

**Subject: Anthropology**

Production of Courseware

**-Content for Post Graduate Courses**

**Paper No. : 08 Human Population Genetics**

**Module : 14 Genetic Polymorphism**



### Development Team

Principal Investigator

Prof. Anup Kumar Kapoor  
Department of Anthropology, University of Delhi

Paper Coordinator

Prof. Gautam K. Kshatriya  
Department of Anthropology, University of Delhi

Content Writer

Amitabh Biswas  
Department of Anthropology, University of Delhi

Content Reviewer

Prof. A.Paparao  
Venkateswara University, Tirupati, Andhra Pradesh

Sri

Description of Module	
<b>Subject Name</b>	Anthropology
<b>Paper Name</b>	08 Human Population Genetics
<b>Module Name/Title</b>	Genetic Polymorphism
<b>Module Id</b>	14

 **Pathshala**  
पाठशाला  
A Gateway to All Post Graduate Courses

## Content

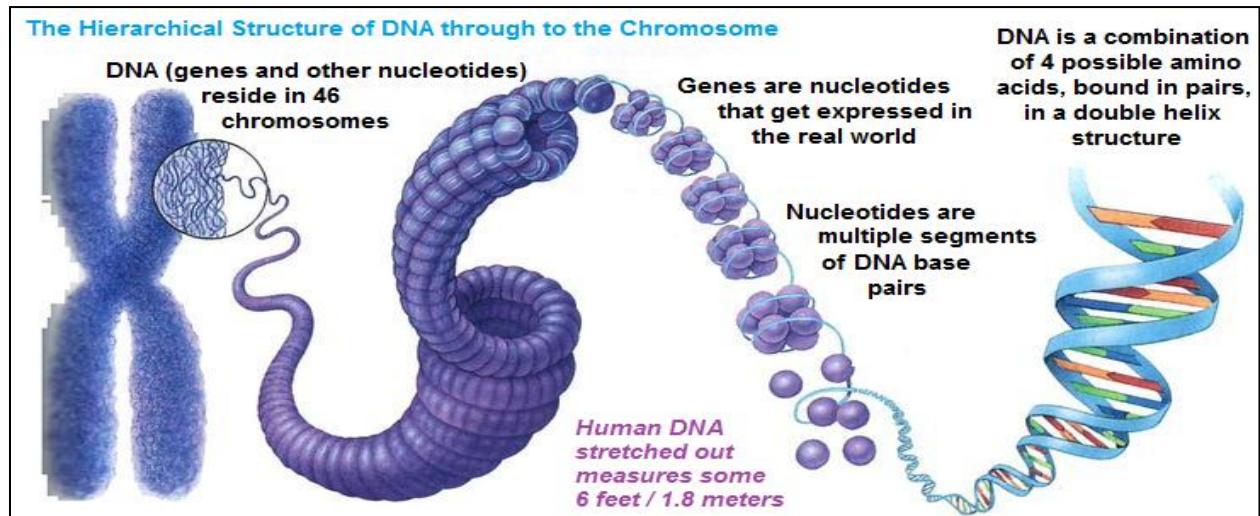
1. Learning Outcomes
2. Introduction
3. Genetic polymorphism
4. Mutation
5. Types of genetic polymorphism
6. Single nucleotide polymorphisms
7. Insertion and deletion polymorphism
8. Copy number variations polymorphism
9. Genotype, Phenotype and Haplotype
10. Significance of genetic polymorphism

### 1. Learning Outcomes

- This study module will help to understand about the genetic polymorphism and their importance.
- It will help to find out the difference in mutation and genetic polymorphism
- It will make you learn the types of Polymorphism, how polymorphism interacts with phenotype.

### 2. Introduction

Before we discuss about the genetic polymorphism, let us brief about the basic about gene and chromosomes. Gene is a functionally inheritable unit which is made of a DNA, representing as a blueprint for the protein development. Genes are located in miniature thread-like structures called chromosomes which are found inside cells. Chromosomes are there in identical sets of two (or pairs) and there are hundreds and thousands of genes in just one chromosome (Fig 1). Chromosomes are numbered according to their sizes, largest chromosome is chromosome 1 and so on. The chromosomes are made of DNA sequences, which is a short form of deoxyribonucleic acid. There are 23 paired set of chromosomes out of which 22 pairs of autologous and Sex chromosomes X or Y which determines the sex whether male or female. They are in pair arrangement because each pair is inherited from each parent (Gangene, 2012).



**Figure 1.** Depiction of Hierarchical structure of DNA through chromosome (Taken from <http://futurehumanevolution.com/what-is-genetic-engineering>).

Every Human cell consist of about 25,000-28,000 genes (Poethig, 2001). Genes hold the information that determines traits, which are features or characteristics that are passed or inherited from parents to offspring's. Genes are responsible for all features and characteristics like skin color, hair color and texture and eye color etc. Total of 25,000 to 28,000 genes are there in the human genome which are bound to be altered. Every Individual carry genetic variations, these variations may have mild or severe effect depending upon the formation of proteins. Genetic variations described as differences between the DNA sequences of individual genomes. Because each of us has two sets of genome (one from paternal and a maternal genome), genetic variation occurs within as well as between individuals. At any genetic locus (DNA region having a unique chromosomal location) the maternal and paternal alleles normally have identical or slightly different DNA sequences (Variations are said to be homozygotes if the both alleles are identical on each chromosome, or heterozygote's if they differ by even a single nucleotide on each chromosome).

### 3. Genetic Polymorphism

Differences in DNA sequences among individuals, groups, or populations from the wild type are known as Genetic variation. Genetic polymorphism word is a combination of the Greek words poly (means multiple) and morph (mean form), in genetics, this term is used to illustrate the compound forms of a single allele/gene that can be present in an individual or population. These include Single Nucleotide Polymorphism (SNPs), copy number variations, insertions, deletions and recombination. Overall, we can define a polymorphism as Variation in DNA sequence that is commonly present in the population. The subjective cut-off point that is defined between a mutation and a polymorphism is 1%. Frequency of variation greater than or equal 1% to be grouped as a polymorphism, the minor common

allele must have a frequency of 1 percent or more in the population. If the frequency is lower than 1% than the allele is considered as a mutation (Ford, 1965).

Some variations are more common than others because variants that lead to human diseases are in general rare in the population because they decrease health fitness. During Evolution, mutant allele eliminated from the line, such mutant allele or variations are termed as mutations. But, not all variations cause diseases. Any novel variant, even if it has neutral or beneficial effect, will commence as a rare mutation (Hedrick, et al., 1986).

Polymorphism usually does not cause any chronic diseases and many of them are found in inter-genic region and are totally neutral. Some polymorphisms may be found within genes, but it may only influence characteristics such as height, hair colour rather than medical traits causing disease. On the other hand, some polymorphism may give disease susceptibility and may also manipulate drug responses and efficacy. A mutation in one population can turn into a polymorphism in another population if it had an advantage and increases in frequency.

Sickle-cell disease is a good example showing advantage of allele over some populations. Sickle cell anemia is caused by a mutation in beta-globin gene that causes a severely blood disorder in Caucasian population. But in certain parts of Africa, nevertheless, same mutant allele is polymorphic because it had resistance to the blood-borne parasite that causes malaria (Ayi et al., 2004).

#### **4. Mutation**

A “Mutation” is defined as any variation in a DNA sequence compared to normal which is abnormal and rare in the population (<1%) and these alterations are also referred as disease-causing change. Mutations originate as a result of changes in our DNA that are not corrected by cellular DNA repair systems. The DNA changes are occasionally induced by radiation and chemicals in our environment, but the great majority arise from endogenous sources. The latter include spontaneous errors in normal cellular mechanisms that regulate chromosome segregation, recombination, DNA replication, and DNA repair and also spontaneous chemical damage to DNA (Loewe, 2008).

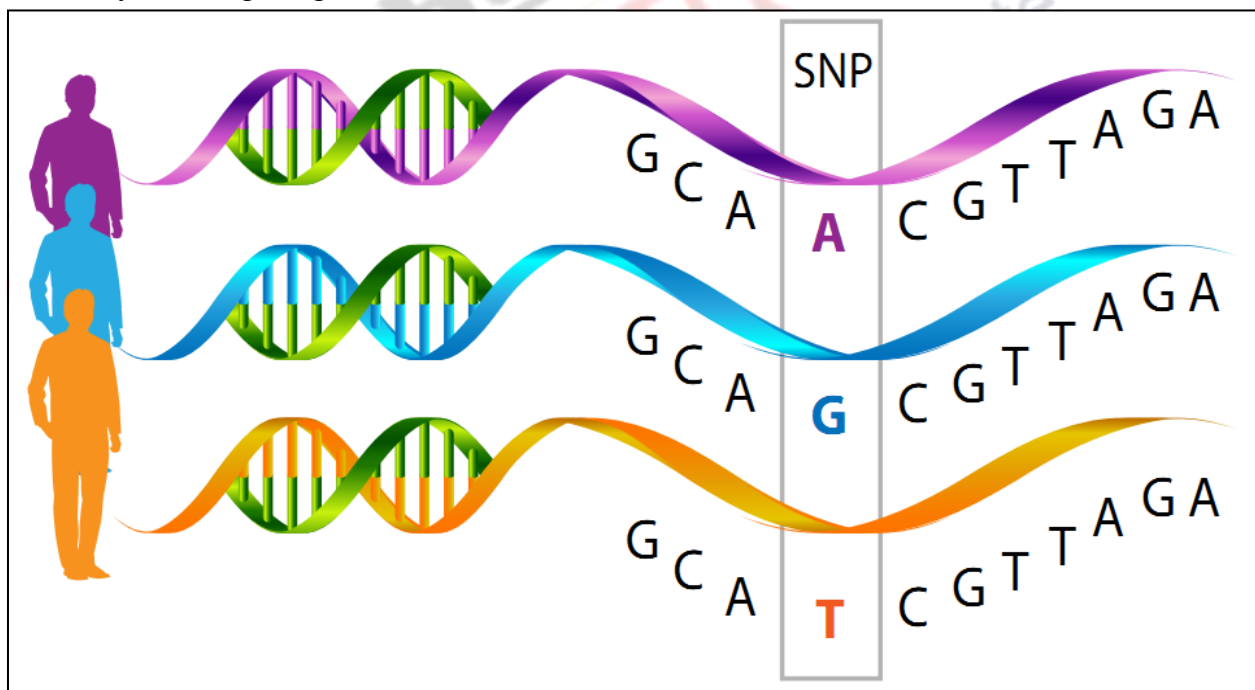
#### **5. Types of Genetic Polymorphism**

Genetic variations occur in DNA sequences in many ways. Genetic polymorphisms in human genomes provide uniqueness. These genetic variations control most of the traits including susceptibility to disease. There are different types of polymorphisms depending upon the number of nucleotide involved and how the variations occurred (Den Dunnen et al., 2001). Genetic polymorphisms are classified under large and small scale polymorphisms. Single nucleotide variations were found to be more prevalent and important form of genetic variation, if Minor allele frequency is more than 1% then it is called Single Nucleotide Polymorphism (SNP). Studies had shown that Copy number variations (CNVs) are present three times more than SNPs. Redon et al. (2006) defined a CNV as segment of DNA (1 kilobase;kb or larger) that is there at a variable copy number in comparison with a reference

genome. Those CNVs which are commonly found (frequency>1%) are considered as CNV polymorphism. An insertion/deletion polymorphism, commonly abbreviated “indel,” is a type of genetic polymorphism in which a specific nucleotide sequence is added (insertion) or absent (deletion) (Rodriguez-Murillo et al., 2013; Weber et al., 2002). Another type of genetic variation is Duplication where there is production of one or more copies of any piece of DNA, including a gene or even an entire chromosome. Translocation is genetic variation known where breakage and removal of a DNA segment from one chromosome, followed by the attachment of segment to a different chromosome.

## 6. Single Nucleotide Polymorphisms

Single nucleotide variations, commonly known as called SNVs are the most common type of genetic variation among people. If frequency is more than 1%, then it is considered as Single nucleotide polymorphism (SNP or snips). SNPs occur usually all over the DNA. On average, SNPs occur in every 300 nucleotides, which mean approximately 10 million SNPs are present in the human genome. These polymorphisms may refer as markers by scientists locate genes that are associated with disease. When SNPs occur within a gene or in a regulatory region near a gene, they may play a more direct role in disease by affecting the gene’s function (Zhao et al., 2012; Sachidanandam et al., 2001).



**Figure2.**Showing SNP in the stretch of DNA. (Taken from <https://neuroendoimmune.wordpress.com/2014/03/27/dna-rna-snp-alphabet-soup-or-an-introduction-to-genetics/>)

In the human genome, the pattern of SNV is not random because of two reasons. First, different mutation rates were found in different DNA regions and different DNA sequences. Like mtDNA has a

higher mutation rate than nuclear DNA (Due to proximity to reactive oxygen species which are generated especially in mitochondria). Excess of C→T substitutions in the human genome because the CG frequently acts as a methylation signal; Cytosine becomes 5-methyl Cytosine, which is prone to deamination to form T (Duret et al., 2009; Zhao et al., 2012).

Other reason for non-randomness in pattern comes from our evolutionary ancestry. Certain nucleotides are polymorphic and surrounded nucleotides that don't have variants or rare variants. In general, the nucleotides found at SNP sites are not particularly susceptible to variation (the germline mutation rate is about  $1.1 \times 10^{-8}$  per generation, approximately 1 nucleotide per 100Mb, and SNPs are steady over evolutionary time). As an alternative, nucleotides at SNP sites mark another ancestral chromosome segments that are frequent in the modern population.

Nucleotide base substitutions are of two types: A transition nucleotide substitution occurs between purines (A, G) or between pyrimidines (C, T). Most common form of SNP constitutes two thirds of all SNPs. Another is transversion substitution which occurs between a purine and a pyrimidine.

Single-nucleotide polymorphisms are present in the coding or non-coding regions of genes, or in the intergenic regions (i.e. regions between genes). Coding region SNPs are of two types, synonymous and non-synonymous SNPs. Synonymous are those polymorphism which do not affect the protein sequence while non-synonymous SNPs change the amino acid sequence of protein. The nonsynonymous SNPs are of two types: missense and nonsense. When the change of a single base pair causes the substitution of a different amino acid is known as missense mutation. Sometimes, amino acid substitution may have neutral effect, or it may cause to be the protein nonfunctional. In genetics, a nonsense mutation is a point mutation in a sequence of DNA that results in a premature stop codon, or a nonsense codon producing truncated, incomplete, and generally nonfunctional protein.

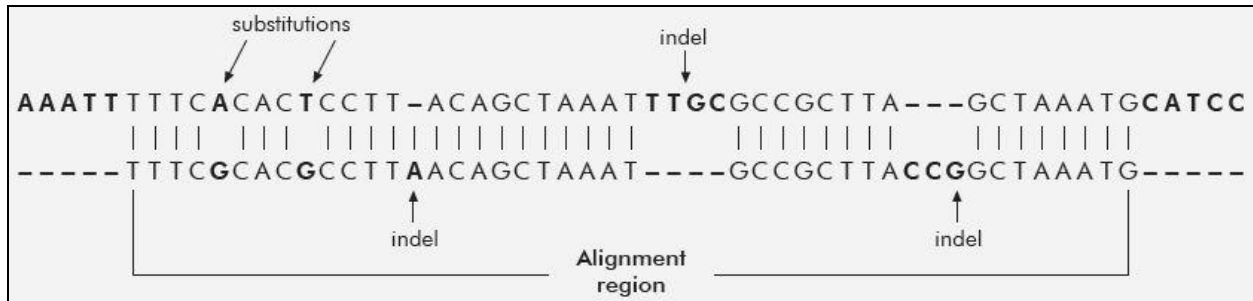
Some SNPs that are present in splicing region may affect gene splicing, transcription factor binding, mRNA degradation, or the non-coding RNA sequences. Gene expressions affected by the SNPs are referred to as an eSNP (expression SNP) and may present in upstream or downstream from the gene.

Scientists are identifying, cataloging, and analyzing small genetic variations among humans that may help to synthesize more specialized and effective medical treatments for specific individuals. Variations are very informative and should be considered while understanding disease mechanism. The challenge for recent researchers is to identify these SNPs that correlate with a particular phenotype of patients. Consistent SNPs could serve as predictive biological markers that may inform about numerous aspects of medical health care, including particular diseases, effectiveness of various drugs and adverse reactions to specific drugs.

## 7. Insertion and Deletion Polymorphism

Insertion and Deletion Polymorphisms are the variations in which a portion of DNA is added or deleted from the human genome. The Number of nucleotides can range from single to kb and its functional

effect can be predicted on the basis of its location. If the Insertion or deletion which is not in multiple of three occurs in a protein coding region, it is also known as a frameshift mutation. Short regions Deletion and insertion mutations can be formed by strand slippage, and deletions or insertions of longer regions are formed usually due to homologous recombination. Mispairing of the template strand leads to formation of new strand during replication causes strand slippage. Denaturing of the newly synthesized strand during DNA synthesis, it can occasionally pair with the wrong sequence. Such mispairing occurs for a single nucleotide or for short, directly repeated sequences. If the template strand loops out, then a deletion will result (Rodriguez-Murillo et al., 2013; Weber et al., 2002).



**Figure 3.** Showing the indels in the DNA sequence (Taken from <http://www.uncommondescent.com/genetics/icc-2013-geneticist-jeff-tomkins-vs-evolutionary-biologist-who-got-laughed-off-stage/>)

Although many studies have been conducted to identify single nucleotide polymorphisms (SNPs) in humans, few studies have been conducted to identify alternative forms of natural genetic variation, such as insertion and deletion (INDEL) polymorphisms. The frequency of insertion/deletion polymorphism in the human genome is about one-tenth the frequency of single nucleotide substitutions. Short insertions and deletions are much more common than long ones. Thus, 90% of all insertions and deletions are of sequences 1–10 nucleotides long, 9% involve sequences from 11 to 100 nucleotides, and only 1% involve sequences greater than 100 nucleotides (Mullaney et al., 2010).

An initial map of human INDEL variation that contains 415,436 unique INDEL polymorphisms was reported in 2006 (Mills et al., 2006). These INDELs were identified with a computational approach using DNA re-sequencing traces that originally were generated for SNP discovery projects. They range from 1 bp to 9989 bp in length and are split almost equally between insertions and deletions, relative to the chimpanzee genome sequence. Five major classes of INDELs were identified, including (1) insertions and deletions of single-base pairs, (2) monomeric base pair expansions, (3) multi-base pair expansions of 2–15 bp repeat units, (4) transposon insertions, and (5) INDELs containing random DNA sequences. On average, 1 INDEL per 7.2kb of DNA was found with some hotspot region with 48-fold increase.

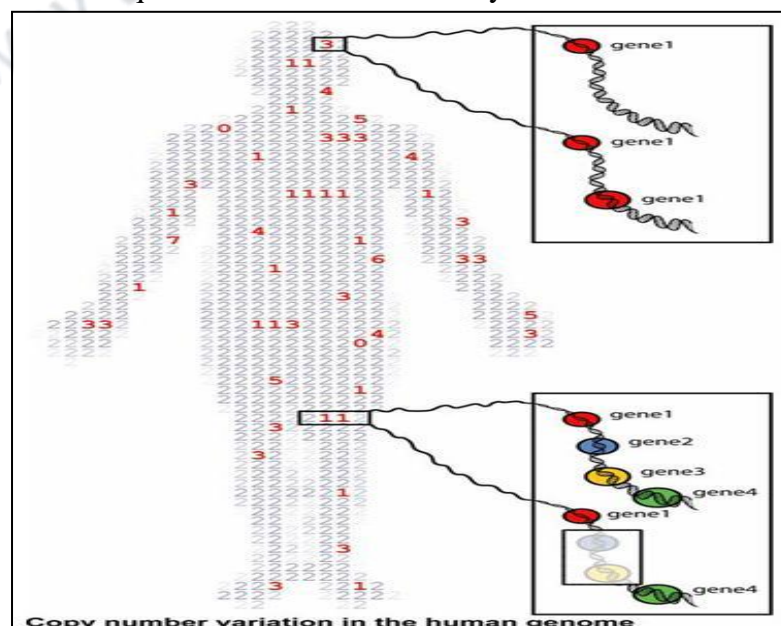


For example, A Common INDEL polymorphism reported in 1990, a polymorphism consisting of the presence or absence of a 250-bp DNA fragment was detected within the angiotensin I-converting enzyme gene (ACE) using the endothelial ACE cDNA probe. This INDEL polymorphism was found to associated with serum level of ACE with shorter allele had higher ACE concentration, Allele frequencies were 0.6 for the shorter allele and 0.4 for the longer allele. This is one of good example of INDEL to show its effect on phenotype (Rigat et al., 1990).

## 8. Copy Number Variations Polymorphism

The number of copies of a particular gene or genetic region in the genotype of an individual referred as gene copy number (also "copy number variants" or CNVs). It was normally consideration that all genes were always present in two copies in a genome (Redon et al., 2006). Nevertheless, recent discoveries have revealed that large segments of DNA, ranging from thousands to millions DNA bases, may vary in copy-number from individual to individual. Such CNVs may include genes leading to dosage imbalances. Like genes that are present in two copies per genome have been found to sometimes present in one, three, or more copies, sometimes genes are found to be missing.

Recent studies reveal that CNVs are three times the SNPs. So it is important to understand the mechanisms of CNV development which may also help us in better understanding of human genome evolution. Medical research will transform with the new global CNV map in mainly four areas (Need et al., 2009). The first and most important aspect is in identifying for genes underlying common diseases. Till date, identification of these genes has not been actually considered the role CNVs in human health. Secondly, Study on familial genetic conditions employed the CNV map. Third, by chromosomal rearrangements, there are thousands of severe developmental defects. The CNV map is used to eliminate variation found in unaffected individuals, helping scientists to aim the DNA region that might be involved in pathogenesis. The generated data will also add to a create more precise and complete human genome reference sequence which can be used by all scientists and researchers.



**Figure 4.** Showing copy number variations in human genome.(Taken from <http://cnv.gene-quantification.info/>)

Investigation of CNVs in the genomes of human and chimpanzee demonstrates the potentially greater role of CNVs in evolutionary change than SNPs (Cheng et al., 2005). Comparisons of the human and chimpanzee genomes revealed that there are twice as many nucleotides involved in CNVs as there are in changes to single nucleotides, 2.7% compared to 1.2%. Cheng et al. (2005) discovered that the majority of CNVs were shared between the human and chimpanzee genomes, but in the human genome approximately one-third of the CNVs were observed unique. To confirm these results, other researchers on the basis of comparison of genomic sequences with comparative genomic hybridization suggested the same results. Other studies have further found that CNVs are also associated to genetic diseases obvious in humans (Stankiewicz & Lupski, 2002).

Because the study of copy number variations is a relatively new area of genetic research, many questions regarding CNVs remain unresolved. Scientists worldwide are actively pursuing research regarding the origin of these structural variations, as well as their contributions to both evolutionary adaptation and human disease. New tools, such as comparative genomic hybridization, should allow scientists to look at CNVs in detail and examine their origin and significance.

When investigators examined the raw sequence data of the human and chimpanzee genomes and estimated that there is 99.9% identicalness between the two species, they focused primarily on differences at the level of single nucleotide polymorphisms. Recent analysis of the structural level — specifically CNVs — has revealed an additional source of variation, and this has led to a revised picture of genomic diversity and a greater appreciation of its dynamic nature.

## 9. Genotype, Phenotype and Haplotype

Genotype, phenotype and haplotype are the most important and basic notions related to genetic variations. It is very important to have a clear understanding about these aspects and the processes of genotyping and haplotyping. Genotype is a precise picture of the genetic profile of an individual. The genetic makeup of an organism as identified by genetic or molecular analysis, i.e. the complete set of genes, both dominant and recessive, possessed by a particular cell or organism. The observable properties of an individual as they have developed under the combined influences of the individual's genotype and the effects of environmental factors is known as Phenotype. Set of SNPs, inherited together in a block referred as haplotype, which help in understanding phenotype association with the DNA sequences regions or genes.

These polymorphisms will help to identify the genetic regions, on the basis of haplotyping of the disease patients and comparing it with the healthy controls. SNPs or INDELS or CNVs modify the phenotype, continuously evolving to remove disease allele and increasing protective polymorphism.

## 10. Significance of Genetic Polymorphism

Genetic polymorphisms are derived from continuous evolution of nature and found in every level of human development. The new horizons for genomics has provided us platform on explaining genetic polymorphisms and raised many challenges for the researchers to be addressed. Genetic polymorphisms includes all types of variations in the DNA sequence, from single base pair substitution (SNPs), deletions or insertions of nucleotides (indels), variable tandem repeats, duplication of gene, rearrangements of nucleotide, the absence or presence of transposable elements, etc. The section of genomes having genetic polymorphism is not large, but most of the functional diversity and adaptation may founded on them. First of all, Genetic polymorphism causing functional effects. In common livestock, 200 diseases are to known to cause by Single base pair DNA polymorphisms (Ibeagha-Awemu et al., 2008), give us frequent cases of how these SNPs affects protein functions to fluctuating degrees. As contribution of genetic polymorphism in functional part makes them an important entity in human health, so more number of cases might be documented in future to provide better view of how these polymorphisms contribute to the molecular systems. Secondly, in context of pathways or networks, genetic polymorphism also play role in different mechanism. For the better understanding of functionality of pathway polymorphisms, many studies on human had provided good examples. Like, in human, Insulin-like growth factor-I (IGF-I) which synthesize polypeptide hormone which promotes normal development and cellular growth, polymorphism in Insulin growth factor type I receptor/IGF-IR and phosphoinositide 3-kinase (PI3KCB) genes disturb the plasma levels of IGF-I (Bonafè et al., 2003). Life longevity may also be influenced by the insulin/IGF-I signal response pathway (Barbieri et al., 2003). Genome wide polymorphisms methodology may help to discover unknown pathways and probable regulators of the pathways by statistical analysis, as presented for the formation of aliphatic glucosinolate.

Advancement in the modern techniques as high-throughput sequencing and generating huge data in parallel have been thought-provoking biologists, complex scenario for statisticians, and har time to computer scientists to generate new paradigms to explain and understand their inquiries in research setups. Inter as well intra collaborations across different disciplines and multifunctional research organization will remain to rule the front bench of biological investigations on genetic polymorphisms and reinforce some key discoveries in the future.

### Summary

- Gene is a functionally inheritable unit which is made of a DNA, representing as as blueprint for the protein development.
- Differences in DNA sequences among individuals, groups, or populations from the wild type are known as Genetic variation.

- Overall, we can define a polymorphism as Variation in DNA sequence that is commonly present in the population. The subjective cut-off point that is defined between a mutation and a polymorphism is 1%.
- Frequency of variation greater than or equal 1% to be grouped as a polymorphism, the minor common allele must have a frequency of 1 percent or more in the population.
- Genetic polymorphisms are classified under large and small scale polymorphisms. Single Nucleotide Polymorphism (SNP), Copy number variations (CNVs), Indels.
- Single nucleotide variations, commonly known as called SNVs are the most common type of genetic variation among people. If frequency is more than 1%, then it is considered as Single nucleotide polymorphism (SNP or snips).
- The pattern of SNV is not random because different mutation rates were found in different DNA regions and different DNA sequences and due to our evolutionary ancestry.
- Coding region SNPs are of two types, synonymous and non-synonymous SNPs. Synonymous are those polymorphism which do not affect the protein sequence while non-synonymous.
- Insertion and Deletion Polymorphisms are the variations in which a portion of DNA is added or deleted from the human genome.
- Short regions Deletion and insertion mutations can be formed by strand slippage, and deletions or insertions of longer regions are formed usually due to homologous recombination.
- 90% of all insertions and deletions are of sequences 1–10 nucleotides long, 9% involve sequences from 11 to 100 nucleotides, and only 1% involve sequences greater than 100 nucleotides.
- The number of copies of a particular gene or genetic region in the genotype of an individual referred as gene copy number (also "copy number variants" or CNVs).
- Medical research will transform with the new global CNV map in mainly four areas i.e. identifying for genes underlying common diseases, Study on familial genetic conditions, understanding severe developmental disorders due to chromosomal rearrangements and mapping ancestry.
- Genetic polymorphisms are contributions from nature and sources of variations at all levels. The postgenomic era permits new perspectives on interpreting genetic polymorphisms and also poses challenges for scientists to answer system questions.