tandem repeat markers), more accurate determination of Hardy-Weinberg equilibrium can be obtained with Markov Chain Monte Carlo methods. Significant deviations from the expected proportions of homozygote and heterozygote classes in a population of case subjects may be caused by association with the disease allele. Lack of consistency with Hardy-Weinberg equilibrium among control subjects should prompt investigation for potential complications, including genotyping errors and population stratification. A final criterion for evaluation of a case-control study is correction for multiple comparisons. This remains a problematic topic requiring additional statistical genetic research. However, an effort to correct for spurious associations, which can result from testing a large number of alleles, is warranted. The multiple comparison issue is especially problematic with markers that have multiple alleles like short-tandem repeat polymorphisms; the conservative Bonferroni approach to use a corrected significance value calculated by multiplication of the observed *P* value by the number of alleles tested. Bonferroni corrections for the total number of alleles at all loci are probably too conservative because the alleles at one locus are not independent of each other and closely linked loci are probably not independent either. A less conservative but more computationally intensive approach is to estimate empirical significance values using simulation approaches.

Genome Wide Studies

Unlike the direct approach of case-control association with candidate genes, genome scanning (screening) is an indirect strategy that does not rely on conjecture. Basically, either affected individuals, usually siblings, from a number of families or families with two or more affected individuals are genotyped with polymorphic DNA markers that cover the entire chromosome complement. A set of about 400 short tandem repeat polymorphic markers that are spaced at about every 10cM is used for most genome scans. This level of resolution has been enhanced with the assembly of about 3000 simple sequence repeat polymorphic markers that are about 1.5cM apart. Single-nucleotide polymorphic sites (SNPs) are preferred for genome scans because they are uniformly distributed about every 300 bases throughout the genome and easily identified with automated equipment. Eventually, sets of SNPs will supersede short tandem repeat polymorphic sequence marker systems.

Furthermore, major landmark attempts that have also been made to study various aspects of human genome, and few are listed below.

Human Genome Project (HGP): A National Institute of Health (NIH, US) initiative started in 1990, HGP was a multinational collaborative project aimed at identifying all the genes in the human DNA and determining the sequence of about 3 billion nucleotide pairs that constitute the human DNA to understand the species' genetic makeup.

First draft was released in 2001 followed by the complete draft in 2003. Some of the main findings from the draft sequence are as follows:

- Total number of genes was estimated at 30, 000.
- The average gene was found to consist of 3000 bp but sizes vary greatly.
- Repeated sequences that do not code for proteins (“junk DNA”) make up at least 50% of the human genome.
- About 1.4 million locations with SNPs were identified.

Findings from HGP are already having profound impact on diverse areas of research including molecular medicine (improved diagnosis of disease, earlier detection of genetic predispositions to disease, rational drug design etc.), bioarchaeology, anthropology, evolution and human migration, DNA forensics (identification), agriculture, livestock breeding etc.

Human Genome Diversity Project (HGDP): HGDP was formally organised in 1993 under Stanford University's Morrison Institute, and was aimed at understanding the diversity patterns worldwide, the contributing factors and the implications of the observed diversity patterns. Findings from the project could also shed light on the origins and migration patterns of the entire human species. HGDP could also aid in understanding the role played by environmental factors in complex human diseases.

HapMap Project: The International HapMap Consortium is an international collaborative venture between Japan, the United Kingdom, Canada, China, Nigeria, and the United States aimed at developing haplotype map of the human genome in a bid to identify genetic determinants of complex diseases. The information made available through the HapMap project is helping researchers find genes that affect health, disease, and individual responses to medications and environmental factors.

Indian Genome Variation (IGV): IGV was the first large scale effort to document and understand the genomic structure of enormously varied Indian populations. The study found high degree of genetic differentiation among the different ethnic groups.

Genetic Testing and Counseling

Frequently the question may arise as to whether the patient has a certain disease for which there is a genetic basis. Often among the 10,000 conditions for which a genetic basis has been identified, the diagnosis can be made from evaluation of personal and family history, physical examination, and conventional laboratory tests. A useful database for identifying these conditions is available on Online Mendelian Inheritance in Man (OMIM) (www.ncbi.nlm.nih.gov/omim). This

catalog is updated regularly and can be searched using multiple terms. The entries provide information about the clinical signs as well as the genetic basis for the condition, if known, including mutations that have been found to cause the condition. To determine whether genetic testing is available for a given condition and to find a laboratory, a useful link is GeneTests, a free online service (www.genetests.org). The entries in this catalog indicate the test menus and contact information for the laboratories, as well as whether the testing is provided on a routine or research basis. A very useful adjunct in the GeneTests Website is GeneReviews, which provides succinct summaries about many genetic conditions and the ways the genetic testing can be used for diagnosing these conditions, including prediction of natural history.

The clinician is likely to encounter many situations in which a genetic test may be useful. Sometimes genetic testing is required from diagnosis when it cannot be made by clinical criteria alone. The fragile X syndrome is the most common genetic form of mental retardation. Although the diagnosis may be suggested by the presence of the characteristic signs—large ears, protruding chin, and large testes—the only way to diagnose fragile X is by genetic testing. For the various forms of spinocerebellar ataxia, there is considerable overlap. Yet, these can be readily distinguished by their specific mutations. Patients with atypical forms of certain diseases may have a negative gold standard test, but positive genetic test. For most patients with cystic fibrosis, the diagnosis can usually be made by a sweat chloride test. However, a number of individuals have been described with pulmonary disease suggestive of this condition for whom the sweat chloride test is normal. For these patients, the diagnosis has been based on observation of mutations in both copies of their CFTR genes.

For some conditions, the signs of disease may not yet have developed, yet on the basis of one's family history, one may want to know about the risk of developing disease. This is true for the person whose parent(s) may have died from Huntington's disease, a progressive neurodegenerative disease or for the person whose mother and sister may have died from breast or ovarian cancer, suggesting a heritable risk. For these individuals, a positive genetic test result will indicate an increased, although not necessarily absolute, risk for developing the disease.

Genetic testing is used for assessing reproductive risks—by testing the parents for carrier status and by testing the fetus. Individuals with a positive family history of genetic disease (usually autosomal recessive or X-linked) or who come from ethnic groups with an increased prevalence of autosomal recessive or X-linked diseases are candidates for carrier screening. Currently, carrier screening for cystic fibrosis, fragile X syndrome, and spinal muscular atrophy is recommended in the United States. For people of Mediterranean, African, or South Asian ancestry, hemoglobinopathy screening is recommended. For individuals of Ashkenazi Jewish ancestry, screening for Tay-Sachs disease, Canavan disease, cystic fibrosis, Gaucher disease, Bloom syndrome, Fanconi anemia, Niemann-Pick disease, familial dysautonomia, maple syrup urine disease, glycogen storage disease, and familial hyperinsulinism is available. An individual who is a carrier for a certain condition may choose not to marry another individual who is a carrier for the same condition. Alternatively, if a carrier couple is identified, they may choose to have prenatal diagnosis to determine whether their fetus is affected with this condition. This can be performed either at 10-11 weeks using the procedure of chorionic villus sampling where a bit of placenta is obtained under ultrasound

guidance. As another option, an amniocentesis can be performed at 15-18 weeks of pregnancy to obtain cells from the amniotic fluid. These couples might also choose to have pre-implantation genetic diagnosis with selection implantation of only those embryos that are deemed unaffected.

Not all genetic testing involves looking for heritable mutations. Sometimes it is used to look for genetic alterations that are confined to a specific population of cells. These alterations may cause certain cells to become cancerous, or if cancerous, to progress to a more aggressive stage. Genetic testing can be used to identify chromosomal translocations between two non-homologous chromosomal segments and in the process diagnose a specific form of leukemia. For example, the translocation between chromosomes 1 and 19 in leukemic cells is diagnostic of the acute promyelocytic form of this disease and the translocation between chromosomes 9 and 22 is diagnostic of the chronic myelogenous form. The expression patterns of RNA transcribed from many genes can be assessed to predict the natural history of the disease. This approach has been used to predict breast cancer outcome and whether more or less aggressive therapies should be used to treat patients.

Individuals might also have genetic tests of identity. These might be voluntary and selected to test specific questions, such as whether they are members of a known patrilineal lineage, such as people with a specific surname. These tests analyze a series of polymorphic genetic markers on the Y chromosome. On the basis of the general pattern of markers, or “haplogroup,” they may be told of the geographic region where their Y chromosome originated. According to the number of markers that match with people who are suspected to be of the same lineage, individuals may be advised about the common ancestors or other people in that lineage. Such testing is also possible for matrilineal lineages by testing mitochondrial DNA markers.

3.9 GENETIC MAPPING OF DISEASE GENE ON HUMAN CHROMOSOME USING POLYMORPHIC MARKERS

In genetic mapping the diseased gene polymorphic markers play a very important role. These markers could be short tandem repeats, variable number of tandem repeats, blood groups, restriction fragment length polymorphism etc. Mapping can be done in a step wise manner.

Collect all the pedigrees where the disease is found. Analyse all the members against various polymorphisms and perform linkage analysis.

Linkage study entails collecting blood cells from members of several two – and three – generation families or from individuals of a large multiple generation family with a specific genetic disorder. The blood can be cultured and cell lines can be maintained large number of polymorphic markers (probes), representing sites from all parts of all autosomes, are used. A two –point (two – locus) LOD score is calculated for each polymorphic locus and the site of the genetic disease from all informative parent offspring combinations and finally the linkage is established. However, genotyping errors can give –ve or +ve LOD score. Hence perfect genotyping is must to get the correct results.

3.10 USE OF POLYMORPHIC MARKERS IN FORENSIC TESTING

Polymorphic markers have great utility in personal identification. As mentioned above no two individuals are alike. These differences are at both phenotypic and genotypic levels. The genetic differences can be identified by testing these markers. This testing is provided by commercial firms that market directly to consumers. Identity genetic analysis may also be involuntary and used for paternity testing of children or fetuses or for identification of forensic samples in murder, assault or rape cases, in which the perpetrator of the crime left a tissue sample of blood, semen, hair, or other tissue type from which DNA can be extracted and the test can be performed. However, it must be kept in mind that there are ethnic differences in the distribution of these markers. Hence every population should have its own genetic profile.

3.11 USE OF POLYMORPHIC MARKERS IN POPULATION STUDIES

Population Diversity Studies

Human genome varies from individual to individual and therefore no two individuals look alike. This was noted long back. Historically, individual variation was studied on the basis of conventional somatoscopic markers. However, with the advancement of technology various genetic markers were discovered and the gene frequency data for studying the evolution of human races was analyzed using these markers. Initially, the classical serological and biochemical markers have played important roles in various types of human population genetic studies. One of the problems that limited their practical utility results from the limited number of possible genotypes at each of such loci. The discovery of hyper variable DNA loci offers the opportunity to ameliorate this problem. It was later realized that comparison of gene frequencies for one or two loci are not reliable since each locus has a different geographical distribution, hence the differences observed may be because of chance factor. Only when a large number of loci are used, the genetic relationship among populations could be drawn successfully. Recent analysis based upon polymorphic markers reveal that inter and intragroup genetic variation may be of a lesser magnitude and may not be of significance if proper markers are not selected and more so if statistical tools used are not highly powerful. However, it is important to record population variation because it is helpful to know the various mechanisms involved in causing variation and it further enhances our knowledge about the molecular basis of disease susceptibility.

3.12 SUMMARY

It is difficult to attribute any functional significance to genetic polymorphisms. However, the non-coding sequences of the genes which are located far away from the functional region of the gene may affect the function of the gene. However, these sequences are otherwise useful in studying population diversity, disease gene mapping, forensic investigations etc. Recently after the advent of microarray genes for many complex disorders have been found by using genome wide association studies.

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

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

Sample Question

- 1) Define polymorphism with few examples.
- 2) What are the evolutionary forces that affect gene frequency of polymorphic markers?
- 3) Give some uses of polymorphic markers.
- 4) What is law of Hardy Weinberg?
- 5) What is genetic testing?
- 6) Describe the utility of studying molecular markers in anthropological genetics.

