GENETIC STRUCTURE OF SUB-ETHNIC GROUPS OF NEWAR POPULATION OF KATHMANDU VALLEY

M.Sc. Thesis

(2014)

Submitted to

CENTRAL DEPARTMENT OF BIOTECHNOLOGY

Tribhuvan University

Institute of Science and Technology

Kirtipur, Kathmandu, Nepal

Nagendra Prasad Awasthi

Exam Roll no: BT 069/068

Registration no. 5-1-61-142-2005

GENETIC STRUCTURE OF SUB-ETHNIC GROUPS OF NEWAR POPULATION OF KATHMANDU VALLEY

M.Sc. Thesis
(2014)

Submitted to

Central Department of Biotechnology
Tribhuvan University
Kirtipur, Kathmandu, Nepal

By

Nagendra Prasad Awasthi

Supervisors

Dr. Tilak R. shrestha
Associated professor
Central Department of Biotechnology
Tribhuvan University, Kritipur

Dr.Krishna Das Manandhar
Professor
Central Department of Biotechnology
Tribhuvan University, Kritipur

Registration no. 5-1-61-142-2005

Recommendations

There are a lots of New, more specific, more reliable molecular technique are developed which insight deeper in the search of ancestry. Some measure to be followed in order to get more specific results:

1. For the study of diversity among the Y-Chromosomal Haplogroup Y-STR (Short Tandem repeats) should be used which gives more deeper diversity in parental lineages.

2. If possible, Genome wide scan should be done which gives detailed genetic structure and diversity of the particular population.

3. The sample size should be large so can be cover large population which results more reliable data that gives more insight about the population.

4. There should be collaboration with Anthropologists which will help to justify further details of genetic diversity of target study.

ABSTRACT

GENETIC STRUCTURE OF SUB-ETHNIC GROUPS OF NEWAR POPULATION OF KATHMANDU VALLEY

Diversity in geographical region, physical environment, flora, and fauna including humans in Nepal has offered suitable platforms to study their origin and genetic relationship. There are very few studies on the populations of Nepal. To understand their genetic structure and affinities, for the first time, we studied Shakya (N = 19), Bajracharya (N = 20) and Udaaya (N=59) population of Newar ethnic group using high resolution Y Chromosomal, Mitochondrial and Autosomal DNA markers. About 10 ml of blood samples were collected from 98 healthy and unrelated individuals, inhabiting Kathmandu Valley with their informed written consent. DNA was extracted from the whole blood with the standard phenol-chloroform method and target regions were amplified using Polymerase chain reaction (PCR). Y-chromosomal Biallelic markers were used to dissect the paternal lineages and whole mitochondrial genome was sequenced and the variations were scored against the revised Cambridge Reference Sequence (rCRS) to trace the maternal lineages. Based on the variations, putative Haplogroups were assigned and were confirmed by coding region markers. Our mtDNA result indicates the dominance of Macrohaplogroup M (90%: includes M2, M3, M5, M9, M30, M33, M35, M38, Z, D, etc.) along with the branches of R (R6, R9) and U (U2 and U7) Haplogroups. Their maternal Genepool was found to harbour around 50% of South Asian, 25% East/Southeast Asian and 20% Central Asian specific Genepool. West Eurasian haplogroup were negligible (5%). Y-chromosome analysis revealed the presence of major haplogroups such as M117-O3a3, M17-R1a, M82-H1, M15-D1 and M124-R2. Overall, South Asian and East/Southeast Asian signature were found prominent among the Newars. The analyses of Autosomal markers (*MYBPC3, LCT, SLC24A, EDAR,* etc.) showed that the genetic structure of Newar population is more or less similar to the Indian Tibeto-Burman populations. In conclusion, the Newar population of Nepal was found to harbour prominent Genepool from South Asia, East/ Southeast Asia. Further, analysis on Y-STRs and high density Autosomal markers will help in tracing the precise origin and affinity of the Newar population of Nepal.

Key Words: Biallelic marker, Haplogroup, Genepool, Autosomal Marker, Y-STR, rCRS

CONTENTS

Chapters Page No.

List of Tables, Figures and Appendices

(A) Tables

(B) Figures

(C) Appendices

LIST OF ABBREVIATIONS

%	-	Percentage
°C	-	Degree centigrade
APS	-	ammonium persulphate
ATP	-	Adenosine 5'-triphosphate
Bis-acrylamide-	-	N, N'-Methylene-bis-acrylamide
bp	-	base pairs
cm	-	centimetre
cpm	-	counts per minute
dATP	-	2'-deoxyadenosine 5'-triphosphate
dCTP	-	2'-deoxycytidine 5'-triphosphate
DDW	-	Double distilled water
dGTP	-	2'-deoxyguanosine 5'-triphosphate
DNA	-	Deoxyribonucleic acid
dNTP	-	2'-deoxynucleotide 5'-triphosphate
ddNTP	-	2',3'-dideoxynucleotide 5'-triphosphate
dTTP	-	2'-deoxythymidine 5'-triphosphate
EDTA	-	Ethylene diamine tetra acetic acid
Et.Br	-	Ethidium bromide
Fig	-	Figure
g	-	Gram
HG	-	Haplogroup
HVRI	-	Hypervariable region I

HVRII	-	Hypervariable region II
KY	-	Kilo Year
kb	-	kilobase
M	-	molarity
mA	-	milli ampere
mg	-	milligram
min.	-	minutes
ml	-	millilitre
mm	-	millimeter
mM	-	millimolar
mtDNA	-	Mitochondrial DNA
N	-	Normality
nm	-	nanometer
NaOH	-	sodium hydroxide
ng	-	nanogram
OD	-	optical density
PAGE	-	Polyacrylamide gel electrophoresis
PCR	-	Polymerase chain reaction
pM	-	picomole
rpm.	-	revolutions per minute
SDDW	-	Sterile Double Distilled water
SDS	-	Sodium dodecyl sulphate
Sec.	-	Seconds

SRY	-	Sex-Determining Region on Y Chromosome
SSC	-	Sodium saline citrate
STR	-	Short Tandem Repeat
TAE	-	Tris-Acetate-EDTA
TBE	-	Tris-Borate-EDTA
TE	-	Tris-EDTA
TEMED	-	N, N, N', N'-tetramethylethelenediamine
Tris	-	Tris (hydroxy methyl) amino methane
U	-	Unit
UV	-	Ultraviolet
V	-	Volts
v/v	-	Volume/Volume
w/v	-	Weight/Volume
µg	-	Microgram
µl	-	Microlitre
µM	-	Micromolar
W	-	Watts
YAP	-	Y-*Alu* polymorphism
YBP	-	Year before Present

CHAPTER 1
INTRODUCTION

1.1 Background

People have always been curious about their origins. In the absence of written records, researcher have made use of information from disciplines as diverse as linguistics, archaeology, physical anthropology, cultural anthropology, history and Paleo-Anthropology to reconstruct their prehistory (Heng, 2009). The most direct account of our past is inferred from the fossil record. Skeletal remains have been instrumental in establishing the evolution of human ancestors in Africa, and they have also provided important information about the evolution of modern *Homo sapiens*. Now a day's use of molecular biology to trace human origin is most important, accurate and sophisticated tool. In this approach genetic variation in the population are used to study of human evolution. This is because our DNA has been passed down to us from our ancestors, accumulating mutations on the way. The DNAs of modern humans are, therefore, different from each other, and these differences, or polymorphisms, provide a record of our relatedness and genetic history.

Human mitochondrial DNA is widely used as tool in many fields including evolutionary biology and population history, medical genetics, genetic genealogy and forensic science. However, most studies of human evolution based on mtDNA sequencing have been confined to the control region, which constitutes less than 7% of the mitochondrial genome (Ruiz-Pesini et al., 2007). Improvements in rapid sequencing technology have moved researchers toward using rapid sequencing of either coding region or whole mtDNA (Abu-Amero et al. 2008). Due to the apparent lack of recombination as it is a semi-autonomously replicating molecule and maternal mode of inheritance mtDNA is used to estimate of the maternal genealogy.

Y-chromosome passes down from father to son, largely unchanged, except by the gradual accumulation of mutations. Different populations often have characteristically different Y chromosome and these studies are likely to make a major contribution to our understanding of the origin of modern humans (Jobling and Smith, 2000). By examining the difference between polymorphic Y-chromosomal markers one can attempt to reconstruct a history of human paternal lineages, population structure and history, genealogy, forensics and the investigation of selective influences in the Y chromosome 95% of the Y chromosome has become a genetic junkyard because it does not recombine. In the Y-chromosome's passage through the generations, changes occur randomly in its junk DNA and so the Y-chromosome of the contemporary populations retains a record of their passage through time. They can reveal the paternal genealogy

of their owners and the relationships between different groups of individuals (Underhill et al., 2007).

Some of the variation in Autosomal gene can be used to study of evolution and genetic variation of particular population. Mostly gene involved in adaptation such as High Altitude Adaptation, alcoholic dependence, pattern of some genetic disease, human skin color gene, gene involve into digestion of particular food etc. The variation on such gene is due to selection and to adapt that particular environment. The variations on such gene are population specific, hence can be used as a marker for the human evolution study (Michael et al., 2010).

1.2 Population Genetics

The origin of variation, transmission of variants generation to generation and the temporal changes that occur in populations because of systematic and random evolutionary forces combine to form the field of population genetics (Daniel, 2008). Allele frequency is the proportion of all variants that is made up of a particular gene variant. It is used to depict the amount of genetic diversity at the individual, population and species level (Chen, 2012).

Allele frequency has been related to genotype frequency by Hardy Weinberg principle, which states that both allele and genotype frequencies in a population remain constant from generation to generation unless specific disturbing influences are introduced (Shriner, 2011). One or more of them is always in effect and so equilibrium is impossible (Cheung et al., 2000). Genetic equilibrium is an ideal state that provides a base line to measure genetic change. Both allele and genotype frequency in a population remain constant that is they are in equilibrium from generation to generation unless specific disturbing influences are introduced (Duvernell et al., 2008). These disturbing influences are mutations, natural selection, random genetic drift and gene flow.

1.2.1 Mutation Study in Population Genetics

Everyone acquires some changes to their DNA during the course of their lives. These changes occur in a number of ways. Sometimes there are simple copying errors that are introduced when DNA replicates itself (Conrad et al., 2011). Other changes are introduced as a result of DNA damage through environmental agents including sunlight, cigarette smoke, and radiation. In multicellular organisms, mutations can be subdivided into germ line mutations, which can be passed on to progeny and somatic mutations, which (when accidental) often lead to the malfunction or death of a cell and can cause cancer. Mutations serve to introduce novel genetic variation that may affect the fitness of the organism (Barton and Knightly, 2002).

1.2.2 Natural Selection

Natural selection comes from differences in survival and reproduction as a result of the environment. Differential mortality is the survival rate of individuals to their reproductive age. Differential fertility is the total genetic contribution to the next generation. Note that, whereas mutations and genetic drift are random, natural selection is not, as it preferentially selects for different mutations based on differential finesse. Natural selection can be subdivided into two categories: Ecological selection and sexual selection. Ecological Selection occurs when organisms that survive and reproduce increase the frequency of their genes in the gene pool over those that do not survive. Sexual selection occurs when organisms that are more attractive to the opposite sex because of their features reproduce more and thus increase the frequency of those features in the gene pool (Anderson, 1995). Mutations that are not affected by natural selection are called neutral mutations. Their frequency in the population is governed entirely by genetic drift and gene flow. It is understood that an organism's DNA sequence, in the absence of selection, undergoes a steady accumulation of neutral mutations. The probable mutation effect is the proposition that a gene that is not under selection will be destroyed by accumulated mutations. This is an aspect of genome degradation (Knightly and Otto, 2006).

1.2.3 Gene Flow

Gene flow also called migration is any movement of genes from one population to another. Gene flow includes lots of different kinds of events, such as pollen being blown to a new destination or people moving to new cities or countries. If genes are carried to a population where those genes previously did not exist, gene flow can be a very important source of genetic variation. Migration of one population into another area occupied by a second population can result in gene flow. Gene flow operates when geography and culture are not obstacles (Su et al., 2003).

1.2.4 Genetic Drift

Genetic drift or allelic drift is the change in the frequency of a gene variant (allele) in a population due to random sampling. The alleles in the offspring are a sample of those in the parents, and chance has a role in determining whether a given individual survives and reproduces. A population's allele frequency is the fraction of the copies of one gene that share a particular form. Genetic drift may cause gene variants to disappear completely and thereby reduce genetic variation (Masel, 2011)

1.2.5 Haplotype

Haplotype is a set of DNA variations, or polymorphisms, that tend to be inherited together. A Haplotype can refer to a combination of alleles or to a set of single nucleotide polymorphisms (SNPs) found on the same chromosome. Information about Haplotypes is being collected by the International HapMap Project and is used to investigate the influence of genes on disease (The International HapMap Consortium, 2003).

1.2.6 Haplogroup

Haplogroup is a group of similar Haplotypes that share a common ancestor having the same single nucleotide polymorphism (SNP) mutation in all Haplotypes. Because a haplogroup consists of similar Haplotypes, it is possible to predict a haplogroup from Haplotypes. Y-chromosome and mitochondrial DNA Haplogroups have different haplogroup designations. Haplogroups pertain to deep ancestral origins dating back thousands of years (The International HapMap Consortium, 2003).

In human genetics, the Haplogroups most commonly studied are Y-chromosome (Y-DNA) Haplogroups and mitochondrial DNA (mtDNA) Haplogroups, both of which can be used to define genetic populations. According to research, Y-DNA is passed solely along the patrilineal line, from father to son, while mtDNA is passed down the matrilineal line, from mother to offspring of both sexes. It is believed that neither recombines, and thus Y-DNA and mtDNA change only by chance mutation at each generation with no intermixture between parents' genetic material.

1.3 Human Genome Diversity Project (HGDP)

Major demographic events like migration, population bottlenecks and population expansion leave genetic imprints where gene frequency of the genome is altered (Thangaraj et, al., 1998). These imprints are passed onto successive generations thus preserving the population history within the population. In general, human beings group themselves into units in such a way that members between units rarely exchange genes due to cultural and geographical barriers. The Human Genome Diversity Project proposed in early nineties is a combined effort preceded by anthropologists, geneticists, doctors, linguists and other scholars from around the world aims at collecting the blood samples from different ethnic populations throughout the world aiming at building up a representative database of human genetic diversity.

The HGD Project was started internationally on mid-September of 1993 and it has 13 countries participating in it. The reason lying behind selecting only tribes for sampling is that they are believed to have been isolated during an evolutionary time, linguistically

and culturally distinct and are often isolated by geographic barriers and thus prove to be best tools for study (www.hagsc.org/hgdp/).

1.3.1 Goals of Human Genome Diversity Project (HGDP)

- ✓ To identify all the approx 20,000-25,000 genes in human DNA, determination of the sequence of the 3 billion chemical base pair that make up human DNA.
- ✓ In silico storage of all DNA database. Improve tools for data analysis.
- ✓ Transfer related technologies to the private sector. Address the ethical, legal and social issues (ELSI) that may arise from the project.
- ✓ To provide information regarding human biological relationship among different groups and human history.
- ✓ To understand the cause and diagnostics of human diseases.

1.3.2 Benefits and Implications

The project reaps fantastic benefits for human kind, some that we can anticipate and other that surprise us. Generations of biologists and researchers have been provided with detailed DNA information that will be the key to understanding the structure, organization and function of DNA in chromosome. The information from HGDP provides information to clarify the origin and biological relationship of specific human populations and the evolution of human being in particular. The variations of frequencies in various populations can reveal how recently they shared a large pool of common ancestors A group of scientists at Stanford University have collaborated on a large study to understand genetic diversity in human populations. Genomic DNA from 1,043 individuals from around the world was analyzed, determining their genotypes at more than 650,000 SNP loci, with the Illumina Bead Station technology. Genomic DNA samples from these fully-consenting individuals were collected by the Human Genome Diversity Project (HGDP), in collaboration with the Centre Etude Polymorphism Humain (CEPH) in Paris. The collection we tested is referred to as the "HGDP-CEPH Human Genome Diversity Cell Line Panel". They represent 51 different populations from Africa, Europe, the Middle East, South and Central Asia, East Asia, Oceania and the Americas. (www.hagsc.org/hgdp/).

1.4 Approaches Used For Tracing Genetic Diversity

The frequency of occurrence of different Haplotypes can be used to distinguish populations and to shed light on the sub-structures within a population, also for inter and intra population variation studies (Cavalli-Sforza et al., 1994).

The incorporation of MtDNA during the 1980s added a powerful tool to the geneticists' tool kit, since mtDNA does not recombine and is transmitted only through female germ line (Stoneking and Soodyall, 1996).

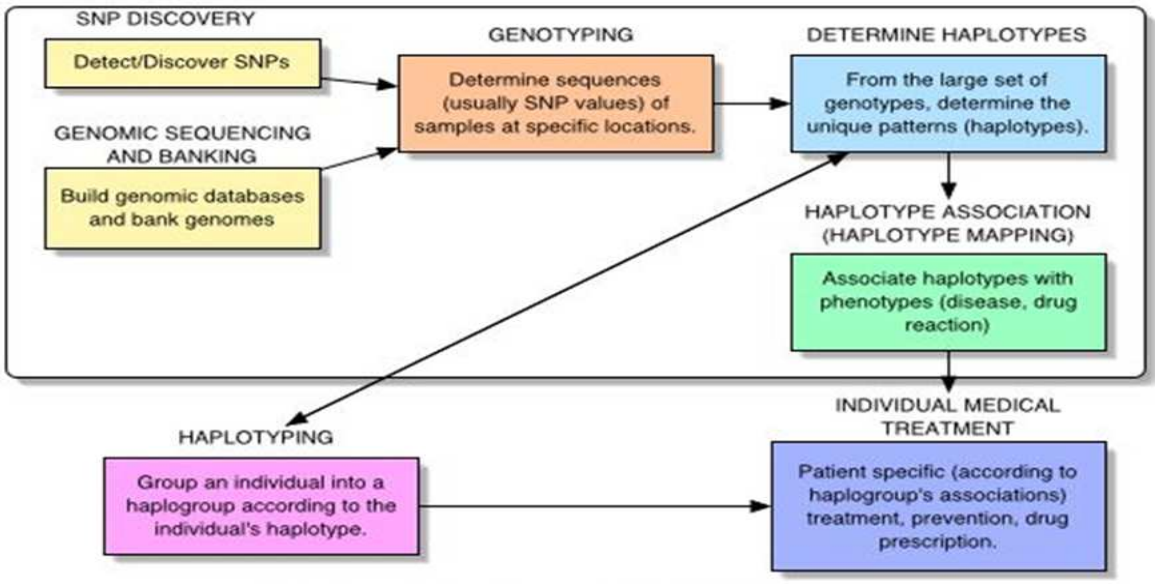


Fig no.1.1 Flowchart of Haplogrouping of Population (Source: www.eupedia.org)

The increasing number of polymorphic markers identified on the Y chromosome has allowed analyzing male lineages, (Hammer and Zegura, 1997). Now day's genome wide analysis of millions of SNPs by using powerful microarray and next generation sequencing used to find out genetic diversity. In this approach Autosomal Gene variation play key role to find out genetic variation among the population.

1.5 Introduction about the Population under Study

Nepal is divided into three regions according to its geopolitical as well as cultural landscape beginning with the mountainous north i.e. Himalayan region, which is located along the southern slopes of the Himalayas and whose people are culturally synonymous to the Tibet. Terai, the southernmost region of Nepal, is predominantly Hindu and consists of low-altitude fertile plains which comprise the northern edge of the Gangetic Plain that extends into North India. Between these two extremes lie the intermediate hills and valleys where a majority of the Nepalese population resides, and whose cultural practices incorporate both Buddhist and Hindu teachings. Nepalese society is the one of the world‘s most diverse and complex consisting about 103 different ethnic group (Nepal census report, 2011). According to historical records, Nepal originally consisted only of Kathmandu Valley, located in the east central hills of present-day Nepal, whose indigenous inhabitants, the Newar, are of postulated Austro-Asiatic, Dravidian, Indo-

Mongoloid and Aryan origin (Regmi, 1969). Today, Newar constitute only 5.5% of the total population and speak the Tibeto-Burman language of Newari or Nepal Bhasa, with the country's official language being the Indo-Aryan language of Nepali, which is closely related to Hindi, the official language of India.

Historically, Nepal was comprised of only the Kathmandu Valley, a region located in the east central hills of present-day Nepal. The Valley witnessed several different waves of migrations attracted by its rich soil, favorable climate, malaria-free zone and trade location (Lewis and Shakya, 1988). Therefore, it is not surprising that the indigenous inhabitants, the Newars, are postulated to be a mixture of Austro-Asiatic, Dravidian, Indo-Mongoloid and Aryan origins (Regmi, 1969). Due to the Valley's strategic position astride trade routes to Tibet, it has a long and distinguished history. Written records (inscriptions) begin in the fifth century A.D. and give evidence of a high and literate civilization derived from the Indian plain. The inscriptions are written in a pure Sanskrit not met with in later periods, but the place-names reveal that the bulk of the population spoke an ancient form of the present day Newars' language, Nepal Bhasa or Newari which is written in Ranjana script. Whereas most of the rest of Nepal remained thinly inhabited and rustic till the modern period, the Kathmandu Valley was able to support a division of labour and a sophisticated urban civilization impossible elsewhere in the Himalayan foothills between Kashmir and Assam.

According to Nepal's 2011 census, the 1,321,933 Newar in the country are the nation's sixth largest ethnic group, representing 5% of total population. 80% of the Newar are Hindu and 20% are Buddhist. There are about 60 castes, on both Buddhist as well as Hindu, covering approximately the same spectrum as the caste system of India.

Bajracharya, Shakya and Udaaya (also called Urey) are Buddhist Newar. Bajracharya belongs to community.

1.5.1 Bajracharya

Bajracharya Means 'Bajra holding priest'. They are also commonly called Guru-ju or Gu-bhaju, the Newari terms related to the Sanskrit term Guru, and translate as 'teacher' or 'priest'. The Bajracharya is the highest ranking of the Newar castes that are born Buddhist. Newar Buddhism is characterized by its extensive and detailed rituals, a rich artistic tradition of Buddhist monuments and artwork like the Chaitya (stupa), Baha and Bahi monastic courtyards, statues, Paubha scroll paintings and Mandala sand paintings, and by being a storehouse of ancient Sanskrit Buddhist texts, many of which are now only extant in Nepal.

1.5.2 Shakya

They are of higher caste among Buddhist Newar equivalent to Bajracharya. The main occupation of Shakya is to make Gold jewelry. They are thought to be migrated from central Asia (Shakya Foundation).

1.5.3 Udaaya (Urays)

Buddhist merchant Newar of Kathmandu in Nepal. They are a prominent community in the business and cultural life of Kathmandu. Udaaya have played key roles in the development of trade, industry, art, architecture, literature and Buddhism in Nepal and the Himalayan region. The Udaaya follow Newar Buddhism and speak Nepal Bhasa as a mother tongue. They are believers in non-violence in personal relations and ritual practices. They are recognized as higher Buddhist caste below Bajracharya and Shakya. They are divided into nine sub-castes that use the following surnames. Each group is associated with a traditional occupation according to the division of labor instituted in the past (Lewis, 1996).

- ✓ Tuladhar (merchant)
- ✓ Sthapit (carpenter, engineer)
- ✓ Tamrakar (worker in copper, gold, silver and brass)
- ✓ Kansakar (worker in bronze, brass and copper)
- ✓ Sikhrākār (roofer)
- ✓ Bania (herbalist)
- ✓ Sindurākār (wood carver)
- ✓ Shilākār (stone carver)
- ✓ Selālikar (confectioner)

1.5.4 Marriage

As being higher caste, Bajracharya only can marry within caste or with higher Buddhist caste such as Shakya, Dhakkhaowa etc. Shakya people also marry with Bajracharya, Dhakkhaowa, and Shakya only. These two populations are endogamous. In case of Udaaya Group, they can marry only within their nine sub-castes. Inter-caste marriage is restricted, if someone do inter-caste marriage, is boycotted from their communities and are prohibited in rituals and cultural programmes. Marriage type is monogamy. Polygamy is not allowed (Allen, 1973).

Fig. 1.2: People of Newar Population [(a) Mr. Karuan Ratna Tuldhar, Pioneer in public Transport, belongs to Udaaya caste (b) a Bajracharya individual (c) Shakya Girl as a Kumari (Considered as Living goddess) (d) a Shakya individual]

1.6 Hypothesis

Newar is the one of the most important ethnic group of Nepal with diverse cultural, linguistic and social heritage, mostly following Hinduism and Buddhism. Their phenotypical characteristics and cultural aspect reveal that they have both Tibeto-Burman and Indo-Aryan heritage. The genetic linkage analysis will provide the information about the genetic imprints of Newar communities and their historical evolutionary prospectus. Within the Newar communities there are several caste having specific phenotype as well as cultural heritage. Even though large numbers of anthropological and sociological studies were done to find out the actual evolution of Newar but exact answer was not found out till date. There resemblance with both Aryan and Mongoloid phenotype suggest that they might have inherited the Haplogroup of both population. Previous studies showed that Newar populations have both haplogroup of Indian subcontinent as well as East Asian. But there cultural and phenotypical diversity within the sub caste may results a different evolutionary pattern. This study will reveal a new insight of the genetic structure and diversity of the Newar communities within their sub-caste and able to traced out the evolutionary and migration pattern.

1.7 Objective

1.7.1 General objective

- ✓ To study the Genetic structure and Genetic diversity of Bajracharya, Shakya and Udaaya Population of Kathmandu Valley.

1.7.2 Specific objectives

- ✓ To collect blood sample and extraction of DNA from the Bajracharya, Shakya and Udaaya male population.
- ✓ To amplify whole MtDNA, Y-Chromosomal marker and Autosomal markers by using PCR and perform sequencing of amplified product.
- ✓ To edit, align and analyze of mtDNA sequence and Y-chromosomal DNA marker and Autosomal DNA marker.
- ✓ To assign the mtDNA haplogroup according to Phylotree.org, Y-chromosomal Haplogroup according to SNP present in Bialleic marker.
- ✓ To analyze the genetic structure, Genetic diversity by using software MEGA V5.1, Alrequin V3.1.1, Network v4, MVSP, SPSS V17.0 etc.

1.8 RATIONALE

Study of genetic structure and diversity of a particular population will help to reveal unanswered quest about the genetic affinity toward different other population, origin, migration and their contemporary genetic makeup. This study adds a new field in the anthropology known as molecular anthropology which gives an instrumentation to study about human history with great accuracy. Such type of study not only helpful for the evolution and origin but also provide deeper insight for the disease pattern which is shared by the population with common genetic structure and history. So, this study will be helpful in the field of Pharmacogenomics.

CHAPTER 2

Review of Literature

2.1 Historical Prospectus of Human diversity and evolution study by using Molecular Technique

At the beginning of the twentieth century, Karl Landsteiner characterized the ABO red blood cell system in 1901. Based on its Mendelian mode of inheritance, provided the first genetic marker for the measure of human genetic variation. The first known study of population genetic variation was survey of ABO blood types by Ludwig and Hanka Hirszfeld's (1919) among military personnel from several different national armies fighting on the Balkan front in World War I. In the following decades, additional human blood groups and hla types were immunologically identified, such as Rhesus (Landsteiner and Levine 1928), revealing allele frequency differences between regional populations and contributing to the formation of a variety of human typologies and classifications that were in some instances misused to support racial ideologies (Pearson et al., 1959; Montagu et al., 1964**).** In the 1960s, many more genetic markers were made available for analysis by the earlier development of protein electrophoresis, a method for separating proteins in an electric field on the basis of their size and charge (Smithies et al., 1955). Most of these proteins represent red blood cell antigens, enzymes and serum proteins, and are collectively referred to as "classical polymorphisms." the first studies of human genetic variation conducted by anthropologists began around this period, such as Frederick Hulse's (1955, 1957) examination of blood variation in native Americans and linguistic barriers to gene flow, and Frank Livingstone's (1958, 1960) research on sickle-cell hemoglobin and malaria. By the 1970s, Anthropological genetics was an active field of research, with an increased focus on the relative effects of evolutionary mechanisms on the patterns of genetic variation within and between local populations and a corresponding shift away from traditional classification and taxonomy of human groups (O'Rourke 2003)**.** This was later complemented by a number of studies examining genetic variation at regional or continental levels, in particular the comprehensive review and multivariate analysis of human classical polymorphisms in the the history and geography of human genes (Cavalli-Sforza et al., 1994), which pioneered the use of synthetic gene frequency maps in order to identify the origins and scope of major human migration events. However, with the advent of new laboratory methods in the 1970s for directly characterizing DNA sequences (i.e., recombinant DNA and sequencing technology), a new class of "molecular markers" began to be increasingly examined by

anthropologists, in particular Restriction Fragment Length Polymorphisms(RFLPs) used in linkage analyses of disease phenotypes (Botstein et al., 1980) and genetic investigations of human origins (Johnson et al., 1983). Molecular markers are more informative than polymorphic blood groups and proteins, with a greater degree of variation generated from an Infinite Sites Model (ISM) of genetic mutation. But it was not until the invention of the polymerase chain reaction (PCR) by Kary Mullis in 1985, a method for rapidly synthesizing many copies of a specific segment of DNA, that researchers were able to cheaply and extensively characterize human genomic variation, and in the process identify a number of genetic systems that are considered well-suited for questions concerning human evolutionary history.

2.2 Human Evolution and Migration

Evolution occurs when there is change in the genes inherited from the parents, in the proportions of different genes in a population. The information contained in genes can change by a process known as mutation. The way particular genes are expressed, how they influence the body or behavior of an organism can also change. Genes affect how the body and behavior of an organism develop during its life, and this is why genetically inherited characteristics can influence the likelihood of an organism‘s survival and reproduction. Evolution does not change any single individual. Instead, it changes the inherited means of growth and development that represent a population. A parent passes adaptive genetic changes to their offspring and ultimately these changes become common throughout a population. As a result, the offspring inherit those genetic characteristics that enhance their chances of survival and ability to give birth, which may work well until the environment changes. Over time, genetic change can alter a species' overall way of life, such as what it eats, how it grows, and where it can live (Stringer, 1994).

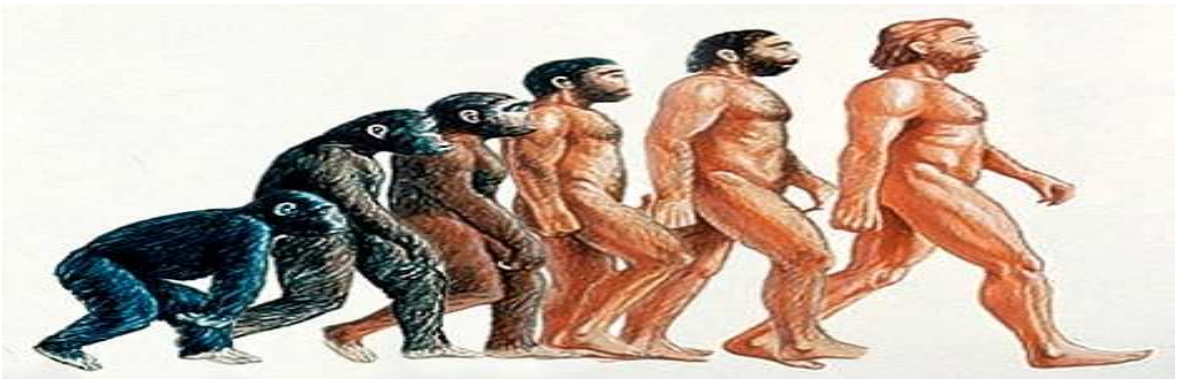


Fig 2.1: Human Evolution Process (Source: www.eupedia.org)

Modern humans are thought to have originated in Africa <200,000 years ago and to have spread throughout the world during the past 100,000 years, but the details of these population movements are poorly understood (Lahr and Foley 1994). Today, all human are classified into a single species called *Homo sapiens sapiens*. In the past there are different species of Hominids: *Homo, Homo habilis*, evolved in East Africa at least Africa <200,000 years ago. *Homo erectus* evolved about <180,000 years ago and by 150,000 years ago had spread throughout the old world (Behar et al., 2008).

2.3 Model of Human Migration

Based on the various evidences available from the past, there are three models of human evolution, which have emerged out of the various research workers in the field.

2.3.1 The Multiregional Model

This model proposes that there was no single geographical origin for humans but that after the radiation of Homo erectus from Africa into Europe and Asia ~0.8-.8 million years before present (YBP), there were independent transitions in regional populations from *Homo erectus* to *Homo sapiens*. The model supported primarily by the continuity of certain morphological traits in the fossil records, which indicates that modern population evolved over a very long period of time in the regions where they are found today. Simultaneously evolution from *Homo erectus* to *Homo sapiens* in dispersed population could have been achieved through extensive gene flow among geographically diverse populations (Smith, 1984; Thorne and Wolpoff, 1992; Wolpoff and Caspari, 2002).

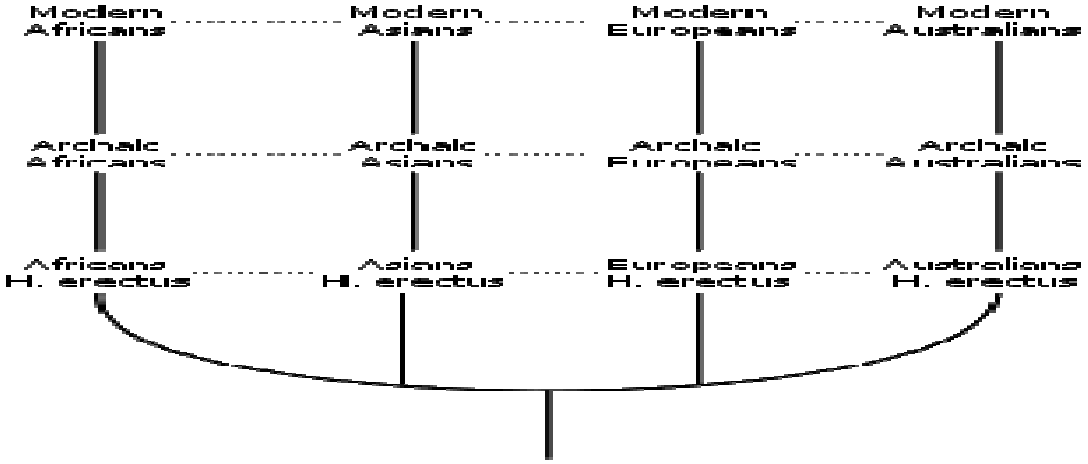


Fig 2.2: Multiregional Model of Human Evolution and Migration (Adapted from: Wolpoff et al., 2000)

2.3.2 The Recent African Origin (Out of Africa)

Genetic studies and fossil evidence show that archaic *Homo sapiens* evolved to anatomically modern humans solely in East Africa, between 200,000 and 150,000 years ago, that members of one branch of *Homo sapiens* left Africa by between 125,000 and 60,000 years ago, and that over time these humans replaced earlier human populations such as Neanderthals and *Homo erectus* (Stringer, 1994).

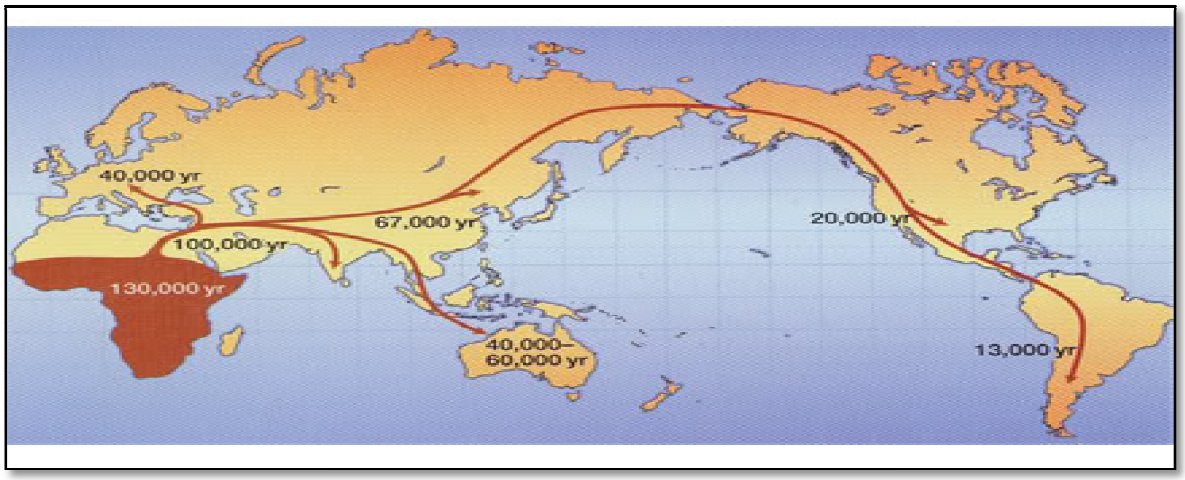


Fig 2.3: Map showing Out of Africa model of Human Migration (www.eupedia.org)

The date of the earliest successful "out of Africa" migration has generally been placed at 60,000 years, although migration out of the continent may have taken place as early as 125,000 years ago according to Arabian archaeology finds of tools in the region. A 2013 paper reported that a previously unknown lineage had been found, which pushed the estimated date for the most recent common ancestor (Y-MRCA) back to 338,000 years ago (Mendez et al., 2013).

2.3.3 The Assimilation (or Hybridization) Model

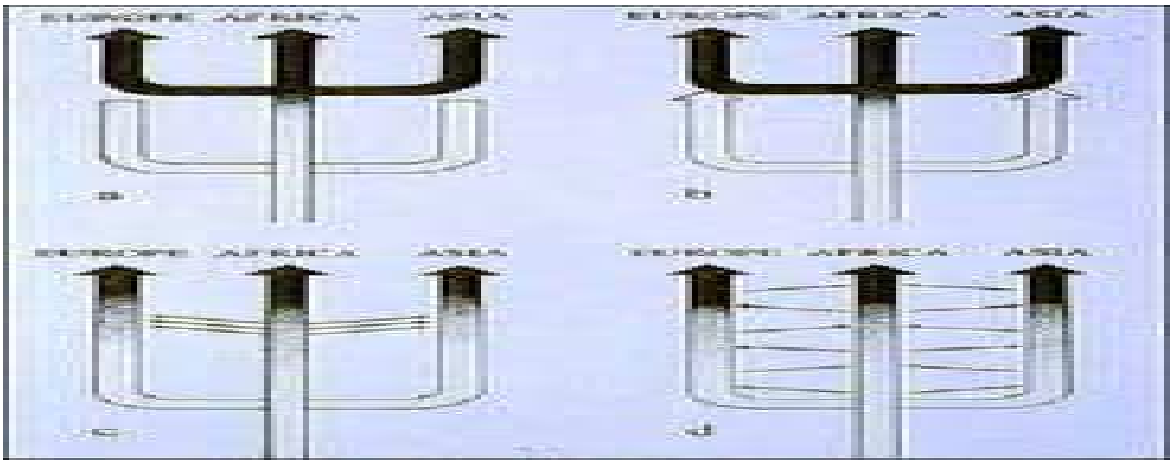


Fig 2.4: The Assimilation or Hybridization Model of Human Evolution and Migration (Source: www.eupedia.org)

It proposes some gene flow between modern humans that migrated from Africa and archaic populations outside Africa. So the evolution of modern humans could have been due to a blending of modern characters derived from African populations with local characteristics in archaic Eurasian populations. The present finding supports "*Out of Africa*" theory (http://anthro.palomar.edu/homo2/mod_homo.htm).

2.4 Markers Used to Study Human Evolution, Migration and Genetic Diversity

2.4.1 Mitochondrial DNA (mtDNA) as a Marker

Human mitochondrial DNA is widely used as tool in many fields including evolutionary biology and population history, medical genetics, genetic genealogy and forensic science. However, most studies of human evolution based on mtDNA sequencing have been confined to the control region, which constitutes less than 7% of the mitochondrial genome (Ruiz-Pesini et al., 2007). The RFLP technique was the most commonly used molecular technique during the 1990s. Today, some researchers still apply the RFLP technique but with finer resolution. However, improvements in rapid sequencing technology have moved researchers toward using rapid sequencing of either coding region or whole mtDNA (Abu-Amero et al. 2008).

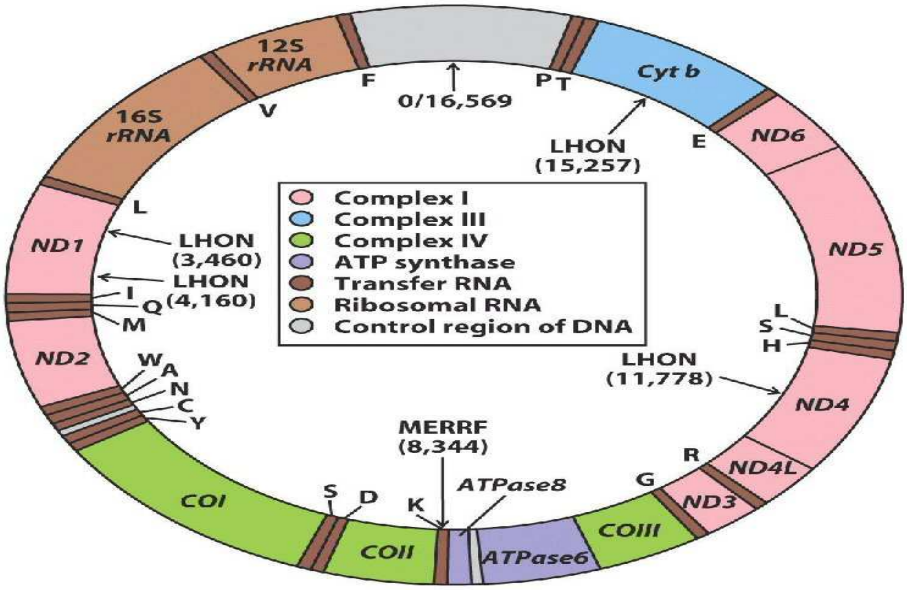


Fig 2.5: Structure of Mitochondrial DNA

The mtDNA contains 37 genes, all of which are involved in the production of energy and its storage in ATP. It encodes 13 mRNAs, 22 tRNAs and 2 rRNAs (Wolstenhome et al., 1992). mtDNA has two strands, a guanine rich heavy (H) strand and a cytosine rich light (L) strand. The heavy strand contains 12 of the 13-polypeptide encoding genes, 14 of the

22 tRNA encoding genes and both rRNA encoding genes. The mtDNA is replicated from two origins. DNA replication is initiated at OH (Origin of heavy chain replication) using an RNA primer generated from the L-strand transcript. H-strand synthesis proceeds two-thirds of the way around the mtDNA, displacing the parental H-strand until it reaches the L-strand origin (OL), situated in a cluster of five tRNA genes. Once exposed on the displaced H-strand, OL folds a stem loop structure and L-strand synthesis is initiated and proceeds back along the H-strand template. Consequently, mtDNA replication is bi-directional but asynchronous (Clayton et al., 1982).

2.4.1.1 Features of Mitochondrial DNA

The analysis of mitochondrial DNA (mtDNA) has been a potent tool in the understanding of human evolution, owing to its characteristics such as:

- ✓ High copy number 1000-10,000 copies per cell (Nass et al., 1969)
- ✓ Apparent lack of recombination as it is a semi-autonomously replicating molecule (Clayton et al., 1982).
- ✓ No repetitive DNA, spacers or introns (Clayton et al., 1982).
- ✓ Maternal mode of inheritance (Giles et al., 1980). So the gene tree is an estimate of the maternal genealogy tells specifically about processes on the female side of the population history.
- ✓ Small size of the molecule and simple genome organization and hence easier to study (Clayton et al, 1982).
- ✓ The mutation rate is not equal across the entire mtDNA molecule; the overall mutation rate in the non-coding control region that carries the genetic signals needed for replication and transcription (bases 16024-576) is about 10 times higher than that of the coding region (bases 577-16023) (Howell et al., 2007).

Since much of this DNA segment is not vital to the survival of the mitochondrion or of the host cell. (Other DNA segments are more vital- mutations could change the nature of the protein formed and gene expression, and therefore mutations could impact the survival of the organism that bears that gene.) By studying the number and variety of base changes within this control region, geneticists can determine the relatedness between individuals. Using the mutation rate within the mitochondrial control region as a "molecular clock," evolutionists can plot the course that hominid evolution has taken place (Ingman et al., 2000).

The rate and pattern of sequence substitutions in the mitochondrial DNA (mtDNA) control region (CR) is of central importance to studies of human evolution. The DNA sequence of the control region is termed hypervariable because it accumulates point mutations at approximately 10 times the rate of nuclear DNA. In the human control

region, the estimates of the rate of substitution were found to range between 2.8 to 5 times the rate of the rest of the mtDNA (Cann et al., 1984). Most of the studies in which control region sequences have been used have focused on intraspecific patterns of variability and Phylogenetic relationships of closely related species, a prominent example being the study of human population history (Cavalli-Sforza et al., 1994). Polymorphic nucleotide sites within this loop are concentrated in two “Hypervariable segments”, HVRI (positions 16024-16383) and HVRII (Wilkinson-Herbots et al., 1996). Hence HVSI and HVSII data can provide useful insights about inter and intra-specific population variations.

Mammalian mitochondrial genes use a slightly different genetic code than nuclear genes, where UGA encode for tryptophan, AUA for methionine, and AGA and AGG for stop codon. Genes take up the majority of the mitochondrial genome (Barrell et al., 1980). A noncoding region of approximately 1,200 nucleotides spans both sides of the arbitrary "0" position of the mitochondrial genome and goes by three terms: control region, D-loop, and hyper variable region. Control region refers to the fact that this region contains the signals that control RNA and DNA synthesis. A single promoter on each DNA strand initiates transcription in each direction, and a single origin initiates replication of each strand. D-loop refers to the early phase of replication, when the first newly synthesized strand displaces one of the parental strands, forming a "bubble" or loop. The DNA sequence of the control region is termed hyper variable, because it accumulates point mutations at approximately 10 times the rate of nuclear DNA (Ingman et al., 2001).

The entire DNA sequence of the human mitochondrial genome 16,569 nucleotides was determined in 1981 (Anderson et, al*.,* 1981) and has been revised (Andrew’s et al*.,* 1999).

2.4.1.2 Female Inheritance

In sexual reproduction, mitochondria are normally inherited exclusively from the mother; the mitochondria in mammalian sperm are usually destroyed by the egg cell after fertilization. Also, most mitochondria are present at the base of the sperm's tail, which is used for propelling the sperm cells; sometimes the tail is lost during fertilization. Paternal sperm mitochondria (containing mtDNA) are marked with ubiquitin to select them for later destruction inside the embryo (Sutovsky et al., 1999).

2.4.1.3 Mitochondrial Eve

Mitochondrial Eve (mt-mrca) is the name given by researchers to the woman who is defined as the matrilineal most recent common ancestor (MRCA) for all currently living humans. Passed down from mother to offspring, her Mitochondrial DNA (mtDNA) is now found in all living humans. However, it is just an assumption. She is believed to have lived about 140,000 years ago in what is now Ethiopia, Kenya or Tanzania (Cann et al., 1987).

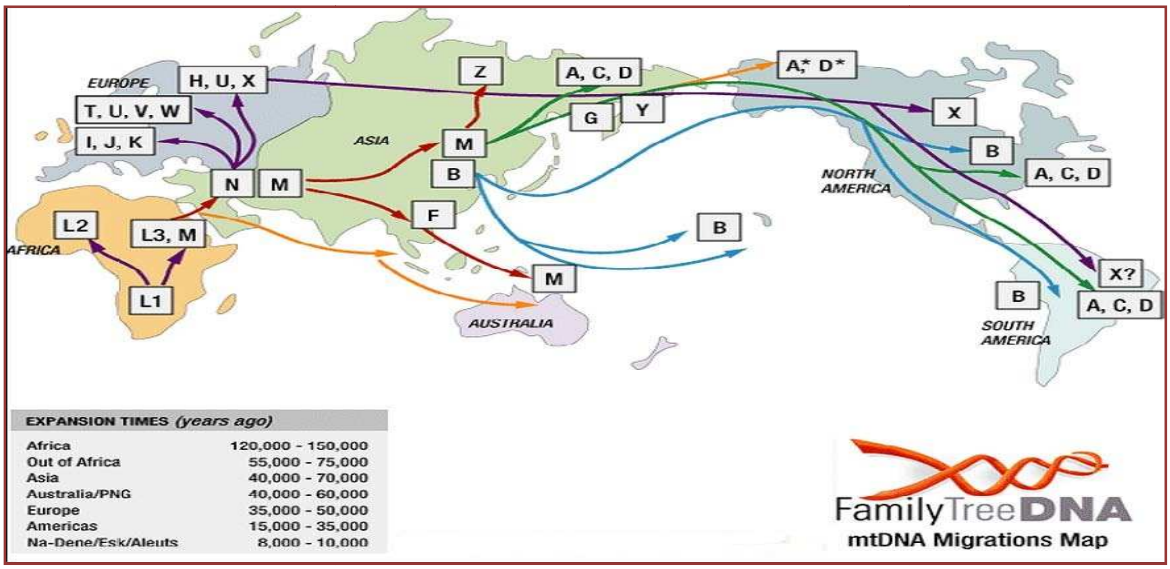


Fig 2.6: Mitochondrial Eve Migration (Source: www.Family treeDNA.com) showing branching of Eve Haplogroup in different descendent Haplogroup migrated a to a particular geography

Haplogroup L3 is an early offshoot from Eve's mitochondrial genetic sequence. L3 appears only in Africa. Around 100,000 years ago two mutations gave rise to two offshoots from L3 the M and N haplogroups from which all subsequent Eurasian lineages are descended 60,000 year before (Quintana-Murciet al. 1999; Mishmar et al., 2003).

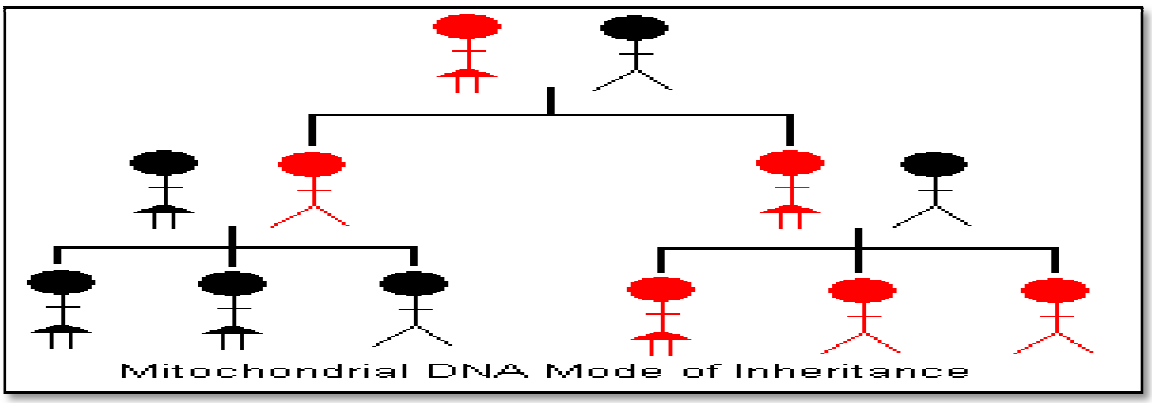
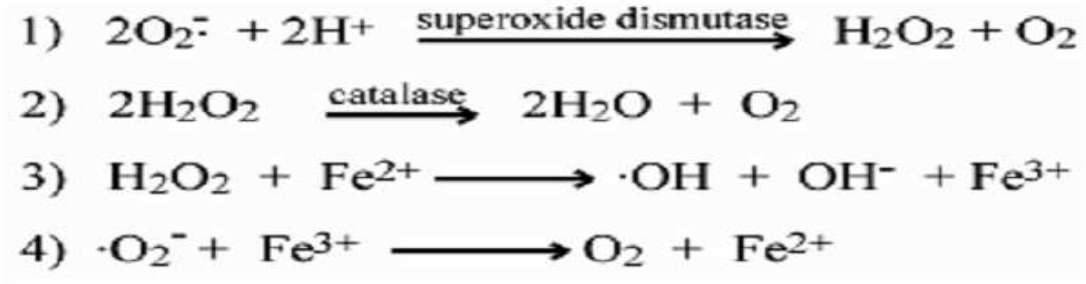


Fig 2.7: Maternal mtDNA inheritance pattern (Source: www.Family treeDNA.com).

2.4.1.4 Chemistry of Mitochondrial Mutation and its role in Human evolution study

The control region is relatively tolerant of a high mutation rate, because binding sites for DNA and RNA polymerase are defined by only short nucleotide sequences (Stoneking et al., 1991). The high mutation rate of mtDNA is almost certainly due to the fact that the mitochondrial genome is located in close proximity to the respiratory machinery of the cell a known source of potent mutagens called oxygen free radicals. The most mutagenic of the reactive oxygen species, hydroxyl (OH) free radical, is generated as a consequence of disabling super oxide to hydrogen peroxide. Hydroxyl radicals react with all types of biologically important molecules - nucleic acids, proteins, sugars, and lipids, producing radicals that undergo further reactions. DNA radicals can react with protein radicals (in Histones) to form cross-links that interfere with chromatin unfolding, DNA repair, replication, and transcription (Starkov, 2008).



Iron ions complex with negatively charged phosphates in the DNA backbone, thus generating hydroxyl radicals within diffusion distance of reactive bonds in deoxyribose and nitrogen bases. Hydroxyl free radicals attack all positions of the deoxyribose sugar, leading to single- and double-stranded breaks in DNA. Hydroxyl radicals also deaminate nucleotides, leading to Point mutations or SNPs - notably C>T, G>C and G>T changes.

The C>T change is termed a transition, because C and T are both pyrimidine nucleotides. The G>C and G>T changes are termed transversions, because a purine nucleotide (G) is converted into a pyrimidine (C or T). The following shows the sequence of events that produce a C>T transition (Ebersberger et al., 2002).

DNA damage by oxygen free radicals suggests an accelerating degradation of mitochondrial function over time. Accumulating mutations in the genes encoding

electron transporters (NADH dehydrogenase, cytochromes, and coenzyme Q) lead to decreased transfer efficiency.

2.4.1.5 Mitochondrial DNA Phylogeny

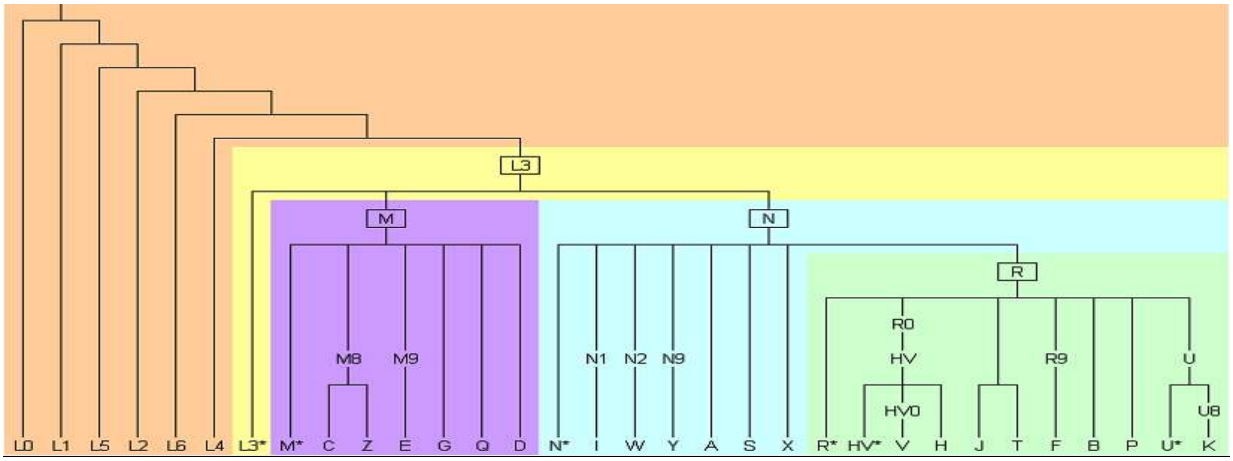


Fig 2.8: Simplified mtDNA phylogeny illustrating the use of alphabetical letters for haplogroup (Source: Palanichamy et al., 2004)

All letters of the alphabet, except O proposed for the designation of a Haplogroup (Palanichamy et al., 2004). The root of the tree is indicated by a star; represent the most recent common matrilineal ancestor of all humans. The Haplogroup L is the most ancient lineages and is African specific indicating the African origin of modern human as well as the out -of-Africa model of migration. Haplogroup L3 gave rise to Macrohaplogroup M, N and R (which is sub clade of N), which encompass all variation observed outside Africa. Nomenclature evolved in such a way that letters C, D, E, G, Q, and Z designate lineages belonging to M; letters A, I, S, W, X, and Y lineages within N; and B, F, HV, H, J, K, P, T, U, and V lineages within R. Haplogroup symbols followed by a star represent all other descendant lineages (besides the ones shown) of a particular clade, for which no unique alphabetical letters were reserved; e.g. N* stands for N5, N12, N13, N14, N21 and N22 (Oven and Kayser, 2008).

2.4.1.6 Maternal Lineages in South Asia

Macrohaplogroup M, which harbors more than 60% of the Indian mtDNA lineage, and to shed light on the origin of its deep rooting Haplogroup based on well-resolved mtDNA Macrohaplogroup M phylogeny, it can be confirmed with the recent studies that, a rapid dispersal of modern human took place in one wave along the Asian coast. HG-M frequency is the highest among tribal and castes groups, particularly in the Austro-Asiatic tribal. Among HG-M individuals, 98.22% belong to sub HG-M* defined by the presence of T at nucleotide 16,223. Sub HG-M2 has the highest

nucleotide diversity in HVS1 indicating that M2 may be the most ancient in India. It occurs in significantly higher ($p < 0.05$) frequencies among tribes (28%) particularly among the AA tribes (32%), than among castes (8.8%). Furthermore, the coalescent time of M2 found in India is estimated to be greater than most East Asian and Papuan branches of HG-M indicating that India was settled early after humankind came out of Africa. These findings imply that the contemporary tribes are descendants of the initial settlers. HG -U is a complex mtDNA lineage whose age is estimated to be 45,000 – 25,000 years. The Phylogenetic structure of major M sub clusters in the Indian subcontinent suggests that the region was settled soon after the African exodus (Thangaraj et al., 2005) and that there has been no extinction or replacement of the initial genetic footprints. The deep roots of M phylogeny clearly as certain the relic of Indian lineages as compared to other M sub lineages suggesting 'in-situ' origin of this sub-haplogroup in South Asia, most likely in India. These deep rooting lineages are not language specific and spread over all the language groups in India. Haplogroup U is common in West Eurasian populations while its three subclades U2a, b and c are present in South Asia and share a deep-rooted coalescence with European founding populations (Kivisild et al., 1999). The pool of mtDNA lineages found in India is partially amalgamated with eastern and western Eurasian mtDNA haplogroup of both ancient and young MRCA (most recent common ancestor) age. Geographically the zone of admixture of West and South Asia n maternal lineages is concentrated towards North-western India (Metspalu et al. 2004). India-specific M sub-haplogroups (i.e. M2, M3, M4, M5, and M6), some autochthonous Haplogroups, including U2a, U2b, U2c, and many unclassified lineages within the nested Macrohaplogroup R and N, have been observed in Indian populations (Palanichamy et al., 2004).

2.4.1.6 Contemporary Studies on Mitochondrial DNA as a tool

The out-of-Africa scenario provided evidence for the precise route by which modern humans left Africa. Two major routes of dispersal have been hypothesized: one through North Africa into the Levant documented by fossil remains, and one through Ethiopia along South Asia, for which little, if any, evidence exists. The geographic distribution and variation of mtDNA can be highly informative in defining potential range expansions and migration routes in the distant past. In 1987, a group of researchers presented data that indicated a recent African origin of modern humans occurred between 140 and 280 thousand years ago (Cann et al., 1987). This study was based on restriction fragment length polymorphism (RFLPs) variation among the mt-DNA of 147 humans from five geographic populations, including Africa, Asia, Australia, Caucus, and New Guinea. Sequencing the control region for use in population genetics has been utilized because

of the high degree of polymorphism that can be examined in this relatively short sequence. Recent studies using mt-DNA sequencing of HVS-I and HVS-II support the Recent Out-of-Africa model (Gunz et al., 2009; Ingman et al., 2000).

Migration routes are influenced by landscape and, consequently, natural passageways acting as corridors played a key-role in the history of human migrations, as exemplified by the Nile River Valley, which connects Africa and Eurasia (Krings et al., 1999), central Asia, which connects west and east Eurasia (Karafet et al., 2001; Wells et al., 2001; Zerjal et al., 2002), and eastern Indonesia, which connects Southeast Asia and Australia/Oceania (Kayser et al., 2000).

The mitochondrial haplogroup M, first regarded as an ancient marker of East-Asian origin, has been found at high frequency in India and Ethiopia, raising the question of its origin. Its variation and geographical distribution suggest that Asian haplogroup M separated from eastern-African haplogroup M more than 50,000 years ago. Two other variants (489C and 10873C) also support a single origin of haplogroup M in Africa. These findings, together with the virtual absence of haplogroup M in the Levant and its high frequency in the South-Arabian peninsula, render M the first genetic indicator for the hypothesized exit route from Africa through eastern Africa/western India. This was possibly the only successful early dispersal event of modern humans out of Africa. (Elson et al., 2001).

Macrohaplogroup M emerged from African macrohaplogroup L3, originated at 60-70 KY ago. 60% of Indian population belongs to this particular haplogroup **(**Quintana-Murciet et al., 1999; Mishmar et al., 2003). Only M1 haplogroup is found in Ethiopia. All the other branches of this macrohaplogroup including M*, C, D, G, E and Z Haplogroups are observed in Asia The lineages M2, M3, M4, M5, M6, M18 and M25 are exclusive to India, with M2 reported to be the most ancient lineage in the sub-continent with an age estimation of 60,000 yrs-75,000 yrs. M7, M8a, M8C, M8Z, M9, E, D, G has been extensively found in East Asian population. Furthermore, the frequencies of these clades among the different geographic, linguistic phyla and social strata have been investigated in detail, yet the fundamental question regarding origin of this super-haplogroup still remains unanswered. While some authors have suggested a southwest Asian origin of M Macrohaplogroup, followed by a back migration to Africa, others support its African ancestry (Rajkumar et al., 2005).

On the basis of complete mtDNA sequencing of 641 individual from 26 tribal population of India, 12 new Haplogroups - M53 to M64 along with Haplogroups M2, M3, M4, M5, M6, M89C9Z, M9, M10, M11, M12-G, D, M18, M30, M33, M35, M37, M38, M39, M40, M41, M43, M45 and M49 were found, which were previously described by control

and/or coding-region polymorphisms. Results indicate that the mtDNA lineages reported in the present study (except East Asian lineages M89C9Z, M9, M10, M11, and M12-G, D) are restricted to Indian region. The deep rooted lineages of Macro-Haplogroup 'M' suggest in-situ origin of these Haplogroups in India and such deep rooting lineages are represented by multiple ethnic/linguist groups. AMOVA shows substantial subdivisions among the tribes (FST= 0.16164). The current Indian mtDNA gene pool was shaped by the initial settlers and was galvanized by minor events of gene flow from the east and west to the restricted zones. Northeast Indian mtDNA pool harbors region specific lineages, other Indian lineages and East Asian lineages. It was found that establishment of an East Asian gene in North East India through admixture rather than replacement. This study is crucial to construct maternal phylogeny and prehistoric dispersals of modern human being in the Indian sub continent (Chandrasekar et al., (2009).

Analysis of HSVI mtDNA sequence variation and 14 Biallelic and five short tandem–repeat Y-chromosome markers of 192 northeast Indians revealed that there is no mtDNA and Y-chromosome admixture between northeast and other Indian groups. Both northeast Indian mtDNAs and Y chromosomes consistently show strikingly high homogeneity among groups and strong affinities to East Asian groups. This is also characterized by a greatly reduced Y-chromosome diversity, which contrasts with extensive mtDNA diversity. The male founder effect was estimated 4,000 YBP during the colonization of northeast India. Therefore, it is concluded that the northeast Indian passageway acted as a geographic barrier rather than as a corridor for human migrations between the Indian subcontinent and East/Southeast Asia, at least within the past millennia and possibly for several tens of thousands years, as suggested by the overall distinctiveness of the Indian and East Asian Y chromosome and mtDNA gene pools (Cordaux et al., 2004).

Study of 127 unrelated male individual from the Dominican Republic by using high-resolution RFLP using restriction endonucleases (AluI, AvaII, BamHI, DdeI, HaeII, HaeIII ,HincII, HinfI, HpaI, MspI, MboI, RsaI, and TaqI) and control-region sequencing revealed that 37% individual harbored L2 haplotypes and remainder belonged to other known African (L1, L3b, L3d, L3e, L3*, and U6), American Indian (A, B, C, and D), and western Eurasian (J and U2) haplogroups. The phylogeny of the L2 complete sequences showed that the two mtDNAs from L2b and L2d seem disproportionately derived compared with those from L2a and L2c. The pattern of nonsynonymous versus synonymous substitutions hints at a role for selection in the evolution of human mtDNA. Regardless of whether selection is shaping the evolution of modern human mtDNAs, the population screening of L2 mtDNAs for the mutations identified by complete sequence study should allow the identification of marker motifs of younger age with more restricted geographic

distributions, thus providing new clues about African prehistory and the origin and relationships of African ethnic groups (Torroni et al., 2001).

One of important polymorphism which is population specific is a deletion of one copy of a 9-bp tandem repeat sequence (CCCCCTCTA) within the non-coding region V. Analysis to date suggests the latter is a valuable anthropological marker for peoples of East Asian origin (Wrischnik et al., 1987). The 9-bp deletion mutation is very common in Asians and populations of Asian ancestry (e.g., Polynesians and Amerinds), and it often has been considered an ethnic specific polymorphism for these populations (Horai and Matsunaga 1986; Hertzberg et al., 1989; Schurr et al., 1990). The 9-bp deletion between the COII and tRNALYS genes is due to the lack of one of two adjacent copies of a 9-bp sequence (5' CCCCCTCTA 3') usually present in human mtDNA (Anderson et al., 1981). Analysis of 500 individual from tribal population of south Asia, 9bp deletion in some Indian tribal population such as Yandis, Siddhis,Maria,Gonds, Irulas, Santhals,Khonda,Dora and Jalaris was found. This suggests that independent origin of 9 bp in some Indian population. Other haplogroup having 9 bp deletions most likely had African and Asian origins implying multiple origin of 9 bp deletion in the south Indian tribes (Watkins et.al., 1999).

Large majority of the mtDNAs harboring the 9-bp deletion constitute a homogeneous group of mtDNA defined by the T-to-C transitions at nucleotide positions (nps) 16189 and 16217 in the D-loop and the lack of the characteristic mutations in the coding regions that define other mtDNA Haplogroups, thus indicating that the 9-bp deletion in those mtDNA is due to a single ancestral mutational event. This mtDNA group has been termed "haplogroup B". (Torroni et al., 1992). However, the same type of high-resolution studies also revealed a very limited number of mtDNA harboring the 9-bp deletion in association with mutations that characterize other mtDNA Haplogroups and that are absent in haplogroup B mtDNA. This finding demonstrates that, in addition to the event that occurred on the ancestral haplogroup B mtDNA, the 9-bp deletion has occurred at other times during human evolution. Unfortunately, because Asians or Asian-derived populations were the most studied, these multiple mutational events were always observed in these populations, generating the impression that the 9-bp deletion events were somehow specific for those populations. However, the 9-bp deletion has been reported recently also in African Pygmies. High-resolution haplotype analysis has revealed that the mtDNAs with the 9-bp deletion are members of an African-specific haplogroup and that the deletion observed in the Pygmies occurred in Africa, independently from those observed in Asia (Chen et al., 1995).

Analysis of mtDNA control region and coding region sequence variation in 98 Altaian Kazakhs from southern Siberia and 149 Barghuts from Inner Mongolia, China revealed that the strong affinity of all Mongolic-speaking populations (Mongolians, Buryats,

Khamnigans, Kalmyks as well as Barghuts) and Turkic speaking Sojots, suggesting their origin from a common maternal ancestral gene pool. Complete mtDNA sequence of 55 samples represent haplogroups R11b, B4, B5, F2, M9, M10, M11, M13, N9a and R9c1. The finding confirms that northern Asian maternal gene pool consists of predominantly post-LGM components of eastern Asian ancestry, though some genetic lineages may have a pre-LGM/LGM origin (Derenko et. al., 2012).

The HVSI/II sequences analysis of mtDNA from 120 individual and 13 whole mt genome of Saudi Arabian population showed greatest similarity to other Arabian Peninsula populations (Bedouin from the Negev desert and Yemeni) and to Levantine populations. Nearly all the main West Eurasian Haplogroups (M,N,H, J, T, K, U) were detected in the Saudi population, including the rare U9 clade. Saudi Arabs had only a minority sub-Saharan Africa component (7%), similar to the specific North African contribution (5%). A small Indian influence (3%) was also detected. The majority of the Saudi-Arab mitochondrial DNA lineages (85%) have a western Asia provenance. Although the still large confidence intervals, the coalescence and phylogeography of (preHV) 1 haplogroup (accounting for 18 % of Saudi Arabian lineages) matches a Neolithic expansion in Saudi Arabia. There result gives the good support of the out Africa migration model (Abu-Amero et al., 2007).

By Sequencing of the HVSI/II regions of Forensic mitochondrial DNA (299 Swedish individuals, 179 different Haplotypes were detected. The genetic diversity was estimated to be 0.9895 (±0.0023), and the random match probability was 1.39 %. The most abundant Haplogroups were HV (including its subhaplogroups H andV) with a frequency of 46.5%, followed by haplogroup U(including its subhaplogroups K) at 27.8 %, haplogroup T at 10.0 % and haplogroup J at 7.0 %, a distribution that is consistent with previous observations in other European populations (Lembring et.al., 2013).

Analysis of high resolution uniparental (mtDNA and Y chromosome) and biparental Autosomal genetic markers among aboriginal Bangladesh tribes(Chakma, Marma and Tripura) speaking Tibeto-Burman language showed high homogeneity among themselves and strong affinities to Northeast Indian Tibeto-Burman groups. However, they show substantially higher molecular diversity than Northeast Indian populations. Unlike Austroasiatic (Munda) speakers of India, observed equal role of both males and females in shaping the Tibeto-Burman expansion in Southern Asia. Moreover, it is noteworthy that in admixture proportion, Tibeto-Burman populations of Bangladesh carry substantially higher mainland Indian ancestry component than Northeast Indian Tibeto-Burmans. Largely similar expansion ages of two major paternal Haplogroups (O2a and O3a3c), suggested that they arose before the differentiation of any language group and approximately at the same time. Contrary to the scenario proposed for colonization

of Northeast India as male founder effect that occurred within the past 4,000 years, it was suggest a significantly deep colonization of this region (Gazi et al., 2013).

South Asia harbors one of the highest levels genetic diversity in Eurasia, which could be interpreted as a result of its long-term large effective population size and of admixture during its complex demographic history. Genome wide data for more than 600,000 SNP markers genotyped in 142 samples from 30 ethnic groups in India and comparing results with other available genome-wide data, Indian populations are characterized by two major ancestry components, one of which is spread at comparable frequency and Haplotype diversity in populations of South and West Asia and the Caucasus. The second component is more restricted to South Asia and accounts for more than 50% of the ancestry in Indian populations. Haplotype diversity associated with these South Asian ancestry components is significantly higher than that of the components dominating the West Eurasian ancestry palette. Modeling of the observed Haplotype diversities suggests that both Indian ancestry components are older than the purported Indo-Aryan invasion 3,500 YBP. Consistent with the results of pair wise genetic distances among world regions, Indians share more ancestry signals with West than with East Eurasians (Metsaplu et al., 2011).

Autosomal, Mitochondrial, and Y-chromosomal DNA study on the 141 unrelated Maldivians individual results 63 different mtDNA Haplotypes that could be allocated to 29 mtDNA Haplogroups, mostly within the M, R, and U clades, 66 different Y-STR Haplotypes in 10 Y-chromosome Haplogroups, predominantly H1, J2, L, R1a1a, and R2. Parental admixture analysis for mtDNA- and Y-haplogroup data indicates a strong genetic link between the Maldive Islands and mainland South Asia. Paternal admixture from West Asia is detected, but cannot be distinguished from admixture from South Asia. Within the Maldives, they find a subtle genetic substructure in all marker systems that were not directly related to geographic distance or linguistic dialect. They found reduced Y-STR diversity and reduced male-mediated gene flow between atolls, suggesting independent male founder effects for each atoll. Detected reduced female-mediated gene flow between atolls confirms a Maldives-specific history of matrilocality (Pijpe et al., 2013).

2.5 Y-chromosome as a Marker

The human Y chromosome is approximately 60 Mb, linear molecule that determines maleness. It is an unusual segment of the human genome since, apart from two small regions in which pairing and exchange take place with the X chromosome, it is male specific and haploid and escapes from recombination. These unique properties of the Y

chromosome have important consequences for its mutation processes, its genes and its population genetics. Y chromosome passes down from father to son, largely unchanged, except by the gradual accumulation of mutations. Different populations often have characteristically different Y chromosome and these studies are likely to make a major contribution to our understanding of the origin of modern humans (Jobling and Smith, 2000). By examining the difference between polymorphic Y-chromosomal markers one can attempt to reconstruct a history of human paternal lineages, population structure and history, genealogy, forensics and the investigation of selective influences in the Y chromosome 95% of the Y chromosome has become a genetic junkyard because it does not recombine. In the Y-chromosome's passage through the generations, changes occur randomly in its junk DNA and so the Y-chromosome of the contemporary populations retains a record of their passage through time. They can reveal the paternal genealogy of their owners and the relationships between different groups of individuals (Bradman and Thomas, 2004)

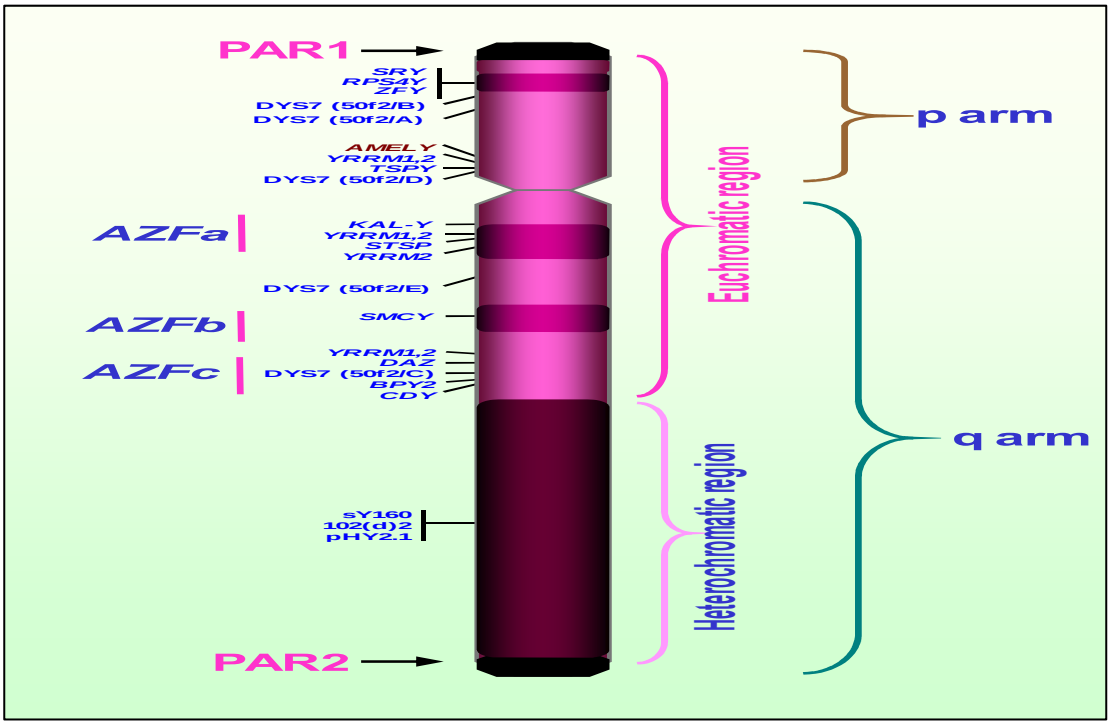


Fig2.9: Y-Chromosome of Human showing different polymorphic region

2.5.1: Features of Y-Chromosome

The Y chromosome has been a potent tool for studying human evolution owing to following characteristics:

- Paternal mode of inheritance as it passes from father to son and thus escapes meiotic recombination.

- Only 3Mb of its length undergoes recombination and thus also referred as non-recombining majority or NRY.
- Haplotypes pass intact from generation to generations and change only by mutation.
- Lower sequence diversity than elsewhere in nuclear genome.
- Using binary polymorphism such as SNPs a unique phylogeny can thus be constructed.
- More susceptible to genetic drift, a useful property for investigating past events.
- Geographical clustering is further influenced by the behavior of men, bearers of Y-chromosome.

2.5.2 Y Chromosomal Changes

Changes that do occur from generation to generation are of four types:

2.5.2.1 Indels

Indels are Insertions or deletions of the DNA at particular locations on the chromosome. One insertion particularly useful in population studies is the YAP, which stands for "Y chromosome Alu polymorphism. Alu is a sequence of approximately 300 bp, which has inserted itself into a particular region of the DNA. There have been some half a million alu insertions in human DNA; YAP is one of the more recent (Hammer, 1994).

2.5.2.2 Snips

They are "single nucleotide polymorphisms" in which a particular nucleotide (an A for example) is changed (perhaps into a G). Stable indels and snips are relatively rare and, in the case of the latter, so infrequent that it is reasonable to assume they have occurred at any particular position in the genome only once in the course of human evolution. Snips and stable Alu's have been termed "unique event polymorphisms" (UEPs)

2.5.2.3 Microsatellites

Are short sequences of nucleotides (such as GATA) specific number of repeats in a particular variant (or allele) usually remains unchanged from generation to generation but changes do sometimes occur and the number may increase or decrease.

It is usually assumed that increases or decreases in the number of repeats take place in single steps, for instance from nine repeats to ten, but whether decreases in number are as common as increases has not been established. Changes in microsatellite lengths occur much more frequently than new UEP arise. What is more, while we can reasonably assume that a UEP has arisen only once, the number of repeat units in a microsatellite may have changed many times along a paternal lineage.

2.5.2.4 Minisatellite

They the repeated sequences are short (often no more than 3 or 4 nucleotides), in minisatellites they are normally 10-60 base pairs long and the number of repeats often extends to several dozen. Changes during the copying process take place more frequently in minisatellites than in microsatellites and the mechanisms may be different in the two cases

In using polymorphisms to study changes over time, there are lots markers, which change at different rates. Perhaps we can think of the UEPs as the hour, the microsatellite polymorphisms as the minute and the Minisatellite as a sweep second of the evolutionary clock. Because most of the Y chromosome does not exchange DNA with a partner, a further benefit of using it to study evolution is that all the markers are joined one to another along its entire length. Such linkage of markers means that a Haplotype constructed from a number of different markers records the evolutionary history of the particular Y chromosome on which they are all located.

Many polymorphic loci scattered over the entire non-recombining part of the Y-chromosome can be identified. Among these polymorphisms, Biallelic markers with a low mutation rate representing unique mutation events (UMEs) in human evolution, such as single base-pair substitutions (Underhill et al., 1997), Alu insertion/ deletion polymorphisms (Hammer, 1994); moderately fast evolving micro satellites or short tandem repeats (STRs) with an average frequency of around 2% per generation (Heye et, al,. 1997; Jobling et, al,. 1999; Kayser et, al,. 2000).

2.5.3 Y chromosome Adam

In human genetics, Y-chromosomal Adam (Y-MRCA) is the patrilineal human most recent common ancestor (MRCA) from whom all Y chromosomes in living men are descended. By analyzing the Y-chromosome DNA from males in all regions of the world, geneticist Spencer Wells has concluded that all humans alive today are descended from a single man who lived in Africa around 60,000 years ago.

Y-chromosomal Adam was the only living male of his time, but he probably co -existed with a large population of human males. None of Y-chromosomal Adam's contemporaries, however, have a direct male line to the present day. Either their lines died out, or they had at least one generation within each line that did not produce sons. The genetic material making up the non-recombining portion of the Y chromosome is effectively haploid and inherited in a patrilineal manner. Therefore, polymorphisms in this region of the nuclear genome are valuable for investigating male-mediated gene

flow and for complementing maternally based studies of mtDNA . A simple polymorphism used for human diversity is insertion of an Alu element on the long arm of the Y chromosome (Persichetti et al., 1992; Hammer 1994; Spurdle et al. 1994)**.** Hammer (1994) referred to as the Y Alu polymorphic (YAP) element, is present at a specific site on the Y chromosome. The frequency of Y chromosomes containing the YAP element

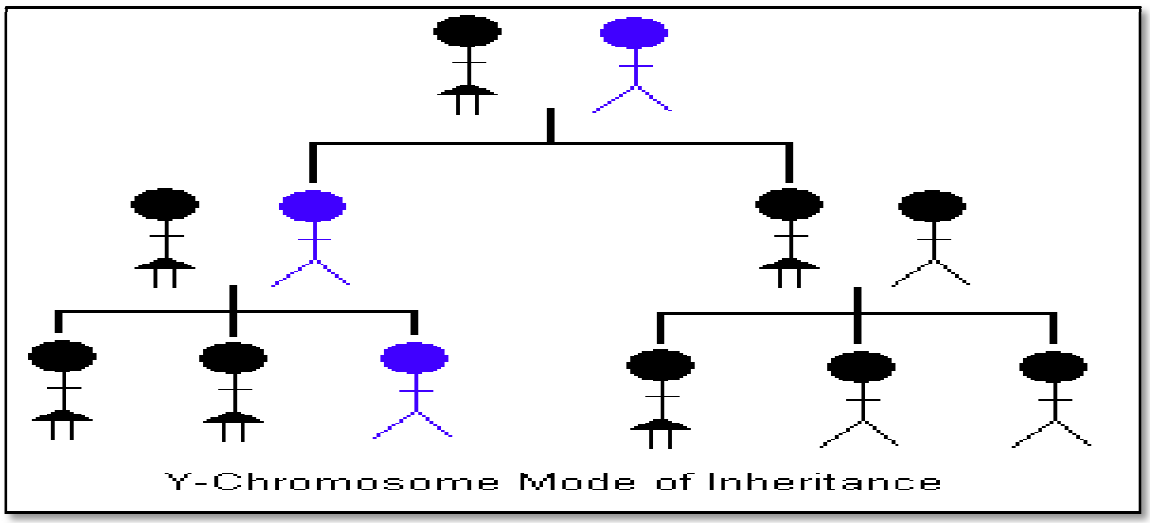


Fig.2.10: Paternal Inheritance of Y-Chromosome (Source: eupedia.org)

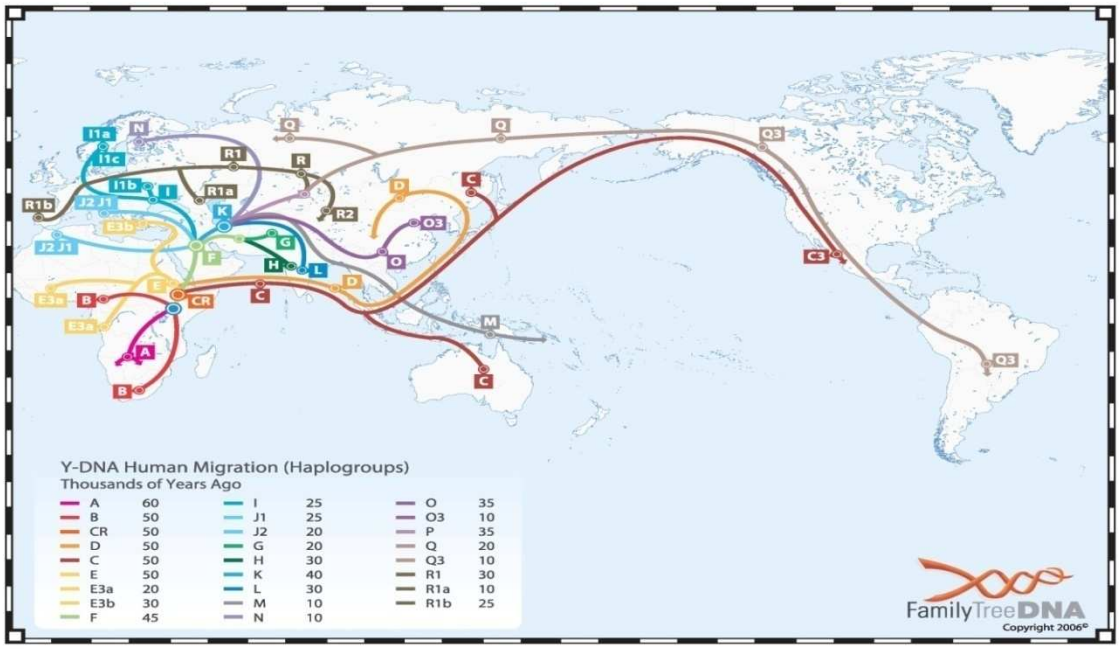


Fig 2.11: Migration route of Y-cromosomal Adam (Source: www.familytreeDNA.com)

(YAP') is highest in sub-Saharan African populations, followed by North African and European populations. Most Asian populations examined so far completely lack the YAP element. An exception to this pattern was the discovery of the YAP element in a small

sample of Japanese subjects (42% and completely absent in Taiwanese). The distribution of The Y Alu polymorphic (YAP) element is confirming the irregular distribution of this polymorphism in Asia (Hammer 1994; Hammer and Horai 1995). This polymorphism raises the possibility of tracing paternal lineages and male-mediated gene flow between largely separated geographical regions.

2.5.4 Some contemporary study on Y-Chromosome from Human genetic Diversity

Y-chromosomal haplogroup also used to trace the entrance of a population from certain geographical region. Analysis of the Y-chromosome diversity of the two Sino-Tibetan populations (Luoba and Deng) from eastern Himalayas region, which is located near the southern entrance through which early modern humans expanded into East Asia, was performed. The Luoba possessed haplogroups D, N, O, J, Q, and R, indicating gene flow from Tibetans, as well as the western and northern Eurasians. The Deng exhibited haplogroups O, D, N, and C, similar to most Sino-Tibetan populations in the east. Short tandem repeat (STR) diversity within the dominant haplogroup O3 in Sino-Tibetan populations showed that the Luoba are genetically close to Tibetans and the Deng are close to the Qiang. The Qiang had the greatest diversity of Sino-Tibetan populations, supporting the view of this population being the oldest in the family. The lowest diversity occurred in the eastern Himalayas, suggesting that this area was an endpoint for the expansion of Sino-Tibetan people. Thus, we have shown that populations with haplogroup O3 moved into the eastern Himalayas through at least two routes (Kang et al., 2011).

A study carried on Bosnia-Herzegovina population by using variation at 28 Y-chromosome biallelic markers, analysed in 256 males (90 Croats, 81 Serbs and 85 Bosniacs). It was revealed that these populations shared high frequency of the “Palaeolithic” European-specific haplogroup I, a likely signature of a Balkan population re-expansion after the Last Glacial Maximum. This haplogroup is almost completely represented by the sub-haplogroup I-P37 whose frequency is, however, higher in the Croats (∼71%) than in Bosniacs (∼44%) and Serbs (∼31%). Other frequent haplogroups are E (∼15%) and J (∼7%), which are considered to have arrived from the Middle East in Neolithic and post-Neolithic times, and R-M17 (∼14%), which probably marked several arrivals, at different times, from eastern Eurasia. Hg E, almost exclusively represented by its subclade E-M78, is more common in the Serbs (∼20%) than in Bosniacs (∼13%) and Croats (∼9%), and Hg J, observed in only one Croat, encompasses ∼9% of the Serbs and∼12% of the Bosniacs, where it shows its highest diversification. By contrast, HG R-M17 displays similar frequencies in all three groups. On the whole, the three main

groups of Bosnia-Herzegovina, in spite of some quantitative differences, share a large fraction of the same ancient gene pool distinctive for the Balkan area (Marjanovic et al., 2005).

In order to investigate the genetic consequences of caste system, which has persisted in Indian Hindu society from 3,500 years, 131 Y-chromosomal binary markers and 16 microsatellites used to analyze male-lineage variation in a sample of 227 Indian men of known caste, 141 from the Jaunpur district of Uttar Pradesh and 86 from the rest of India. It was found striking evidence for male substructure: in particular, Brahmins and Kshatriyas (but not other castes) from Jaunpur each show low diversity and the predominance of a single distinct cluster of Haplotypes. These findings confirm the genetic isolation and drift within the Jaunpur upper castes, which are likely to result from founder effects and social factors. In the other castes, there may be either larger effective population sizes, or less strict isolation, or both (Jerjal et al., 2006).

To trace the origin and historic expansion of Austro-Asiatic groups of India, Y-chromosome SNP and STR data of the 1222 individuals from 25 Indian populations was analysed, covering all the three branches of Austro-Asiatic tribes, viz. Mundari, Khasi-Khmuic and Mon-Khmer. Finding was compared with previously published data on 214 relevant populations from Asia and Oceania. Results suggest a strong paternal genetic link, not only among the subgroups of Indian Austro-Asiatic populations but also with those of Southeast Asia. However, maternal link based on mtDNA is not evident. The results also indicate that the haplogroup O-M95 had originated in the Indian Austro-Asiatic populations ~65,000 yrs BP (95% C.I. 25,442 – 132,230) and their ancestors carried it further to Southeast Asia via the Northeast Indian corridor. Subsequently, in the process of expansion, the Mon-Khmer populations from Southeast Asia seem to have migrated and colonized Andaman and Nicobar Islands at a much later point of time. These find have good agreement with linguistic study of such population (Kumar et. al., 2007).

India's vast coastal rim played an important role in the dispersal of modern humans out of Africa but the Karnataka state, which is located on the southwest coast of India, remains poorly characterized genetically. High-resolution analyses of Y-chromosome single nucleotide polymorphisms (Y-SNPs) and 17 short tandem repeat (Y-STR) studied on two Dravidian populations, namely Lingayat (N=101) and Vokkaliga (N=102), who represent the two major communities of the Karnataka state, revealed that the majority of the Lingayat and Vokkaliga paternal gene pools are composed of four Y-chromosomal haplogroups (H, L, F* and R2) that are frequent in the Indian subcontinent. The high level of L1-M76 chromosomes in the Vokkaligas suggests an agricultural expansion in the region, while the predominance of R1a1a1b2-Z93 and J2a-M410

lineages in the Lingayat indicates gene flow from neighboring south Indian populations and West Asia, respectively. Lingayat (0.9981) also exhibits a relatively high haplotype diversity compared to Vokkaliga (0.9901), supporting the historical record that the Lingayat originated from multiple source populations. In addition, we detected ancient lineages such as F*-M213, H*-M69 and C*-M216 that may be indicative of genetic signatures of the earliest settlers who reached India after their migration out of Africa (Chennakrishnaiah et al., 2013).

Various genetic data such as Y-chromosomal, MtDNA of genome wide- Autosomal SNPs showed that there is some extent of presence of Central South Asian as well as West Eurasian genetic component in mainland East Asia. High-resolution genotyping of Y chromosome phylogeny of 3,826 male individual showed that there are four dominant haplogroups(total of 93.21%); O-M175, D-M174, C-M130, and N-M231, in both southern and northern East Asian populations, which is consistent with the proposed southern route of modern human origin in East Asia. However, there are other haplogroups (6.79% in total) (E-SRY4064, C5-M356, G-M201, H-M69, I-M170, J-P209, L-M20, Q-M242, R-M207, and T-M70) detected primarily in northern East Asian populations and were identified as Central-South Asian and/or West Eurasian origin based on the phylogeographic analysis. In particular, evidence of geographic distribution and Y chromosome short tandem repeat (Y-STR) diversity indicates that haplogroup Q-M242 (the ancestral haplogroup of the native American-specific haplogroup Q1a3a-M3) and R-M207 probably migrated into East Asia via the northern route. The age estimation of Y-STR variation within haplogroups suggests the existence of postglacial (; 18 Ka) migrations via the northern route as well as recent (; 3 Ka) population admixture. This studied reveal that although the Paleolithic migrations via the southern route played a major role in modern human settlement in East Asia, there are ancient contributions, though limited, from WE, which partly explain the genetic divergence between current southern and northern East Asian populations (Zhong et. al., 2011).

Analysis of 32 Y-chromosome SNPs and 17 Y-STRs in 607 males from nine populations (Munda, Birhor, Oraon, Paharia, Santhal, Ho, Lachung, Mech and Rajbanshi) residing in East and Northeastern India revealed high frequency of the O2a haplogroup in Austroasiatic tribes and high haplotype diversity within specific haplogroups demonstrating a lesser degree of admixture of these populations with neighbouring populations in eastern India. In addition, the presence of O3a haplogroups in Sino-Tibetan populations reflects the influx from Southeast Asia during the demographic expansion through the Northeastern corridor. The study suggested that the majority of the male gene flow of Austroasiatic tribes occurred during the late Pleistocene period. The results suggest gene flow from Southeast Asia to Northeast India, albeit more

significantly among Tibeto-Burman than Austroasiatic-speaking populations (Borkar et al., 2011).

The majority of lineages observed in contemporary European populations fall into the following main haplogroups: E, G, I, J, N and R. Typically, >50% of men in Europe are affiliated with Macrohaplogroup R and also widespread in Western, Central and Southern Asia. R1b-M412 appears to be the most common Y-chromosome haplogroup in Western Europe (>70%), while being virtually absent in the Near East, the Caucasus and West Asia (Jobling et al., 2003; Chiaroni et a.,l 2009).

Study of 718 male belongs to 12 ethnic groups of Pakistan using of 18 binary polymorphisms and 16 multiallelic, short-tandem-repeat (STR) loci from the non-recombining portion of the human Y chromosome, 11 stable haplogroups and 503 combination binary marker/STR haplotypes were indentified. Haplogroup frequencies were generally similar to those in neighboring geographical areas, and the Pakistani populations speaking a language isolate (the Burushos), a Dravidian language (the Brahui), or a Sino-Tibetan language (the Balti) resembled the Indo-European–speaking majority. Nevertheless, median-joining networks of haplotypes revealed considerable substructuring of Y variation within Pakistan, with many populations showing distinct clusters of haplotypes. These patterns can be accounted for by a common pool of Y lineages, with substantial isolation between populations and drift in the smaller ones. Few comparative genetic or historical data are available for most populations, but the results can be compared with oral traditions about origins. The Y data support the well-established origin of the Parsis in Iran, the suggested descent of the Hazaras from Genghis Khan's army, and the origin of the Negroid Makrani in Africa, but do not support traditions of Tibetan, Syrian, Greek, or Jewish origins for other populations (Qamar et al., 2002).

2.6 Autosomal Marker

They are DNA marker found in the 22 chromosome (autosome) other than sex (XY) chromosome. Unlike Y and Mitochondrial DNA they inherited biparentally. Unlike Y chromosome (male ancestry) or mitochondrial (female ancestry) markers, autosomal markers are distributed throughout all chromosomes and represent accumulated inheritance from all our ancestors .They can be used as a important marker for study of genetic diversity and human evolution. There is the very important role of environmental condition on the Evolutionary process. Only those who can adapt that particular environment survive at that condition. The adaptation towards the particular

environment is lead by the genetic change in particular gene that is responsible to cope with particular environmental condition.

2.6.1 SLC24A5 (rs 1426654)-G/A Marker

SLC24A5 belongs to a family of solute carrier family 24 (potassium-dependent sodium/calcium exchanger). The human SLC24A5 gene has a role in skin pigmentation (Rees, 2004)**.** Lamason et al., (2005) investigated polymorphism in the human SLC24A5 gene to evaluate the role of the gene on human skin pigmentation. They reported that, in 98.7 to 100% of European American population, a G-to-A transition at amino acid 111 in exon 3 of the SLC24A5 gene (rs1426654) results in an alanine-to-threonine substitution. The threonine is associated with lighter skin pigmentation among European-Americans and among admixed African Americans and African Caribbeans. Skin pigmentation in humans is largely determined by the quantity and distribution of the pigment melanin, which is packed in melanosomes and then transferred from melanocytes to the surrounding epidermal keratinocytes. Human melanin is primarily composed of two distinct polymers: eumelanin (brown/ black) and pheomelanin (yellow/red), which differ in their physical properties and chemical composition. In addition to the amount and type of melanin, other factors such as the size, shape, number, and cellular distribution of melanosomes also contribute to the variation in skin color (Thody et al., 1991).

Mallick et al., (2013) studied in Indian population which reveal that the there is significant association of rs1426654 SNP with skin pigmentation, explaining about 27% of total phenotypic variation. Polymorphism in different ethnic populations across the Indian subcontinent showed that the presence of the derived-A allele, although the frequencies vary substantially among populations. It was also found that the geospatial pattern of this allele is very complex, but most importantly, reflects strong influence of language, geography and demographic history of the populations. Time of coalescence of the light skin associated allele at 22–28 KYA was estimated. This study confirms that this gene has been a target for positive selection among Europeans as well as populations of the Middle East, Central Asia, Pakistan and North India but not in South India.

2.6.2 EDAR (Ectodysplasin-A receptor, rs3827760-T/C)

The cross-sectional area of East Asian hair fibers averages about 30% larger than that of Africans and 50% larger than that of Europeans. The variant may have been selected for in East Asians either because thicker hair was beneficial in the cold north Asian climate. It could be used as a marker for East Asian ancestry. Hair morphology differs dramatically between human populations: people of East Asian ancestry typically have a

coarse hair texture, with individual fibers being straight, of large diameter, and cylindrical when compared to hair of European or African origin. Ectodysplasin-A receptor (EDAR) is a cell surface receptor of the tumor necrosis factor receptor (TNFR) family involved in the development of hair follicles, teeth, and sweat glands. In this signaling process the cell surface receptor ectodysplasin-A receptor (EDAR; MIM]604095) is engaged by its ligand ectodysplasin-A (EDA), causing recruitment of EDAR associated death domain (EDARADD) and ultimately activation of the transcription factor nuclear factor NF-kB. EDAR recruitment of EDARADD is mediated by their C-terminal death domains. Mutations that abolish the action of this pathway impair hair follicle formation in mouse and human (Botchkarev and Fessing, 2005; Courtois and Smahi, 2006; Thesleff and Mikkola, 2002).

Analyses of genome-wide polymorphism data from multiple human populations suggest that EDAR experienced strong positive selection in East Asians. It is likely that a nonsynonymous SNP in EDAR, rs3827760, was the direct target of selection as the derived p.Val370Ala variant is seen at high frequencies in populations of East Asian and Native American origin but is essentially absent from European and African populations. (Chunyan et al, 2008).

2.6.3 EPHX1 Gene Polymorphism

Microsomal epoxide hydrolase (mEH) is a smooth endoplasmic reticulum enzyme that is expressed relatively in most tissues of many species. The human mEH enzyme is an important phase II biotransformation enzyme, and it catalyzes the hydrolysis of different epoxides and reactive epoxide intermediates into less reactive and more water-soluble dihydrodiols, which are then excreted from the body. It catalyses the hydrolysis of arene and alkine oxides to form trans-dihydrodiols. mEH is a protective enzyme involved in general oxidative defense against a number of environmental and endogenous stimulants including endotoxin exposure (Rodd et al., 2007).

The human EPHX1 gene lies in the chromosomal region 1q42.1. A number of functional variants exist for the EPHX1 gene. Two single nucleotide polymorphisms (SNPs) have been described in the coding region of the EPHX1 gene that produces two protein variants. Genotype combinations that are associated with “slow” enzyme activity will lead to inefficient metabolizing of reactive oxygen species generated by endotoxin exposure and may eventually disturb the normal physiological activities. Previous in vitro studies have suggested that the 113His– 139His haplotype of both Tyr113His and His139Arg polymorphisms are associated with decreased enzyme activity (Hartsfield et al.,1998).In vitro expression studies of cDNA revealed that the T/C mutation, which changes the tyrosine residue at codon 113 to histidine of exon 3, decreased EPHX

enzymatic activity by approximately 40%, whereas a change of histidine to arginine at codon 139 of exon 4 increased EPHX activity by approximately 25% (Hassett et al., 1994). These polymorphisms are thought to be linked with protein stability. EPHX1polymorphisms have also been studied with respect to several other disease end points and have been variously associated with colorectal polyp formation, lung cancer, orolaryngeal cancer, and sensitivity to 1,3-buadiene. Variations in enzyme activity for EPHX as a result of such polymorphisms may lead to altered individual susceptibility to diseases like pre-eclampsia and the HELLP syndrome (Petra et al., 2001).

2.6.4 High altitude adaptation gene

The Tibetan Plateau is the world's highest plateau with very low temperature. However, Tibetans have lived on the plateau for tens of thousands of years and adapted to the high-altitude environment better than other populations. Tibetans exhibit many biological features in common with other high-altitude mammalian species (such as antelopes and pigs), including absence of chronic mountain sickness (CMS), thin-walled pulmonary vascular structure, and high blood flow; all these phenotypes are highly correlated with physiological responses to low oxygen concentration in the air, which facilitate uninterrupted oxygen-processing and the up-regulation of erythropoiesis and angiogenesis to allow for more efficient oxygen utilization (Monge et al., 1991). Human adaptation to high-altitude environment is believed to a result of advantageous genetic mutation and selective pressure. Three recent studies have identified several genes that play important roles in high-altitude adaptation, including EGLN1, PPARA, and EPAS1 (Simonson et al., 2010).

This suggests that EPAS1 is potentially under positive selection only in Tibetans. The EGLN1 gene within this region is also involved in the response to hypoxia and potentially be the target of positive selection. EPAS1 and EGLN1 play central roles in the activation of hypoxia-inducible genes and homeostasis of HIF under hypoxia and normoxia (Lofstedt et al., 2007)**.**

Tibetans possess biological characteristics or phenotypes unique to people who live at high altitudes. These characteristics include adaptation to hypoxia, the absence of CMS, and high offspring survival rate. Adaptation to hypoxia is mediated by the hypoxia inducible factor (HIF) complexes which consist of a (HIF-1a, HIF-2a) and b subunits (HIF-1b).EPAS1 gene (encode HIF-2a) had undergone positive selection in Tibetans, but not HIF-1a despite its involvement with most of hypoxia-inducible genes (Hu et al., 2003). HIF-1a is highly conserved and serves as a ‘master regulator’ of cellular and systemic oxygen homeostasis (Beall, 2007). Unlike HIF-1a, which is universally expressed, EPAS1 predominantly expressed in highly vascularzed tissues such as the lung and placenta

(Sood et al., 2006). HIF-2a can escape degradation at near-normoxic conditions but HIF-1acannot. Furthermore, unlikeHIF-1a, which responds to acute hypoxia, EPAS1 plays an important role in prolonged hypoxia, a condition with exactly the same symptoms as high altitude hypoxia.

Another candidate gene under positive selection is EGLN1/ PHD2, which is a member of the 2-oxoglutarate-dependent dioxygenase superfamily and a sensor for low oxygen levels (Bruick and McKnight, 2001). Under normal oxygen levels, HIF-a proteins are modified by prolyl hydroxylases (PHDs), resulting in the subsequent proteasomal degradation of HIF (D'Angelo et al., 2003). Interestingly, although HIFa stability is regulated by PHDs, PHD2 is subject to feedback up-regulation in aHIF1a-dependent, butHIF2a-independent, manner (Aprelikova et al., 2004). In the process of Tibetans' adaptation to high-altitude hypoxia, both HIF-2a and its degradation regulator EGLN1 had undergone positive selection. However,HIF-1a, as the up-regulator of PHD2, had not. Our data have demonstrated that the Proto-Tibeto-Burman people form a major clade within East Asian populations, and that the major migration route into the Tibetan Plateau was via the Hengduan Mountain valleys. We also found that the Yi population, and not the Han populations used in previous studies, was a more appropriate reference for exploring the adaptation of Tibetans to high-altitudes.

2.6.5 Lactase persistence

LCT gene encodes a single protein having both lactase activity and phlorizin hydrolase activity collectively called as Lactase phlorizin hydrolase (LPH) (Lacey et al., 1994). Lactase persistence varies in frequency in different human populations, being most frequent in northern Europeans and certain African and Arabian nomadic tribes, who have a history of drinking fresh milk. Selection is likely to have played an important role in establishing these different frequencies since the development of agricultural pastoralism approximately 9,000 years ago (Harvey et al., 1995). DNA polymorphisms in the lactase gene with linkage disequilibrium over a 70-kb region showed that the 'element' responsible for the lactase persistence/nonpersistence polymorphism in humans is cis acting to the lactase gene and that lactase persistence is associated, in Europeans, with the most common 70-kb lactase Haplotype, A. Genetic drift was important in shaping a general pattern of non-African Haplotype diversity, with recent directional selection in northern Europeans for the Haplotype associated with lactase persistence (Romereo et al, 2012). There is no difference in the sequence of the LCT gene in adults with high or low levels of intestinal lactase. It was found that some cases of adult-type hypolactasia may reflect an unusual mechanism involving a developmentally regulated change in protein processing or targeting which does not occur in subjects with lactase persistence, affects specifically LPH and not the other

brush border glycosidases. It does not depend for its specificity on an LPH sequence different from that found in lactase persistence. Multiple independent mutations in a 100-bp region—part of an enhancer— approximately 14-kb upstream of the LCT gene are associated with this trait in Europeans and pastoralists from Saudi Arabia and Africa (Romero et al., 2012). Four causative single-nucleotide polymorphisms (SNPs) approximately 13.9 kb upstream from LCT gene at the 100bp enhancer region: -13910C/T (rs4988235), -13907C/G (rs41525747), -13915T/G (rs41380347) and -14010G/C, have been subsequently identified as the candidate cis-acting elements based on genotype–phenotype association analyses and functional experiments. The derived allele of -13910*T is associated with LP in European and some Central Asian populations. This allele is identified in Indian populations and is inferred to make substantial contribution to LP in India. The rest three alleles, -13907*G, -13915*G and-14010*C are responsible for the LP in some African and Middle Eastern populations. The variation of that LP allele with population suggests that this allele emerged independently in different ethnic /geographical population (Romero et al., 2012).

In the Tibetan Plateau, milk and milk products (for example, from domestic yaks) are important ingredients of the daily diets for Tibetans, especially for the herders (Dong et al., 2003). To investigate genetic variants for LP in Tibetans, they sequenced a region of 321 bp (position -14044 to -13724 upstream LCT) covering the five previously reported single-nucleotide polymorphisms (SNPs): -13907C/G (rs41525747),-13910C/T (rs4988235), -13915T/G (rs41380347),-14010G/C and -22018G/A (rs182549), which are associated with the LP in populations from a vast region surrounding Tibet. The five SNPs were nearly absent in Tibetan populations, suggesting LP likely to have an independent origin in Tibetans rather than to be introduced via gene flow from neighboring populations. Three novel SNPs (-13838G/A,-13906 T/A and-13908C/T) are identified in Tibetans. In particular,-13838G/A might be functional as it is located in the binding motif for HNF4athat acts as a transcription factor for intestinal gene expression (Peng et al., 2012).

2.6.6 MYBPC-3 (myosin binding protein C, cardiac): 25 bp deletion

It encodes cardiac myosin binding protein C (cMyBP-C), a key constituent of the thick filaments localized to doublets in the C-zone of the A-band of the sarcomere. By binding to myosin, titin and actin, cMyBP-C contributes to the structural integrity of the sarcomere and regulates cardiac contractility in response to adrenergic stimulation (Previs et al., 2012). Because individuals who have heritable Cardiomyopathies with

cMyBP-C defects have a disorganized sarcomeric structure and late-onset symptoms,MYBPC3has emerged as a candidate gene for increased risk of heart failure through either hypertrophic or dilated Cardiomyopathies (HCM or DCM). The 25-bp deletion is the one of the most important genetic variant that leads to Cardiomyopathies. Its prevalence was found to be high (4%) in populations of Indian subcontinental ancestry. The 25-bp deletion also observed in Pakistan, Sri Lanka, Indonesia and Malaysia, (all heterozygote) but was absent from other population. Thus, the deletion is a common variant in individuals from South Asia, present in Southeast Asia, but undetectable elsewhere. It was conclude that the 25-bp deletion, a commonMYBPC3variant in South Asians, is associated with chronic risk of heart failure. The delayed symptoms, mild hypertrophy and influence of secondary risk factors pose a lifelong threat to carriers. Haplotype analysis determined that the common 25-bp deletion likely arose approximately 33,000 years ago on the Indian subcontinent **(**Dhandapany et al., 2009).

2.6.7 Transforming Growth factor (TGF-β1)

Is a multi-functional cytokine that plays an important role in breast carcinogenesis.TGF-β1 is a potent inhibitor of proliferation of epithelial, endothelial and hematopoietic cells and acts as a tumor suppressor. TGF-β1 has dual role in carcinogenesis with tumor suppressive effects in epithelial cells, but tumor invasion and metastasis promoting effects during later stages of carcinoma progression (Critical role of TGF-β signaling in cancer progression, particularly breast cancer, is well documented. Polymorphisms in the gene encoding this multifunctional cytokine have been found to influence its expression; however, their influence of breast cancer risk is not clear) (Oh et al., 2000).

'TT' genotype at +29 C>T locus increased breast cancer risk, irrespective of ethnicity. On the other hand, the presence of 'CC' genotype at +74 G>C locus reduced breast cancer risk significantly. The polymorphic status of +29C>T was comparable among the Indo-Europeans, Dravidians, and Tibeto-Burman. It was found that the Tibeto-Burman populations did not exhibit any polymorphism at +74G>C locus, which matches with Chinese populations. However, the Brahmins of Nepal (Indo-Europeans) showed polymorphism. So it was conclude that TT genotype at codon 10 increases breast cancer risk, while 'CC' genotype at codon 25 reduces the risk. The protective effect of codon 25 polymorphism is more selective.

2.7 Human Evolution and Genetic Study on Nepalese Population

There are some research takes place to trace origin and migration of different population of Nepal. One of the most important studies was taken out in Tharus.

Fornarino et al., (2009) did high resolution analysis of and Y-chromosomal and compared with Indians (Hindus from Tarai and New Delhi and tribal's from Andhra Pradesh) to allowed the identification of three principal components: East Asian, West Eurasian and Indian, the last including both local and inter-regional sub-components. Many mitochondrial-DNA and Y-chromosome lineages are shared or derived from ancient Indian Haplogroups, thus revealing a deep shared ancestry between Tharus and Indians. Interestingly, the local Y-chromosome Indian component observed in the Andhra-Pradesh tribal's is present in all Tharu groups, whereas the inter-regional component strongly prevails in the two Hindu samples and other Nepalese populations. The complete sequencing of mtDNA from unresolved Haplogroups also provided informative markers that greatly improved the mtDNA phylogeny and allowed the identification of ancient relationships between Tharus and Malaysia, the Andaman Islands and Japan as well as between India and North and East Africa. Overall, this study gives a paradigmatic example of the importance of genetic isolates in revealing variants not easily detectable in the general population. Super-Haplogroup M (55.7%) and, to a lesser extent, R (39.3%) are the most represented in the dataset. The M lineages were predominant (>50%) in all populations with highest values in the Tharu and Andhra Pradesh samples (75–88% and 76%, respectively). By contrast, the R lineages were present at higher frequencies among Hindus (43.7%) than among the Tharu and the Andhra Pradesh tribal (19.1% and 24.1%) with a few overlaps in the haplogroup distribution. The N(xR) lineages were observed only in three Hindus (4.9%). The 9-bp polymorphism was found exclusively in the Tharus, associated with three different haplogroups: the deletion (6.4%) with haplogroups B5a (eight subjects) and M33 (three subjects), and the insertion (one subject –0.6%) with haplogroup M38.

To investigate Himalayan range act as barrierr for gene flow or not, Gyaden et al., 2007 study the genetic structure of Newar (N= 66), Tamang (N= 45), Kathmandu (N =77) and Tibetan (N = 156) by using 103 Biallelic Y-chromosomal markers and 20 Y-STR loci. The haplogroup O predominated with average frequency 35.46% followed by Haplogroup R with 24.70 % frequency. Haplogroup D is restricted to the Tibetans, accounting for 50.6% and completely absent Nepalese Population. Haplogroup D is characterized in Tibetans by subclades D1-M15 (28.2%), D3-P47 (18.6%), and D*-M174 (3.8 %). Other important Haplogroups include haplogroup H, found in Newar (6.1%), Kathmandu (11.7%), and Tibet (1.9%) but absent in Tamang. Admixture analysis further supports this, since Newar and Kathmandu both show high admixture proportions from India (43.4% and 48.6%, respectively), whereas the opposite is observed in Tamang, which has a high Southeast Asian contribution (66.2%) and null contribution from India. The CA also suggests strong Indian influence, since Kathmandu and Newar group closer to the Indian

populations. In addition, the CA graph displays close genetic association of Central Asian populations (Kyrgyz and Karalkapak) with Newar and Kathmandu groups, in agreement with the admixture results (56.6% and 51.4%, respectively). Yet the presence of haplogroup O3a5-M134 representatives in Nepal indicates that the Himalayas have been permeable to dispersals from the east. These genetic patterns suggest that this cordillera has been a biased bidirectional barrier. The mtDNA analysis showed that Newar and Kathmandu harbors several deep-rooted Indian specific Lineages including M2, R5, U2 , U7 etc. A total of 75 different Haplogroups and sub-Haplogroups were observed . Among the four groups, Kathmandu exhibits the highest degree of heterogeneity, with 44 lineages, followed by Tibet (39), Newar (26) and Tamang (22).. While the majority of mtDNA diversity in Tibet (96.1%) and Tamang (66.7%) is represented by East Asian-specific Haplogroups, including those from Northeast [i.e., A, D, G, and M8 (M8a, C, and Z)] and Southeast (i.e., F, M9, and M13) Asia, Kathmandu, Newar and, to a lesser extent, Tamang, display a considerable proportion of South Central Asian lineages (49.3%, 36.4% and 17.8%, respectively). These South Central Asian-specific markers include M2, M3, M4"67, M5,M30, M33, M34, M35, M38, M43, R2, R5, R6, R30, U2 and U4, whereas the West Eurasian mtDNA component in Kathmandu (9.1%), Tamang (15.6%) and Newar (18.2%) is characterized by a combination of Haplogroups HV, H, J, N1e, T and U7 (Gaydan et al., 2013).

The observed high frequency of the East Eurasian paternal lineages in Nepal, suggest that these lineages were introduced into Nepal from Tibet directly; however, possibility is that the East Eurasian genetic components might introduced from northeast India where abundant East Eurasian maternal lineages have been detected. To solve this dilemma Wang et al., (2012) collect the blood from 246 unrelated populations from 3 different geographical regions and compare their result with 43 different populations from neighboring region of Nepal, India and Tibet. They found that, the majority (96.34%; 237/246) of the Nepalese mtDNA could unambiguously be allocated into the defined Haplogroups of East Eurasian (36.59%; 90/246), South Asian (51.63%; 127/246) and West Eurasian ancestries (8.13%; 20/246). They completely sequencing of mtDNA genomes that revealed virtually all of these samples in fact belong to the already defined haplogroups, such as M3, M5, M18, M30, M35, M43, D4, R8 and M60. They also defined a novel haplogroup characterized by variations 9266 and 11827, which was named M81. They also perform various bioinformatics and stastical analysis. The phylogenic analysis showed the affinity between the East Eurasian haplogroups identified in the Nepalese and those from the Tibetan, northeast and northwest Indian populations. Principle component analysis plot of the 43 populations, which was constructed based merely on the East Eurasian lineages. Among the five Nepalese populations under study, three

clustered with the Tibetans. They calculated Fst value with the populations from its neighboring regions, the smallest genetic distance was observed between the Nepalese and the Tibetans. By taking the Tibetans and northern Indians as the parental populations, the results of the admixture estimation analysis revealed that the Tibetans made major contribution to virtually all Nepalese populations. They further compared the Phylogenetic affinity of the East Eurasian lineages observed in Nepalese (including Haplogroups A11, C, G2a, M9a, F1c and Z) with those from the neighboring regions, Tibet, northeast and northwest India, by means of median networks. On the basis of the constructed networks several features could be observed:
(1) Nepalese share some basal or internal Haplotypes with the Tibetans.
(2) Nepalese harbor a number of unique Haplotypes at the terminal level, most of which branched off directly from the nodes occupied almost exclusively by the Tibetan lineages. (3) Only few Haplotypes are shared sporadically between the Nepalese and the northern Indians.

They also found some Nepalese specific haplogroup suggesting de novo differentiation in Nepal. By estimating time interval of two clades of haplogroup G2a and M9a1a2, they reveal the age of these haplogroup is 5.7 KYA. In Tibet the East Asian maternal genetic component was introduced around 8KYA. These data revealed clearly that East Asian genetic component was introduced in Nepal about 6 KYA from Tibet.

CHAPTER 3

MATERIALS AND METHODOLOGY

3.1 Blood Collection

10 ml of intravenous blood samples from 98 healthy and unrelated individuals belonging to Bajracharya (n-20), Shakya (n=19) and Udaaya (n=59) Caste of Newar Ethnic group from Kathmandu valley, Nepal was collected in vacutainers containing EDTA as an anti coagulant with their informed written concern. The samples were brought to Center for Cellular and Molecular Biology (CCMB), Hyderabad in an icebox and stored at 4°C, until DNA was extracted from the sample.

Fig 3.1: Sample Collection Area (Kathmandu Valley)

3.2 Reagents used

The following solutions were prepared, sterilized and used in different experiments:

- ✓ Reagent A (Lysis buffer 1): 10mM Tris HCl (pH 8.0), 320 mM Sucrose, 5mM $MgCl_2$, 1%Triton X.
- ✓ Reagent B (Lysis Buffer II): 400mM Tris HCl, 60mM EDTA, 150mM NaCl, 1%SDS added after autoclaving.

- ✓ Tris Saturated Phenol: Phenol, 0.1% 8-Hydroxy Quinoline, 0.5 M Tris HCl (pH 8.0), 0.1 M Tris HCl (pH 8.0).
- ✓ Chloroform: Isoamylalcohol (24:1): 24ml of Chloroform was added to 1ml Isoamylalcohol.
- ✓ Reagent C: 5M Sodium per chlorate (not to be autoclaved).
- ✓ T.E. Buffer (100ml): 10mM Tris HCl (pH 7.5), 1mM EDTA (pH 8.0).
- ✓ 20% SDS: 20g of SDS dissolved in 80ml DDW at 65°C, volume made up to 100ml.
- ✓ 70% Ethanol: 70ml of absolute alcohol in 30 ml DDW.
- ✓ 10X TAE Buffer: 48.4g Tris base, 20ml 0.5M EDTA (pH 8.0), 11.402ml glacial acetic acid, mixed and volume made up to 1 liter.
- ✓ 6x loading Dye: 0.125g of Bromophenol Blue, 0.125g of Xylene Cyanol FF, 15 ml of glycerol. Diluted with DDW to make up volume to 50ml.
- ✓ Ethidium Bromide: 10mg of Ethidium Bromide in 1 ml DDW. Stored in dark bottles.

3.3 Reagents for PCR

- ✓ PCR Master Mixture (Takara™): 500mM KCl, 100mM Tris (pH 8), 15mM $MgCl_2$, 0.1% gelatin,25mM $MgCl_2$, 2.5mM dNTPs,
- ✓ Molecular Grade water(Takara™)

3.4 Reagents for cycle sequencing and processing

- ✓ ExoSap (Bioserve™)
- ✓ Big Dye ™, 50% HiDi Formamide™, 70% ethanol.
- ✓ 3M Sodium Acetate: Dissolved 24.612g of Sodium Acetate in 80ml DDW, pH was adjusted to 5.2 with Conc. HCl

3.5 Instruments Used

- ✓ Centrifuge (Eppendorff 5810R, Biofuge, Remi R8C, Astec)
- ✓ Thermo Scientific NanoDrop™ 1000 Spectrophotometer
- ✓ PCR machines (MJ Research PTC 200, Gene Amp 9600 Perkin Elmer, verity Thermal cycler, Eppendorf Thermal Cycler)

- ✓ Electrophoresis apparatus (Pharmacia Biotech EPS600, Hoefer power pack)
- ✓ Gendoc (Syngene bioimaging system) by using software SnapGene for Syngene.
- ✓ Vortex
- ✓ ABI PRISM® 3730xl DNA Analyzer: Automatically analyzes DNA molecules labeled with multiple fluorescent dyes. It consists of a charge couple device (CCD) camera and a powerful Macintosh computer that includes software for data collection and data analysis. After samples are loaded onto the system's vertical gel, they undergo electrophoresis, laser detection, and computer analysis. Electrophoretic separation can be viewed on-screen in real-time.

3.6 Sequence Analysis Software

3.6.1 Sequencing Analysis Software™

Two software packages automatically process gel files or raw sample files to analyze sample files with base calls matching sequence peaks. Sequencing Analysis Software™ Ver. 3.4.1 with free Factura™ software Ver. 5.2 is used for analysis of data for 310 and 3700 genetic analyzers running on a Mac® OS platform. Sequencing Analysis Software™ v3.7 with Factura™ software is used for analysis of data from 310, 377, 3100 and 3700 genetic analyzers. Both Sequencing Analysis Software™ v3.4.1 and v3.7 are powered by multiple base caller algorithms to perform signal processing and classification of peaks from raw data collected from ABI PRISM® Genetic Analyzers. The result yields accurate sequence data with electropherograms that can be viewed by Sequencing Analysis Software™ or Edit View software.

3.6.2 Auto Assembler version 2.0

This is a sequence assembly program and can handle at least 1000 sequences of 500 bp. It allows on-screen alignment of chromatograms. The manufacturer claims that the software has no known limitations or bugs. It certainly has the nice feature of lining up all the chromatograms under each other making analysis easier.

3.6.3 ARLEQUIN VER 3.1 (Software for Population Genetics Data Analysis)

It is population genetic software able to handle large sample of molecular data (RFLP's, DNA sequences, micro satellites), while retaining the capacity of analyzing conventional genetic data (standard multi locus data or more allele frequency data). The analysis ARLEQUIN can perform on the data fall into two main categories: intra-population and inter population methods.

The intra-population methods include Molecular Diversity, Mismatch distribution, Haplotype frequency estimation, Linkage disequilibrium, Hardy-Weinberg equilibrium etc. The inter-population methods include search for shared Haplotype between populations, AMOVA (Analysis of Molecular Variance), pair wise genetic distances, Exact test for population differentiation, Assignment test of genotypes. The mismatch distribution of the chosen population alone from main.htm of Alrequin run folder *.res was copied and pasted in MS Excel and smooth line graph of mismatch distribution was drawn with this data (Escoffier, 2004).

3.6.4 NETWORK 4.X Phylogenetic Network Analysis Software

The network methods are designed for non-recombining DNA Haplotypes, RNA or amino acid sequences. Multi-state data typically are amino acid sequences, and also DNA sequences containing nucleotide positions with more than two different nucleotides. In contrast, STR data are generally binary (if a single-repeat mutation mechanism has generated the STR alleles). Multi-state data can be analyzed only by the Median-Joining (MJ) network method. Binary data typically are STRs and closely related DNA sequences within a species (Forster et al. 2000). This software is freely available on http://www.fluxus-engineering.com.

3.7 Protocol for DNA Isolation from Blood Samples

Then DNA was isolated by using standard Phenol: Chloroform Method. Four volumes of reagent A was added to 10ml of blood sample and centrifuged at 3600 rpm for 10 min. The supernatant was discarded and the pellet was disturbed. Then 5ml of reagent B and 1.25ml of reagent C were added and shaken thoroughly. Then 3ml of Tris-Phenol and 3ml Chloroform-Iso amyl alcohol (24:1) were added, mixed gently and centrifuged at 3000 rpm for 10 min. The two layers were formed. The bottom layer was organic phase & the upper layer was aqueous phase. The aqueous phase was taken in another tube using blunt tips. To the aqueous aliquot 3ml chloroform was added, mixed well and centrifuged at 3000rpm for 5min. The aqueous phase was separated into fresh tube and two volumes of ice-cold Isopropyl alcohol was added to precipitate DNA. The DNA was taken in the sterile Eppendorff tube carefully and 80% ethanol was added. The DNA was given a short spin and the ethanol was removed by inverting on tissue paper. Then it was air-dried. 80–100µl TE solution was added to the Eppendorff containing DNA. For dissolving DNA, the Eppendorff were kept in water bath at 55^0C for 1-2 hours and then stored at 4^0C, in cold room (Thangaraj et al., 2005).

3.8 Quantification of DNA

The extracted DNA was quantified by the spectrophotometer method followed by checking in 0.8% Agarose (Axygen™) gel.

3.8.1 NanoDrop method

The Thermo Scientific NanoDrop™ 1000 Spectrophotometer measures 1 ul samples with high accuracy and reproducibility. The full spectrum (220nm-750nm) spectrophotometer utilizes a patented sample retention technology that employs surface tension alone to hold the sample in place. This eliminates the need for cumbersome cuvette and other sample containment devices and allows for clean up in seconds. In addition, the NanoDrop 1000 Spectrophotometer has the capability to measure highly concentrated samples without dilution (50X higher concentration than the samples measured by a standard cuvette spectrophotometer).

The ratio of absorbance at 260 to 280nm indicates the purity of the sample. This ratio of DNA solutions should range from 1.7 to 1.8. The presence of impurities like proteins or phenol tends to decrease this ratio.

To quantify the DNA, 1µl of Mili-Q water was taken and kept in the well to clean the well. Then 1µl of TE-buffer was taken and kept in the well and measure the blank. Finally 1µl of the DNA solution was taken and kept in to the well and the concentration was measured.

After Measurement, The genomic DNA was diluted to make 10ng/µl final Concentration by adding required volume of Mili-Q water.

3.8.2 Agarose Gel Electrophoresis

Agarose gel electrophoresis is an efficient technique to separate DNA molecule according to their molecular weights in the same manner as a sieve. 0.8 gm of Agarose (Axygen™) was dissolved in 100 ml 0.5X TAE buffer in a 250ml conical flask and was boiled to dissolve Agarose completely. 0.7µl Ethidium bromide was added from stock solution to make a final concentration of 0.5µg/ml. Gel was cooled down poured onto a gel tray and was allowed to set. Samples DNA were loaded and electrophoresis was carried out at a constant voltage of 80V. After 30 Min. of run halfway the gel was observed under Gendoc (Syngene bioimaging system) by using software SnapGene for Syngene.

3.9 Protocol for Genetic Analysis

3.9.1 Amplification and Sequencing Whole Mitochondrial DNA (mtDNA) Sequence

To analyze mitochondrial haplogroup present in the studied sample whole mtDNA was amplified and sequenced. 24 different primer set were used to amplify the whole mtDNA sequence. Then PCR master mix was prepared in a 1.5ml tube and then equally dispensed into PCR plates containing 1.5µl (10ng/µl) of the DNA template of the individual samples. The PCR was carried out in Thermal Cycler. PCR products were electrophoreses at 120V in 2% Agarose (Axygen™) gel. The PCR products were then visualized under Gel-documentation. On obtaining a single band devoid of any primer-dimer bands the PCR products were then sequenced. The list of mtDNA primer was tabulated in the table (Appendix no.1).

Table 3.1: PCR 'reaction mix' for the mitochondrial marker used (PCR Master Mix was obtained from Takara™ along with Mili-Q water)

Template DNA (10ng/µl)	**PCR master mix**	**Forward Primer**(10pm ol/ µl)	**Reverse Primer**(10p mol/ µl)	**Mili-Q**	**Total**
1.5 µl	5 µl	0.15 µl	0.15 µl	3.2 µl	10 µl

Following Programme was set up in Thermal Cycler (Applied Biosystem, USA)

Table 3.2: PCR conditions for Mitochondrial DNA Amplification

Stages	**Steps**	**Temperature**	**Time**	**No. of Cycle**
1.	Initial Denaturation	95˚C	5 min	
2.	Denaturation	95˚C	1 min	35
	Annealing	58˚C	30 sec	
	Extension	72˚C	2 min	
3.	Final Extension	72˚C	7 min	
4.	Hold	4˚C	∞	

3.9.2 Amplification and Sequencing Y-Chromosomal SNP Markers

The various SNP markers used to study Y-SNP Polymorphism. Using the result, the SNP tree was drawn to show the group in which the chosen population fall. Each of the individual SNP marker, the nucleotide change and position along with the primer sequence used are tabulated in the table (Appendix no.2)

PCR mix was prepared and Amplification of the Y-Chromosomal Marker was done.
Table 3.3: PCR mix for Y-Chromosomal Marker amplification

Template DNA (10ng/µl)	**PCR Master Mix**	**Forward Primer(10pmol/ µl)**	**Reverse Primer(10pmol/ µl)**	**Mili-Q**	**Total**
1.5 µl	5 µl	0.15 µl	0.15 µl	3.2 µl	10 µl

Following Programme was set up in Thermal Cycler (Applied Biosystem, USA)

Table 3.4: PCR conditions for Y-Chromosomal Marker

Stages	**Steps**	**Temperature**	**Time**	**No. of Cycle**
1.	Initial Denaturation	95˚C	5 min	
2.	Denaturation	95˚C	1 min	35
	Annealing	55˚C	30 sec	
	Extension	72˚C	2 min	
3.	Final Extension	72˚C	7 min	
4.	Hold	4˚C	∞	

3.9.3 Autosomal Markers

3.9.4 Single Nucleotide Polymorphism (SNP) markers

The various SNP markers used to study Polymorphism for Autosomal markers. Using the result, the frequency distribution for each marker was done for the chosen population. For each of the individual SNP marker, the nucleotide change and position along with the primer sequence used are tabulated in the (Appendix No.4)

Table 3.5: The PCR condition and PCR reaction mix for the different Autosomal SNP Marker

Autosomal Marker	**PCR MasterMix**	**Forward Primer** (10pmol/ µl)	**Reverse Primer**(10pmol/ µl)	**Mili-Q**
EPAS 1	5.0 µl	0.15 µl	0.15 µl	3.20 µl
PARA	5.0 µl	0.15 µl	0.15 µl	3.20 µl
SLC	5.0 µl	0.15 µl	0.15 µl	3.20 µl
TGF	5.0 µl	0.15 µl	0.15 µl	3.20 µl
EGLN	5.0 µl	0.15 µl	0.15 µl	3.20 µl
EDAR	5.0 µl	0.15 µl	0.15 µl	3.20 µl
EPHX13	5.0 µl	0.15 µl	0.15 µl	3.20 µl
EPHX14	5.0 µl	0.15 µl	0.15 µl	3.20 µl
MYBPC	5.0 µl	0.15 µl	0.15 µl	3.20 µl
LCT	5.0 µl	0.15 µl	0.15 µl	3.20 µl

Following Programme was set up in Thermal Cycler (Applied Biosystem, USA)

Table 3.6: PCR conditions for EPAS1, PARA, EGLN-1, MYBPC, LCT, EPHX-13, EPHX-14, SLC24A5 Markers

Initial Denaturation		**Denaturation**		**Annealing**		**Extension**			**Final Extension**	
Temp	**Time**	**Temp**	**Time**	**Temp**	**Time**	**Temp**	**Time**	**Cycles**	**Temp**	**Time**
95˚C	5 min	95˚C	1 min	55˚C	30sec	72˚C	2min	35	72˚C	7min

Table: 3.7: PCR condition for the Autosomal Marker EDAR

Initial Denaturation		**Denaturation**		**Annealing**		**Extension**			**Final Extension**	
Temp	**Time**	**Temp**	**Time**	**Temp**	**Time**	**Temp**	**Time**	**Cycles**	**Temp**	**Time**
95˚C	5 min	95˚C	1min	65˚C	30sec	72˚C	2min	35	72˚C	7min

Table 3.8: PCR condition for the TGF-β1 Marker

Initial Denaturation		**Denaturation**		**Annealing**		**Extension**			**Final Extension***	
Temp	**Time**	**Temp**	**Time**	**Temp**	**Time**	**Temp**	**Time**	**Cycles**	**Temp**	**Time**
95˚C	5 min	95˚C	1min	56˚C	30 sec	72˚C	2min	35	72˚C	7 min

3.10 Protocol for Sequencing the PCR Product

Before sequencing we have to clean up unwanted DNA or nucleotide present in the PCR product. The enzyme Shrimp Exonuclease i.e. ExoSap is used to digest such unwanted nucleotide in the product. This enzyme degrades the single stranded DNA and dephosphorylete free nucleotide present in the sample. Thus double stranded PCR Amplicon remains intact and gives better sequence.

Table 3.9: ExoSap programme for PCR product

Incubation	**Enzyme Inactivation**	**Hold**
37^0C for 15 min	80^0C for 15 min	4^0C

3.11 Cycle sequencing of the amplified products

PCR products of mtDNA and Y-Chromosomal markers (SNPs) and Autosomal marker respectively were directly sequenced using the ABI Prism 3700 DNA analyzer. The amount of dNTPs and the concentration of the primers during the PCR were optimized. A master mix was prepared in a 1.5ml tube and then dispensed equally into Micro Amp 96 well plate. The PCR of the Micro Amp plates containing the DNA and mixture was carried out in the Gene Amp 9600 Thermocycler (Perkin-Elmer). The PCR was carried out for 3hours after optimizing the conditions.

Table 3.10: Sequencing PCR condition & reaction mixture for PCR products of Mitochondrial and Y-chromosomal Marker and Autosomal Marker

Requirement	Volume	Steps	Condition	Temperature	Cycle
Big Dye™	0.25 µl	1.	Initial Denaturation	96˚C for 30 sec	1
Sequencing Buffer (5x)	1.55 µl	2.	Denaturation	96˚C for 10 sec	32
Primer R(10 pmol/ µl)	0.075 µl		Annealing	55˚C for 7sec	
Mili Q	2.13 µl		Extension	60˚C 4 min	
DNA (10ng/µl)	0.8 µl	3.	Hold	4˚C	

3.12 Processing of Sequence Plate

After Cycle Sequencing plate processing was done. 3ml absolute alcohol was added to 120µl of 3N sodium acetate (pH 5.2) in a tube. The tube was mixed thoroughly. 25µl of the above mixture was added in each well of the plate. The plate was centrifuged at 4000rpm for 20min in Eppendorff (5810R) centrifuge at 25ºC. The plate was then inverted to remove the supernatant. 100µl of fresh 80% ethanol was added to each well and again centrifuged at 4000rpm for 10min.The plate was once again inverted and placing filter paper and giving a pop spin for few seconds at 750rpm removed alcohol. The plate was covered properly with fresh foil. At the time of sequencing, 10µl of 50% HiDye™ formamide was added to all the wells. The sample plates were kept and run in the ABI Prism® 3730 xl DNA Analyzer (for sequencing).

Fig 3.2 DNA Sequencer (ABI 3730xl DNA analyzer)

3.13 DNA Sequence and Data Analysis

Thus obtained sequences were base-called by using Sequence analysisv5.2. Then sequences were aligned with reference sequence and edited them by using Auto-assembler V2.0 in Mac platform. To find out the variation in Y-Chromosomal Biallelic Marker and Autosomal SNP Marker, obtained sequence were aligned with reference sequence and SNP position was traced out and variation was noted. If there is variation in the Y-Chromosomal Marker, it defines particular haplogroup. In case of mtDNA sequence, after edit, mutation on every position was scored against revised Cambridge Sequence (rCRS). Then by using Global Human mitochondrial phylotree, Each Haplogroup for Particular set of Haplotype were assigned. Then by using different stastical software package Genetic Diversity data were analyzed.

CHAPTER 4

RESULT

4.1 DNA Isolation and Quantification

DNA was isolated, diluted and quantified by Nanodrop method. Then Agarose gel electrophoresis was done by using 0.8% Agarose gel. Then gel was checked in Gel documentation machine by using GeneSnap software.

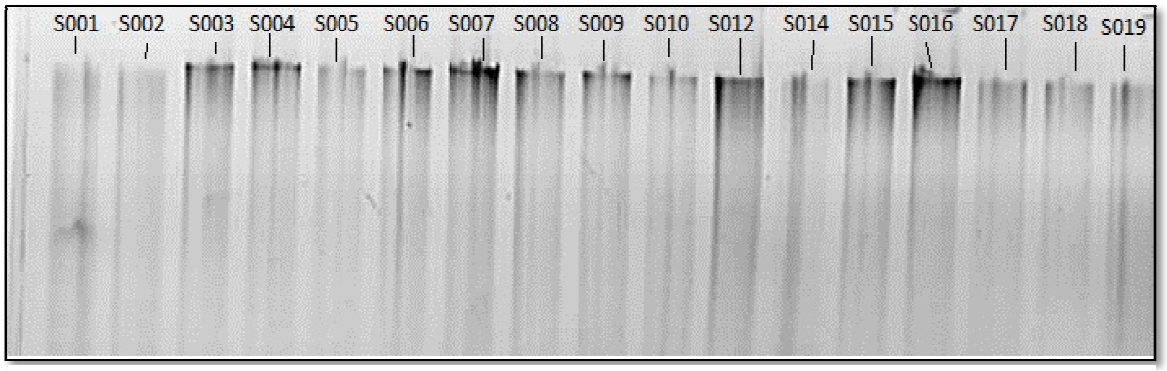


Fig 4.1: Diluted genomic DNA of some Sample

PCR product of Mitochondrial, Y-chromosomal and Autosomal Marker ware checked into 2% Agarose gel.

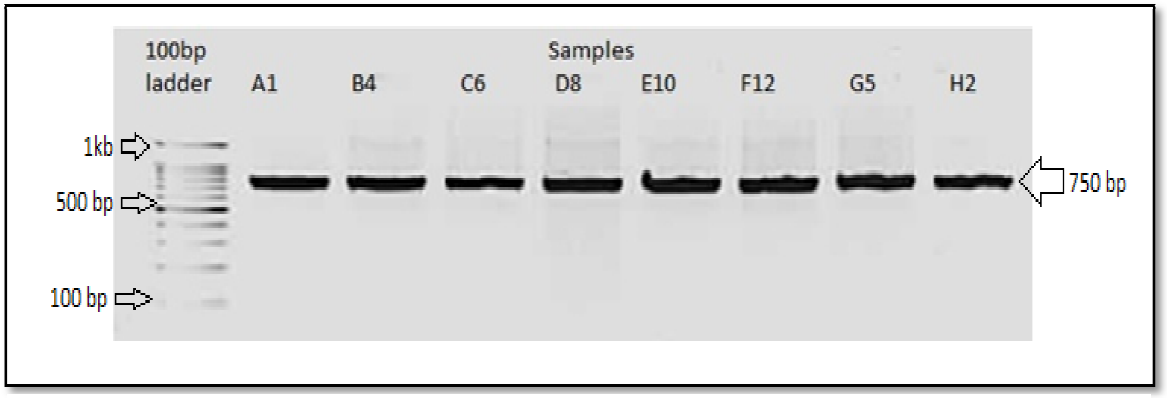


Fig 4.2: Mitochondrial DNA PCR product.

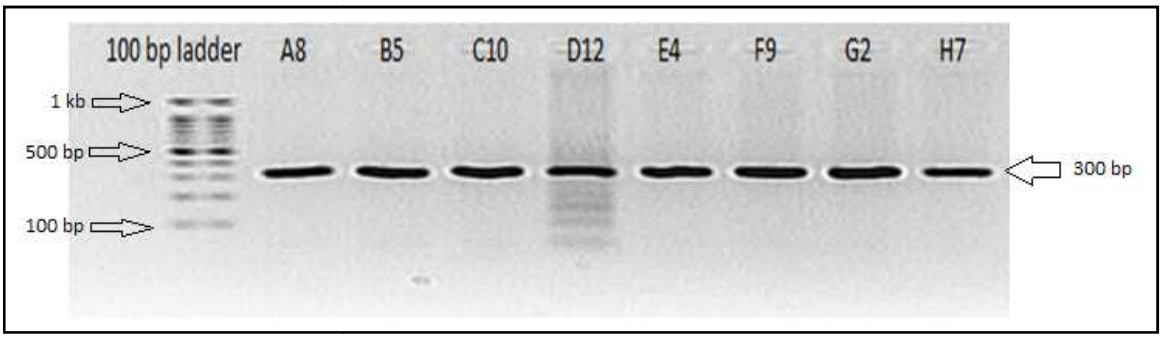


Fig.4.3: Y-chromosomal DNA PCR product.

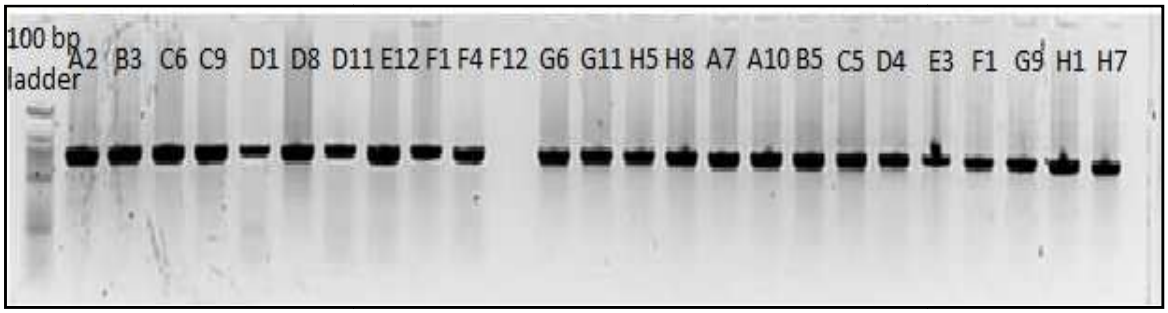


Fig 4.4: Gel Picture of Autosomal Marker. 100 bp ladder is used to compare the size of Amplicon. The sample F12 is absent in the gel means there no amplification takes place for that sample.

4.2 Analysis of molecular genetic marker

To reveal the genetic affinities and diversity of Bajracharya, Shakya and Udaaya sub-caste of Newar Ethnic group from Kathmandu valley, used three sets of genetic marker viz. The uniparentally inherited Mitochondria DNA and Y-chromosomal in addition to Autosomal marker which are associated with High Altitude adaptation, Skin Color Variation, Lactase Persistence, Cardiomyopathies, Breast Cancer and Hair morphology.

4.2.1 Mitochondrial DNA Marker

Whole mitochondrial DNA of 16569 bp of 98 individual from Shakya, Bajracharya and Udaaya sub caste of Newar Population by using 24 mitochondrial DNA primer set were sequenced. Base calling of obtained sequence was performed by sequence analysis var 5.1 software. Then pair wise alignment and editing of the sequence was done by using software Autoassembler in Mac platform. All the mutation present in the whole mitochondrial DNA sequence was scored against with revised Cambridge Reference Sequence (rCRS).

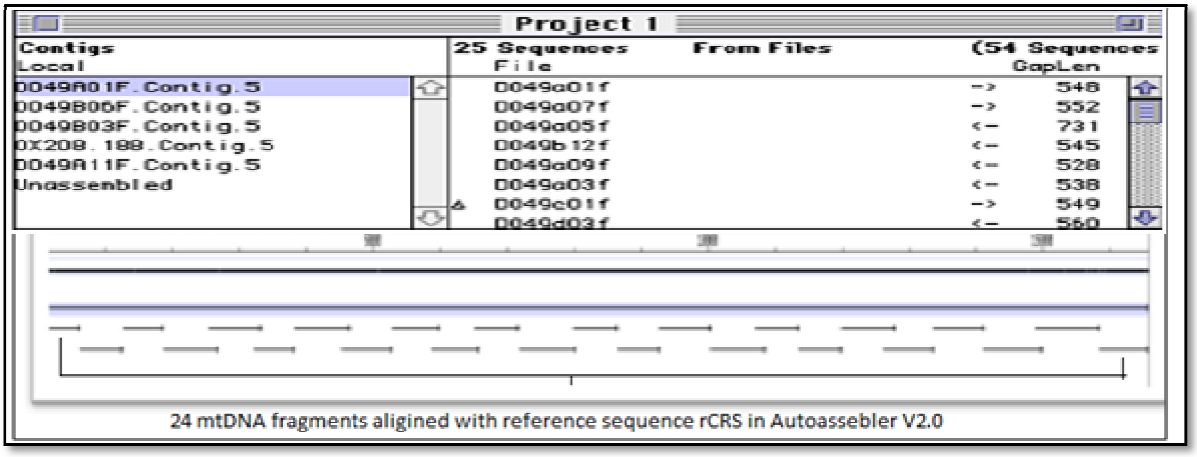


Fig 4.5: Complete mtDNA sequence amplified with 24 sets of overlapping primers and aligned with rCRS

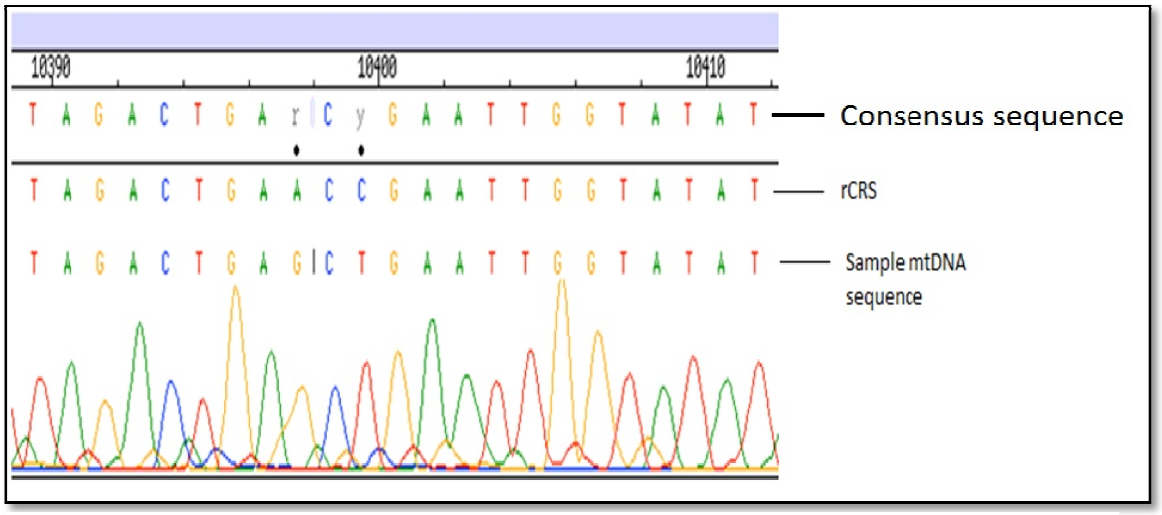


Fig 4.6: mtDNA variation at position 10398 (A/G) and 10400 (C/T), aligned with rCRS (Revised Cambridge Reference Sequence)

4.2.1.1 Mitochondrial Haplogroup

4.2.1.1.1 Mitochondrial Haplogroup Present in the entire sample

Mitochondrial haplogroup were identified according to mutation present in studied sample by using global human mitochondrial Phylotree.org (Van Oven, 2014).

Table 4.1: mtDNA Haplogroup present in all samples (HG refers to the Haplogroup)

Population	HG	Population	HG	Population	HG	Population	HG
S001	Z	B007	Z7	U012	M2	U037	G2b2a
S002	U2c	B008	M3	U013	M30d	U038	M5a
S003	Z3a1a	B009	A4	U014	M30b	U039	Z3a1a
S004	U7	B010	A5	U015	M5c2	U040	G2b
S005	M3c1a	B011	M3c1a	U016	M30d1	U041	Z3a
S006	Z3a1a	B012	Z3a	U017	Z3a1a	U042	D4
S007	M5c2	B013	R6a	U018	M5c2	U043	F1d
S008	M5	B014	R6a	U019	D4	U044	Z3a1a
S009	U7	B015	F1d	U020	Z3a1a	U045	M3

S010	Z7	B016	M5b2b	U021	M5c2	U046	M2
S012	D5a	B017	T2a1a	U022	F1d	U047	F1d
S014	M3c1a	B018	U2c1	U023	Z3a1a	U048	F1d
S015	D4	B019	D4	U024	M5b2b	U049	M33a1a
S016	Z3a1a	B020	D4b2	U025	F1d	U050	M3
S017	M	U001	Z3a1a	U026	M2b2	U051	Z3a1a
S018	M5a	U002	M35b	U027	F1g	U052	M3
S019	R6	U003	M35b	U028	F1a'c'f	U053	F2
S020	U7	U004	M5b2b	U029	W3a	U054	M5a
S021	D4	U005	D4a	U030	M33a1a	U055	Z3a1a
B001	M2a	U006	G2	U031	M9a1b1	U056	M30b
B002	Z3a1a	U007	M30	U032	Z3a1a	U057	M5c2
B003	F1d	U008	M34a	U033	F2	U058	G2a1d1
B004	M38a	U009	M5c2	U034	D4	U059	M30b
B005	M	U010	W	U035	M3c1a		
B006	R6a	U011	Z3a1a	U036	M30d		

According to SNP's Present in the sample mtDNA haplogroup were assigned e.g. Haplogroup Z3a1a was defined by Mutation T152C, G207A, T489C, A6752G, A8913G, T9090C, G9713A, T10208C, C10400T, T11075C, T13620T, T14783C, G15043A, T15784C, G15928A, C16260T, C16185T. Thus, 48 different sub-Haplogroups were found in all three populations. They all together make 22 Haplogroups. In the Shakya population have least number of sub-haplogroup i.e. 13 were present where in Bajracharya and Udaaya have 18 and 29 respectively.

4.2.1.1.2 Mitochondrial Haplogroup Frequency

Total 22 different mitochondrial Haplogroups were found. The mitochondrial Haplogroup Z was found to be most frequent constituting about 19.38% where as

Haplogroup A4, A5, D5, M9, M34, M38 were found to be least frequent constituting 1.02 % only.

Table 4.2: Mitochondria Haplogroup Frequency

HG	Frequency	HG	Frequency
A4	0.0102	**M9**	0.0102
A5	0.0102	**M30**	0.0714
D4	0.08162	**M33**	0.0204
D5	0.0102	**M34**	0.0102
F1	0.09182	**M35**	0.0204
F2	0.0204	**M38**	0.0102
G2	0.0408	**R6**	0.0408
M	0.0204	**U2**	0.0306
M2	0.0408	**U7**	0.0306
M3	0.0816	**W**	0.0204
M5	0.13262	**Z**	0.19386

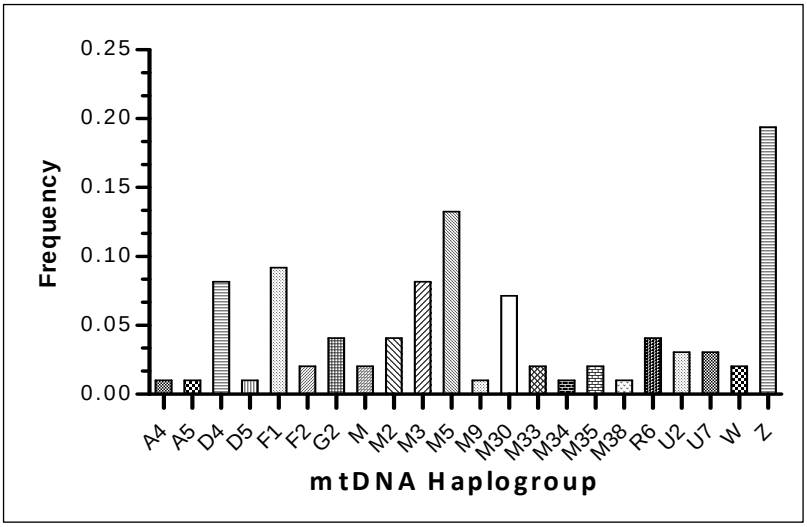


Fig 4.7: Graph of Mitochondrial Haplogroup Frequency

4.2.1.1.3 Comparative frequency of Mitochondrial Haplogroup Bajracharya, Shakya and Udaaya population

Table 4.3: Mitochondrial Haplogroup frequency of Shakya, Bajracharya and Udaaya population

Haplogroup	Bajracharya	Shakya	Udaaya	Haplogroup	Bajracharya	Shakya	Udaaya
A4	0.05	0	0	**M33**	0	0	0.0339
A5	0.05	0	0	**M34**	0	0	0.0169
D4	0.1	0.1053	0.0676	**M35**	0	0	0.0339
D5	0	0.0526	0	**M38**	0.05	0	0
F1	0.1	0	0.118	**R6**	0.15	0.0526	0
F2	0	0	0.039	**T2**	0.05	0	0
G2	0	0	0.0676	**U2**	0.05	0.0526	0
M	0.05	0.0526	0	**U7**	0	0.1578	0
M2	0.05	0	0.0508	**W**	0	0	0.038
M3	0.1	0.1053	0.0677	**Z**	0	0.0526	0
M5	0.05	0.1578	0.1565	**Z3**	0.1	0.1578	0.1863
M9	0	0	0.0169	**Z7**	0.05	0.0526	0
M30	0	0	0.1139				

In Bajracharya, Shakya and Udaaya population 14, 12 and 14 different haplogroup were identified. When comparing the frequency of mtDNA Haplogroup within three populations, Haplogroup Z3 is most frequent in Shakya and Udaaya having 15.78% and

18.63% whereas most frequent Haplogroup in Bajracharya was found to be R6 with 15% frequency.

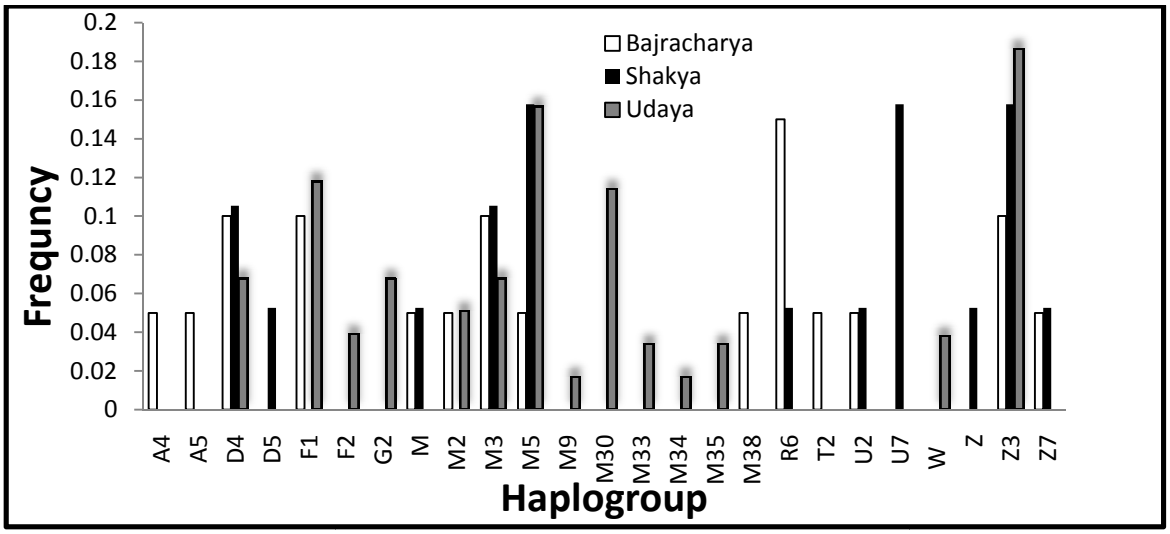


Fig 4.8: Graph representing comparative Haplogroup frequency among Bajracharya, Shakya and Udaaya Population

4.2.1.1.4 Region wise Genepool Contribution in Bajracharya, Shakya and Udaaya Population

Table 4.4 Region wise Genepool contribution in Bajracharya, Shakya and Udaaya

Geographical region	Bajracharya	Shakya	Udaaya
South Asia	0.3	0.3157	0.4736
Central Asia	0.15	0.263	0.1863
West Eurasia	0.25	0.263	0.038
East Asia	0.3	0.1579	0.3091

South Asia = India, Nepal, Pakistan, Bangladesh, Srilanka Maldives
East Asia = Tibet, China, Korea, Japan etc
Central Asia = Siberia, Magnolia, Tajikistan, Kazakhstan Turkmenistan, etc
West Eurasia = Europe, Middle East, Caucasia etc

Table 4.5: Distribution of different country according to geographical region (Source: Gaydan et al., 2007)

Region wise Genepool contribution in Bajracharya, Shakya and Udaaya population was dominantly by South Asia constituting 30%, 31.47% and 47.56% respectively. In Shakya population Central Asian and West Eurasian Genepool contributes by 26.3% and 15.79% by East Asia. In Bajracharya population Central Asian, West Eurasian and East Asian Genepool contribution was found to be 15%, 25% and 30% respectively. Similarly in Udaaya population Genepool contribution from central Asia, West Eurasia and East Asia was found to be 18.63%, 3.9%, and 30.91% respectively.

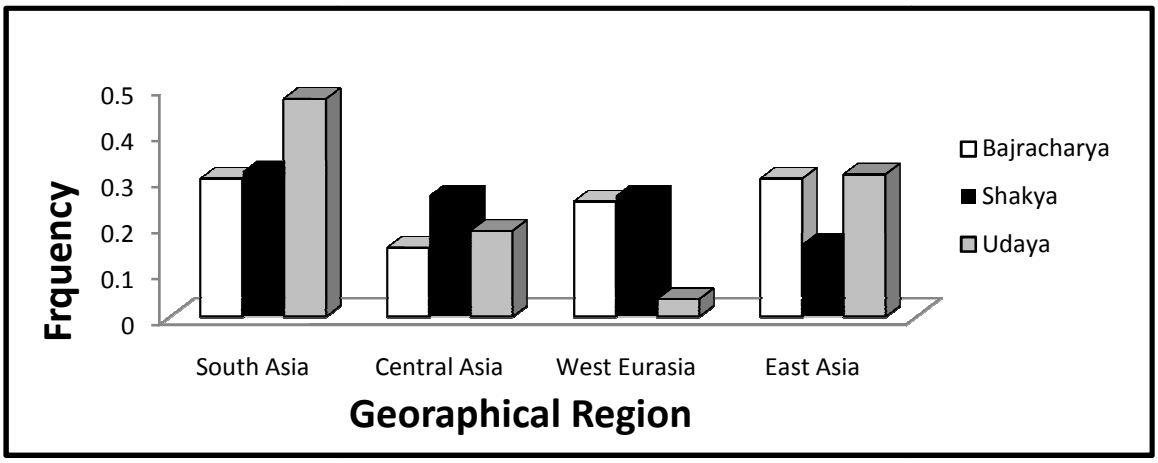


Fig 4.8: Region wise gene pool contribution in Bajracharya, Shakya and Udaaya population

4.2.1.5 Principal component analysis of Bajracharya, Shakya and Udaaya population based on mtDNA haplogroup frequency

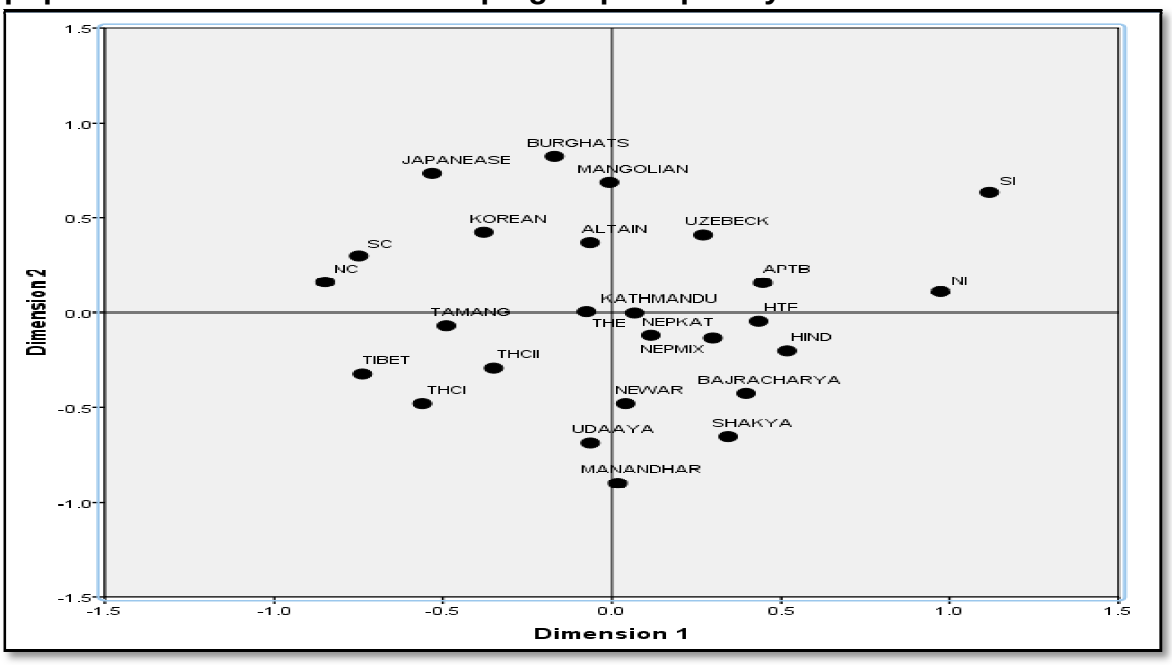


Fig 4.9: Principle Component Analysis (PCA) plot of Mitochondrial Haplogroup (Where HIND = Hindu India, HTE = Hindu Tarai, NC = North China, NI = North India, SC = South China, SI = South India TAP = Tribe of Andra Pradesh, THE = Tharu Eastern, THCI = Tharu Chitwan I, THCII = Tharu Chitwan II, NEPKAT = Nepal Kathmandu and NEPMIX = Nepal Mix)**.**

The Principle Component Analysis (PCA) was done by using software SPSS var 16.1 according to mtDNA Haplogroup frequencies with reference to previously studied neighboring population. Principle Component Analysis transformed the large no of linear

data of different variable into a simplified data metrics called components which gives better idea about correlation of different variable (Abdi and Williams, 2010).

4.2.1.1.7 Mitochondrial Phylogenetic tree of Newar Population according to Haplogroup

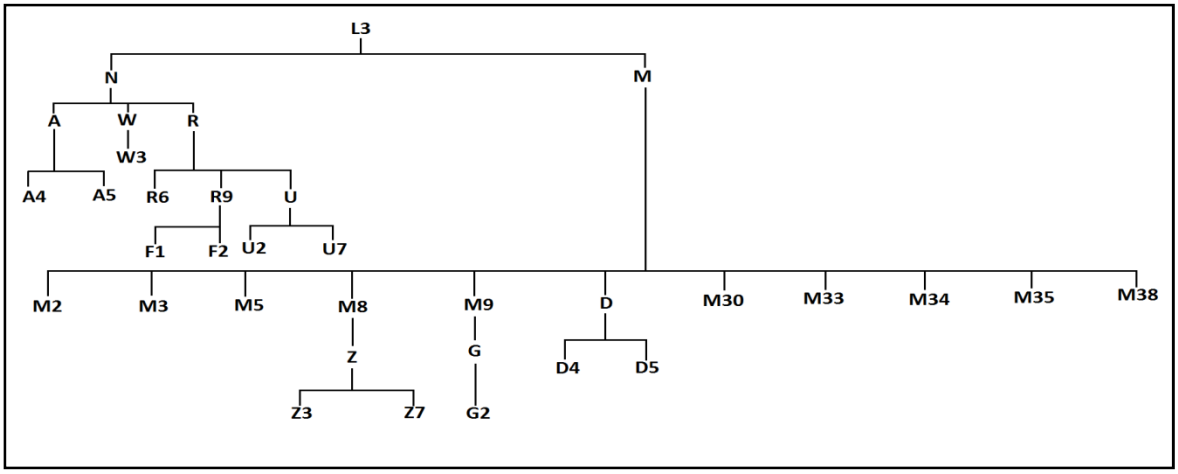


Fig 4.10: Mitochondrial Phylogenetic tree according to mt Haplogroup

L3 was the African haplogroup which was first migrated out of Africa. The Macrohaplogroup M which is restricted to Asia, most frequently in South Asia and Eurasian Macrohaplogroup N were directly descendent from L3. In our studied population most of the haplogroup belongs to Macrohaplogroup M. This tree showed that how the haplogroup present in our studied population were descendent from their ancestor haplogroup.

4.2.1.1.7 Genetic Diversity of Bajracharya, Shakya and Udaaya population

Table 4.6: Genetic Diversity indices of Shakya, Bajracharya and Udaaya population

Genetic Diversity Parameter	**Bajracharya (N= 20)**	**Shakya (N = 19)**	**Udaaya (N = 59)**
Nucleotide Diversity	0.002014/±0.00102	0.001726/±0.00088	0.002038/± 0.001021
Mean no. of Pairwise Difference	33.375012/±15.1906	28.602339/±13.9600	34.513747/± 15.24738
No. of Transition	172	137	460
No. of Transversion	15	11	144
Total no. of polymorphic sites	187	148	460

The Genetic Diversity parameters were calculated by using of Software Alrequin V 3.2. Analysis was carried out by using whole mitochondrial DNA sequence. Genetic diversity was analyzed by calculation Pair wise Distance, Nucleotide Diversity. Nucleotide Diversity measures the overall polymorphism in the DNA sequence of the all sample in a particular population. Mean No. of Pair wise Difference gives the idea about how the sequences of a particular population were correlated and they give an overview of diversity present in DNA sequence.

4.2.1.1.8 Neutrality Test

Table 4.7: Tajima's Test Statics of Shakya, Bajracharya and Udaaya Population

Population	M	S	Ps	θ	π	D
Shakya	19	147	0.008873	0.002539	0.001726	-1.337222
Bajracharya	20	186	0.011228	0.003165	0.002010	-1.513741
Udaaya	59	455	0.027466	0.005911	0.002076	-2.318596

Abbreviations: m = number of sequences, S = Number of segregating sites, p_s = S/m, θ = p_s/a_1, π = nucleotide diversity, and D is the Tajima test statistic.

The purpose of the test is to distinguish between a DNA sequence evolving neutrally and one evolving under a non-random process, including directional selection or balancing selection, demographic expansion or contraction, genetic hitchhiking, or introgression. Tajima's statistic computes a standardized measure of the total number of segregating sites in the sampled DNA and the average number of mutations between pairs in the sample. The two quantities whose values are compared are both method of moment's estimates of the population genetic parameter theta, and so are expected to equal the same value. If these two numbers only differ by as much as one could reasonably expect by chance, then the null hypothesis of neutrality cannot be rejected. Otherwise, the null hypothesis of neutrality is rejected. (Tajima et al., 1986). A negative Tajima's D value signifies an excess of low frequency polymorphism to relative to expectation indicating population size expansion through bottleneck or genetic drift.

4.2.1.1.9 Fu's Fs test of Neutrality

Table no. 4.8 Fu's Fs Test Statics of Shakya, Bajracharya and Udaaya Population

Population	M	Fs	Fs P-value
Bajracharya	20	-4.82709	0.0220
Shakya	19	-4.38565	0.0220
Udaaya	59	-24.17214	0.0000

Abbreviations: m = number of sequences

Whereas negative value of Fu's Fs test signifies an excess number of alleles, as would be expected from a recent population expansion. Fu's simulations suggest that FS is a more sensitive indicator of population expansion and genetic hitchhiking than Tajima's D.

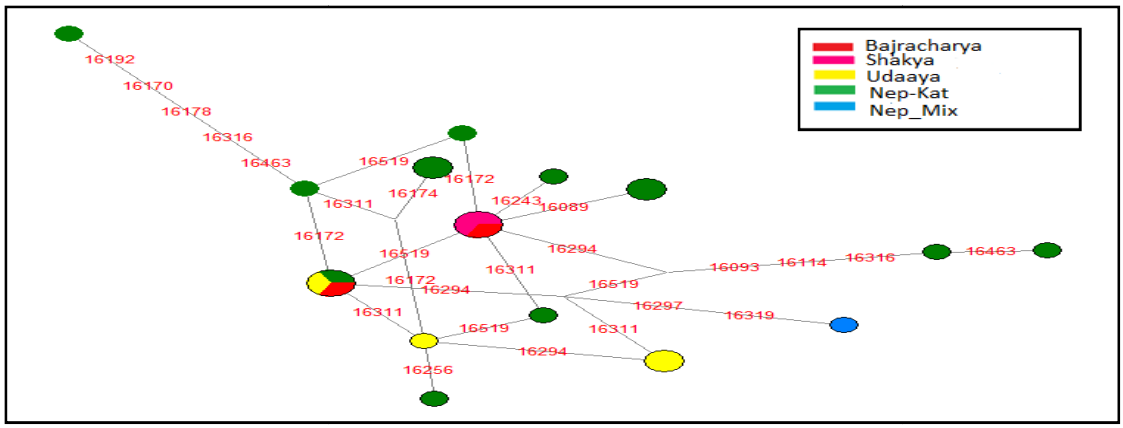
4.1.1.10 Phylogenetic Network of Different haplogroup Present In Bajracharya, Shakya and Udaaya population

The Phylogenetic Network provides the clue of the diversity of the population within a haplogroup. Besides this the length of the branch resembles with the frequency of Haplotype present in the given or the given sample. The size of the branch which is either population is directly proportional to the frequency of the population or haplogroup. The tree can be used for the estimation of coalescence time. The MJ tree was calculated by using software Network v 4.6.1.

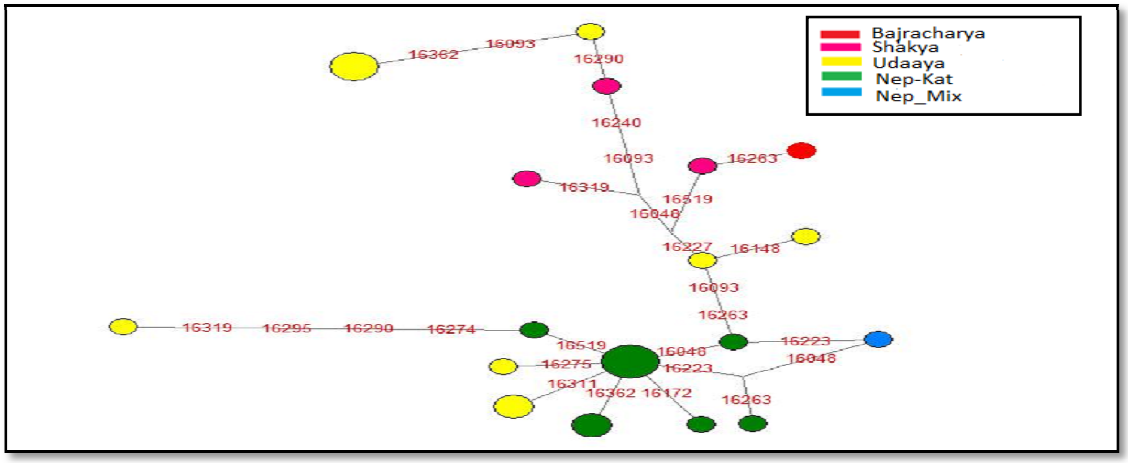
Haplogroup D4: Nep_Kat showed highest diversity having higher number of Haplotype such as 16192, 16178, 16316, 16463 which were absent in other population. Bajracharya, Shakya and Nep_Mix have least diversity. Udaaya have comparable diversity.

Haplogroup M5: Udaaya and Nep_kat showed highest diversity where as Bajracharya population have least no of braches means least diversity. Comparing with population size Shakya showed higher Haplotype diversity.

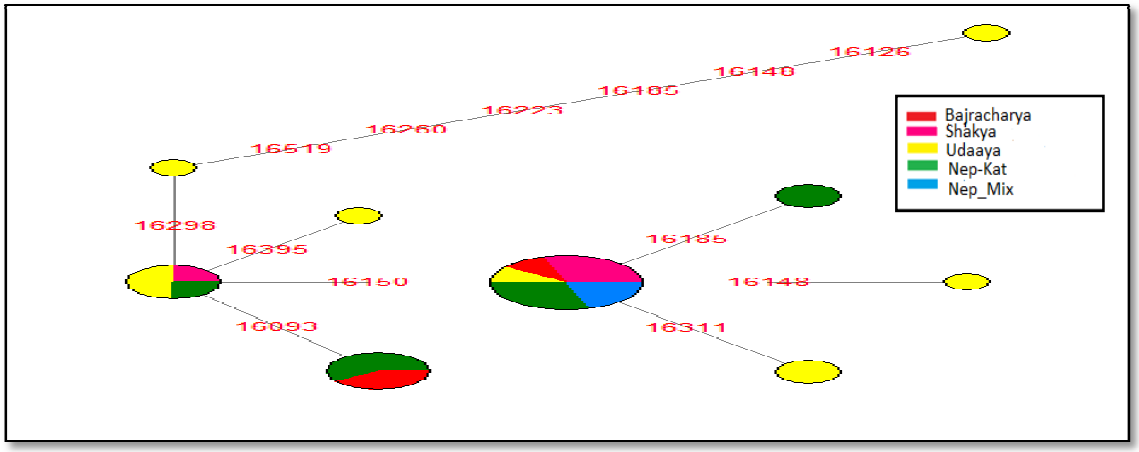
Haplogroup Z: Udaaya populations have highest number of branches with more Haplotype which clearly suggest the higher diversity. The Haplotype 16150 was common in all population.



Haplogroup D



Haplogroup M5



Haplogroup Z

Fig 4.11: Median joining Phylogenetic Network based on HVR1 of Different haplogroup Present in Bajracharya, Shakya and Udaaya population compared with Nepal_Kat and Nepal_Mix population (Wang et al., 2012).Three most commonly present haplogroup were selected for the calculation MJ tree. The sample size if Haplogroup D4, M5 and Z in Bajracharya (N= 2, 1 and 3), Shakya (N= 2, 3 and 4), Udaaya (N=4, 10 and 9), Nepal_Kat (N= 13, 10 and 10) and Nepal_Mix (N = 1, 1 and 2) respectively.

4.2.2 Y-Chromosomal Marker

The SNP data gives a comprehensive idea about where our study population belongs to in the world Phylogenetic tree. A total of 98 male individual from Shakya, Bajracharya and Udaaya sub caste of the Newar population were analyzed by using 20 Biallelic SNP markers. These Biallelic markers were universally accepted for the identification of paternal lineages of the particular population. Y-Chromosome Consortium 2002 develops the Nomenclature of Each of the SNP and related Haplogroup (Appendix no.4).

4.2.2.1 Frequency of Y-Chromosomal Haplogroup found in all population

Table 4.9: Frequency of Y-chromosomal Haplogroup

HG	Frequency
NO	0.01
H1a	0.02
P	0.04
F*	0.05
R1a	0.08
K*	0.08
J2a	0.09
R2	0.12
D1	0.12
O3ac1	0.38

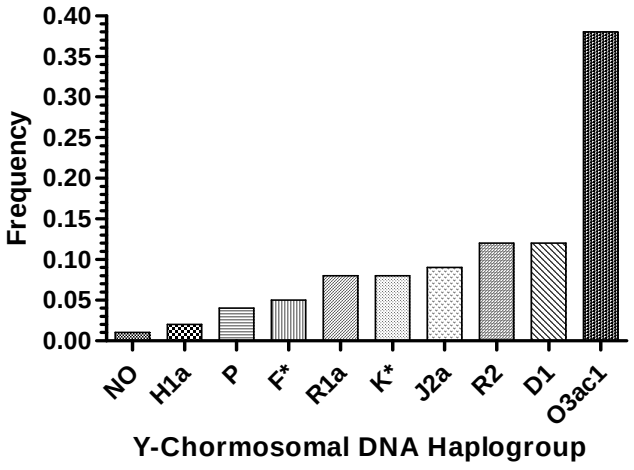


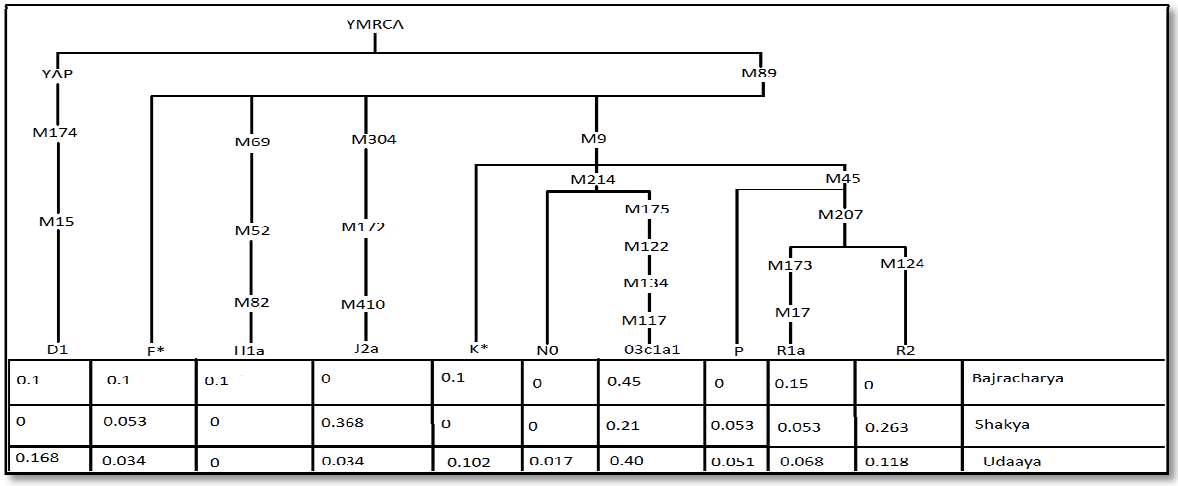
Fig 4.12: Graph of Y-chromosomal Haplogroup

In all three population, 13% population remained unresolved completely where as Haplogroup O3a3c1-M117 was found to be most frequent constituting 38%. Haplogroup R2-M124 and D1-M15 made equal contribution having frequency 12% of each Haplogroup. The frequency of Haplogroup NO-M214, H1a-M82, P-M45, R1a-M17, J2a-M410, was found to be 1%, 2%, 4%, 8%, and 9% respectively. The Haplogroup F-M89 (5%) and K-M9 (8%) were not completely resolved.

4.2.2.2 Comparative Y-Chromosomal Frequency in Bajracharya, Shakya and Udaaya Population

Total 10 haplogroup were identified by using 20 Bialleic Marker. In Bajracharya only 6 haplogroup were present where as in Shakya and Udaaya, 6 haplogroup and 9 haplogroup respectively. The haplogroup O3a3c1-M117 was found to be most frequent in Bajracharya and Udaaya with 45% and 40.71 % frequency. Haplogroup J2a-M410 was most frequent in Shakya population with frequency 36.8%.

Table 4.10: Y-chromosomal Haplogroup Frequency in Shakya, Bajracharya and Udaaya Population



0.1	0.1	0.1	0	0.1	0	0.45	0	0.15	0	Bajracharya
0	0.053	0	0.368	0	0	0.21	0.053	0.053	0.263	Shakya
0.168	0.034	0	0.034	0.102	0.017	0.40	0.051	0.068	0.118	Udaaya

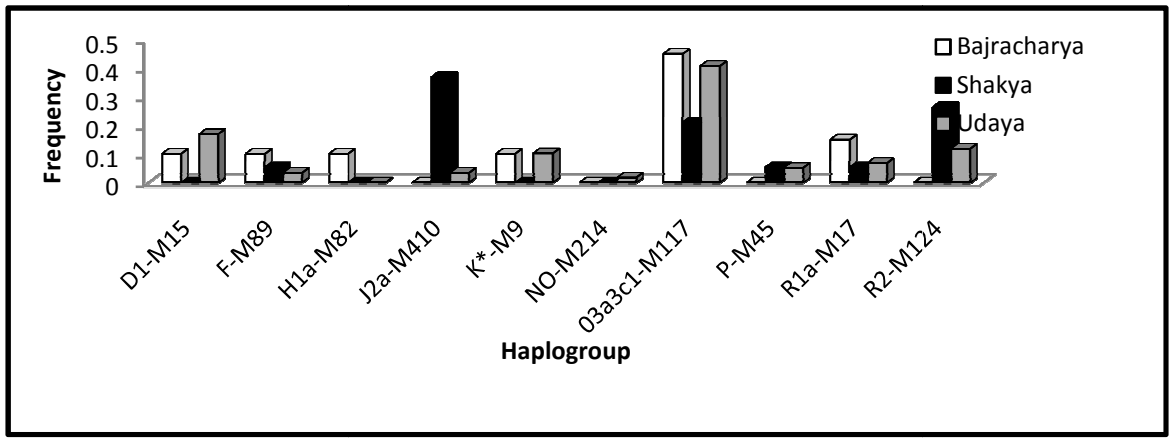


Fig 4.13: Y-chromosomal Haplogroup Frequency in Bajracharya, Shakya and Udaaya Population

4.2.2.3 Region wise Y-chromosomal Genepool contribution in Bajracharya, Shakya and Udaaya population

Table 4.11: Region wise Genepool Contribution in Bjracharya, Shakya and Udaaya population

Geographic Region	Population		
	Bajracharya	Shakya	Udaaya
South Asia	0.2	0.263	0.152
East/South East Asia	0.55	0.211	0.576
West Eurasian	0.15	0.0526	0.0847
Central Asia	0	0.421	0.0847
Undefined	0.1	0.0526	0.102

East Asia contribute highest level of gene pool in Bajracharya and Udaaya population with frequency 55% and 57.6% where as in Shakya population highest contribution was from central Asia with frequency 42.1%. In Bajracharya Population there was no contribution from central Asia. 5.26%, 10% and 10.2% remained unresolved in Shakya, Bajracharya and Udaaya population. In Shakya population Genepool contribution from East Asia was found to be 21.1%. Similarly the Central Asian gene pool in Udaaya population was found to be 8.47%. South Asian Genepool in Shakya, Bajracharya and Udaaya population was found to be 26.3%, 20% and 15.2% respectively. West Eurasian Genepool contribution was found 5.26%, 15% and 8.47% in Shakya, Bajracharya and Udaaya Population respectively.

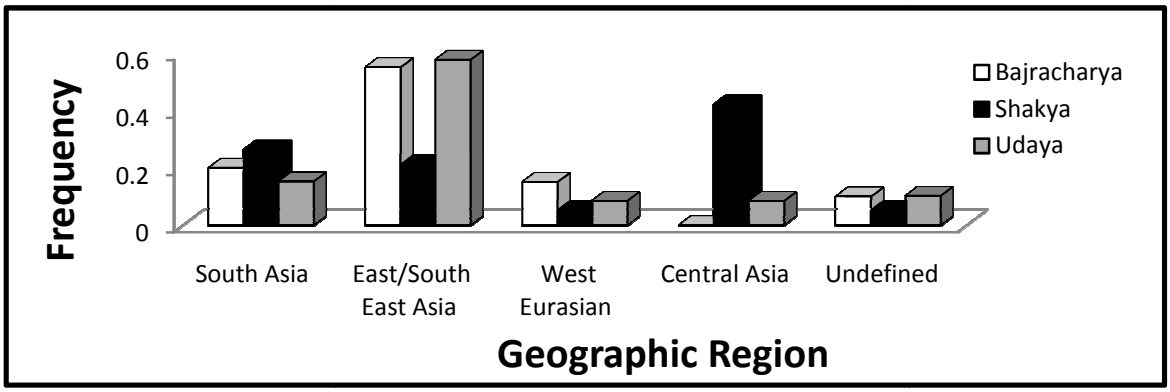


Fig 4.14: Region wise gene pool contribution in Bajracharya, Shakya and Udaaya population

4.2.2.4 Principal component analysis of Bajracharya, Shakya and Udaaya population based on Y chromosomal haplogroup frequency

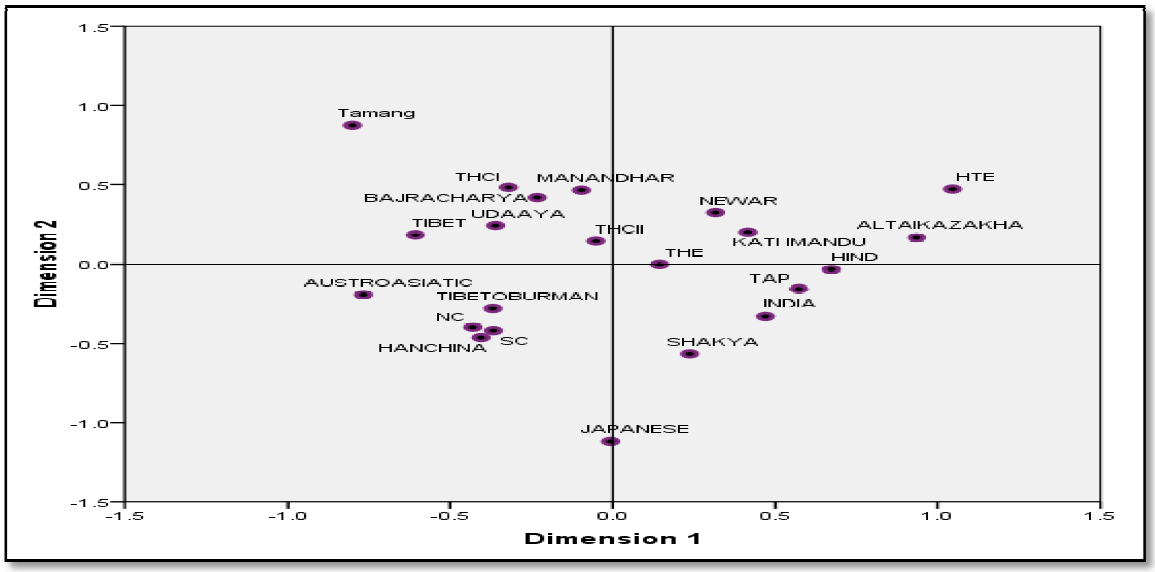


Fig 4.15: PCA plot of Newar population based on Y-chromosomal haplogroup (Where HIND = Hindu India, HTE = Hindu Tarai, NC = North China, SC = South China, TAP = Tribe

of Andra Pradesh, THE = Tharu Eastern, THCI = Tharu Chitwan I, THCII = Tharu Chitwan II).

Principle Component Analysis (PCA) was performed based on Y- Chromosomal haplogroup Frequency taking different neighboring population as a reference population. Analysis was carried out by using software Stastical Package for Social Science (SPSS 16.0). Bajracharya, Shakya and Udaaya population was considered as a single population during analysis. The studied population was lies with Tharu population and Tibetan population.

4.2.3 Autosomal Marker

Autosomal Markers are crucial for the population migration and evolution study. There may occur variation in the genotype according to geography and environment to adapt that particular condition. These markers provide good insight for the evolutionary aspects of human being. There are hundreds of different markers among some very important maker are high altitude adaptation maker, skin color variation, alcohol dependency, lactase persistence and population specific disease related markers. The high altitude markers are PPARA, EPAS-1, and EGLN-1. Other marker type are skin color variation marker such as SLC24A5, Hair morphology marker such as EDAR, Lactase Persiatance polymorphism, Epioxide hydrolase gene polymorphism marker (EPHX-13 and EPHX-14) and some population specific disease marker such as TGF-b1 polymorphism in breast cancer and Cardiomyopathies related 25 bp deletion in MYBPC-3 gene.

4.2.3.1 High altitude adaptation, Skin color, and Hair morphology polymorphism

The EPSA-1, PARA and EGLN-1 are hypoxia related high altitude markers and reflect Tibetan specific Haplotype (Peng et al., 2011). After PCR and sequencing, The SNP for the EPSA-1, PARA and EGLN-1 were traced by Auto assembler by comparing with reference sequence. The mutant allele is positively selected for the high altitude. EPAS-1(rs13419896) is located at chromosome 2 have allele G/A where G is ancestral. This polymorphism is intronic variants of HIF-2α gene. In highlander there is non-coding exon variant is found in EGLN-1 gene. EGLN-1(rs2275279) have variation A/T where A is ancestral allele. It located at chromosome 2:23727094. PPARA has the variation T/C where C is Ancestral found in chromosome no.22:46463431. It is non coding exon variants.

The skin color variation SLC24A5 has the variation A/G where G is ancestral allele. This is missense variant and located in chromosome 15:484264484. EDAR (rs3827760) is associated hair morphology or hair thickness. The variation is C/T where TT allele is ancestral. This is missense variants and located in chromosome 2:10951360.

Table 4.12: Allele Frequency different adaptation marker

Marker	Allele	Population		
		Bajracharya	Shakya	Udaaya
EPAS-1rs13419896(G/A)	GG	0.45	0.579	0.6101
	GA	0.55	0.421	0.3738
	AA	0	0	0.017
PARA(rs6520015-C/T)	CC	0.15	0.1052	0.0508
	CT	0.3	0.4736	0.4067
	TT	0.55	0.421	0.5423
EGLN-1(rs2275279)-A/T	AA	0.55	0.7894	0.6949
	AT	0.45	0.2106	0.305
	TT	0	0	0
SLC24A4(rs1426654)-G/A	GG	0.15	0.2621	0.1864
	GA	0.55	0.5789	0.6101
	AA	0.3	0.1579	0.2033
EDAR(rs3827760)-T/C	TT	0.65	0.5263	0.4915
	TC	0.25	0.421	0.4406
	CC	0.1	0.0526	0.0677

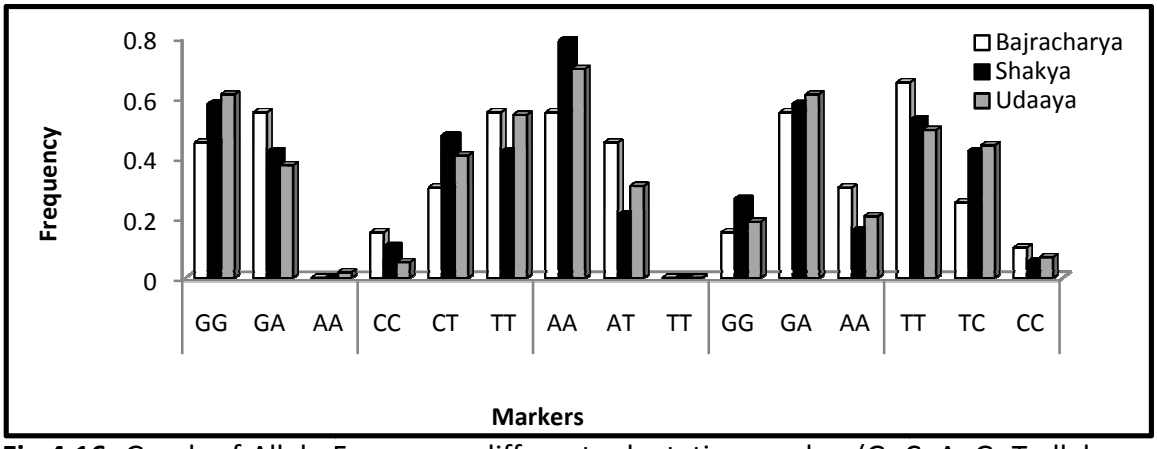


Fig 4.16: Graph of Allele Frequency different adaptation marker (G, C, A, G, T allele are ancestral to EPAS-1, PARA, EGLN-1, SLC24A5 and EDAR marker respectively.)

4.2.3.2 Lactase Persiatance Marker

Four causative single-nucleotide polymorphisms (SNPs) of Lactase persistence are approximately 13.9 kb upstream from LCT gene at the 100bp enhancer region: -13910C/T (rs4988235), -13907C/G (rs41525747), -13915T/G (rs41380347) and -14010G/C, have been subsequently identified as the candidate cis-acting elements based on genotype–phenotype association analyses and functional experiments.

Table 4.13: Allele Frequency Lactase persistence marker

Marker	Allele	Population		
		Bajracharya	Shakya	Udaaya
LCT(rs41525747)- C/G	CC	1	1	1
	CG	0	0	0
	GG	0	0	0
LCT(rs4988235)-C/T	CC	0.9	0.9473	0.949
	CT	0	0	0
	TT	0.1	0.0526	0.05.1
LCT(229)-C/T	CC	1	1	1
	CT	0	0	0
	TT	0	0	0
LCT(rs41525747)-A/C	AA	1	1	0.966
	AC	0	0	0.034
	CC	0	0	0

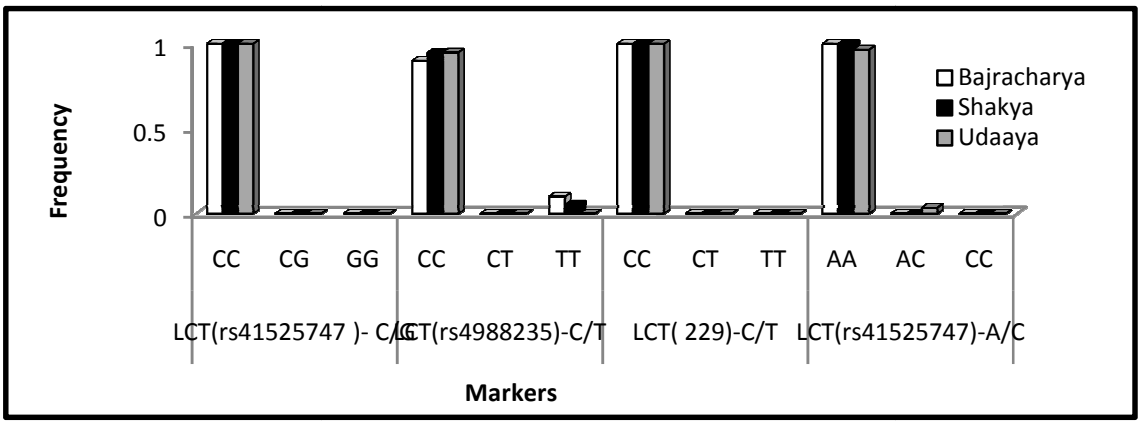


Fig 4.17: Graph of Lactase persistence Bjracharya, Shakya and Udaaya Population (C, C, C, A, are Ancestral Allele in rs41525747, rs4988235, LCT-229, and rs415257472 respectively).

4.2.3.3 Alcohol Dependence marker (EPHX-13(rs1051740)-T/C and EPHX-14(rs2234922)-G/A)

Two single nucleotide polymorphisms (SNPs) have been described in the coding region of the EPHX1 gene that produces two protein variants. The 113His– 139His haplotype of both Tyr113His and His139Arg polymorphisms are associated with decreased enzyme activity (Hartsfield et al., 1998).In vitro expression studies of cDNA revealed that the T/C mutation, which changes the tyrosine residue at codon 113 to histidine of exon 3, decreased EPHX enzymatic activity by approximately 40%, whereas a change of histidine to arginine at codon 139 of exon 4 increased EPHX activity by approximately 25% (Hassett et al., 1994). These polymorphisms are thought to be linked with protein stability.

Table 3.14: Allele Frequency EPHX marker

Marker	Allele	Population		
		Bajracharya	Shakya	Udaaya
EPHX-13(rs1051740)-T/C	TT	0.35	0.3684	0.3728
	TC	0.5	0.421	0.3228
	CC	0.15	0.2166	0.305
EPHX-14(rs2234922)-G/A)	GG	0.04	0.0525	0.0508
	GA	0.3	0.2631	0.3728
	AA	0.65	0.6842	0.5762

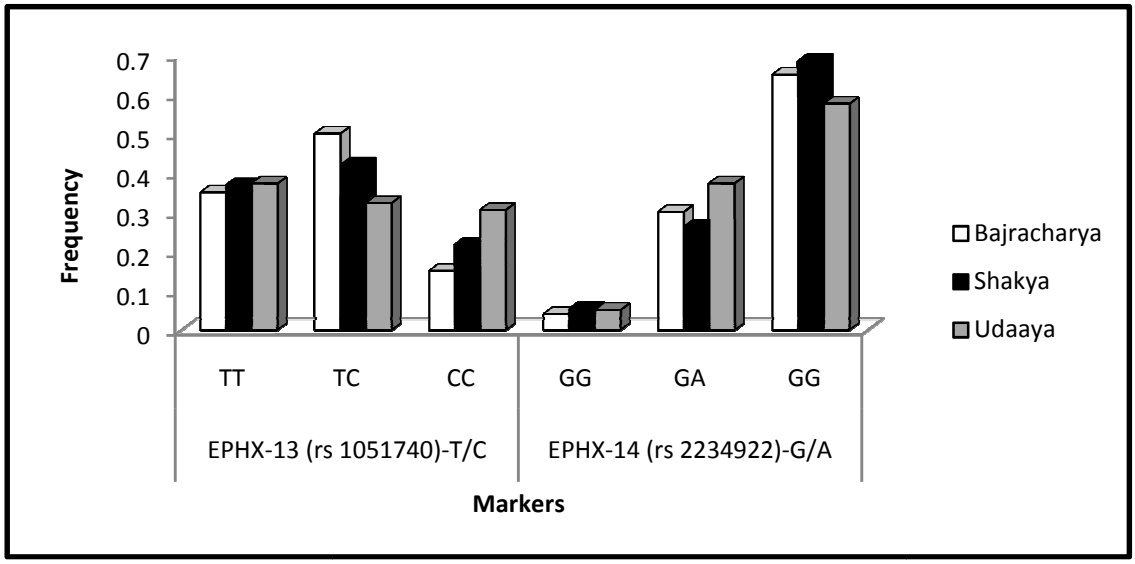


Fig 4.18: Graph Allele Frequency EPHX marker (T and G are the ancestral allele of EPHX13 and EPHX-14 Marker.)

4.2.3.4 Transforming growth factor (TGF-b1) and MYBPC

Table 4.15 Allele Frequency TGF-β1 and MYBPC-3 marker

Marker	Allele	Population		
		Bajracharya	Shakya	Udaaya
TGF-β1 (rs1982073)-C/T	CC	0.2223	0.2106	0.1525
	CT	0.4444	0.2106	0.4745
	TT	0.3333	0.5789	0.3389
TGF-β1 (rs1800471)-G/C	GG	0.95	1	0.9298
	GC	0.05	0	0.0754
	CC	0	0	0
MYBPC-3 (25bp deletion)	ND	1	1	1
	D	0	0	0

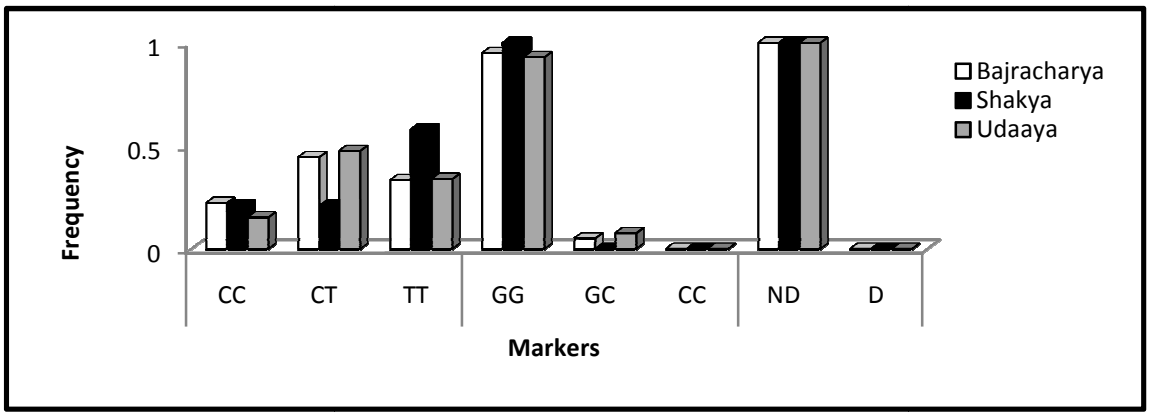


Fig 4.19: Graph of Allele Frequency TGF-β1 and MYBPC-3 Marker (C and G are the ancestral Allele of T TGF-β1 (rs1982073) and TGF-β1 (rs1800471)-G/C. ND refers No Deletion whereas D refers Deletion)

'TT' genotype at +29 (TGF-β1 (rs1982073-C/T) locus increased breast cancer risk irrespective of ethnicity. On the other hand, the presence of 'CC' genotype at +74 (TGF-β1 (rs1800471)-G/C loci reduced breast cancer risk significantly. Frequency of TT allele in Bajracharya, Shakya and Udaaya was found to be 22.23% , 21.06 % and 15.25% where as frequency of CC allele of TGF-β1 (rs1800471)-G/C variation is absent in all three population.

The 25bp deletion in MYBPC3 gene is found in South Indian population with high frequency and completely absent in Tibeto-Burman population (Dhandapany et al., 2009). In our population, this allele is completely absent. It means that Tibeto-Burman population and our population might be genetically closure than other Indian population.

CHAPTER 5

Discussion

Nepal has vast diversity anywhere in the world. Despite of small population size and geographical area it harbors hundreds of different ethnic group. Nepal is the melting pot of all major human races viz. Indo-Aryan, Tibeto-Burman, Austro-Asiatic as well as Dravidian. Kathmandu is the capital city an ancient city where development of Nepalese civilization occurred before 3,000 year ago. Newar are considered as the first inhabitant of the valley where they develop their own civilization, culture, tradition. Newars are postulated to be a mixture of Austro-Asiatic, Dravidian, Indo-Mongoloid and Aryan origins (Regmi, 1969.) Their unique Aryan-Mongloid phenotype suggests their complex Genepool. In order to resolve unanswered quest about genetic structure, affinities, diversity and migration pattern, of contemporary Newar people especially from Buddhist caste Shakya, Bajracharya and Udaaya, three different set of DNA marker viz. mtDNA, Y-chromosomal and Autosomal Marker were analyzed.

5.1 Mitochondrial Haplogroup Diversity

The Mitochondrial DNA undergoes uniparental inheritance so that there is no recombination occurs. Due to this reason, mtDNA can be used to trace maternal genetic lineages of the human population from different geographical region and there diversity (Brown et al., 1981). The whole mtDNA sequence provides deep information about the maternal lineages. Whole mitochondrial DNA of 16,659 bp was analyzed by comparing with rCRS (Cambridge reference Sequence) (Anderson et al., 1981) and Haplogroup grouping was done.

Total 22 different Mitochondrial Haplogroups were indentified. The frequency of Haplogroup F2, M33, M, W were found to be 2.04% each, U2 and U7 was 3.06%, R6, G2, M2 were 4.08%, D4 and M3 were 8.16%, F1, M5, M30, and Z3 were found 9.182%, 13.262%,7.414%,and 19.386% respectively. A4, A5, D5, M9, M34, M35, M38 were least frequent, consisting 1.02% each of the Haplogroup. Haplogroup A4, A5, F1 and F2 are descendent from Macro-Haplogroup N and are most prevalent in East Asia (Kong et al., 2003). U2, U7, R6 and W descendent from R Haplogroup most commonly found in West Eurasia (Palanichamy et al., 2004). M2, M3, M5, M30, M34, M35, M38 are found in South Asia (Thangraj et al., 2005). Haplogroup A4, A5, D4, G2 are East Asian specific (Kivisild et al., 2002). M38 and G2 are found in the Tharu population of Nepal (Fornarino et al., 2009). The Haplogroup Z is most frequent consisting total 19.386%.

Haplogroup A which was originated in Central Asia 50,000YBP was most frequent in Tibetan (21% = A4), Korean (14.6% = A5) Japanese (12%) (Derenko et al., 2007). The haplogroup A4 and A5 was found only in Bajracharya population with frequency 5% each, suggested that Bajracharya Share the common ancestral between Tibetan, Korean and Japanese in terms of haplogroup A.

The Haplogroup D which was descendent of Super-Haplogroup M was originated in East Asia before 40,000-60,000 YBP (Vile et al., 2008). The frequency of Haplogroup D4 in Japanese (Hokkaido) was 41.5% and in Korean (Seoul) was 39.8% (Derenko et al., 2007). The frequency of Haplogroup D5 in some Tibetan population exceeds more than 55% (Wen et al., 2004). In Xiuan Han Chinese population the frequency of Haplogroup D4 and D5 made 30%. Frequency of Haplogroup D4 and D5 in Shakya, Bajracharya and Udaaya population was found to be (10.53%, 5.26%), (10%, 0 %) 6.76%, 0%) respectively. The Haplogroup D4 was found to be common in all studies population where as D5 was found only in Shakya.

The time of origin of Haplogroup F was 43,000 YBP (Soares et al., 2009) in East Asia descendent from haplogroup R9. The Frequency of Haplogroup F in Indian Tibeto-Burman population was found to be 31% (Derenko et al., 2009) and in Tharu of Chitawan was 6% (Fornarino et al., 2009). Haplogroup F was completely absent in Shakya population where as in Bajracharya the frequency was 10 % (F1 haplogroup) and in Udaaya 11.8% for F1 and 3.9% for Haplogroup F2.

The Haplogroup G which was descendent from M12;G was originated 35,000 YBP (Soares et al., 2009) in East Asia. This Haplogroup was frequently found Tharu Chitawan and Tharu Morang constituting 23.3% and 12.5% (Fornarino et al., 2009), Newar (Wang et al., 2012). This Haplogroup was only present in Udaaya population constituting 3.9%.

Super Haplogroup M and its descendent Haplogroup M2, M3, M5, M9, M30, M33, M34, M35, M38 were originated in south Asia from 60,000 YBP and after on. They all together contribute more than 60% of South Asian Maternal Lineages (Thangraj et al., 2005). Haplogroup M3, M9, M30, M33 were commonly found in Newar (Wang et al., 2012) and Tharu Population (Fornarino et al., 2009).

The Haplogroup R6 was originated in West Eurasia which was found in Indian subcontinent with very low frequency (Metsplasu et al., 2004). In Bajracharya and Shakya population, the frequency of this Haplogroup was found to be 15% and 5.26% respectively. This result reveals the some of the maternal population in Bajracharya was migrated from West Eurasia via Western India.

T2 is the West Eurasian Haplogroup which was originated in Mesopotamia (Now Iraq) about 25,149 YPB (Sengupta et al., 2005). This Haplogroup is distributed Africa, Europe

and Middle East. It Accounts 10% of all European matrineal lineage (Kivilisid et al., 2004). In Bajracharya population this Haplogroup accounts 5%. This means that Genepool in Bajracharya was also contributed by European population.

The West Eurasian Haplogroup U2 was found in Shakya and Bajracharya population with frequency 5.26% and 5% respectively. Haplogroup U7 was found only in Shakya population with frequency 15.78%. The Haplogroup U2 was most frequent in south Asia. The Frequency of Haplogroup U7 in Gujarati population, Pakistani and overall India was found to be 12%, 5%, and 2% respectively. This Haplogroup reveals some of the population was first migrated from Europe to Indus valley and West India, from there they migrated to Nepal and accumulated with different other population which were migrated from other geographical region to make a single ethnic group.

The Haplogroup W was originated in West Eurasia about 23,900 YBP and has highest frequency in Northern Pakistan with 10% frequency (Metspalu et al 2007). This Haplogroup is also common in some European population such as Finish (9.6%), Hungarian (5.2%) and Macedonian (4%) respectively (eupedia.org). This Haplogroup was only found in Udaaya population with 3.8% frequency. This result suggested that there might be genetic affinity between Udaaya populations in some European population. Gene pool in Udaaya population was also contributed by the European population in some extent.

The Haplogroup Z was the most frequent in studied population. The frequency of this Haplogroup is more than 15%. Z, Z3 and Z7 sub-haplogroup were present. Among them Z3 is most frequent. Haplogroup Z was originating in Central Asia and distributed throughout East Asia with higher frequency. The result of comparative study showed that Haplogroup Z was most frequent only in Shakya and Udaaya population where as R6 haplogroup was most frequent in Bajracharya. The Z3 is the most frequent Haplogroup which is subclades of Haplogroup Z. Similar trend was found in the Newar and Kathmandu population (Gaydan et al., 2013). Presence of Haplogroup Z in the studied population with higher frequency reveals Central Asia Serve as main contributor of maternal genetic component. Frequent Presence of haplogroup Z in East Asia and Tibetan population suggest that migration of the haplogroup to East Asia then cross to the Himalayan Range and entered into the Kathmandu valley.

Frequent presence of Haplogroup U in Shakya population and R6 in Bajracharya population contributes higher Genepool from West Eurasia in both Shakya and Bajracharya where as in Udaaya population West Eurasian component is negligible. The haplogroup U7 was most common on Shakya population. The frequency of this haplogroup in Newar and Tamang is relatively high (Gaydan et al., 2013). This result can

reveal there is affinity between Tamang and Newar in some extent, especially with Shakya.

The presence different mt Haplogroup in the studied population which were originated and distributed into the different geographical region of the world gives an idea about how the maternal genetic component were migrated from different region of world. Then during the course of time these different population were accumulated to form an ethnic group having common culture, religion etc.

Region wise Genepool contribution in Shakya, Bajracharya and Udaaya population was dominantly by South Asia constituting 31.47%, 30% and 47.56% respectively. In Shakya population Central Asian and West Eurasian Genepool contributes by 26.3% and 15.79% by East Asia. In Bajracharya population Central Asian, West Eurasian and East Asian Genepool contribution was found to be 15%, 25% and 30% respectively. Similarly in Udaaya population Genepool contribution from central Asia, West Eurasia and East Asia was found to be 18.63%, 3.9%, and 30.91% respectively. Contribution of maternal gene pool in the Shakya, Bajracharya and Udaaya is highest from South Asia. But in Shakya population after South Asia, Central and West Eurasian genetic component were prevalent and East Asia contributes Least. Where as In Bajracharya, East Asia contributes after South Asia then followed by Central Asia. In Udaaya Population after South Asian genetic components, East Asian then Central Asian component were prevalent where as West Eurasian contributes least.

The Median Joining tree of three common haplogroup D4, M5, and Z showed that Udaaya population have more diverse in haplogroup D4 and Z where Bajracharya showed in haplogroup M5. The reference population Nepal_Kat showed more diversity in all the haplogroup. This might be due to larger sample size in Udaaya and Nepal_Kat population. The next region behind such result might be due to Nepal_Kat was not taken uniformly. There population component was not exactly mentioned. In Kathmandu valley is inhabited by large number ethnic group. Therefore, MJ phylotree showed more no of nodes in Nepal_Kat population.

Principle component analysis (PCA) reveals there is close genetic affinity between Bajracharya, Shakya, Udaaya, Manandhar, Newar and Kathmandu population (Gaydan et al., 2013), Terai Hindu and Indian Hindu (Fornarino et al., 2009), Nepal_Kat and Neoal_Mix (Wang et al., 2012) because theses population shares common Haplogroup. But, Newar and Kathmandu population does not reveal which ethnic group/caste was taken in the study, so it can not show exact affinity with our studied population. Most of Haplogroup were shared by these reference populations and our population so, they showed more affinity. These population lies in same Quadrate means they share common Haplogroup with similar frequencies. Indian population such as South Indian

and North Indian (Thangraj et al., 2005) have larger distance with our studies population but lies in dimension 2, means these population Share some haplogroup but there frequency is far difference. Some population such as Tharu, Tamang and our studies population lies in dimension 1, means they share some Haplogroup with similar frequencies.

5.1.1 Genetic Diversity Indices Based on Mitochondria DNA sequence

Udaaya have the more diversity then other two populations. The number of Haplogroup and there frequency also showed much more diversity in the Udaaya population. It might be due to Udaaya group consist of nine more subgroup where intermarriage between them is practiced. Due to this factor they have more diversity then others. The Nucleotide diversity is lowest in the Shakya with 0.001726 where as Bajracharya and Udaaya have 0.02014 and 0.02038. In Newar population it was found to be least with compare Tamang, Tibetan population with 0.0016 (Gaydan et al., 2013). This indicated there is lower diversity in the Shakya as well as Newar population then Bajracharya, Udaaya population.

5.1.2 Test of Neutrality

Tajima D and Fu's Fs test are neutrality Test. They give the idea about neutral theory of evolution. In the studied population the value of Tajima D test is negative that signifies an excess of low frequency polymorphism to relative to expectation indicating population size expansion after population bottleneck or purifying selection (Tajima 1969). If there is purifying selection, mutations will occur and accumulate at silent sites, but they aren't likely ever to become very common. Thus, there are likely to be lots of segregating sites, but not much heterozygosity. Whereas negative value of Fu's Fs test signifies an excess number of alleles, as would be expected from a recent population expansion. Fu's simulations suggest that FS is a more sensitive indicator of population expansion and genetic hitchhiking than Tajima's D.

The Tajima D value in low in the Shakya where as slightly high in the Udaaya population. This indicates that there is more expansion of the population means the population is undergoes more diversified in the Bajracharya and Udaaya then Shakya. This fact is further supported by Fu's Fs value. The value is very higher in Udaaya group. These differences might be due to different size of population. Such discrepancy was already reported (Derenko et al., 2007).

5.2 Y- Chromosomal Haplogroup Diversity

Y-chromosomal Biallelic SNP markers present in the Non-recombining Region (NRY) used to infer the Paternal Lineages. Total 20 different Biallelic markers were typed. Total 10 different Haplogroup were found.

The Haplogroup K*-M9 was originated 47,000 YBP in West Asia. This Haplogroup is also not a defining Haplogroup itself (Hammer et al., 2006). This Haplogroup was absent in Shakya population where as in Bajracharya and Udaaya frequency was 10% and 10.2% respectively.

Haplogroup NO-M214 was originated in North Eastern Asia about 41,000 YBP (Rootsi et al., 2004). This Haplogroup was absent in Shakya and Bajracharya population where as in Udaaya frequency was 1.7%. Means Udaaya population has affinity to North East population then other.

Haplogroup J2a-M410 was originated in West Eurasia about 41,000 YBP (Semino et al., 2004). This haplogroup is distributed throughout Caucasia, Greece, Central Asia and Levant. The frequency in Ingush population, Cyprus, Greece was found to be 88%, 56%, 37% and 23%. In Indian upper caste frequency of this Haplogroup was found to be 23% (Eupedia.org). In Shakya population this Haplogroup accounts 36.8% and in Udaaya 3.4% where as complete absent in Bajracharya population. The haplogroup J2 was most frequent around the central Asia. The hypothesis that there was migration of Shakya is from Central Asia. This haplogroup have good agreement with the hypothesis.

The most frequent Haplogroup in the studied population was O3a3c1-M117. This Haplogroup was originated in China about 30,000 YBP (shi et al., 2006). It is the terminal mutational Haplogroup of O3a3c-M134. The frequency of this Haplogroup in Tibetan, Tamang, Newars and Kathmandu population was found to be 28.8%, 86.5%, 21.20%, and 16% respectively (Gaydan et al.,2007). This Haplogroup was also common in Tharu constituting 33.3% (Fornarino et al., 2009). The Frequency of this Haplogroup in Shakya, Bajracharya and Udaaya Population was found to be 21%, 45%, and 40.7% respectively.

The Haplogroup P-M45 was originated in Central Asia about 27,000-41,000 YBP and commonly distributed throughout Central Asia (Derenko et al., 2006). Frequency of this Haplogroup in Turans population was 34.5%, in Altains was 28.3% (shoo et al., 2006). The frequency of this Haplogroup in Shakya and Udaaya population was found to be 5.3% and 5.1 % where as in Bajracharya this Haplogroup was absent.

The Haplogroup R1a-M17 referred as R-M168 was found in South Asia, West Asia and throughout Europe (Underhill et al., 2009). The frequency exceeds more than 60%in Indian Brahmin population (Sengupta et al., 2005). The frequency of this Haplogroup in Punjabi population was found to be 47% (Kivisild et al., 2003), Hindus from Tarai Nepal

was 69% (Fornarino et al., 2009), In Germany 20-30% and in Norway 20% (Keyser et al., 2005). In Shakya, Bajracharya and Udaaya population frequency was found to be 5.3%, 15% and 6.8% respectively.

Haplogroup R2-M124 is a South Asian specific Y-chromosomal haplogroup which was originated in Central Asia 12,000 YBP. 90 % of this Haplogroup was prevalent only in Indian subcontinent. The frequency of this haplogroup in Upper caste Indian population, Tibeto-Burman Indian, Astro-Asiatic, Dravidian was found to be 16.28%, 5.75%, 10.94%, 13.79% respectively (Sengupta et al., 2005). The frequency of this Haplogroup in Shakya and Udaaya was found to be 26.3% and 11.8% respectively. In Bajracharya this haplogroup was absent.

The South Asian haplogroup H1a was originated 25,000-45,000 YBP. The frequency of H1a in Newar, Kathmandu population, Tarai population and Andra-pradesh Tribes was 6%, 12% , 10.66%, 27.5% respectively and this Haplogroup is completely absent in Tibetan population (Gaydan et al., 2007). Similarly frequency of this Haplogroup in North Indian population was found to be 13.99% and this is the most frequently present Romani people of Europe which suggest the gene flow in Europe from South Asia (Rai et al., 2012). The frequency of haplogroup H1a in Bajracharya population was found to be 10% where this haplogroup was completely absent in Shakya and Bajracharya population.

The East Asian specific Haplogroup D1-M15 was originated 50,000YBP in East Asia or Central Asia and descendent from haplogroup D-M174. The frequency of this Haplogroup in Tibetan population was found to be 28.2% (Gaydan et al., 2009). The frequency of this haplogroup in Bajracharya and Udaaya population was found to be 10% and 16.9% respectively where as absent in Shakya population.

East Asia contribute highest level of gene pool in Bajracharya and Udaaya population with frequency 55% and 57.6% where as in Shakya population highest contribution was from central Asia with frequency 42.1%. In Bajracharya Population there was no contribution from central Asia. 5.26%, 10% and 10.2% remained unresolved in Shakya, Bajracharya and Udaaya population. In Shakya population Genepool contribution from East Asia was found to be 21.1%. Similarly the Central Asian gene pool in Udaaya population was found to be 8.47%. South Asian Genepool in Shakya, Bajracharya and Udaaya population was found to be 26.3%, 20% and 15.2% respectively. West Eurasian Genepool contribution was found 5.26%, 15% and 8.47% in Shakya, Bajracharya and Udaaya Population respectively.

Overall East Asia was found main Genepool contributor in Newar population. Among them Bajracharya have some different pattern. Their Genepool was from Central Asia. This result clearly suggested that Central Asian and East Asian population were migrated

toward Mongolian and Tibetan grassland and then they cross the Himalayan range and settled into Kathmandu valley. Some populations were migrated from West Eurasia to India and along with South Asian population they migrated toward the hill region and settled into Kathmandu valley. Then all population merged and development of new ethnic group.

Principle component analysis (PCA) of the Y-chromosomal Haplogroup reveals that there is close genetic affinity between Bajracharya, Udaaya, Tharu-CI and Tibetan It's due to the presence of Haplogroup O3ac1 in all population with high frequency. Shakya population was clustered with Indian and Central Asian population such as Altai-Kazak and Uzbek. It might be due to presence of haplogroup J2a in central Asian and R2 in Indian population with high frequency.

5.3 Variation in Autosomal Marker

Autosomal markers are distributed throughout all chromosomes and represent accumulated inheritance from all our ancestors. They can be used as an important marker for study of genetic diversity and human evolution. The adaptation towards the particular environment is lead by the genetic change in particular gene that is responsible to cope with particular environmental condition.

EPAS-1(rs13419896) is located at chromosome 2 have allele G/A where G is ancestral. This polymorphism is intronic variants of HIF-2α gene. Udaaya have the highest frequency of AA allele (1.7%) where as in Shakya and Bajracharya this allele is absent. The next important hypoxia related high altitude gene is EGLN-1. In highlander there is non-coding exon variant is found in EGLN-1 gene. EGLN-1(rs2275279) have variation A/T where A is ancestral allele. It located at chromosome 2:23727094. The TT allele is absent in all three studied population. PPARA has the variation T/C where C is Ancestral found in chromosome no.22:46463431. It is non coding exon variants. The frequency of CC allele is 42.1%, 55% and 54.23% in Shakya, Bajracharya and Udaaya population respectively. The presence of high altitude genotype is almost rare. This clearly suggests that there is no any positive selection for high altitude adaptation in Newar population.

The skin color variation SLC24A5 has the variation A/G where G is ancestral allele. This is missense variant and located in chromosome 15:484264484. In European population frequency of the AA allele is about 100% which is solely responsible white skin color and also found in high frequency in south Asian population. It has no role in the white skin coloration of East Asian population. It means that it is highly population specific. The frequency of AA allele is 15.79%, 30%, 20.33% in Shakya, Bajracharya and Udaaya population respectively.

EDAR (rs3827760) is associated hair morphology or hair thickness. The variation is C/T where TT allele is ancestral. This is missense variants and located in chromosome 2:10951360. The frequency of CC allele in Shakya, Bajracharya and Udaaya population was found to be 5.26%, 10% and 6.77% respectively. There no any significant level of frequency was seen. This means that studied population did not share this variation with East Asian population.

In some human population the activity of enzyme Lactase continued till adulthood and condition is called as Lactase persistence. There are different allele are responsible for this phenotype in different population. Theses studied variations were mostly found among African, West Eurasian and South Asian population with high frequency. rs4988235)-C/T is the West Eurasian polymorphism. The frequency of TT allele in Indo-European population is 18.4% where as in 0.8% in Tibeto-Burman population. LCT (rs41525747) - C/G is found mostly in African population (Romero et al., 2012).

The allele TT and GG or other polymorphism are absent in our studied population. It suggests that Newar have either no Lactase persistence or attributed by other unraveled polymorphism. In Tibetan population LP is attributed by other novel SNPs such as 13838G/A, 13906T/A and 13908C/T (Peng et al., 2012). These results suggest the close affinity between East Asian population and Newar population in terms of LP.

The alteration of enzyme activity results into Alcohol Dependency. EPHX-13(rs1051740)-T/C is associated with change of Tryosine to histidine at codon 113. It results into decrease of enzyme activity by approximately 40%. EPHX-14(rs2234922)-G/A polymorphism attributes to change of Histidine to Arginine at codon 139 results into increased enzyme activity by 25%. The T allele is risk allele seems having dominant genetic effect over C allele because both CC and CT genotype had decreased risk for the Alcohol Dependency with compared TT genotype (Bhasker et al., 2012). The frequency of TT, CT and CC allele is around 35%, 32% to 50% and 21 to 30% respectively in three studied population. In south Indian population frequency is 64%, 15% and 23 % where as in Bhotia population 21.6%, 54.1% and 24.3% respectively (Bhasker et al., 2012). The pattern of polymorphism is similar to Bhotia i.e. Tibeto-Burman population. It might be due to Alcohol is vital part of their ritual in Newar as well as most of the Tibeto-Burman Population.

TT genotype of TGF-β1 (rs1982073)-C/T variation is associated with increase breast cancer risk and CC genotype of TGF-β1 (rs1800471)-G/C variation is associated with reduced breast cancer risk. The frequency of CC significantly high in North Indian population where is completely absent in Tibeto-Burman population. Frequency of TT allele in Shakya, Bajracharya and Udaaya was found to be 21.06%, 22.23% and 15.25%

where as frequency of CC allele of TGF-β1 (rs1800471)-G/C variation is absent in all three population. These results have good agreement with Tibeto-Burman population.

The 25bp deletion in MYBPC3 gene is found in South Indian population with high frequency and completely absent in Tibeto-Burman population (Dhandapany et al., 2009). In our population, this allele is completely absent. It means that Tibeto-Burman population and our population might be genetically closure than other Indian population.

The analysis of Mitochondrial DNA haplogroup and its diversity, Y-chromosomal Marker and Autosomal Marker reveal that the Newar Ethnic group was not made by single Ancestor population. The ethnic group is the result of accumulation different population within a single ethnic group. Even within the caste genetic diversity is very high. The trend of maternal lineage and paternal lineage showed that they originated different geographical region. The male founder population and female founder population migrated from different geographical region and admixed to make a new ethnic group. Autosomal Marker also has different affinity with different population. Lactase Persiatance result showed more affinity with Tibetan population where as Skin color variation showed that they had more affinity toward South Asian/West Eurasian population. These data suggested that during course of time Newar population develop a unique adaptation type to their living environment. There habitation is different from Tibetan as well as other population such as South Indian or West Eurasian so they adapt different environment. This indicates they have a long history of developing their genetic history.

CHAPTER 6

Summary and Conclusion

6.1 Summary

Under this study, The Genetic Structure and Diversity of Shakya (N=19), Bajracharya (N= 20) and Udaaya (N= 59) population of Newar Ethnic group from Kathmandu Valley was done. 10 ml of blood was collected from 98 samples. Then DNA was isolated by standard Phenol: Chloroform Method. Whole Mitochondrial DNA was amplified using 24 sets of specific primer as well as Y-Chromosomal and Autosomal Marker was amplified by their specific marker. Sequencing of Each amplicon was done. Then editing and alignment of sequence was performed. For Y-Chromosomal and Autosomal Marker the position of SNP was searched by using Sequence Analysis V5.2 Software. Editing, alignment for Mitochondrial DNA sequence, were done by using rCRS as a reference sequence on the Auto-assembler v2.1 software. Each Mutation ware noted. According to the mutation present on the mtDNA, Haplogroup was assigned by using global Human Phlyotree.org (Van Oven, 2013). In the Y-chromosomal if SNP defined mutation was found in the given sample was considered as Particular Haplogroup.

Frequency of each Haplogroup present on population was calculated. The most frequent mtDNA Haplogroup was found to be Z3 with 18% frequency and Y-Chromosomal DNA was found to be O3-M117 with frequency 38.587%. Genetic Diversity indices such as Nucleotide Diversity, Haplotype Diversity, Tajima Neutrality and Fu's FS neutrality were calculated by using population genetics software Alrequin V3.1. This result showed that there was higher level of diversity among and within the population. The result of neutrality of test shows there was more diversification, indicates that there these population was expanding from recent genetic drift or the get cluster together. Then Regionwise-genepool was calculated. The highest Genepool contribution was from South Asia for maternal lineages and East Asia for Paternal Lineages. Then genotype frequency of Autosomal marker was performed. The result showed that there is no any exact relationship with any particular population such as Tibetan, Indian. Some of Autosomal genotype resembled with Tibetan population such as lactase Persiatance, TGF-β, MYBPC3-25bp deletion where as some with Indian Such as EDAR, SLC24A5 , High altitude adaptation marker (EPAS-1, EGLN-1 and PARA) etc. Mostly affinity was seen with Indian Tibeto-Burman population.

6.2 Conclusion

The study on the Shakya, Bajracharya and Udaaya caste of Newar population by using Mitochondrial, Y-chromosomal and Autosomal Marker reveal the very high genetic diversity among and within the population. The south Asia serve as highest contributor of maternal genetic component where as East Asia serve as main contributor for paternal Lineages. In Shakya population highest paternal Genepool was contributed from central Asia.

This result indicates that there is a genetic admixes between male founder population and female founder population. That admixture gives rise to a new population with highly diverse in terms of their genetic component. Autosomal Marker showed that there unique was genetic diversity that does not completely resemble with any other population. Some marker showed Tibeto-Burman type genetic structure where as some showed Indo-European. This also confirmed that there is genetic admixture from both the population. The high genetic diversity may be due to If the population is older and accumulating more mutation and if there is gene flow from other population had takes place. It can be concluded that the ethic group was formed by accumulating different population within a single ethic group with common language, culture etc.

This type of research will open new verse of anthropology, where Molecular technique can be use to trace out the evolutionary and migratory history of human being. Even descendent from single common ancestor, now there may be hundreds of different haplogroup are present in different geographical area that defines structure of particular population. One haplogroup defines one population. Hence Newar constitute hundreds of different population to make a single ethnic group.

References

- Abdi H, and Williams LJ, (2010)." Principal component analysis." Wiley Interdisciplinary Reviews: Computational Statistics, 2: 433 459. doi:10.1002/wics.101.
- Abu-Amero KK, González AM, Larruga JM, Bosley TM, and Cabrera VM (2007)" Eurasian and African mitochondrial DNA influences in the Saudi Arabian population" BMC Evolutionary Biology2007, 7:32 doi:10.1186/1471-2148-7-32
- Anderson S, Bankier AT, Barrell BG, De Bruijn MHL, Coulson AR, Drouin J, Eperon IC, et al (1981) Sequence and organization of the human mitochondrial genome. Nature 290:457-465
- Aprelikova O, Chandramouli GV, Wood M, Vasselli JR, Riss J, et al. (2004) Regulation of HIF prolyl hydroxylases by hypoxia-inducible factors. J Cell Biochem 92: 491–501
- Arredi B, Poloni ES, Paracchini S, Zerjal T, Fathallah DM, Makrelouf M, Pascali VL, Novelletto A and Smith CT (2004) " A Predominantly Neolithic Origin for Y-Chromosomal DNA Variation in North Africa" Am. J. Hum. Genet. 75:338–345, 2004
- Ballinger SW, Schurr TG, Torroni A, Gan Y-Y, Hodge JA, Hassan K, Chen K-H, et al (1992) Southeast Asian mitochondrial DNA analysis reveals genetic continuity of ancient mongoloid migrations. Genetics 130:139-152.
- Barrell BG, Anderson S, Bankier AT, De-Bruijn MLA, Chen E, Coulson AR, Drouin J, Epperon IC, Nierlich DP, Roe BA, Sanger F, Schreier AJ, Smith H, Staden R, and Young IG, (1980) "Different pattern of codon recognition by mammalian mitochondrial tRNAs" Proc. Natd. Acad. Sci. USA Vol. 77, No. 6, pp. 3164-3166
- Behar DM, Villems R, Soodyall H, Blue-Smith J, Pereira L, Metspalu E, Scozzari R, Makkan H, Tzur S,Comas D. 2008b. "The dawn of human matrilineal diversity". Am J Hum Genet 82(5):1130-1140.
- Botchkarev VA, Fessing MY. 2005. Edar signaling in the control of hair follicle development. J Investig Dermatol Symp Proc 10:247–251.
- Bandelt H-J, Macaulay V, and Richards M, (2006) "Human mitochondrial DNA and the evolution of homo sapiens" Berlin; New York: Springer. xiv, 271 p. p.

- Basu Mallick C, Iliescu FM, Mols M, Hill S, Tamang R, et al. (2013) ”The Light Skin Allele of SLC24A5 in South Asians and Europeans Shares Identity by Descent". PLoS Genet 9(11): e1003912. doi:10.1371/journal.pgen.1003912
- Beall CM, (2007) "Detecting natural selection in high-altitude human populations" Respir Physiol Neurobiol 158: 161–171.
- Bermisheva MA, Viktorova TV, and Khusnutdinova EK, (2003) "Polymorphism of human mitochondrial DNA" Russian Journal of Genetics 39(8):849-859.
- Bobillo MC, Zimmermann B, Sala A, Huber G, Röck A, Bandelt HJ, Corach D, Parson W, (2009) "Amerindian mitochondrial DNA haplogroups predominate in the population of Argentina: towards a first nationwide forensic mitochondrial DNA sequence database” Int J Legal Med (2010) 124:263–268 doi: 10.1007/s00414-009-0366-3
- Borkar M, Ahmad F, Khan F, Agrawal S, (2011) "Paleolithic spread of Y-chromosomal lineage of tribes in eastern and northeastern India." Ann Hum Biol. 2011 Nov; 38(6):736-46. doi: 10.3109/03014460.2011.617389.
- Brown GG, Castora FJ, Frantz SC, and Simpson MV, (1981) " Mitochondrial DNA Analysis of DNA Sequence Data, polymorphism: evolutionary studies on the genus Rattus. Ann. NY Acad. Sci. 361 135-153
- Bruick RK, McKnight SL, (2001) "A conserved family of prolyl-4-hydroxylases that modify HIF". Science 294: 1337–1340
- Budowle B, Wilson MR, DiZinno JA, Stauffer C, Fasano MA, Holland MM, and Monson KL, (1999) "Mitochondrial DNA regions HVI and HVII population data" Forensic Science International 103(1):23-35.
- Cann RL, Stoneking M, and Wilson AC, (1987) "Mitochondrial DNA and Human Evolution" Nature 325(6099):31-36.
- Cann RL, Wilson AC (1983) Length mutations in human mitochondrial DNA. Genetics 104:699-711
- Cann RL, Stoneking M, Wilson AC, (1987), "Mitochondrial DNA and human evolution", *Nature* 325 (6099):3136,Bibcode:1987Natur.325.31C, doi:10.1038/325031a0 .
- Cann, RL, (2001) "Genetic clues to dispersal in human populations: retracing the past from the present" Science291: 1742–1748.

- Carlson CS, Thomas DJ, Eberle MA, Swanson JE, Livingston RJ, Rieder MJ, Nickerson DA, (2005) "Genomic regions exhibiting positive selection identified from dense genotype data" Genome Res 15:1553–1565.
- Cavalli-Sforza Ll, Menozzi P, and Piazza A (1994) "The History and Geography of Human Genes". New Jersey, Usa: Princeton University Press.
- Chennakrishnaiah S, Perez D, Gayden T, Rivera L, Regueiro M, Herrera RJ.(2013) "Indigenous and foreign Y-chromosomes characterize the Lingayat and Vokkaliga populations of Southwest India"Gene. 2013 Sep 10;526(2):96-106. doi: 10.1016/j.gene.2013.04.074.
- Chen Y-S, Torroni A, Excoffier L, Santachiara-Benerecetti AS, Wallace DC (1995) Analysis of mtDNA variation in African populations reveals the most ancient of all human continentspecific haplogroups. Am J Hum Genet 57:000-000
- Chandrasekar A, Kumar S, Sreenath J, Sarkar BN, Urade BP, et al. (2009) "Updating Phylogeny of Mitochondrial DNA Macrohaplogroup M in India: Dispersal of Modern Human in South Asian Corridor" PLoS ONE 4(10): e7447. doi: 10.1371/journal.pone.0007447
- Chiaroni J, Underhill PA, Cavalli-Sforza LL: Y (2009) "Chromosome diversity, human expansion, drift, and cultural evolution" Proc Natl Acad Sci USA2009; 106: 20174–20179.
- Chunyan Mou, Helen A. Thomason, Pamela M. Willan, Christopher Clowes, W. Edwin Harris, Caroline F. Drew, Jill Dixon, Michael J. Dixon,1, and Denis J. Headon (2008) "Enhanced Ectodysplasin-A Receptor (EDAR) Signaling Alters Multiple Fiber Characteristic to Produce the East Asian Hair Form" HUMAN MUTATION 29(12), 1405 1411, 2008
- Conrad DF, Keebler JE, DePristo MA, Lindsay SJ, Zhang Y, Casals F, Idaghdour Y., Hartl CL, Torroja C, Garimella KV, Zilversmit M, Cartwright R, Rouleau GA, Daly M, Stone EA, Hurles ME, Awadalla P, (2011). "Variation in genome-wide mutation rates within and between human families". *Nature Genetics* 43 (7): 712–4. doi:10.1038/ng.862
- Cordaux R, Weiss G, Saha N, and Stoneking M (2004) "The Northeast Indian Passageway: A Barrier or Corridor for Human Migrations?" Mol. Biol. Evol.21 (8):1525–1533. 2004 doi:10.1093/molbev/msh151

- Cutbush M, Mollison Pl, And Parkin Dm (1950) "A New Human Blood Group" Nature 165: 188-189.
- Cruciani F, Trombetta B, Massaia A, Destro-Bisol G, Sellitto D,and Scozzari R, (2011) "A Revised Root for the Human Y Chromosomal Phylogenetic Tree: The Origin of Patrilineal Diversity in Africa" The American Journal of Human Genetics88, 814–818, June 10, 2011
- Clayton DA, Walbrec MW, (1981) "Sequence and properties of the human KB cell and mouse L cell D-loop regions of mitochondrial DNA". Nucl. Acids Res. 9: 5411-5421.
- D'Angelo G, Duplan E, Boyer N, Vigne P, Frelin C, (2003) "Hypoxia up-regulates prolyl hydroxylase activity: a feedback mechanism that limits HIF-1 responses during reoxygenation" J Biol Chem 278: 38183–38187.
- Dhandapany PS et al., (2009) "commonMYBPC3(cardiac myosin binding protein C) variant associated with cardiomyopathies in South Asia"Nature Genetics doi:10.1038/ng.309
- Derenko M, Malyarchuk B,, Denisova G, Perkova M, Rogalla U, Grzybowski T, Khusnutdinova E, Dambueva I, Zakharov I, (2012) " Complete Mitochondrial DNA Analysis of Eastern Eurasian Haplogroups Rarely Found in Populations of Northern Asia and Eastern Europe" PLoS ONE 7(2): e32179. doi:10.1371/journal.pone.003217
- Di Giacomo F, Luc F, Popa LO, Akar N, Anagnou N, Banyko J, Brdicka R, Barbujani G, et al (2004). "Y chromosomal haplogroup J as a signature of the post-neolithic colonization of Europe". *Human Genetics*. 115 (5): 357–71.
- Dong, SK, Long RJ, & Kang M Y, (2003) "Milking and milk processing: traditional technologies in the yak farming system of the Qinghai-Tibetan Plateau, China" Int. J. Dairy. Techol 56, 86–93.
- Ebersberger I, Metzler D, Schwarz C, Pääbo S (June 2002). "Genomewide comparison of DNA sequences between humans and chimpanzees". *Am. J. Hum. Genet.*70 (6): 1490–7. doi:10.1086/340787
- Excoffier L, Laval G, & Schneider S, Arlequin (version 3.1) (2005) "an integrated software package for population genetics data analysis" .Evol. Bioinform. Online1, 47–50.

- Fornarino S, Pala M, Battaglia V, Maranta R, Achilli A, Modiano G, et al., (2009) "Mitochondrial and Y-chromosome diversity of the Tharus (Nepal): a reservoir of genetic variation" BMC Evol. Biol.9, 154.
- Gayden T, Mirabal S, Cadenas, A M, Lacau H, Simms, T M, Morlote, D, et al., (2009) "Genetic insights into the origins of Tibeto-Burman populations in the Himalayas". J. Hum. Genet.54, 216–223.
- Gazi NN, Tamang R, Singh VK, Ferdous A, Pathak AK, Singh M, Anugula S, Veeraiah P, Kadarkaraisamy S, Yadav BK, Reddy AG, Rani DS, Qadri SS, Singh L, Chaubey G, Thangaraj K, (2013) "Genetic structure of Tibeto-Burman populations of Bangladesh: evaluating the gene flow along the sides of Bay-of-Bengal"
- Gunz P, Bookstein FL, Mitteroecker P, Stadlmayr A, Seidler H, and Weber GW, (2009) "Early modern human diversity suggests subdivided population structure and a complex out-of-Africa scenario" Proceedings of the National Academy of Sciences of the United States of America 106(15):6094-6098.
- Hammer MF, (1994) "A recent insertion of an Alu element on the Y chromosome is a useful marker for human population studies" Mol Biol Evol 11:749-761
- Hammer MF, and Horai S, (1995) "Y Chromosomal DNA Variation and the Peopling of Japan" Am. J. Hum. Genet. 56:951-962, 1995
- Hammer et al. (2005) "Dual origins of the Japanese: common ground for hunter-gatherer and farmer Y chromosomes," The Japan Society of Human Genetics, 2005
- Hassett C, Aicher L, Sidhu JS, Omiecinski CJ, (1994) "Human microsomal epoxide hydrolase: Genetic polymorphism and functional expression in vitro of amino acid variants" Hum Mol Genet 1994; 3(3):421–428.
- Hartsfield JK, Sutcliffe MJ, Everett ET, Hassett C, Omiecinski CJ, Saari JA, (1998) "Assignment of microsomal epoxide hydrolase (EPHX1) to human chromosome 1q42.1 by in situ hybridization" Cytogenet Cell Genet 1998; 83(1–2):44–45.
- Harvey CB, Pratt WS, Islam I, Whitehouse DB, Swallow DM, (1995) "DNA polymorphisms in the lactase gene: linkage disequilibrium across the 70-kb region" Europ. J. Hum. Genet. 3: 27-41, 1995
- Heng HH, 2009). "The genome-centric concept: resynthesis of evolutionary theory". *BioEssays* 31 (5): 512–25. doi:10.1002/bies.200800182

- Hertzberg M, Mickleson KNP, Serjeantson SW, Prior JF, Trent RJ, (1989) "An Asian-specific 9-bp deletion of mitochondrial DNA is frequently found in Polynesians" Am J Hum Genet 44:504-510
- Howell N, Elson JL, Howell C, Turnbull DM, (2007) "Relative rates of evolution in the coding and control regions of African mtDNA" Mol Biol Evol 24(10):2213-2221.
- Horai S, Kondo R, Nakagawa-Hattori Y, Hayashi S, Sonoda S, Tajima K (1993) "Peopling of the Americas, founded by four major lineages of mitochondrial DNA" Mol Biol Evol 10:23-47
- Horai S, Matsunaga E (1986) "Mitochondrial DNA polymorphisms in Japanese. II. Analysis with restriction enzymes of four or five base pair recognition" Hum Genet 72:105-117
- Hu CJ, Wang LY, Chodosh LA, Keith B, Simon MC (2003) "Differential roles of hypoxia-inducible factor 1alpha (HIF-1alpha) and HIF-2alpha in hypoxic gene regulation" Mol Cell Biol 23: 9361–9374.
- Human Mitochondrial DNA Variation" HUMAN MUTATION Mutation in Brief 1039, 30:E386-E394, (2008)
- Hunter Rl, And Markert Cl (1957) "Histochemical Demonstration Of Enzymes Separated By Zone Electrophoresis In Starch Gels" Science 125: 1294-1295.
- Ingman M, Kaessmann S, Pääbo U, Gyllesten U, (2000) "Mitochondrial genome variation and the origin of modern humans,". Nature, 408, 708-713
- Ingman M, Gyllensten U, (2001) "Analysis of the complete human mtDNA genome: methodology and inferences for human evolution". J Hered. 2001 Nov-Dec;92(6):454-61
- Jobling MA, and Tyler-Smith C, (1995) " Fathers and sons: the Y chromosome and human evolution" doi:10.1016/S0168-9525(00)89144-1
- Jobling MA, Tyler-Smith C (2003) "The human Y chromosome: an evolutionary marker comes of age" .Nat Rev Genet 2003; 4:598–612.
- Kang L, Lu Y, Wang C, Hu K, Chen F, Liu K, Li S, Jin L, Li H and The Genographic Consortium (2011) "Y-chromosome O3 Haplogroup Diversity in Sino-Tibetan Populations Reveals Two Migration Routes into the Eastern Himalayas" doi: 10.1111/j.1469-1809.2011.00690.

- Karafet T, Mendez F, Meilerman MB, Underhill PA, Zegura S L, Hammer, Michael F, (2008) "New binary polymorphisms reshape and increase resolution of the human Y chromosomal haplogroup tree" *Genome Research* 18 (5): 830–8. doi:10.1101/gr.7172008.
- Kivisild T, Tolk HV, Parik J, Wang Y, Papiha SS, Bandelt HJ, Villems R, (2002). "The emerging limbs and twigs of the East Asian mtDNA tree". Mol Biol Evol 19:1737–1751.
- Kivisild T, Reidla M, Metspalu E, Rosa A, Brehm A, Pennarun E, Parik J, Geberhiwot T, et al., (2004). "Ethiopian mitochondrial DNA heritage: tracking gene flow across and around the gate of tears". *The American Journal of Human Genetics* 75 (5): 752–770. doi:10.1086/425161.
- Knightly PD, and Otto SP, (2006). "Interference among deleterious mutations favors sex and recombination in finite populations". *Nature* 443 (7107): 89–92.doi:10.1038/nature05049
- Kong QP, Yao YG, Sun C, Bandelt HJ, Zhu CL, and Zhang YP, (2008) " Phylogeny of East Asian Mitochondrial DNA Lineages Inferred from Complete Sequences" Am. J. Hum. Genet. 73:671–676, 2003
- Kumar V, Reddy NS, Babu JP, Rao TN, Langstieh BT, Thangaraj K, Reddy AG, Singh L and Reddy BM, (2007) "Y-chromosome evidence suggests a common paternal heritage of Austro-Asiatic populations" BMC Evolutionary Biology 2007, 7:47 doi:10.1186/1471-2148-7-47
- Lah MM, and Foley R, (1994) "Multiple dispersals and modern human origins" Evol. Anthropol.3:48–60
- Lamason RL, Mohideen MA, Mest JR, Wong AC, Norton HL, Aros MC, Jurynec MJ, Mao X, Humphreville VR, Humbert JE, Sinha S, Moore JL, et al., (2005) " SLC24A5, a putative cation exchanger, affects pigmentation in zebrafish and humans" Science 310: 1782-1786, 2005.
- Landsteiner K, And Levine M, (1928) "On The Inheritance Of Agglutinogens Of Human Blood Demonstrable By Immune Agglutinins" Journal Of Experimental Medicine 43: 223-224.
- Lembring M, van Oven M, Montelius M, Allen M, (2013) "Mitochondrial DNA analysis of Swedish population samples" Int J Legal Med. 2013 Nov;127(6):1097-9.

- Lacey SW, Naim HY, Magness RR, Gething MJ, Sambrook JF. 1994. Expression of lactase-phlorizin hydrolase in sheep is regulated at the RNA level. Biochem J. 302(Pt 3):929–935.
- Lewis, Todd T. (January 1996). "Notes on the Uray and the Modernization of Newar Buddhism".*Contributions to Nepalese Studies*. Retrieved 31 January 2011. Page 110.
- Lewis TT, and Shakya DR, (1988) "Contributions to the history of Nepal: eastern Newar diaspora settlements" Contrib Nepal Stud 15:25–65
- Lofstedt T, Fredlund E, Holmquist-Mengelbier L, Pietras A, Ovenberger M, et al., (2007) "Hypoxia inducible factor-2alpha in cancer" Cell Cycle 6: 919–926.
- Marjanovic D, Fornarino S, Montagna S, Primorac D, Hadziselimovic R, Vidovic S, Pojskic N, Battaglia V, Achilli A, Drobnic K, Andjelinovic S, Torroni A, and Semino O,(2005) " The Peopling of Modern Bosnia-Herzegovina: Y-chromosome Haplogroups in the Three Main Ethnic Groups" doi: 10.1111/j.1529-8817.2005.00190.
- Masel J, (2011). "Genetic drift". *Current Biology* 21(20): R837R838. doi: 10.1016/j. cub.2011.08.007
- Mendez F, Krahn T, Schrack B, Krahn AM, Veeramah K, Woerner A, Fomine F, Leypey M, Bradman N, Thomas M, Karafet T, Hammer M, (2013). "An African American Paternal Lineage Adds an Extremely Ancient Root to the Human Y Chromosome Phylogenetic Tree". The American Journal of Human Genetics. doi: 10. 1016/j. ajhg. 2013.02.002
- Metspalu M, Romero IG, Yunusbayev B, Chaubey G, Mallick CB, Hudjashov G, Nelis M, Mägi R, Metspalu E, Remm M, Pitchappan R, Singh L, Thangaraj K,Villems R, Kivisild T. '(2011) "Shared and unique components of human population structure and genome-wide signals of positive selection in South Asia"AmJ Hum Genet. 2011 Dec 9;89(6):731-44. doi: 10.1016/j.ajhg.2011.11.010.
- Michael R, Speicher S, Antonarakis E, Arno GM, (2010) "Vogel and Motulsky's Human Genetics: Problems and Approaches" Springer. p. 606. ISBN 978-3-540-37653-8.
- Mishmar D, Ruiz-Pesini E, Golik P, Macaulay V, Clark AG, Hosseini S et al., (2003) "Natural selection shaped regional mtDNA variation in humans" Proc. Natl. Acad. Sci.100, 171–176.

- "Modern Humans Came Out of Africa, "Definitive" Study Says". News. nationalgeographic.com. 2010-10-28. Retrieved 2011-05-14
- Monge C, Leon-Velarde F, (1991) "Physiological adaptation to high altitude: oxygen transport in mammals and birds" Physiol Rev 71: 1135–1172.
- Monreal AW, Ferguson BM, Headon DJ, Street SL, Overbeek PA, Zonana J (1999) "Mutations in the human homologue of mouse dl cause autosomal recessive and dominant hypohidrotic ectodermal dysplasia" Nat Genet 22:366–369 Montagu A (1964) Concept Of Race In The Human Species In The Light Of Genetics. Journal Of Heredity 32: 243-247.
- Nei M, and Roychoudhury AK, (1993) "Evolutionary relationships of human populations on a global scale" Mol. Biol. Evol. 10:927–943
- Novelletto A, (2007) "Y chromosome variation in Europe: continental and local processes in the formation of the extant gene pool" Ann Hum Biol 2007;34: 139–172.
- O'rourke D, (2003) "Anthropological Genetics In The Genomic Era: A Look Back And Ahead" American Anthropologist 105: 101-109.
- Oven MV and Kayser M,(2008) "Updated Comprehensive Phylogenetic Tree of Global Human Population
- Oh SP, Seki T, Goss KA, Imamura T, Yi Y, Donahoe PK, Li L, Miyazono K, Dijke P, Kim S, Li E, (2000). "Activin receptor-like kinase 1 modulates transforming growth factor-β1 signaling in the regulation of angiogenesis". *Proc. Natl. Acad. Sci.* 97 (6: 2626–31. doi:10.1073/pnas.97.6.2626
- Pakendorf B, Stoneking M, (2005) "Mitochondrial DNA and human evolution" AnnuRev Genomics Hum Genet 6:165-183.
- Passarino G, Semino 0, Modiano G, Santachiara-Benerecetti AS, (1993) "COII/tRNALy, intergenic 9-bp deletion and other mtDNA markers clearly reveal that the Tharus (southern Nepal) have Oriental affinities" Am J Hum Genet 53:609-618
- Pearson R, (1959) "Blood Groups And Race" London: Clair Press.
- Peng MS, He JD, Zhu CH, Wu SF, Jin JQ. and Zhang YP, (2012) "Lactase persistence may have an independent origin in Tibetan populations from Tibet, China" Journal of Human genetics, doi:10.1038/jhg.2012.41(2012)

- Pijpe J, Voogt AD, Oven MV, Henneman P, J. van der Gaag K, Kayser M, and Knijff PD,(2013) "Indian Ocean Crossroads: Human Genetic Origin and Population Structure in the Maldives" AMERICAN JOURNAL OF PHYSICAL ANTHROPOLOGY 151:58–67
- Petra L M Zusterzeel, WilbertHMPeters, Willy Visser, Kristel J M Hermsen, Hennie M J Roelofs, Eric A P Steegers (2001) " A polymorphism in the gene for microsomal epoxide hydrolase is associated with pre-eclampsia" Med Genet 2001;38:234–237
- Previs MJ, Previs SB, Gulick J, Robbins J, Warshaw DM, (2012)" Molecular mechanics of cardiac myosin-binding protein C in native thick filaments." Science 337: 1215-1218, 2012.
- Qamar R, Ayub Q, Mohyuddin A, Helgason A, Mazhar K, Mansoor M, Zerjal T, Tyler-Smith C, and Mehdi SQ, (2002) "Y-Chromosomal DNA Variation in Pakistan" Am. J. Hum. Genet. 70:1107–1124, 2002
- Quintana-Murci L, Chaix R, Wells RS, et al., (2004). "Where west meets east: the complex mtDNA landscape of the southwest and Central Asian corridor" *American Journal of Human Genetics* 74 (5): 827–45. doi :10.1086/383236.
- Quintana-Murci L, Semino O, Bandelt H-J, Passarino G, McElreavey K, and Santachiara-Benerecetti AS, (1999) "Genetic evidence of an early exit of Homo sapiensfrom Africa through eastern Africa". Nat. Genet.23, 437–441
- Rajkumar R, Banerjee j, Hima BG, Trivedi R, and Kashyap VK, (2005)" Phylogeny and antiquity of M Macrohaplogroup inferred from complete mtDNA sequence of Indian specific lineages" BMC Evolutionary Biology 2005, 5:26 doi:10.1186/1471-2148-5-26
- Reid GBR, Hetherington R, (2010) "The climate connection: climate change and modern human evolution" Cambridge, UK: Cambridge University Press. p. 64.ISBN 0-521-147239
- Richards M, MacAulay V, Hickey E, Vega E, Sykes B, Guida V, Rengo C, Sellitto D, et al., (2000). "Tracing European Founder Lineages in the Near Eastern mtDNA Pool". *The American Journal of Human Genetics* 67 (5): 1251–76. doi:10.1016/S0002-9297(07)62954-1
- Regmi DR, (1969) "Ancient Nepal" K.L. Mukhopadhyay, Calcutta.

- Rootsi S, et al*.,* (2007) "A counter-clockwise northern route of the Y-chromosome haplogroup N from Southeast Asia towards Europe". *European Journal of Human Genetics* 15 (2):204–211.doi:10.1038/sj.ejhg.5201748.
- Ruiz-Pesini E, Lott MT, Procaccio V, Poole JC, Brandon MC, Mishmar D, Yi C, Kreuziger J, Baldi P, Wallace DC, (2007) 'An enhanced MITOMAP with a global mtDNA mutational phylogeny" Nucleic Acids Res 35:D823-D828
- Sahoo S, (2006). "A prehistory of Indian Y chromosomes: Evaluating demic diffusion scenarios". *Proceedings of the National Academy of Sciences* 103 (4): 843.doi:10.1073/pnas.0507714103.
- Semino O, Magri C, Benuzzi, G, Lin AA, Al-Zahery N, Battaglia V, (2004) "Origin, Diffusion, and Differentiation of Y-Chromosome Haplogroups E and J: Inferences on the Neolithization of Europe and Later Migratory Events in the Mediterranean Area". *AJHG*. 74 (5): 1023–34.
- Sengupta S, Zhivotovsky LA, King R, Mehdi SQ, Edmonds CA, Chow T, Lin AA, Mitra M, Sil SK, Ramesh A, Rani MVU, Thakur CM, Cavalli-Sforza, LL, Majumder PP, Underhill PA, (2006). "Polarity and temporality of high-resolution y-chromosome distributions in India identify both indigenous and exogenous expansions and reveal minor genetic influence of Central Asian pastoralists". *The American Journal of Human Genetics* 78(2): 202–221. doi:10.1086/499411.
- Simonson TS, Yang Y, Huff CD, Yun H, Qin G, et al., (2010) "Genetic evidence for high-altitude adaptation in Tibet" Science 329: 72–75.
- Simms TM, Wright MR, Hernandez M, Perez OA, Ramirez EC, Martinez E, Herrera RJ, (2012) " Y-chromosomal diversity in Haiti and Jamaica: contrasting levels of sex-biased gene flow" Am J Phys Anthropol. 2012 Aug;148(4):618-31. doi: 10.1002/ajpa.22090
- Smithies O, (1955) "Zone Electrophoresis In Starch Gels: Group Variation In The Serum Proteins Of Normal Human Adults" Biochemical Journal 61: 629-641.
- Sood R, Zehnder JL, Druzin ML, Brown PO (2006) "Gene expression patterns in human placenta" Proc Natl Acad Sci U S A 103: 5478–5483
- Stringer CB, (1994). "Evolution of Early Humans". In Steve Jones, Robert Martin, David Pilbeam. The Cambridge Encyclopedia of Human Evolution. Cambridge:

Cambridge University Press. p. 242. ISBN 978-0-521-32370-3. ISBN 978-0-521-46786-5

- Su H, Qu LJ, He K, Zhang Z, Wang J, Chen Z., Gu H, (2003). "The Great Wall of China: a physical barrier to gene flow?" *Heredity* 90 (3): 212–9.doi:10.1038/sj.hdy.6800237
- Thangaraj K, Chaubey G, Kivisild T, Rani DS, Singh VK, Ismail T, et al., (2008) "Maternal footprints of Southeast Asians in North India" Hum. Hered.66, 1–9 .
- The International HapMap Consortium (2003). "The International HapMap Project". *Nature* 426 (6968): 789–796. doi:10.1038/nature02168
- Torroni A, Chen Y-S, Semino 0, Santachiara-Beneceretti AS, Scott RC, Lott MT, Winter M, et al (1994a) "mtDNA and Y-chromosome polymorphisms in four Native American populations from southern Mexico" Am J Hum Genet 54:303-318
- "Torroni A, Miller JA, Moore LG, Zamudio S, Zhuang J, Droma T, Wallace DC (1994) Mitochondrial DNA analysis in Tibet: implications for the origin of the Tibetan population and its adaptation to high altitude" Am J Phys Anthropol 93:189-199
- Torroni A, Neel JV, Barrantes R, Schurr TG, Wallace DC (1994d) "A mitochondrial DNA "clock" for the Amerinds and its implications for timing their entry into North America" Proc Natl Acad Sci USA 91:1158-1162
- Torroni A, Schurr TG, Cabell MF, Brown MD, Neel JV, Larsen M, Smith DG, et al., (1993a) "Asian affinities and continental radiation of the four founding Native American mtDNAs" Am J Hum Genet 53:563-590
- Torroni A, Rengo C, Guida V, Cruciani F, Sellitto D, Coppa A, Calderon FL, Simionati B, Valle G, Richards M, Macaula V, and Scozzari R (2001)"Do the Four Clades of the mtDNA Haplogroup L2 Evolve at Different Rates" Am. J. Hum. Genet. 69:1348–1356, 2001
- Torroni A, Schurr TG, Yang C-C, Szathmary EJE, Williams RC, Schanfield MS, Troup GA, et al., (1992) "Native American mitochondrial DNA analysis indicates that the Amerind and the Nadene populations were founded by two independent migrations" Genetics 130:153-162

- Torroni A, Sukernik RI, Schurr TG, Starikovskaya YB, Cabell MF, Crawford MH, Comuzzie AG, et al (1993) "mtDNA variation of aboriginal Siberians reveals distinct genetic affinities with Native Americans" Am J Hum Genet 53:591-608
- Underhill PA, Kivisild T, (2007) "Use of Y chromosome and mitochondrial DNA population structure in tracing human migrations". Annu Rev Genet 2007;41:539–56
- Underhill PA, Myres NM, Rootsi S, Metspalu M, Zhivotovsky LA, King RJ, Lin AA, Cheryl-Emiliane T., et al., (2009). "Separating the post-Glacial coancestry of European and Asian Y chromosomes within haplogroup R1a". *European Journal of Human Genetics* 18 (4): 479–84. doi:10.1038/ejhg.2009.194
- Van Oven M, Kayser M, (2009). "Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation". *Human Mutation* 30 (2): E386–94. doi:10.1002/humu.20921.
- Wang H.W, Li Y.C ,Sun F, Zhao M, Mitra B, Chaudhuri T.K, Regmi P, Wu S.F, Kong QP, and Zhang YP, (2012) "Revisiting the role of the Himalayas in peopling Nepal: insights from mitochondrial genomes" Journal of Human Genetics(2012)57,228–234; doi:10.1038/jhg.2012.8
- Ward RH, Frazier BL, Dew-Jager K, Paabo S (1991) "Extensive mitochondrial diversity within a single Amerindian tribe" Proc Natl Acad Sci USA 88:8720-8724
- Watkins WS, Bamshad M, Dixon ME, Bhaskara RB, Naidu JM, Reddy PG, Prasad BV, Das PK, Reddy PC, Gai PB, Bhanu A, Kusuma YS, Lum JK, Fischer P, Jorde LB (1999) "Multiple origin Of the mtDNA 9bp deletion in the population of south India. Am J,Phys, Anthropoloy 109:147-158
- Wilkinson-Herbots H, Richards M, Forster P, Sykes B, (1996) "Site 73 in hypervariable region II of the human mitochondrial genome and the origin of European populations". Ann Hum Genet 60: 499–508
- Wrischnik LA, Higushi RG, Stoneking M, Erlich HA, Arnheim N, Wilson AC (1987) "Length mutations in human mitochondrial DNA: direct sequencing of enzymatically amplified DNA" Nucleic Acids Res 15:529-542
- Wrischnik LA, Higuchi RG, Stoneking M, Erlich HA, Arnheim N, and Wilson AC, (1987) "Length mutations in human mitochondrial DNA; direct sequencing of enzymatically amplified DNA" Nucleic Acids Res. 15:529-542

- Wolpoff MH, Hawks J, Caspari R, (2000) "Multiregional, not multiple origins" .*American Journal of Physical Anthropology* 112 (1): 129–36. doi:10.1002/(SICI)1096-8644(200005)112:1<129::AID-AJPA11>3.0.CO;2-K
- Xue Y, Zerjal T, Bao W, Zhu S, Shu Q, Xu J, Du R, Fu S, et al., (2005). "Male Demography in East Asia: A North-South Contrast in Human Population Expansion Times" *Genetics* 172 (4):24319. doi:10.1534/genetics.105.054270.
- Zerjal T, Pandya A, Thangaraj K, Ling EY, Kearley J, Bertoneri S, Paracchini S, Singh L, Tyler-Smith C, (2006) "Y-chromosomal insights into the genetic impact of the caste system in India" Hum Genet. 2007 Mar; 121(1):137-44. Epub 2006 Oct 31.
- Zhong H, Shi H, Qi XB, Duan ZY, Tan PP, Jin L, Su B, and Ma RZ, (2011) " Extended Y Chromosome Investigation Suggests Postglacial Migrations of Modern Humans into East Asia via the Northern Route" Mol. Biol. Evol.28(1):717–727. 2011 doi:10.1093/molbev/msq247

Websites

- http://phylotree.org/tree/main.htm
- http://www.mitomap.org/
- http://empop.org/
- http://www.hagsc.org/hgdp/
- http://www.ncbi.nlm.nih.gov/
- https://www.familytreedna.com/
- http://www.ensembl.org/
- http://www.ncbi.nlm.nih.gov/SNP/
- http://www.fluxus-engineering.com/
- http://omim.org/

Appendices

Appendix 1: List of Mitochondrial Primers

Primer Name	Forward Primer	Reverse Primer	Length	
			F	R
1	ctcctcaaagcaatacactg	tgctaaatccaccttcgacc	20	20
2	cgatcaacctcaccacctct	tggacaaccagctatcacca	20	20
3	ggactaacccctataccttctgc	ggcaggtcaatttcactggt	23	20
4	aaatcttaccccgcctgttt	aggaatgccattgcgattag	20	20
5	tacttcacaaagcgccttcc	atgaagaatagggcgaaggg	20	20
6	tggctcctttaacctctcca	aaggattatggatgcggttg	20	20
7	actaattaatcccctggccc	aatggggtgggttttgtatg	20	20
8	ctaaccggctttttgccc	acctagaaggttgcctggct	18	20
9	gaggcctaacccctgtcttt	attccgaagcctggtaggat	20	20
10	ctcttcgtctgatccgtcct	agcgaaggcttctcaaatca	20	20
11	acgccaaaatccatttcact	cgggaattgcatctgttttt	20	20
12	acgagtacaccgactacggc	tgggtggttggtgtaaatga	20	20
13	tttccccctctattgatccc	gtggccttggtatgtgcttt	20	20
14	cccaccaatcacatgcctat	tgtagccgttgagttgtggt	20	20
15	tctccatctattgatgagggtct	aattaggctgtgggtggttg	23	20
16	gccatactagtctttgccgc	ttgagaatgagtgtgaggcg	20	20
17	tcactctcactgcccaagaa	ggagaatgggggataggtgt	20	20
18	tatcactctcctacttacag	agaaggatataattcctacg	20	20
19	aaacaacccagctctccctaa	tcgatgatgtggtctttgga	21	20
20	acatctgtacccacgccttc	agaggggtcagggttgattc	20	20
21	gcataattaaactttacttc	agaatattgaggcgccattg	20	20
22	tgaaacttcggctcactcct	agctttgggtgctaatggtg	20	20
23	tcattggacaagtagcatcc	gagtggttaatagggtgatag	20	21
24	caccatcctccgtgaaatca	aggctaagcgttttgagctg	20	20

Appendix 2: List of Y-Chromosomal Primers

PRIMER	REGION/STS	SIZE	SITE	MUTATION	FORWARD PRIMER (5'-3')	REVERSE PRIMER (5'-3')
M9	G10.35a	340 bp	68	C->G	gcagcatataaaactttcagg	gcttgagcaaagttaggtttt
M82	B9.t18	328 bp	##	-2bp	ctgtactcctgggtagcctgt	aagaacgattgaacacactaactc
M172	DFFRY Ex45	345 bp	##	T->G	ttgaagttacttttataatctaatgctt	ataatttattactttacagtcacagtgg
M45	B9.12	352 bp	##	G->A	gctggcaagacacttctgag	aatatgttcctgacaccttcc
M17	G10.47a	333 bp	68	4G->3G	ctggtcataacactggaaatc	tgaacctacaaatgtgaaact
M124	G3.27c	393 bp	##	C->T	tggtaaactctacttagttgcctttt	cagcgaattagattttcttgc
M11	G10.10	222 bp	44	A->G	tctctctgtctgtctctccctcc	gagcataaacaagaacttactgagc
M175	UTY1exon 07	444 bp	84-88	-5bp	ttgagcaagaaaaatagtaccca	ctccattcttaactatctcaggga
M70	B9.62b	257 bp	45	A->C	ggttatcatagcccactatactttg	atctttattccctttgtcttgct
M89	B9.95	527 bp	##	C->T	agaagcagattgatgtcccact	tccagttaggagatcccctca
M170	DFFRYEx 08	405bp	##	A->C	tgcttcacacaaatgcgttt	ggtcttaaatgttgaaagtaattgg
M130	RSP4YC711T	205bp	41	C->T	tatctcctcttctattgcag	gtgtttttttccccctttgtgg
M117	G3.25b	429bp	142-145	4bp (ATCT) deletion	aagtatgacttatgaagtacgaagaaa	attcagttagattttacaatgagca
M15	G10.16	295 bp	##	9bp(TTGTACAGAGA) insertion	acaaatcctgaacaatcgc	tgcatgtggttaaaatttcc
Z93		338bp	##	G->A	aacaaagcatcatcatcaaaggc	catgattcgttatgacctgc
Z95		429bp	##	C->T		
M410				A->G	caatcattgaccttaagtctgagtccc	actggatacctttctaggaagaattg
P47		363bp	##	C->T	ctgatgttgcagtgttgagc	acacagccaaataccagtcg

Appendix 3: List of Autosomal Primer

Marker	SNP	F-primer	R-primer
EPAS1	G/A	cctgttccctcctcctttttt	Tcttttccccttggagacct
PARA	T/C	aaggaaagaaccctttttcac	Taatttttcactcactcgcc
SLC	G/A	ctcacctacaagccctctgc	tagtgccctgttgtccatcc
TGF	C/T	cgcgcccatctaggttatttcc	ggtgacctccttggcgtagtag
EDAR	T/C	gtaggtcttagccccac	Catccagccgctcaatc
EGLN	A/T	aagttaccctccaccccatc	agatccccaagcaatcacag
EPHX13	T/C	cttgtgctctgtccttcccatccc	gaaggctgttctcatgacatacatcc
EPHX14	A/G	tctggtgccagagcctgaccgtgc	atggaacctctagcagccccgtacc
MYBPC	25bp deletion	gtttccagccttgggcatagtc	aggacaacggagcaaacccc

Appendix 4: Y-chromosomal Haplogroup

POP	M117	M124	M17	Z93	Z95	M82	M172	M410	M89	M174	M214	P47	M15	M9	M201	M214	M175	M231	M170	M45	HG
S001	A	D																			R2
S002	D	A																			O3a3C1
S003	D	A																			O3a3C1
S004	D	A																			O3a3C1
S005	A	A	A			A	A		D					A	A				A		F*
S006	A	D																			R2
S007	A	D																			R2
S008	A	A	A			A	A		D					D	A	A				D	P
S009	A	A	A			A	D	D													J2a
S010	A	A	A			A	D	D													J2a
S012	D	A																			O3a3C1
S014	A	A	A			A	D	D													J2a
S015	A	A	A			A	D	D													J2a
SO16	A	A	A			A	D	D													J2a
S017	A	A	A			A	D	D													J2a
S018	A	A	D	A	A	A															R1a
S019	A	A	A			A	D	D													J2a
S020	A	D																			R2
S021	A	D																			R2
B001	A	A	D	D	D	A															R1a

B002	A	A	A			A	A		A	D	A		D								D1
B003	A	A				A			A	D	A		D								D1
B004	A	A	D	A	A	A															R1a
B005	A	A	D	A	D	A															R1a
B006	A	A	A			A	A		D					A	A				A		F*
B007	D	A																			O3a3C1
B008	D	A																			O3a3C1
B009	D	A																			O3a3C1
B010	A	A	A	A			A		D					D	A	A					K*
B011	A	A	A			A	A		D					D	A	A					K*
B012	A	A	A			A	A		D					A	A				A		K*
B013	A	A	A			D															H1a
B014	D	A	A			A	A														O3a3C1
B015	D	A																			O3a3C1
B016	D	A																			O3a3C1
B017	D	A																			O3a3C1
B018	A	A	A			D	A		D					A	A		D				H1a
B019	D	A																			O3a3C1
B020	D	A																			O3a3C1
U001	A	A	A			A	A		D					D	A	A					K*
U002	A	A	A			A	A		D					D	A	D		A			NO
U003	A	A	A			A	A		A	D	A		D								D1
U0	A	A	A			A	A		A	D	A	A	D								D1

04																				
U005	D	A																		O3a 3C1
U006	A	A	A			A	A		D					D	A	A				K*
U007	D	A															D			O3a 3C1
U008	D	A																		O3a 3C1
U009	A	A	A			A	A		D					D	A	A				D	P
U010	A	D																		R2
U011	A	D																		R2
U012	D	A																		O3a 3C1
U013	A	A	A			A	A		D					D	A	A				A	K*
U014	A	A	A			A	A		D					D	A	A				A	K*
U015	A	A	A			A	A		A	D	A	A	D								D1
U016	D	A																			O3a 3C1
U017	A	A	A			A	A		A	D		A	D								D1
U018	D	A																			O3a 3C1
U019	A	A	A			A	A		D					D	A	A				D	P
U020	D	A																			O3a 3C1
U021	A	D																			R2
U022	A	A	D		D	A	A		D					D	A						R1a
U023	D	A																			O3a 3C1
U024	A	A	A			A	D	D													J2a
U025	A	D																			R2
U026	A	D																			R2

U027	A	A	A			A	A		A	D	A	A	D								D1
U028	A	A	A			A	A		A	D	A	A	D								D1
U029	A	A	A			A	A		A	D	A	A	D								D1
U030	A	A	A			A	D	D													J2a
U031	D	A																			O3a 3C1
U032	D	A																			O3a 3C1
U033	A	A	A				A		D					D	A	A					K*
U034	A	A	A			A	A		D						A	A		A			F*
U035	D	A																			O3a 3C1
U036	D	A																			O3a 3C1
U037	D	A																			O3a 3C1
U038	A	A	D		D	A	A		D					D	A						R1a
U039	D	A																			O3a 3C1
U040	A	D																			R2
U041	A	A	A			A	A		D					D	A	A				D	P
U042	D	A																			O3a 3C1
U043	D	A																			O3a 3C1
U044	A	A	A	A	A	A	A		D					A	A	A			A		F*
U045	A	A	A			A	A		D					D	A	A				A	F*
U046	D	A																			O3a 3C1
U047	A	A	D	D	D	A															R1a
U048	D	A				A															O3a 3C1
U0	D	A																			O3a

49																					3C1
U050	D	A																			O3a 3C1
U051	A	D																			R2
U052	A	A	A			A	A		A	D	A	A	D								D1
U053	A	A	D	D	D	A															R1a
U054	D	A																			O3a 3C1
U055	A	A	A			A	A		A	D	A	A	D								D1
U056	A	A	A			A	A		A	D	A	A	D								D1
U057	D	A																			O3a 3C1
U058	D	A																			O3a 3C1
U059	D	A																			O3a 3C1

A = ancestral, D = Derived

Appendix 5: Whole Mitochondrial DNA Sequence (Total 98 whole mtDNA (16569 bp) were sequenced).

>S001 (Z7)

gatcacaggtctatcaccctattaaccactcacgggagctctccatgcatttggtattttcgtctgggggtgtgcacgcgatagcattgcgagacgctggagccggagcaccctatgt
cgcagtatctgtctttgattcctgcctcatcctattatttatcgcacctacgttcaatattacaggcgaacatacttactaaagtgtgttaattaattaatgcttgtaggacataataataa
caattgaatgtctgcacagccgctttccacacagacatcataacaaaaaatttccaccaaacccccctcccccgcttctggccacagcacttaaacacatctctgccaaaccccaaa
aacaaagaaccctaacaccagcctaaccagatttcaaattttatcttttggcggtatgcacttttaacagtcaccccccaactaacacattattttcccctcccactcccatactactaat
ctcatcaacacaaccccgcccatcctacccagcacacacacaccgctgctaaccccataccccgaaccaaccaaacccccaaagacacccccccacagtttatgtagcttacctcctca
aagcaatacactgaaaatgtttagacgggctcacatcaccccataaacaaataggtttggtcctagcctttctattagctcttagtaagattacacatgcaagcatccccgttccagtg
agttcaccctctaaatcaccacgatcaaaagggacaagcatcaagcacgcagcaatgcagctcaaaacgcttagcctagccacaccccccacgggaaacagcagtgattaacctttta
gcaataaacgaaagtttaactaagctatactaaccccagggttggtcaatttcgtgccagccaccgcggtcacacgattaacccaagtcaatagaagccggcgtaaagagtgttttta
gatcacccctccccaataaagctaaaactcacctgagttgtaaaaaactccagttgacacaaaatagactacgaaagtggctttaacatatctgaacacacaatagctaagaccca
aactgggattagataccccactatgcttagccctaaacctcaacagttaaatcaacaaaactgctcgccagaacactacgagccacagcttaaaactcaaaggacctggcggtgctt
catatccctctagaggagcctgttctgtaatcgataaaccccgatcaacctcaccacctcttgctcagcctatataccgccatcttcagcaaaccctgatgaaggctacaaagtaagcg
caagtacccacgtaaagacgttaggtcaaggtgtagcccatgaggtggcaagaaatgggctacattttctaccccagaaaactacgatagcccttatgaaacttaagggtcgaaggt
ggatttagcagtaaactgagagtagagtgcttagttgaacagggccctgaagcgcgtacacaccgcccgtcaccctcctcaagtatacttcaaaggacatttaactaaaaccccctacg
catttatatagaggagacaagtcgtaacatggtaagtgtactggaaagtgcacttggacgaaccagagtgtagcttaacacaaagcacccaacttacacttaggagatttcaacttta
acttgaccgctctgagctaaacctagccccaaacccactccaccttactaccagacaaccttagccaaacatttacccaaataaagtataggcgatagaaattgaaacctggcgca
atagatatagtaccgcaagggaaagatgaaaaattataaccaagcataatatagcaaggactaacccctataccttctgcataatgaattaactagaaataactttgcaaggagag
ccaaagctaagacccccgaaaccagacgagctacctaagaacagctaaaagagcacacccgtctatgtagcaaaatagtgggaagatttataggtagaggcgacaaacctaccga
gcctggtgatagctggttgtccaagatagaatcttagttcaactttaaatttgcccacagaaccctctaaatccccttgtaaatttaactgttagtccaaagaggaacagctctttggac
actaggaaaaaaccttgtagagagagtaaaaaatttaacacccatagtaggcctaaaagcagccaccaattaagaaagcgttcaagctcaacacccactacctaaaaaatcccaa
acatataactgaactcctcacacccaattggaccaatctatcaccctatagaagaactaatgttagtataagtaacatgaaaacattctcctccgcataagcctgcgtcagattaaaac
actgaactgacaattaacagcccaatatctacaatcaaccaacaagtcattattaccctcactgtcaacccaacacaggcatgctcataaggaaaggttaaaaaaagtaaaaggaa
ctcggcaaatcttaccccgcctgtttaccaaaaacatcacctctagcatcaccagtattagaggcaccgcctgcccagtgacacatgtttaacggccgcggtaccctaaccgtgcaaa

ggtagcataatcacttgttccttaaatagggacctgtatgaatggctccacgagggttcagctgtctcttacttttaaccagtgaaattgacctgcccgtgaagaggcgggcatgacac
agcaagacgagaagaccctatggagctttaatttattaatgcaaacagtacctaacaaacccacaggtcttaaactaccaaacctgcattaaaaatttcggttggggcgacctcgga
gcagaacccaacctccgagcagtacatgctaagacttcaccagtcaaagcgaactactatactcaattgatccaataacttgaccaacggaacaagttaccctagggataacagcgc
aatcctattctagagtccatatcaacaatagggtttacgacctcgatgttggatcaggacatcccgatggtgcagccgctattaaaggttcgtttgttcaacgattaaagtcctacgtga
tctgagttcagaccggagtaatccaggtcggtttctatctacnttcaaattcctccctgtacgaaaggacaagagaaataaggcctacttcacaaagcgccttcccccgtaaatgatat
catctcaacttagtattatacccacacccacccaagaacagggtttgttaagatggcagagcccggtaatcgcataaaacttaaaactttacagtcagaggttcaattcctcttcttaac
aacatacccatggccaacctcctactcctcattgtacccattctaatcgcaatggcattcctaatgcttaccgaacgaaaaattctaggctatatacaactacgcaaaggcccc aacgt
tgtaggcccctacgggctactacaacccttcgctgacgccataaaactcttcaccaaagagcccctaaaacccgccacatctaccatcaccctctacatcaccgccccgaccttagctc
tcaccatcgctcttctactatgaaccccctccccatacccaaccccctggtcaacctcaacctaggcctcctatttattctagccacctctagcctagccgtttactcaatcctctgatca
gggtgagcatcaaactcaaactacgcccctgatcggcgcactgcgagcagtagcccaaacaatctcatatgaagtcaccctagccatcattctactatcaacattactaataagtggct
cctttaacctctccacccttatcacaacacaagaacacctctgattactcctgccatcatgaccctggccataatatgatttatctccacactagcagagaccaaccgaacccccttcg
accttgccgaaggggagtccgaactagtctcaggcttcaacatcgaatacgccgcaggccccttcgccctattcttcatagccgaatacacaaacattattataataaacaccctcacc
actacaatcttcctaggaacaacatatgacgcactctcccctgaactctacacaacatattttgtcaccaagaccctacttctaacctccctgttcttatgaattcgaacagcataccccc
gattccgctacgaccaactcatacacctcctatgaaaaaacttcctaccactcaccctagcattacttatatgatatgtctccataccattacaatctccagcattccccctcaaaccta
agaaatatgtctgataaaagagttactttgatagagtaaataataggagcttaaaccccctttatttctaggactatgagaatcgaacccatccctgagaacccaaaattctccgtgcca
cctatcacaccccatcctaaagtaaggtcagctaaataagctatcgggcccataccccgaaaatgttggttataccccttcccgtactaattaatcccctggcccaacccgtcatctactc
taccatctttgcaggcacactcatcacagcgctaagctcgcactgatttttacctgagtaggcctagaaataaacatgctagcttttattccagttctaaccaaaaaaatgaaaccctcg
ttccacagaagctgccatcaagtatttcctcacgcaagcaaccgcatccataatccttctaatagctatcctcttcaacaatatactctccgggcaatgaaccataaccaatactacca
atcaatactcatcattaataatcataatggctatagcaataaaactaggaatagccccctttcacttctgagtcccagaggttacccaaggcgcccctctaacatccggcctgcttcttc
tcacatgacaaaaactagcccccatctcaatcatataccaaatctctccctcactaaacgtaagccttctcctcactctctcaatcttatccatcatagcaggcagttgaggtggattaa
accaaacccagctacgcaaaatcttagcatactcctcaattacccacataggatgaataatagcagttctaccgtacaaccctaacataaccattcttaatttaactatttatattatcc
taactactaccgcattcctactactcaacttaaactccagcaccacgaccctactactatctcgcacctgaaacaagctaacatgactaacacccttaattccatccaccctcctctccct
aggaggcctgccccgctaaccggctttttgcccaaatgggccattatcgaagaattcacaaaaaacaatagcctcatcatccccaccatcatagccaccatcaccctccttaacctct
acttctacctacgcctaatctactccacctcaatcacactactccccatatctaacaacgtaaaaataaaatgacagtttgaacatacaaaacccaccccattcctccccacactcatc
gcccttaccacgctactcctacctatctccccttttatactaataatcttatagaaatttaggttaaatacagaccaagagccttcaaagccctcagtaagttgcaatacttaatttctgta
acagctaaggactgcaaaaccccactctgcatcaactgaacgcaaatcagccactttaattaagctaagcccttactagaccaatgggacttaaacccacaaacacttagttaacag
ctaagcaccctaatcaactggcttcaatctacttctcccgccgccgggaaaaaaggcgggagaagccccggcaggtttgaagctgcttcttcgaatttgcaattcaatatgaaaatca
cctcggagctggtaaaaagaggcctaacccctgtctttagatttacagtccaatgcttcactcagccatttttacctcacccccactgatgttcgccgaccgttgactattctctacaaacc
acaaagacattggaacactatacctattattcggcgcatgagctggagtcctaggcacagctctaagcctccttattcgagccgagctgggccagccaggcaaccttctaggtaacga
ccacatctacaacgttatcgtcacagcccatgcatttgtaataatcttcttcatagtaatacccatcataatcggaggctttggcaactgactagttcccctaataatcggtgcccccgat
atggcgtttcccgcataaacaacataagcttctgactcttacctccctctctcctactcctgctcgcatctgctatagtggaggccggagcaggaacaggttgaacagtctaccctccc
ttagcagggaactactcccaccctggagcctccgtagacctaaccatcttctccttacacctagcaggtgtctcctctatcttaggggccatcaatttcatcacaacaattatcaatataa
aacccccctgccataacccaataccaaacgcccctcttcgtctgatccgtcctaatcacagcagtcctacttctcctatctctcccagtcctagctgctggcatcactatactactaacaga
ccgcaacctcaacaccaccttcttcgaccccgccggaggaggagaccccattctataccaacacctattctgatttttcggtcaccctgaagtttatattcttatcctaccaggcttcgga
ataatctcccatattgtaacttactactccggaaaaaaagaaccatttggatacataggtatggtctgagctatgatatcaattggcttcctggggtttatcgtgtgagcacaccatata
tttacagtaggaatagacgtagacacacgagcatatttcacctccgctaccataatcatcgctatccccaccggcgtcaaagtatttagctgactcgccacactccacggaagcaatat
gaaatgatctgctgcagtgctctgagccctaggattcatctttctttttcaccgtaggtggcctgactggcattgtattagcaaactcatcactagacatcgtactacacgacacgtactac
gttgtagctcacttccactatgtcctatcaataggagctgtatttgccatcataggaggcttcattcactgatttcccctattctcaggctacaccctagaccaaacctacgccaaaatcc
atttcactatcatattcatcggcgtaaatctaactttcttcccacaacactttctaggcctatccggaatgccccgacgttactcggactaccccgatgcatacaccacatgaaacatcct
atcatctgtaggctcattcatttctctaacagcagtaatattaataatttttcatgatttgagaagccttcgcttcgaagcgaaaagtcctaatagtagaagaaccctccataaacctgga
gtgactatatggatgccccccaccctaccacacattcgaagaacccgtatacataaaatctagacaaaaaaggaaggaatcgaacccccccaaagctggtttcaagccaaccccatg
gcctccatgacttttttcaaaaaggtattagaaaaaccatttcataactttgtcaaagttaaattatagcctaaatcctatatatcttaatggcacatgcagcgcaagtaggtctacaaga
cgctacttcccctatcatagaagagcttatcacctttcatgatcacgccctcataatcattttccttatctgcttcctagtcctgtatgccctttttcctaacactcacaacaaaactaactaa
tactaacatctcagacgctcaggaaatagaaaccgtctgaactatcctgcccgccatcatcctagtcctcatcgccctcccatccctacgcatcctttacataacagacgaggtcaacg
atccctcccttaccatcaaatcaattggccaccaatggtactgaacctacgagtacaccgactacggcggactaatcttcaactcctacatacttcccccattattcctagaaccaggcg
acctgcgactccttgacgttgacaatcgagtagtactcccgattgaagcccccattcgtataataattacatcacaagacgtcttgcactcatgagctgtccccacattaggcttaaaaa
cagatgcaattcccggacgtctaaaccaaaccactttcaccgctacacgaccgggggtatactacggtcaatgctctgaaatctgtggagcaaaccacagtttcatgcccatcgtcct
agaattaattcccctaaaaatctttgaaatagggcccgtatttaccctatagcaccccccctaccccctctagagcccactgtaaagctaacttagcattaacctttttaagttaaagatt
aagagaaccaacacctctttacagtgaaatgccccaactaaatactaccgtatggcccaccataattacccccatactccttacactattcctcatcacccaactaaaaatattaaaca
caaactaccacctacctcccctcaccaaagcccataaaaataaaaaattataacaaaccctgagaaccaaaatgaacgaaaatctgttcgcttcattcattgcccccacaatcctagg
cctacccgccgcagtactgatcattctatttccccctctattgatccccacctccaaatatctcatcaacaaccgactaatcaccacccaacaatgactaatcaaactaacctcaaaaca
aatgataaccatacacaacactaaaggacgaacctgatctcttatactagtatccttaatcatttttattgccacaactaacctcctcggactcctgcctcactcatttacaccaaccacc
caactatctataaacctagccatggccatcccctatgagcgggcgcagtgattataggctttcgctctaagattaaaaatgccctagcccacttcttaccacaaggcacacctacacc
ccttatccccatactagttattatcgaaaccatcagcctactcattcaaccaatagccctggccgtacgcctaaccgctaacattactgcaggccacctactcatgcacctaattggaag

cgccaccctagcaatatcaaccattaaccttccctctacacttatcatcttcacaattctaattctactgactatcctagaaatcgctgtcgccttaatccaagcctacgttttcacacttct
agtaagcctctacctgcacgacaacacataatgacccaccaatcacatgcctatcatatagtaaaacccagcccatgacccctaacaggggccctctcagccctcctaatgacctccg
gcctagccatgtgatttcacttccactccataacgctcctcatactaggcctactaaccaacacactaaccatataccaatggtggcgcgatgtaacacgagaaagcacataccaagg
ccaccacacaccacctgtccaaaaaggccttcgatacgggataatcctatttattacctcagaagttttttttcttcgcaggatttttctgagccttttaccactccagcctagccccctaccc
cccaactaggagggcactggcccccaacaggcatcaccccgctaaatcccctagaagtcccactcctaaacacatccgtattactcgcatcaggagtatcaatcacctgagctcacca
tagtctaatagaaaacaaccgaaaccaaataattcaagcactgcttattacaattttactgggtctctattttaccctcctacaagcctcagagtacttcgagtctcccttcaccatttcc
gacggcatctacggctcaacatttttttgtagccacaggcttccacggacttcacgtcattattggctcaactttcctcactatctgcttcatccgccaactaatatttcactttacatccaa
acatcactttggcttcgaagccgccgcctgatactggcattttgtagatgtggtttgactatttctgtatgtctccatctattgatgagggtcttactcttttagtataaatagtaccgttaa
cttccaattaactagttttgacaacattcaaaaaagagtaataaacttcgccttaattttaataatcaacaccctcctagccttactactaataattattacattttgactaccacaactc
aacggctacatagaaaaatccacccccttacgagtgcggcttcgaccctatatccccgcccgcgtcccttttctccataaaattcttcttagtagctattaccttcttattatttgatctaga
aattgccctccttttacccctaccatgagccctacaaacaactaacctgccactaatagttatgtcatccctcttattaatcatcatcctagccctaagtctggcctatgagtgactacaa
aaaggattagactgagctgaattggtatatagtttaaacaaaacgaatgatttcgactcattaaattatgataatcatatttaccaaatgcccctcatttacataaatattatactagca
tttaccatctcacttctaggaatactagtatatcgctcacacctcatatcctccctactatgcctagaaggaataatactatcgctgttcattatagctactctcataaccctcaacaccca
ctccctcttagccaatattgtgcctattgccatactagtctttgccgcctgcgaagcagcggtgggcctagccctactagtctcaatctccaacacatatggcctagactacgtacataac
ctaaacctactccaatgctaaaactaatcgtcccaacaattatattactaccactgacatgactttccaaaaaacacataatttgaatcaacacaaccacccacagcctaattattagc
atcatccccctactatttttttaaccaaatcaacaacaacctatttagctgttccccaacctttttcctccgaccccctaacaaccccccctcctaatactaactacctgactcctaccctcac
aatcatggcaagccaacgccacttatccagtgaaccactatcacgaaaaaaactctacctctctatactaatctccctacaaatctccttaattataacattcacagccacagaactaa
tcatattttatatcttcttcgaaaccacacttatccccaccttggctatcatcacccgatgaggcaaccagccagaacgcctgaacgcaggcacatacttcctattctacaccctagtag
gctcccttcccctactcatcgcactaatttacactcacaacaccctaggctcactaaacattctactactcactctcactgcccaagaactatcaaactcctgagccaacaacttaatat
gactagcttacacaatagcttttatagtaaagatacctctttacggactccacttatgactccctaaagcccatgtcgaagcccccatcgctgggtcaatagtacttgccgcagtactctt
aaaactaggcggctatggtataatacgcctcacactcattctcaacccccgacaaaacacatagcctacccctccttgtactatccctatgaggcataattataacaagctccatctg
cctacgacaaacagacctaaaatcgctcattgcatactcttcaatcagccacatagccctcgtagtaacagccattctcatccaaaccccctgaagcttcaccggcgcagtcattctca
taatcgcccacggacttacatcctcattactattctgcctagcaaactcaaactacgaacgcactcacagtcgcatcataatcctctctcaaggacttcaaactctactcccactaatag
ctttttgatgacttctagcaagcctcgctaacctcgccttacccccccactattaacctactgggagaactctctgtgctagtaaccacgttctcctgatcaaatatcactctcctacttaca
ggactcaacatactagtcacagccctatactccctctacatatttaccacaacacaatggggctcactcacccaccacattaacaacataaaaccctcattcacacgagaaaacaccc
tcatgttcatacacctatcccccattctcctcctatccctcaaccccgacatcattaccgggttttcctcttgtaaatatagtttaaccaaaacatcagattgtgaatctgacaacagagg
cttacgacccctttatttaccgagaaagctcacaagaactgctaactcatgcccccatgtctaacaacatggctttctcaacttttaaaggataacagctatccattggtcttaggcccca
aaaattttggtgcaactccaaataaaagtaataaccatgcacactactataaccaccctaaccctgacttccctaattcccccccatccttaccaccctcgttaaccctaacaaaaaaaaa
ctcataccccccattatgtaaaatccattgtcgcatccacctttattatcagtctcttccccacaacaatattcatgtgcctagaccaagaagttattatctcgaactgacactgagccaca
acccaaacaacccagctctccctaagcttcaaactagactacttctccataatattcatccctgtagcattgttcgttacatggtccatcatagaattctcactgtgatatataaactcag
acccaaacattaatcagttcttcaaatatctactcattttcctaattaccatactaatcttagttaccgctaacaacctattccaactgttcatcggctgagagggcgtaggaattatatcc
ttcttgctcatcagttgatgatacgcccgagcagatgccaacacagcagccattcaagcaatcctatacaaccgtatcggcgatatcggtttcatcctcgccttagcatgatttatccta
cactccaactcatgagacccacaacaaatagcccttctaaacgctaatccaagcctcaccccactactaggcctcctcctagcagcagcaggcaaatcagcccaattaggtctccacc
cctgactcccctcagccatagaaggccccaccccagtctcagccctactccactcaagcactatagttgtagcaggaatcttcttactcatccgcttccacccccctagcagaaaatagc
ccactaatccaaactctaacactatgcttaggcgctatcaccactctgttcgcagcagtctgcgcccttacacaaaatgacatcaaaaaaatcgtagccttctccacttcaagtcaact
aggactcataatagttacaatcggcatcaaccaaccacacctagcattcctgcacatctgtacccacgccttcttcaaagccatactatttatgtgctccgggtccatcatccacaacct
taacaatgaacaagatattcgaaaaataggaggactactcaaaaccatacctctcacttcaacctccctcaccattggcagcctagcattagcaggaataccttttcctcacaggtttct
actccaaagaccacatcatcgaaaccgcaaacatatcatacacaaacgcctgagccctgtctattactctcatcgctacctccctgacaagcgcctatagcactcgaataattcttctc
accctaacaggtcaacctcgcttcccccacccttactaacattaacgaaaataaccccacccctactaaaccccattaaacgcctggcagccggaagcctattcgcaggatttctcattac
taacaacatttcccccgcatccccctccaaacaacaatccccctctacctaaaactcacagccctcgctgtcactttcctaggacttctaacagccctagacctcaactacctaaccaa
caaacttaaaataaaatccccactatgcacatttttatttctccaacatactcggattctaccctagcatcacacaccgcacaatcccctatctaggccttcttacgagccaaaacctgcc
cctactcctcctagacctaacctgactagaaaagctattacctaaaacaatttcacagcaccaaatctccacctccatcatcacctcaacccaaaaaggcataattaaactttacttcct
ctctttcttcttcccactcatcctaaccctactcctaatcacataacctattcccccgagcaatctcaattacaatatatacaccaacaaacaatgttcaaccagtaactactactaatca
acgcccataatcatacaaagcccccgcaccaataggatcctcccgaatcaaccctgacccctctccttcataaattattcagcttcctacactattaaagtttaccacaaccaccacccc
atcatactctttcacccacagcaccaatcctacctccatcgctaaccccactaaaacactcaccaagacctcaacccctgaccccatgcctcaggatactcctcaatagccatcgctg
tagtatatccaaagacaaccatcattcccctaaataaattaaaaaaactattaaacccatataacctccccaaaattcagaataataacacacccgaccacaccactaacaatca
atactaaaccccataaataggagaaggcttagaagaaaaccccacaaaccccattactaaacccacactcaacagaaacaaagcatacatcattattctcgcacggactacaacc
acgaccaatgatatgaaaaaccatcgttgtatttcaactacaagaacaccaatgaccccaatacgcaaaattaaccccctaataaaactaattaaccactcattcatcgacctcccca
ccccatccaacatctccgcatgatgaaacttcggctcactccttggcgcctgcctgatcctccaaatcaccacaggactattcctagccatgcactactcaccagacgcctcaaccgcct
tttcatcaatcgcccacatcactcgagacgtaaattatggctgaatcatccgctaccttcacgccaatggcgcctcaatattctttatctgcctcttcctacacatcggacgaggcctata
ttacggatcatttctctactcagaaacctgaaacatcggcattatcctcctgcttgcaactatagcaacagccttcataggctatgtcctcccgtgaggccaaatatcattctgaggggc
cacagtaattacaaacttactatccgccatcccatacattgggacagacctagttcaatgaatctgaggaggctactcagtagacagtcccaccctcacacgattctttaccttttcactt
catcttacccttcattattgcagcctagcagcactccacctcctattcttgcacgaaacgggatcaaacaaccccctaggaatcacctcccattccgataaaatcaccttccacccttta
ctacacaatcaaagacgccctcggcttacttctcttccttctctccttaatgacattaacactattctcaccagacctcctaggcgacccagacaattataccctagccaaccccttaaac

accccctccccacatcaagcccgaatgatatttcctattcgcctacacaattctccgatccgtccctaacaaactaggaggcgtccttgccctattactatccatcctcatcctagcaataa
tccccatcctccatatatccaaacaacaaagcataatatttcgcccactaagccaatcactttattgactcctagccgcagacctcctcattctaacctgaatcggaggacaaccagta
agctacccttttaccatcattggacaagtagcatccgtactatacttcacaacaatcctaatcctaataccaactatctccctaattgaaaacaaaatactcaaatgggcctgtccttgt
agtataaactaatacaccagtcttgtaaaccggagatgaaaacctttttccaaggacaaatcagagaaaaagtctttaactccaccattagcacccaaagctaagattctaatttaaa
ctattctctgttctttcatggggaagcagatttgggtaccacccaagtattgactcacccatcaacaaccgctatgtatttcgtacattactgccagccaccatgaatattgtacggtacc
ataaatacttgaccacctgtagtacataaaaacccaatccacatcaaaacccccctccctatgcttacaagcaagtacagcaatcaaccttcaactatcacacatcaactgcaactcca
aagccacccctcacccactaggatatcaacaaacctacccacccttaacagtacatagtacataaagccatttaccgtacatagcacattacagtcaaatcccttctcgcccccatgg
atgacccccctcagataggggtcccttgaccaccatcctccgtgaaatcaatatcccgcacaagagtgctactctcctcgctccgggcccataacacttgggggtagctaaagtgaact
gtatccgacatctggttcctacttcagggtcataaagcctaaatagcccacacgttccccttaaataagacatcacgatg

Appendix 6: Format of Consent Form

Genetic Struture of Newar Population of Kathmandu Valley

Consent Letter

Study No : .....

Participant's Name : .....

We would like your participation in this research as we need your blood sample to study the genetic linkage of Newar population. The purpose of this letter is to ensure your right to decide whether you want to participate in the research or not. You have the full right to ask questions if you are confused about the procedure. We will take 10ml of your blood .The whole procedure requires 10 minutes. The collected blood samples will be taken to India; Centre for Cell and Molecular Biology, Hyderbaad where further research work will be done. Confidentiality will be maintained regarding your identity. After the use of the blood samples in the current research, the result could be retained in Genetic Database for future use. While withdrawing blood you will not be harmed in any way.

Advantage:

There is no such definite personal advantage for being involved in this research. However, from this research the Genetic Data Base could be established.

Confidentiality

The results for this research could be published, but your identity will not be revealed.

Agreement for Self Participation

You are participating for this research according to your will. You can withdraw from this research at any time without any hesitation. By signing below you are agreeing that you

have read, have listened and your queries have been answered. Thus you allowed for 10 ml blood collection entirely by your own wish.

Participant's signature : .....

Participant's name : .....

Address : .....

Date : .....

Contact number : .....

Only for illiterate

I verify that I have read out and described all details to the above mentioned participant Mr..... I am ensured that he has understood all requirements, he was given chance to ask questions and he has agreed to participate in this research. I verify that the finger print below is of the participant.

Field worker's signature : .....

Field worker's name : .....

Date : ....
Participant's finger print

Appendix 7: Reference population to be studied

1. Tamang	Gaydan et al., 2007 and 2013	11. Hindu Tarai	Fornarino et al., 2009
2. Newar		12.AP Tribes	
3. Kathmandu		13.S. and N. India	Thangraj et al., 2005
4. Tibet		14.Nepal_Kat& Mix	Wang et al., 2012
5.North China	Deng et al., 2005	15.Burghats	Derenko et al., 2008
6. South China		16.Altain	
7.Han China		17.Uzbeck	
8. Tharu Eastern	Fornarino et al., 2009	18.Korean	
9.Tharu CI and CII		19.Japanease	
10. Hindu India		20.Mongloian	