CHAPTER 1

INTRODUCTION

1. Introduction

Human genome project and genetic research has generated an enormous amount of data about the genetic differences among the world and Indian populations. Investigations of these differences have transformed the understanding of the origin and nature of the human diseases and their connection with the genetic background (**Bamshad *et. al*., 2004., cavalla- Sforza, 1994**). Geneticists have studied the genetic variation among the individuals forming a species (**Cavalla- Sforza and Feldman, 2003., Jorde and Wooding, 2004**). Different human traits like hair, eye color and others vary among individuals in many populations due to the genetic variations present in the individuals. These traits contribute to the understanding of the extent of variations in humans. The first genetic variation study is of ABO blood groups which were described long time ago (**Landsteiner, 1927**).

These studies were then extended to other blood group systems and to other markers to show that the different human populations have different proportion of blood groups. The first site of the staggering magnitude of genetic variation came later in the beginning of 50s and coming to full development in 60s when individual differences for proteins could be systematically studied. Different protein chains vary considerably in their amino acid compositions and serve different functions. The same protein structure may show small strictly inherited differences between individuals. In case of sickle cell anemia it was first observed that replacement of one amino acid by another in hemoglobin molecule was the determinant of the hereditary disease. The amino acid change in sickle cell amino acid involved change in the electric charge of the hemoglobin molecule. Electrophoretic analysis has been developed and helped to detect the variation in the proteins (**Glazko, 2005**).

Recent trends mainly focused on the DNA variations. Changes in the DNA sequences in different human populations helped to find out the variations among the populations. It is a very unique tool to study variations at the DNA level.

1.1 Evolutionary process affecting population diversity

Anthropologist tried to construct evolutionary relationships and history on the basis of genetic background and for many years by the analysis of cephalic index (percentage of skull breadth to

length) introduced in the middle of last century. However, with a single trait two populations of different origin could well turn out to be more or less identical. The anthropometric traits are not well established because the characters are under the control of inheritance and variations are short term and under control of environmental changes. Single gene frequency study is not very authentic as that the frequency for one gene in different human populations may not change during many years (**Itoh, 2002**). By studying more than one gene and many alleles of different genes. It will definitely reconstruct the human evolutionary history (**Jorde and Wooding, 2004)**.

Different statistical analysis such as Univariate for single trait and multivariate analysis for multiple genes helped to combine information from the population data. Multivariate analysis is useful for understanding the evolutionary process such as migration, random genetic drift etc. The construction of human evolution including migration, mutation, genetic drift and natural selection is difficult and challenging. Results from the genetic data should be compared with the existing data to infer some conclusion.

Different genetic types of different proteins and genes exist and one needs to count individuals carrying it. These proportions vary from population to population because they change over time (**Nei, 2005**). The change is very slow but incessant over generations. Therefore, the primary interest is to understand the evolutionary process.

Population is a local group of organism of the same species that interbreed. A group of individuals within whom marriages are constructed is called a Mendelian population. The genetic information carried by a population is called its gene pool. Gene pool transfers from one generation to the next. The gene pools of a new generation have different alleles or gene frequencies than their parental pool. The changes in allele frequency can cause changes in phenotypic frequency which causes evolutionary change. Members of the same local group are very close than the group separated from geographic barriers or other. These results in the change of behavior occur at the micro or macro level. The evolution also changed the genotypic and phenotypic constitution of the population (**Joaquim, 2004**).

To understand the genetic structure of human populations one needs to calculate the gene frequencies or the allele frequencies at different loci. Gene frequencies are the estimates of the relative frequencies of alleles and are of widest application in human evolution. The gene

frequency estimation is observed by the phenomenon of sampling fluctuation and misclassification of genotypes which can be solved by using right markers and more sophisticated software or statistical analysis tools. In any population, the genotype frequencies among zygotes are determined in large part by the patterns in which genotypes of previous generation come together to form mating pairs. In random mating, genotypes form mating pairs in the proportions expected from random collisions. It is solved by the theory of Hardy-Weinberg equilibrium (HWE). For a gene with two alleles A and a in a random mating, the expected genotype of AA, Aa and aa are given by P^2, 2pq and q^2 respectively (where p and q are the allele frequencies of A and a, with p+q= 1). Statistical tests of HWE are often based on the χ^2 test, but this test is relatively weak in detecting departures from expected frequencies especially those caused by admixture of subpopulations differencing in allele frequency. The gene frequencies can be changed by several factors like mutation, genetic drift, gene flow, natural selection etc (**Cavalla- Sforza, 2003**).

Gene is being extensively studied now-a-days and the alternative form of this is called allele. This allele is produced due to mutations in the populations (**Roychoudhury, 1988**). The geographic distribution of a particular allele may give information on the place of origin of the genetic changes that generated it. Mutation occurs in both coding and non coding region and causes polymorphism in the DNA. The geographic distribution of a particular allele may give information on the place of origin of genetic changes that generated it. Correlation of distributions of gene frequencies with environmental parameters at the geographic level has been instrumental in the discovery of specific genetic adaptations (**Nei, 2005**). The proportion of allele varies from place to place but greatest variation observed in case of large distances.

Random genetic drift is also one of the major factors that change the gene frequencies among populations. Frequencies change over time because of the accumulation of random sampling errors while passing from one generation to other. When a small population migrates to a new place, the sampling error is large and allele frequencies in new population may be different from the parent population. In course of time the errors are decreased and constitute a very different frequency. This is called ―founder effect‖ (**Strachan *et. al*., 1996**).

Gene flow is a process by which interbreeding among certain groups of individuals results in those populations which become similar to each other. Mating can be either positive assortative

or negative assortative. Positive assortative mating increases the homozygosity. While mating is carried out between individuals with different genotypes then it is called negative assortative (C**avalla- Sforza, 1998**) which results in high heterozygosity in populations.

Natural selection is a shifting process. Organisms that can adapt in particular changing environment are at an advantage than the others. It is influenced by the organism‘s biology but also with the interaction of biology with environment (**Sternberg, 2004**).

Adaptation and migration are also two factors that change the genetic diversity of human populations. It is possible that much of the variation seen among the groups of human populations indirectly resulted from the pattern of expansion and migrations accompanied by genetic drift. Over time frequencies of DNA variants changed only in terms of the total DNA composition but changed enough to produce differences (**Harding *et. al*., 2000**).

1.2 Genetic tools for studying human genetic diversity

Advancement of various techniques in molecular biology enabled us to study the human genetic diversity and the association with different diseases. Previously, DNA markers were that of functional genes. However, these were very less informative. In recent times development of DNA markers has been discovered along with more robust DNA typing technology. These markers are neutral and not affected by selection pressure. These markers are also polymorphic, making them more informative in studying genetic diversity among human populations. The different genetic tools are as follows-

1.2.1 Short tandem repeats (STR)

STR DNA markers are short fragments of DNA that are commonly used in forensic identification. These markers are useful in studying human genetic diversity because of their polymorphism. STR marker analysis estimates the exact number of repeating units in DNA. Single nucleotide polymorphisms are characterized by single base changes in DNA sequence. SNPs are highly stable and preserved across populations and help for genetic diversity study.

1.2.2 Mitochondrial DNA (mtDNA) and Y chromosome markers

Mitochondrial DNA markers are essentially variations in DNA sequences. mtDNA is highly stable allowing lineage analysis among the populations across time and different genetic geographic areas. mtDNA, SNPs are very effective markers for analyzing genetic diversity and disease association study among populations.

Y chromosome markers are exclusively paternally inherited, allowing genetic analysis of the male lineage. As mtDNA is maternally inherited it allows genetic analysis of maternal line.

Along with the MHC and KIR, another marker came into focus that is TLR or Toll like receptors. This marker is mainly of innate immunity genes. Gene- environmental interaction has been described with the help of this marker.

1.3 Population of Indian Sub- continent

Indian subcontinent is located between 8 degree N to 37-degree N latitude and 68 degree to 97 degree longitude. This country assembled over 100 million people in the country with all the different populations and their different cultural background (**Cann, 2001**). The populations reside in this country are mixed between the western Caucasians and the Oriental in the East (**Chakraborty, 1992., Balakrishnan, 1996., Agarwal & Arundhati, 1999., Jaini, 2002., Chhaya, 2005**).

Investigation of the genetic diversity among the present human populations can be useful in reconstructing concepts regarding population diversity and migration routes and also in identifying the ancestral populations. The Indian subcontinent not only exhibits enormous morphological, cultural, and linguistic diversity but also stands only second to Africa in its genetic richness (**Cann, 2001**). Therefore, people of the Indian continent have been and continue to be of interest for investigation in different areas, all aimed at exploring their vast genetic wealth. Moreover, the continent has served as a major corridor for the dispersal of modern humans that started from Africa about 10,000 years ago (**Majumdar, 1998**). Thus, India occupies a center stage in human evolution.

Various socio-cultural practices have led to a unique gene pool of the human population of India and thus there is a need to study these unique populations at genetic level (**Srivastava, 2007)**. It

has been reported by the population geneticists that unique marital patterns, such as endogamous caste groups, could be one such contributory factors for the creation of unique gene pool. India, on the other hand also has undergone several historical invasions, as a result of which a significant admixture might have taken place. The most salient feature of Indian population is that in India socio-cultural barriers also plays an important role in determining the gene flow between populations (**Balakrishnan, 1978**), thereby establishing the diversity at the genetic level. This could have led to the formation of numerous close gene pools, which has remained virtually undisturbed for many generations that may have disturbed the original gene pool of the Indian population.

North Bengal region has got full of cultural and linguistic variations. The presence of various ethnic populations in this region signifies the variation in the gene pool of the populations. In the present study such four ethnic populations namely Rajbanshi, Gurkha, Muslim, Rabha were chosen to reveal the gene-environment interaction of ten human TLR genes among the populations.

Therefore, in the present study the objectives were formulated as mentioned below.

OBJECTIVES OF THE STUDY

1. To study the distribution and frequency of TLR genes of some human populations of North Bengal Region of West Bengal.

2. To study the heterogeneity among the local population(s) and trace their phylogenetic relationships.

3. To correlate the association of Rheumatoid arthritis with TLR genes if any.

4. To study the association of typhoid fever with TLR genes if any.

5. To study the association of HIV positive (+ve) patients with TLR genes if any.