MHC Class II Antigen Variation Study in Selected Populations of Northern Parts of India

THESIS SUBMITTED TO THE

UNIVERSITY OF NORTH BENGAL,

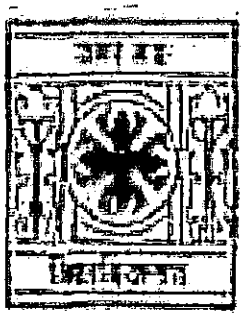
RAJA RAMMOHAN PUR

FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY
IN
SCIENCE

BY

Sanjeev Kumar Srivastava





University of North Bengal,
Dist. Darjeeling, West Bengal,
India

Sanjay Gandhi Post Graduate
Institute of Medical Sciences,
Lucknow-226014, India

2007

Ref.

591.2920954
S774m

198547
29 AUG 2007

STOCKTAKING-2011

In memory of my loving brother who left for heaven....

UNIVERSITY OF NORTH BENGAL
DEPARTMENT OF ZOOLOGY

P.O. North Bengal University
Dist. Darjeeling, West Bengal
India, Pin -734430

CERTIFICATE

This is to certify that the thesis entitled "**MHC Class-II Antigen Variation Study in Selected Populations of Northern Parts of India**" is the original investigative study performed by **Mr. Sanjeev Kumar Srivastava**, M.Sc. under my guidance and preceptorship and has not been submitted for a degree or diploma of any university. He has carried out the work during the period2003-2006 and has fulfilled the requirements of the degree of Doctor of Philosophy in Science (Zoology) of University of North Bengal.

He is conversant with techniques and literature cited in the dissertation and carried out the work thoroughly. In character and disposition Mr. Sanjeev Kumar Srivastava is fit to submit the thesis for Ph.D. degree.

Suraksha Agrawal

Prof. Suraksha Agrawal
Professor & Head,
Department of Medical Genetics
SGPGIMS, Luconow
(Co-Investigator)

T. K. Chaudhuri · 8/1/07

Prof. T. K. Chaudhuri
Professor
Department of Zoology
University of North Bengal
(Principal Investigator)

Phone: (0353) 2582124 Fax: (0353) 2581546
e-mail: Zoology_nbu@rediffmail.com Visit us at: www.nbu.ac.in

Acknowledgements

All these years of my PhD program seems like a little moment. Extended, standing still. Just yesterday as if. The help, cooperation and support of many associated directly or indirectly enabled me to complete this arduous task. Many gave me a helping hand and it is my duty and privilege to express my deep sense of appreciation for them.

I will remain ever grateful to Director, Sanjay Gandhi Post Graduate Institute of Medical Sciences, Lucknow, for allowing me not only to pursue my research goals but also develop a scientific attitude. I thank the Institute administration for providing the necessary infrastructure for research. I sincerely wish to express my thankfulness to the Vice Chancellor and Dean, University of North Bengal, Siliguri for selecting me as a PhD student and allowing me complete this grueling task.

I feel the limitation of words from inside in trying to express my feelings for my supervisor Dr. Tapas Kumar Chaudhuri, Professor, Department of Zoology, University of North Bengal Siliguri. Throughout my tenure, in different phases I always got the spontaneous help from him. I express my gratitude and sincere thankfulness to him for successfully completing my dissertation work with his intellectual guidance and constant inspiration.

I feel fortune to accomplish my research work under the dynamic guidance of my co supervisor Dr. Suraksha Agrawal, Head Department of Medical Genetics, Sanjay Gandhi Post Graduate Institute of Medical Sciences, Lucknow. Words are inadequate to express my heartful gratitude to her for illuminative guidance, constant inspiration, patience, encouragement, temperament and kindness to support me during ups and downs in the course of the study. I shall always remain indebted for her constant Motherly touch.

I am thankful to Dean sir, Dr. Prof. J. Pal (Head Dept of Zoology.), Prof.A. K. Chakravarty, Prof. D.C. Dev, Prof. A. Mukhopadhaya, Dr. S. Barat and Dr. M. Bahadur, of Zoology Department, University of North Bengal, for their valuable suggestions during my registration. I am also thankful to the non-teaching staff for their kind cooperation.

I fail to gather appropriate words to express my special thanks to my University Collegues Monojit Deobnath Da and Bikash Mitra because without their help and cooperation this task could not have been possible. I remember the days from my enrollment, registration sample collection, practical help etc. during my stay at sliguri when these people cooperated just like anything. I really enjoyed with

them. I am also thankful to Shikta Bannerji, Mayuk and Dr Anandita for their valuable suggestions.

I would be failing in my duties if I do not express my thanks to my friend Faisal Khan for all his concern, encouragements, guidance and assistance provided during my entire period. He not only helped and taught me like a teacher but also acted like brother.

There are two persons, which need a separate mention, Vivek Pratap Singh and Piyush Tripathi. I would like to thank Vivek Pratap Singh for his ever smiling greetings wherever and whenever we talked and discussed during these years, for his love, help, care and encouragements that I always felt whenever I was with him. I would appreciate Piyush Trilpathi who continuously helped me in the technicalities of the experiments and will always be thankful to him for his continous mental support Words are not adequate to thank Piyush for all he has done for me.

I thank Indian Council of Medical Research, New Delhi for providing me financial assistance during the tenure of this dissertation.

I was lucky to have colleagues who were always ready to help at times of need. I fondly remember the way Sudha di, Padma, Fahim and Meenal helped me during the laboratory work and analysis. Words are really not enough for the support of Sudha di, I remember how she taught me Dot Blot and setting up PCR. I am also thankful to Dr Himanshu, Manorama, Atul, Gaurav, Ariz, Tabrez, Abhishekh and Ankur for their constant support.

I do not know how to express my thanks to Neera didi and Sanjay Bhaiya. I will always remember the support, care and assistance of Neera didi. She helped me like an elder sister. Without the help of Sanjay Bhaiya this PhD thesis could not have been possible. He always guided me right things with sincerety and love. I would also like to express my thanks to Mishra ji and Shashank bhaiya for their kind cooperation. I am am also thankful to K C Ji for his help. I am also very much thankful to Anand Ji and Dua Ji for their secretarial help.

This long journey would have been impossible without my friends. Special thanks to my best friends Sandeep, Sanjeev, Vijay and Manish. Sandeep always lend me his hands at the time of need and lift my morale at the times of despair. Vijay and Sanjeev were always there with their undemanding care and spontaneous love and always took out time in spite of his being engrossed in their own priorities, both professional and personal. The togetherness that filled our hearts let us bear all the pressures and failures of life with smiles. The innocent regards tendered by Manish would remain memorable.

God cannot be with everyone all the times. That's why he made parents. A child can never thank his parents enough. He owes his very existence to them. My efforts will remain incomplete if I do not express my gratitude and indebtness to my parents whose blessings stood by me at all odds and for their constant inspiration and encouragements and always patting at times of achievements and always correcting me at times of needs. It would have been impossible to reach the present destination without their support. I owe a lot to my brother Manish for his constant care and affection. I thank my cousin brothers and sisters Jonty, Sonu, Monu, Himanshu, Guria and Shalu for their love, which always gave me a reason to smile.

I pay heartful thanks to my wife Keerti for her love and concern. During my ups and downs she was the person who gave me the reasons to smile. It would have been difficult without her to complete my task without her constant concern and care. Worda are really not enough for my son Aryan whenever I feel depressed, I used to see his smile and that make me refreshing again Mention of my family is incomplete without my Father in law and Mother in law. I specially thanks to Ankit, Pratiksha, Shruti and Smriti. I thank them for their everlasting love and care.

I am thankful to the God Almighty, for putting such wonderful people in my life and bestowing His Warmth and Affection through them.

Last but not least, I sincerely like to dedicate this thesis to my late elder brother Rupesh.

Sanjeev Srivastava
8.1.07
Sanjeev Kumar Srivastava

Contents

		Pages
1	**Introduction**	1 – 26
2	**Review**	27 – 55
3	**People**	56 – 87
4	**Materials and Methods**	88 – 117
5	**Observations**	118 – 143
6	**Discussion**	144 -196
7	**Summary**	197 – 200
8	**Bibliography**	201 - 222
	Appendices (A-I)	223 -249
	Publications	250

Abbreviations

ddH_2O	Double distilled water.
DNA	Deoxyribonucleic Acid.
EDTA	Ethylene di-amine tetra acetic acid.
EtBr	Ethidium bromide.
F_{ST}	Inter population inbreeding coefficient.
F_T	Intra population inbreeding coefficient.
gms	Grams.
G_{ST}	Coefficient of population differences.
HLA	Human leucocyte antigens.
HPLC	High pressure liquid chromatography grade.
IAM	Infinite allele model.
Kd	Kilo daltons.
LMP	Low molecular proteases.
$MgCl_2$	Magnesium chloride.
MHC	Major histocompatibility complex.
min	Minutes.
ml	Millilitre.
mM	Millimolar.
mtDNA	Mitochondrial DNA.
N	Normality.
NaCl	Sodium chloride.
NJ	Neighbour-Joining method.
O.D.	Optical density.
P^{32}	Radioactive isotope of phosphorus.
PAGE	Polyacrylamide gel electrophoresis.
PCR	Polymerase chain reaction.
RFLP	Restriction fragment length polymorphism.
rRNA	Ribosomal ribonucleic acid.
SDS	Sodium dodecyl sulphate.
SMM	Stepwise mutation model.
SNP	Single nucleotide polymorphisms.
SSO	Single strand oligonucleotide.
SSOP	Single strand oligonucleotide probes.
STR	Short tandem repeats.
t RNA	Transfer ribonucleic acid.
TAP	Transporter associated with antigen processing.
TBE	Tris-borate EDTA buffer.
TFC	Transferrins.
U	Units.
UPGMA	Unpaired group mean analysis.
X	Times concentrated.
µg	Micrograms.
µl	Microlitre.
φX	Phage phi X-174.

Introduction

Introduction

Human genetics research has generated enormous amount of data about the genetic differences among individuals and groups. Investigation of these differences has transformed our understanding of the origins and nature of human diseases (**Cavalli-Sforza, 2005; Bamshad *et al.*, 2004, Collins *et al.*, 2003**). Since long geneticists are involved in the Human genetic variation studies among the individuals forming a species, but the remarkable extent of this variation was not appreciated until about 25 years ago (**Cavalli-Sforza and Feldman, 2003; Jorde and Wooding, 2004**). Conspicuous human traits like hair and eye color clearly vary from one individual to the other in many populations; these differences are easily perceived by the layman, as are variation in height, weight, body build, and facial traits, which are also genetically determined to some extent. Their hereditary transmission, however, is complex, and these traits contribute little to understand the extent of variation. The first example of clear-cut genetic variation is that of ABO blood groups which was described at the beginning of the century (**Landsteiner, 1927**). Dissimilarities between individuals regarding ABO blood-group variation are due to small chemical differences between molecules found at the surface of red blood cells.

These studies were soon extended to other blood-group systems, and a body of data began to accumulate showing that different human populations have different proportions of blood groups. However, the first glimpse of the staggering magnitude of genetic variation came later beginning in the 1950s and coming to full development in the 1960s when individual differences for proteins could be systematically studied. A protein is a large molecule made of a linear sequence of components called amino acids; different proteins vary considerably in their amino-acid composition and serve very different functions. The relationship between structure and function has been demonstrated for many proteins. The same protein

may show small, strictly inherited differences between individuals. The first example was observed in the protein hemoglobin, in which the replacement of a specific amino acid by another was shown to determine a hereditary disease known as sickle-cell anemia. This first case of molecular pathology was detected by subjecting the protein to an electric field with a procedure called electrophoresis. The amino-acid replacement involved in sickle-cell anemia causes a change in the electric charge of the hemoglobin molecule, which allows the separation of normal and sickle-cell hemoglobin's. Electrophoretic analysis has since been further developed and has helped detect a great deal of variation in proteins. It is now known that the majority of the tens of thousands of different proteins found in an organism exist in more than one form, so that some individuals may have one form of the protein, whereas others may have another form (**Glazko G,2005**).

Protein variation is still the tip of the iceberg. Only when the analysis could be carried out at the level of the hereditary material itself, deoxyribonucleic acid (DNA), could the full extent of individual genetic variation begin to emerge. This technique became widely available only in the 1980s, and although comparisons of segments of DNA in different individuals are still rare, they are becoming more common. They are, however, adequate to convince us that there is much more variation at the DNA level than was suspected when only proteins and blood groups could be analyzed.

Population diversity and Evolutionary Process

Anthropologists tried to reconstruct evolutionary relationships and history on the basis of a single character or gene since long. A favorite for over 100 years was the cephalic index (the percentage of skull breadth to length) introduced shortly before the middle of the last century. However, with a single trait, two populations of different origin could well turn out to be more or less identical. Anthropometric traits of this kind also have another very serious drawback: there is no guarantee that the character is completely under the control of biological inheritance and the variations observed could be due to short-term response to environmental changes. Every gene frequency varies over time in ways that can be considered, at least superficially,

nearly random. Therefore, it is not surprising that populations having clearly different evolutionary histories may show similar gene frequencies (**Itoh T, 2002**). This drawback can be avoided if one cumulates the information from more than one gene. As one increase the number of genes considered simultaneously, the probability that a similar confusion takes place becomes more and more remote. In 1963 it was shown that even with as few as 20 alleles from five genes one could successfully attempt a reconstruction of human evolution. Further larger number of alleles were discovered due to which the picture become clearer (**Jorde and Wooding, 2004; Castro de Guerra *et al.*, 1999**).

Several methods allow us to combine the information from many genes into appropriate statistical indices. They are usually called multivariate to distinguish them from those using single traits or genes (univariate).

Multivariate analysis is especially useful for understanding evolutionary forces that tend to operate in a parallel fashion on all genes: migration and random genetic drift (the random fluctuation of gene frequencies in time, to be further explained later). These and other methods are applied to the existing data with the aim of extracting information of genetic and evolutionary interest. The reconstruction of human evolution, including the fissions, the major migrations, and the understanding of the roles of mutation, drift, and natural selection is often difficult and challenging. Results from genetic data should be compared with relevant knowledge from other fields, in particular, paleo- anthropology, prehistory, history, the geographic and ecological setting, and the cultural evidence that comes indirectly from linguistic studies.

If we know that there exist different genetic types of a specific protein or other strictly inherited character, we can count individuals carrying one type or the other and establish the proportions of that type in the population being examined. These proportions vary from one population to another because they change over time in each population in a relatively unpredictable manner (**Nei 2000**). The change in proportions of these types over time is the evolutionary process itself. It proceeds slowly but incessantly over generations. The analysis of populations living today in

different places gives us a cross section in time of this continuing process, which is inevitably diverse in the various parts of the inhabited Earth.

So the primary interest is in understanding this evolutionary process. The first task is to describe the existing variation, using a variety of techniques to test the relevant evolutionary models. Initially the interest was restricted to aboriginal populations, which were already living in the area of study in A.D. 1492. After this time, geographic discoveries stimulated the expansion and migrations of the economically more advanced populations all over the planet. Some movement took place before A.D. 1492, but at a smaller scale. Ordinarily, populations that migrated after that date have mixed only partially with earlier residents and were easily recognizable on the basis of physical appearance and historical and social knowledge **(Underhill 2004; Carmelli and Cavalli-Sforza,1979).**

A population is a local group of organisms of the same species that normally interbreed. The word population does not refer to an entire species; it refers instead to a group of organisms of the same species living within a sufficiently restricted geographical area where members can potentially mate with each other provided they are of opposite sex. A group of individuals within whom marriages are contracted is called a Mendelian population. The set of genetic information carried by a population is known as its gene pool. The gene pool of a new generation is descended from the parental generation but for several reasons, including chance, the gene pool of the new generation may have different allele frequencies than the parental pool. Overtime, the changes in allele frequency can cause changes in phenotypic frequency. The long-term effect of changes in allele frequency is evolutionary change. Members of the same local group are more closely related to each other than members of groups who live in different geographical areas and people who live in the same geographical region are more similar than those separated by geographical barriers. Populations grow and interact with one another through competition and predation. These factors can influence behavior ecology and evolution, either at micro or macro level. The evolution of such population results in to the change in the genetic constitution of the population followed by alterations in

the genotypic and phenotypic constitutions (**Joaquim Fort 2004; Castro de Guerra *et al.*, 1999**).

To achieve the above goals the important step in the understanding of the genetic structure of human populations consists of estimating frequencies of alleles at different loci. Gene frequencies are the estimates of the relative frequencies of alleles and are of widest application in the studies of structural dynamics and evolution of natural (particular in human) populations. Information about the relative frequencies of different alleles is of paramount importance in the study of population structure. However, the estimation of gene frequencies is plagued by the phenomenon of sampling fluctuations and misclassification of genotypes however this can be solved by using right markers and more robotic techniques. After the selection of the markers it is important to use more robust statistical tools. In any population, the genotype frequencies among zygotes are determined in large part by the patterns in which genotypes of the previous generation come together to form mating pairs. In random mating, genotypes form mating pairs in the proportions expected from random collisions. For a gene with two alleles A and a in a random-mating population, the expected genotype frequencies of AA, Aa, and aa are given by p2, 2pq, and q2, respectively, where p and q are the allele frequencies of A and a respectively, with p + q=1. The expected genotype frequencies with random mating constitute the Hardy-Weinberg equilibrium (HWE). The rate at which the HWE frequencies are attained depends on the life history of the organism. In an organism with non overlapping generations, such as an annual plant, each generation is separated in time from the preceding and the following generation; in this case, the Hardy-Weinberg frequencies are attained in one generation of random mating provided that the allele frequencies are equal in both the sexes. In an organism with non overlapping generations, the approach to HWE is gradual. Statistical tests of HWE are often based on the χ^2 test, but this test is relatively weak in detecting departures from the expected frequencies, especially those caused by admixture of subpopulations differing in allele frequency.

One of the principal implications of the HWE is that the allele frequencies and the genotype frequencies remain constant from generation to generation, hence

genetic variation is maintained. Another major implication is that, when an allele is rare, the population contains many more heterozygote for the allele.

Nonrandom association between the alleles of different genes is measured by the linkage disequilibrium parameter D. Random association between alleles of different genes is called linkage equilibrium, and it is indicated by D, when D=0, the alleles are said to be in linkage disequilibrium. Ordinarily, unless there is some countervailing process that maintains linkage disequilibrium between two genes, D is expected to go to zero at a rate determined by the recombination fraction between the genes that recombine with a frequency r, D decreases by the fraction r in each generation. Significant linkage disequilibrium is usually found in natural populations for genes that are tightly linked, for genes that are within or near an inverted segment of chromosome. Significant linkage disequilibrium can also result from admixture of two or more subpopulations differing in allele frequencies. The gene frequencies can be changed and the factors that shape the contemporary genetic diversity include Mutations, Random genetic drift, Gene flow or genetic exchange, Natural selection **(Cavalli-Sforza,2003; Templeton,2002).**

Mutations

The geographic representation of the data of living population, a segment of DNA endowed with a specific function called gene is being studied extensively these days and for each gene, analysis of the alternative form of gene called alleles is carried out, and these alleles are a result of mutations occurring in a populations. Geographic maps of an allele are useful for understanding facts specific to that allele, including its evolutionary history and the effects of evolutionary factors like mutation and natural selection **(Roychoudhury, 1988).** The geographic distribution of a particular allele may give information on the place of origin of the genetic change (mutation) that generated it. Mutation results in the rise of both functional (i.e., coding) and nonfunctional (i.e., noncoding) polymorphisms Geographic maps of an allele are useful for understanding facts specific to that allele, including its evolutionary history and the effects of evolutionary factors like mutation and natural selection. The geographic distribution of a particular allele may give information on

the place of origin of the genetic change (mutation) that generated it. Correlations of the distributions of gene frequencies with environmental parameters at the geographic level have been instrumental in the discovery of specific genetic adaptations **(Nei 2005).**

It is well established that the proportion of an allele varies considerably from place to place, but usually there is little difference between neighboring populations so that the greatest variation is observed at large distances. It is thus possible to prepare geographic maps representing these proportions for a particular allele (also called allele frequencies, or simply, gene frequencies) when a sufficient number of populations have been tested.

Allele and genotype frequencies change in unexpected ways in successive generations with dispersive processes such as reduced population size and migration. Conversely, the direction of alleles and genotypes changes over time due to population size, birth rate, disease, migration and adoption to environmental factors. The sickle-cell anemia gene was the first example, because its geographic distribution showed a correlation with that of malaria **(Haldane,1954).** Angiotensin gene (AGT), which, encodes important component of rennin-angiotensin blood pressure regulation pathway. In many populations it had been found that 235T a variant of AGT is associated with 10-20% increase in the risk of developing hypertension **(Kunz & Kreutz *et al.*, 1997)**. This is found in 90% frequency in some African populations and 30% in European populations **(Nakajima T *et al.*, 2004).**

Random genetic drift

Sometimes populations may change in frequency over time as a result of the accumulation of random sampling error in the passing on of alleles from generation to generation. When a very small number of individuals migrate and start a new population, the sampling error (random genetic drift) is very large, and allele frequencies in the new population may be very different from those in the parent population. Allele frequencies fluctuate from one generation to the next because of the randomness of the transmission process until, by chance, one of the alleles is lost and the other is fixed. As the new population grows over a few

generations, the magnitude of the sampling error per generation decreases and the new population will continue to have very different frequencies from the parent population. This extreme form of random genetic drift is referred to as a "founder effect," **(Strachan *et al.*, 1996)** because the population expanded from very few founders with a relatively restricted gene pool. For example, available evidence suggests that a small group of individuals left Africa and over time allele frequencies changed markedly from those in the African populations left behind **(Nakajima *et al.*, 2004)**.

Gene flow or genetic exchange

Gene flow is a process by which interbreeding among certain groups of individuals results in those populations becoming increasingly similar to each other. Two populations that start off quite different genetically, if they mate, can produce offspring that represent the genes present in both of the original populations. Matting can be either positive assortative mating or negative assortatve mating positive assortative mating increases homozygosity in the population because of mating between individuals with identical genotypes, while if mating is carried out between individuals with different genotypes then it is called negative assortatve mating and this will result in high heterozygosity in a population **(Cavalli-Sforza *et al.*, 1999)**.

Natural selection

By natural selection organisms with gene patterns that are adaptive to a given environment become more prevalent over time. For example, organisms that can adapt to changing climatic patterns are at an advantage over those that adapt only with great difficulty. Natural selection is a constantly shifting process. It is influenced not only by an organism's biology, but also by the interaction of that biology with environmental conditions **(Sternberg,2004)**.

Variation in Heritability within a Given Population

Heritability is not a fixed value for a given attribute **(Herrnstein & Murray, 1994)**. Heritability depends on many factors, but the most important one is the range of environments. Because heritability represents a proportion of variation, its value will

depend on the amount of variation. Herrnstein in 1973 pointed out, if there were no variation in environments, heritability would be perfect, because there would be no other source of variation. If there is wide variation in environments, however, heritability is likely to decrease. When one speaks of heritability, one needs to remember that genes always operate within environment contexts. All genetic effects occur within a reaction range such that, inevitably, environment will have differential effects on the same genetic structure. The reaction range is the range of phenotypes (observable effects of genes) that a given genotype for any particular attribute can produce, given the interaction of environment with that genotype. For example, genotype sets a reaction range for the possible heights a person can attain, but childhood nutrition, diseases, and many other factors affect the adult height realized. Moreover, if different genotypes react differently to environmental variation, heritability will show differences depending on the mean and variance in relevant environments **(Lewontin,1974)**. Thus, the statistic is not a fixed value.

Adaptation and Migration

In zones with more intense exposure to sunlight, darker skin puts individuals at an adaptive advantage. The melanin that acts as a pigmentation to produce darker skin better protects individuals against the damage that large amounts of ultraviolet radiation can cause to the skin, this radiation increases susceptibility to skin cancer, especially melanoma, a form of skin cancer that easily can become fatal. In zones with weaker exposure to sunlight, lighter skin is an advantage. People rely on sunlight to produce active vitamin D3 in the capillaries. Lighter skin allows greater bodily production of vitamin D3. Deficiencies in vitamin D3 can cause rickets in children and osteoporosis in adults. There is as yet no conclusive evidence of positive selection for light coloration. Instead, evidence to date may indicate that light pigmentation in climates distant from the equator represents a lessening of the selective factors that lead to dark pigmentation near the equator rather than to any particular factors leading to lighter pigmentation **(Harding *et al.*, 2000)**

It is probable that much of the variation seen among groups of humans indirectly resulted from the pattern of expansion and migrations accompanied by

random genetic drift. Over the years, frequencies of DNA variants changed only slightly in terms of total DNA composition but changed enough to produce differences, many of which we still do not fully understand.

Models of modern human evolution

There are three main theories for the evolution of modern humans.

Multiregional model

The multiregional model proposes that there was no single geographical origin for modern humans but that, after the radiation of *HOMO ERECTUS* from Africa into Europe and Asia ~800,000–1.8 million years before present (yr BP), there were independent transitions in regional populations from *H. erectus* to *Homo sapiens*. This model is supported primarily by the continuity of certain morphological traits in the fossil record (for example, the robust cheekbones observed in *H. erectus* fossils from Southeast Asia and in modern Australian aborigines), which indicate that modern populations evolved over very long periods of time in the regions where they are found today.

Simultaneous evolution from *H. erectus* to *H. sapiens* in dispersed populations could have been achieved through extensive gene flow between populations, requiring a large effective population size to sustain gene flow among geographically diverse populations. The RAO model proposes that all non-African populations descend from a *H. sapiens* ancestor that evolved in Africa 100,000–200,000 yr BP. This ancestor then spread throughout the world, replacing archaic *Homo* populations (for example, the Neanderthals). The fossil record supports this model, as the earliest modern human fossils were found in Africa and the Middle East, dating to 90,000–120,000 yr BP **(Figure 1.1)**.

Recent African origin (RAO) model

The RAO model predicts that all genetic lineages derive from a recent common African ancestor and that non-African populations should carry a subset of the genetic variation present in modern African populations.

Assimilation model

The assimilation (or hybridization) model proposes that gene flow between the early human populations was not equal over time and space. This model allows for some gene flow between modern humans that migrated from Africa and archaic populations (for example, the Neanderthals) 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. This model predicts that the modern gene pool derives from variable contributions of genes from archaic African and non-African populations.

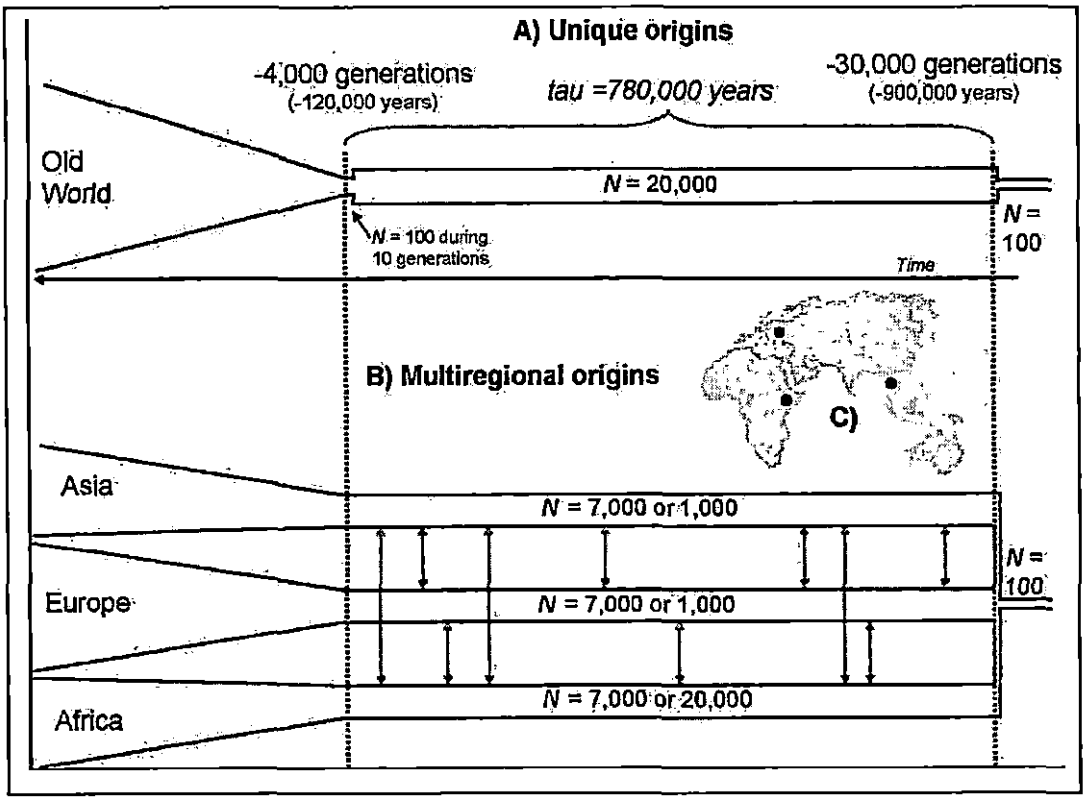


Figure 1.1 Demography and time line of the different simulated evolutionary models. (*A*) Unique origin (UO) model. In this model, 30,000 generations ago, a small population (N = 100 genes) went through a demographic expansion after a first speciation event. Then, 4000 generations ago, a range expansion followed a bottleneck of 10 generations to mimic a second speciation event. The large population preceding the speciation and range expansion can be considered to be a large subdivided population. (*B*) Multiregional evolution (ME) model. As in *A*, a small population went through a speciation event and instantaneously colonized the three continents 30,000 generations ago. For 26,000 generations the continents harbored relatively large populations and exchanged occasional migrants. Then, 4000 generations ago, three range expansions were initiated from the three different origins shown in C

The Origin of Modern Humans

Information about the history of our species comes from two main sources: the paleo-anthropological record and historical inferences based on current genetic differences observed in humans. The existing fossil evidence suggests that anatomically modern humans evolved in Africa, within the last approx 200,000 years, from a pre-existing population of humans (**Klein, 1999**). Although it is not easy to define anatomically modern human in a way that encompasses all living humans and excludes all archaic humans (**Lieberman *et al.*, 1992, 2003**), the physical characteristics of anatomical modernity include a high rounded skull, facial retraction, and a light, as opposed to heavy and robust, skeleton (**Lahr, 1996, 1998**). Early fossils with these characteristics have been found in eastern Africa and have been dated to approx 160,000– 200,000 years ago (**White *et al.*, 2003; McDougall *et al.*, 2005**). At that time, the population of anatomically modern humans appears to have been small and localized (**Harpending *et al.*, 1998**). Much larger populations of archaic humans lived some where in the Old World, including the Neanderthals in Europe and an earlier species of humans, *Homo erectus,* in Asia (**Swisher *et al.*, 1994**). Fossils of the earliest anatomically modern humans found outside Africa are from two sites in the Middle East and date to a period of relative global warmth, approx 100,000 years ago, though this region was rein habited by Neanderthals in later millennia as the climate in the northern hemisphere again cooled (**Lahr and Foley, 1998**).

Groups of anatomically modern humans appear to have moved outside Africa permanently sometime 160,000 years ago. One of the earliest modern skeletons found outside Africa is from Australia and has been dated to approx 42,000 years ago (**Bowler *et al.*, 2003**), although studies of environmental changes in Australia argue for the presence of modern humans in Australia 155,000 years ago (**Miller *et al.*, 1999**). To date, the earliest anatomically modern skeleton discovered from Europe comes from the Carpathian Mountains of Romania and is dated to 34,000–36,000 years ago (**Trinkaus *et al.*, 2003**). Existing data on human genetic variation support and extend conclusions based on the fossil evidence. African populations

exhibit greater genetic diversity than do populations in the rest of the world, implying that humans appeared first in Africa and later colonized Eurasia and the Americas (**Tishkoff & Williams 2002; Yu *et al.*, 2002; Tishkoff &Verrelli 2003**). The genetic variation seen outside Africa is generally a subset of the variation within Africa, a pattern that would be produced if the migrants from Africa were limited in number and carried just part of African genetic variability with them (**Cavalli-Sforza & Feldman 2003**). Patterns of genetic variation suggest an earlier population expansion in Africa followed by a subsequent expansion in non-African populations, and the dates calculated for the expansions generally coincide with the archaeological record (**Jorde *et al.*, 1998**). Studies of mtDNA (**Ingman *et al.*, 2000**), the Y chromosome (**Underhill *et al.*, 2000**), portions of the X chromosome (**Kaessmann *et al.*, 1999**), and many (though not all) autosomal regions (**Harpending & Rogers 2000**) support the "Out of Africa" account of human history, in which anatomically modern humans appeared first in eastern Africa and then migrated throughout Africa and into the rest of the world, with little or no interbreeding between modern humans and the archaic populations they gradually replaced (**Tishkoff *et al.*, 2000; Stringer 2002**). However, several groups of researchers suggests that humans bearing modern traits emerged several times from Africa, over an extended period, and mixed with archaic humans in various parts of the world (**ZieTkiewicz *et al.*, 1998**). As a result, they say, autosomal DNA from archaic human populations living outside Africa persists in modern populations, and modern populations in various parts of the world still bear some physical resemblance to the archaic populations that inhabited those regions (**Wolpoff *et al.*, 2001**). Studies of mtDNA from archaic and modern humans and Y-chromosomes suggest that any surviving genetic contributions of archaic humans outside Africa must be small, if they exist at all (**Krings *et al.*, 1997; Nordborg 1998; Takahata *et al.*, 2001; Serre *et al.*, 2004**). The observation that most genes studied to date coalesce in African populations points toward the importance of Africa as the source of most modern genetic variation, perhaps with some subdivision in the ancestral African population (**Satta and Takahata 2002**). In addition to having higher levels of genetic diversity, populations in Africa tend to have lower amounts of linkage

disequilibrium than do populations outside Africa, partly because of the larger size of human populations in Africa over the course of human history and partly because the number of modern humans who left Africa to colonize the rest of the world appears to have been relatively low (**Gabriel *et al.*, 2002**). In contrast, populations that have undergone dramatic size reductions or rapid expansions in the past and populations formed by the mixture of previously separate ancestral groups can have unusually high levels of linkage disequilibrium (**Nordborg & Tavare, 2002**). Sequence data for hundreds of loci from widely distributed worldwide populations eventually may clarify the population processes associated with the appearance of anatomically modern humans (**Wall, 2000**), as well as the amount of gene flow among modern humans since then.

Genetic tools for studying history

The advent of various techniques in molecular biology enabled the application of genetics to the study of human evolution giving rise to the fields of molecular evolution and molecular anthropology. In early applications, genetic or DNA markers were that of functional genes. These DNA markers were relatively less informative and were subjected to selection pressures. More recently, more informative DNA markers were discovered along with more robust DNA typing technology. These recent DNA markers are neutral, not being part of functional genes, are therefore not affected by selection pressures and are stable across populations and generations. These DNA markers are also highly polymorphic making them more informative in studying genetic variation between and amongst human populations.

Short Tandem Repeat (STR)

STR DNA markers (autosomal STRs and Y chromosome STRs) are short fragments of DNA that are commonly used in forensic human identification. However, because of high polymorphism, these markers are useful in studying human genetic diversity. Single Nucleotide Polymorphism (SNP) DNA markers (autosomal SNPs and Y chromosome SNPs) are characterized by single base

changes in DNA sequence. SNPs are highly stable and preserved across populations and generations allowing analysis of genetic diversity across time and geography.

Mitochondrial DNA (mtDNA)

Mitochondrial DNA markers are essentially variations in DNA sequences. Similar to SNPs, mtDNA is highly stable allowing lineage analysis among populations across time and different geographical areas. There are also certain features unique to these markers. STRs, SNPs and mtDNA are highly effective in analyzing degraded samples making them suitable to analyze anthropological samples (human remains, ancient DNA).

Y chromosome markers

Y chromosome markers (STRs and SNPs) are exclusively paternally inherited, allowing evolutionary genetic analysis of the male lineage. As a complement, mtDNA is maternally inherited allowing genetic analysis of the maternal line.
HLA / MHC being highly polymorphic is another marker of choice nowadays used widely in population genetic study.

Major Histocompatibility Complex

The Major Histocompatibility Complex (MHC) is unique in that it is the most polymorphic system in the human genome and the only system to display functional polymorphism (**Spinola, 2005; Marsh, 2000**). Due to its high polymorphism, tight linkage among the loci and non random association of alleles this system has become interesting from perspective of population genetics. All the regions of MHC are known to be highly polymorphic, constituting several closely linked loci each with large number of genes that can be further split into many allelic types differing in their nucleotide sequences. Therefore the importance of this system in the study of polymorphism and their significance in population selection and survival and in providing clues to mechanism of generation as well as maintenance of this variability within the populations is immense.

With the advent of DNA based molecular typing techniques the polymorphisms in various allelic families of the MHC have also been revolutionized The number of alleles increases with the different typing methods like from serology to RFLP and IEF, followed by Polymerase chain reaction methods of PCR-SSP and SSOP, to Reverse Blot Hybridization to more refined Sequence based typing to the field of Micro array technology.

Apart from being an invaluable tool for population genetic studies, MHC polymorphism has important role in transplantation and disease associations. HLA associations have also helped to define syndromes of disease categories having common / shared pathogenic mechanism, like ankylosing spondylitis and related spondylo-arthropathies are associated with HLA-B27. It has been studied that HLA associations with infectious and autoimmune disease shows susceptibility and protective alleles in populations of different ethinic origins (**Hill ,1991; Carrington M,1999; Bowness P,2002**).

HLA associations with diseases vary in different populations. Disease predisposing genes and their molecular subtypes could help to determine and predict the incidence of the diseases in some populations. It is therefore important to have a population based database of HLA alleles and their frequencies of prevalence in healthy individuals so that disease predisposing influence of a particular phenotype could effectively be assessed in the populations.

Two individuals differ in only 0.2% of the genome leading to diversity, which is very important for natural selection and survival. Restricted distribution of alleles and subtypes of globally prevalent alleles within populations has helped to assign the ethnic origin of alleles. Alleles predominantly found in selected populations of similar ethnicity can thus be characterized as typically belonging to particular ethnic group and referred to as "Oriental allele" or a "Caucasoid allele". This distribution of subtypes of globally distributed allelic families has helped to trace the ethnicity and lineages of populations. Several theories of origin of human population and their routes of dispersal have been suggested. Of these, the Origin from Africa theory and subsequent migrations to the east and the west coupled with large number of natural selection processes and phenomenon has been the most

controversial as well as best supported by evidences (**Hill *et al.*, 1992; Diamond 1994; De Knijff L 2001**). Being a functionally polymorphic system, investigations into the distribution of MHC alleles in world populations are very important in this regard since the MHC genetic makeup of each of these populations would reflect interplay of both the basic genetic origin and effects of natural phenomenon such as founder effect and environmental selection. Differences in the prevalence of HLA alleles in different populations in varied environmental conditions could be utilized to assess the role of each of these alleles in conferring survival advantage to human populations.

Indian Subcontinent

Indian population is particularly relevant lying in a geographical transition zone between the western Caucasians and the Orientals in the East. Because of the historical, racial admixture, the population has become a 'melting pot of various races'. It shows immense heterogeneity in terms of cultures, languages, customs, religions and other factors that are known to divide people in groups. Particularly the population in the North has an extensive history of frequent migrations and invasions from both the east as well as the west of the subcontinent and constant amalgamation of these populations. Due to extensive history of admixture in this population, the gene pool is hypothesized to have been influenced by genetic contributions from various races. Previous studies on the MHC and other gene systems in this population provide evidence for the historical documentation of constant admixture (**Chhaya 2005; Jaini , 2002; Agrawal & Arundhati , 1999; Su B , 1999; Balakrishnan 1996; Mehra *et al.*, 1986; Mittal 1982; Shankarkumar , 1999, Chakraborty 1992**).

Indian subcontinent is located between 8 degree N to 37 degree N latitude and 68 degree to 97 degree longitude. It is an assemblage of more than one billion individuals that constitute 1/5th of the total world population. There are many reasons for the researchers to explore the gene pool of Indian populations; Geography is one of the important factors (**Cann 2001; Lahr & Foley 1998**). Second is extensive gene flow through series of migrations and invasions, resulting in the distribution of the

198547
29 AUG 2007
LIBRARY

contemporary genetic variation across different geographic locations of India and creation of large genetic diversity. All these factors resulted into formation of various religions, cultures, communities, castes and linguistic groups (**Johnson *et al.*, 1992**) that resulted into the structuring of population and formation of various endogamous groups. Broadly it could be classified as Dravidians and Aryans. The Dravidians were considered as the original inhabitants of India who were driven south wards following invasions by Aryans from north-west during second and third millennium BC. Aryans were migrated from Iranian plateau. The colour of their skin was light and dark both; they have dark black hair and brown eyes. Language was mostly Sanskrit or derived Sanskrit. They introduced highly elaborated caste system in India with divisions into priests *(Brahmin), warriors (Kshatriya), Traders (Vaishya)* and inferior craftsman *(Sudra)* (**Mehra *et al.*, 1986**) .The broader four groups have been subdivided into smaller groups which marry within themselves. As a result of this the entire population has been divided into a large number of groups. The Indian caste system does not permit large-scale inter-caste, inter-religious and inter-ethnic marriages whereby the gene pool of each caste has evolved over the times and might have been fixed (**Jensen 1991**). These groups follow strict endogamy, which has resulted in a great deal of variation in the mating patterns, all of which invariably result in a wide genetic diversity (**Malhotra & Vasulu,1993; Naipaul,1992**).

Historically, it is known that various invasions have caused admixturing of Negrito, Negroid, Protoaustraloid, Mongoloid and Europoid elements in the Indian populations. Further, the population is also subdivided into four linguistic families i.e. i) Austro-Asiatic ii) Dravidian iii) Indo-European and iv) Tibeto-Burman. Under these four linguistic families about 325 languages are spoken in India. As per 1971 census there were 73.8% speakers of Indo-European languages, 24.2% of Dravidian languages, 0.8% of Tibeto-Burman family and only 1.2% Austro-Asiatic speakers. The population of India has been subjected to successive waves of immigrations and invasions. In the second millennium BC, waves of Indo-European speakers invaded the Indian subcontinent and imposed both religion and hierarchical caste system.. In 325-327 BC Alexander's Macedonian army invaded from both west and east with incursions into Punjab followed by Muslims in 1200-1500 AD. At that time the

existing religion was Hinduism, with numerous modified forms, e.g. Buddhism, Jainism and the animistic religions practiced by the tribal populations. During the Muslim rule spanning three centuries a large section of the population was converted to Islam. This led to another religious division within the same caste population. Around 500 years ago, the birth of Sikhism took place, but was confined largely to the North -Western parts of India, among the Punjabi speaking population. . Later empires of Muslims declined and the Portuguese and Dutch fragmented the whole kingdom into smaller kingdoms that were then first occupied and then Britishers occupied it from 1600-1947. This was the period when Christianity arrived. Most of the section of the population then became a follower of Christianity. Due to invasions and massive population movements in India its population structure subdivided into caste and tribes **(Gonem A,1996)**.

The emerging pattern of social organization based on endogamy and system of marriage rules among subdivided populations provides a clear picture of the biological composition with wide genetic diversity. Another important dimension of the Indian populations, especially among small populations, is that it offers potential opportunities for the operation of micro-evolutionary forces, which bring rapid changes in gene frequency of certain genetic traits **(Malhotra & Vasulu 1993)**.

All these factors must have affected the Indian populations and might have resulted into the unique gene pools. There is a need to study these unique populations at genetic level **(Agrawal,2004, Agrawal *et al.*, 2005)**. To study this it is required to investigate the populations for a set of genetic characters. It is of great interest to see whether the populations can be grouped in some way on the basis of inter relationships. The similarities and differences between the populations within such groups and those between populations in different groups can then be used to get some idea about the nature of forces that have prevailed in the evolutionary history of the populations. The differences in the populations can be studied using various statistical measures like breeding value (A), coefficient of inbreeding (F), coefficient of kinship (f), heritability (h^2), population mean (M), effective population size (N_e), Variance (V), rate of Inbreeding (ΔF), change of gene frequency (Δq), genetic distances, and phylogenetic relationships etc. It does not matter which statistical index

is being used, since different indices are all highly correlated. All of these are functions of differences in allele frequency. The essential feature of all these indices lies in representing the populations by points in a multi-dimensional space. They also allow the measurement of distance geometrically **(Balakrishnan,1988)**.

The study of populations at genetic level involves the identification of different alleles through observation of the expressed traits or outward, physical manifestation of a gene, called the phenotype. Mendelian genetics allowed population geneticists to identify the heritable form of a gene (genotype) including individual variants (alleles). Advances in molecular genetics – laboratory research on the chemical structure encoding the genes (DNA) facilitated identification of single genes at the molecular or biochemical level. Regardless of the method used to identify genes and their alleles, allelic data obtained on different populations is analyzed by statistical analyses of allele frequencies to understand and make prediction about gene flow in populations, past, present and future.

An important step in the understanding of genetic structure of human populations consists of estimating frequencies of alleles at different loci. Gene frequencies are the estimates of the relative frequencies of alleles and are of widest application in the studies of structural dynamics and evolution of natural (particular in human) populations. Information about the relative frequencies of different alleles is of paramount importance in the study of population structure. However, the estimation of gene frequencies is plagued by the phenomenon of sampling fluctuations and misclassification of genotypes however this can be solved by using right markers and more robotic techniques.

A population is a local group of organisms of the same species that normally interbreed. The word population does not refer to an entire species; it refers instead to a group of organisms of the same species living within a sufficiently restricted geographical area that members can potentially mate with each other provided they are of opposite sex. Members of the same local group are more closely related to each other than members of groups who live in different geographical areas and people who live in the same geographical region are more similar than those separated by geographical barriers.

Populations grow and interact with one another through competition and predation. These factors can influence behavior ecology and evolution, either at micro or macro level.

Barriers to mating have partitioned the world's population into subpopulations, which are geographically distinct **(Comas *et al.*, 1998).** Subpopulations within a population may be completely isolated, which provides the information that there is no migration among subpopulations. They may be incompletely isolated, which is more often the case. However, the populations that live in geographical proximity do not always interbreed. Additional barriers may be language and cultural affiliations (tribe, country, etc). In practice, none of these barriers are absolute and it is rarely possible to define a population unit that is sharply distinguished from a geographically adjacent one. For the most part, genetic features that distinguish one population gradually merge into the next.

Genome diversity study in humans not only quantify the genomic variations within and between populations but also can reconstruct the evolutionary history of populations using the genomic data and trace the human migration and correlate the genomic diversities and affinities with archaeological, linguistic, epidemiological and cultural histories with a view to reconstruct the story of human evolution **(Nitai Pada Bhattacharyya,1999; Partha P Majumdar, 1999).**

During the process of evolution, new form of gene are naturally introduced by a process known as mutation and existing form of gene may be lost by process of natural selection and random genetic drift. When the humans grouped themselves in such a way that member within the group can interbreed but the exchange of the genes was rare because of cultural, geographic barriers and biological evolution of these groups. The extent of divergence is correlated with the time of such independent evolution of the populations. The expected genetic divergence between two such independent populations evolving for a known period of time can be calculated **(Crow and Kimura, 1970; Nei ,1987).**From the estimate of the genetic distance between the two populations, their time of divergence can be estimated and can provide useful information for reconstructing the evolutionary history. From about the turn of this century, studies based on diversity and affinities based primarily on

polymorphisms in expressed genes have been conducted globally as well as in India which yielded many interesting results on our origins (**Cavalli-Sforza *et al.*, 1994**). Due to the effects of natural selection on naturally occurring variants of expressed genes, the amount of variability at such gene loci between individuals is limited. With the introduction of molecular genetics techniques; regions of the DNA that do not code for expressed proteins can be studied in individuals. Such regions have the advantage that naturally occurring Alleles in these Loci are not under differential effects of natural selection and hence evolve as neutral alleles. Such regions are highly polymorphic which is extremely useful for population genetic studies .In the recent past very large number of highly polymorphic loci have been discovered throughout the human genome, thus it is possible now to study multiple loci within a short length of the genome, thus enabling construction of haplotypes. Haplotypic variation provides much more information for population movements and disease susceptibility, than variation at individual unlinked loci (**Partha P Majumdar, 1999**).

New alleles and haplotypes appear in the population because of mutations that occur in the germ line of individual organisms. This means that many genes are polymorphic, two or more alleles being present in the population as a whole, each with its own allele frequency. Allele frequencies change over time, due to natural selection and random genetic drift. However, the most important step in understanding the genetic structure of populations consists of screening studies aimed at estimating frequencies of alleles at different loci and interpreting their relevance to the particular phenomenon. A relative frequency of different alleles provides the information about importance in the study of population structure. Before knowing the population divergence it is important to understand a short history of mankind

Starting about 500 years ago large-scale population movements started to mix world populations on a scale not encountered previously. Some of these migrations are:

1. Europeans expanded into America, Australia and South Africa.
2. West Africans were transported to the Caribbean and North America through the slave trade. Afro-Caribbeans have further migrated to the British Isles.

3. Migrations from the Indian subcontinent to the British Isles
4. Migration from the far east to the USA

As a result of these migrations and admixtures the populations have been partially but not fully isolated. Hence, the genetic structures of various populations have undergone number of complexities and fission. The present study is an attempt to study the six North Indian populations of Uttar Pradesh and three North Eastern populations of North Bengal at genetic level by using highly polymorphic HLA class II antigens markers. This set of genetic markers would be helpful to provide distinct profiles, which might be expected if different evolutionary processes promoted their molecular differentiation.

Utility of population studies

The critics of genetic variation studies put forward the *'common disease-common variant'* (CD/CV) hypothesis, which states that the common genetic diseases are affected by common disease susceptibility alleles (or variants) at few loci that exist at high frequency across ethnically diverse populations (**Chakravarti, 2001; Reich, 2001**). These alleles probably arose before population differentiation and are common across populations. Supporters of CD/CV hypothesis cite that all population specific or frequent diseases like high frequency of C28Y-HFE allele and hematochromatosis; and that of Δ508-CFTR alleles and cystic fibrosis among northern Europeans are because these diseases are monogenic disorders. Therefore, it has been quoted that although substantial genetic variation is there but it is present in all populations. On the contrary, the neutral genetic variation aids biomedical research in at least three ways:

(i) Molecular sub-classification of the diseases can be done on the basis of genetic profile. High frequency of hemoglobin HbS allele, variant of glucose-6-phosphate de hydrogenase and sickle-cell anemia are found among sub-Saharan Africans (**Tishkoff *et al*, 2001; Luzzato and Mehta, 1995**). However, the same disease with the underlying mutations is also found in Hispanics and inhabitants of northwestern India (**Braun, 2002**) and central Greece (**Braun, 2002; Kevles, 1995**). Therefore, labeling the disease only on the basis of ethnic affiliation or

phenotypic occurrence can be wrong interpretation and could possess serious health consequences. The concept of genetic ancestry is a much better indicator than race or ethnicity to determine that whether one carries the marker of a genetic disease. It has also been reported that differential effects of risk allele is seen in people with different genetic ancestry like homozygous APOE4 Asian individuals have ~5 fold higher risk of developing Alzheimer's disease than homozygous Africans (**Farrer *et al*, 1997**).

(ii) Information about an individual genetic ancestry can be used to improve medical diagnosis and treatment. The genetic differences among ethnic groups often cause differences in drug responses. The null allele of CYP2D6, a drug-metabolizing enzyme (DME) that encodes a member of cytochrome P450 family (**Weinshilboum, 2003**) renders the gene product inactive to an extent that homozygous null allele individuals experience little or no analgesic effect. This null allele occurs in a frequency of 10% among north European ancestry and therefore they do not experience an analgesic effect from the prodrug codeine (**Bradford, 2002**). On the contrary, about 98% Arabs are able to transform codeine into the active form morphine (**Mclellan ,1997**).

(iii) Incorporation of population genetic structuring in study designs of association studies. Undetected population stratification in case-control studies could lead to false positive associations (**Pritchard *et al*, 1999**) therefore, genetic knowledge of population sub-structuring and stratification is an essential requirement for proper selection of controls and for identifying disease pre-disposing alleles that may differ across ethnic groups.

Overall, knowledge of the genetic ancestry or origin of population sub-group, and information of population diversity, sub-structuring, stratification and phylogenetic relationship is a key in biomedical research and worldwide efforts are going on to replace the proxy tags of geography, ethnicity, race or caste by an accurate genetic profile. However, the major hurdle is in the reliability of genetic markers to infer the correct genetic origin of a sub population and the level till which a marker can resolve the genetic-sub-structuring i.e. it can differentiate between an Indian and an African or it can even resolve a north Indian from a south Indian.

Search of such genetic tags require information about the genetic ancestry and presence of genetic structuring in contemporary populations. More knowledge about of various evolutionary and socio-cultural factors that have shaped the present day genetic diversity is required, more data on admixed population is desired and more identification of population specific private marker, private alleles or haplotypes are required.

However, more vision, planning and better approach are required to conduct such studies in Indian sub-continent due to its complex structure as illustrated above. Each study should be able to raise a specific question and selection of populations and markers should comply with the need of that particular study, as there are several aspect of Indian gene pool which has albeit got detected but not completely understood like:

(i) More defined studies on the Tribal and Dravidian populations based on markers additional to the one used in Caucasian populations are required to reveal the exact composition of Indian gene pool specially to know about the pre-Indo Aryan admixture.

(ii) Well defined geographical and cultural groups should be chosen to infer their role on Indian gene pool like north east population are best suited for studying East Asian gene flow while north Indian populations are most suitable for studying the effect of Indo Aryans and Muslims migrations on Indian gene pool.

(iii) Similarly, studies on more endogamous groups can reveal the structuring of castes in genetic context. Moreover as described above caste groups of India observe strict endogamy hence it is important to analyze the effect of endogamy on genetic structuring of Indian populations.

Over all, these studies will define the pattern and distribution of genetic variation in Indian population and will aid in assessing the level of genetic sub-structuring and correct genetic ancestry in different endogamous and tribal groups. Furthermore, such studies along with offering an unsullied elucidation on human genetic diversity of Indian population will also help in tracing the missing block of ancestral human settlers that will form the connecting link of standard model of human evolution.

The present study is an attempt to study the six North Indian populations namely *Kayastha, Rastogies, Mathurs, Vaish, Shia and Sunni* of Uttar Pradesh and 3 populations of North Eastern region of India namely *Lachung*, *Mech*, and *Rajbanshi* at genetic level. The genetic marker selected for the present study is highly polymorphic class two antigens which will provide an opportunity to search for an explanation of numerous unanswered enigma concerning the effect of the stringent social fabric on the genetic makeup of North Indian populations and completely isolated populations of North East India. We have made an attempt to compare these populations among themselves and also with other Indian and world populations. The main objective of this study was to understand the effect of social and geographical barriers on Indian population sub structuring.

Review of Literature

Review of Literature

In 1986 Human genome Project was launched by Charles DeLisi . Goal of this initiative was to understand the human genome. He said that "Knowledge of the human genome is as necessary to the continuing progress of medicine and other health sciences as knowledge of human anatomy has been for the present state of medicine."

Human genome project was completed on 14 April 2003, In October 2004, researchers from the International Human Genome Sequencing Consortium (IHGSH) of the HGP announced a new estimate of 20,000 to 25,000 genes in the human genome **(Fiers 2004).** Previously 30,000 to 40,000 genes had been predicted, while at the start of project the estimated Number of gene was approx 2,000,000. The goals of the original HGP were not only to determine all 3 billion base pairs in the human genome, but also to identify all the genes in this vast amount of data.

It is anticipated that detailed knowledge of the human genome will provide new avenues for advances in medicine and biotechnology. Like genetic tests that can show predisposition to a variety of illnesses, including breast cancer, disorders of hemostasis, cystic fibrosis, liver diseases and many others. Also, the etiologies for cancers, Alzheimer's disease and other areas of clinical interest are considered likely to benefit from genome information and possibly may lead in the long term to significant advances in their management **(Farrer *et al.*, 1997; Barnhart & Benjamin 1989).** The Human Genome Diversity Project, also aims at mapping the DNA that varies between human ethnic groups. In the future, HGDP could possibly expose new data in disease surveillance, human development and anthropology. HGDP could unlock secrets behind and create new strategies for managing the susceptibility of ethnic groups to certain diseases. It could also show how human populations have adapted to these Susceptibilities **(Barnhart & Benjamin 1989).** The analysis of similarities between DNA sequences from different organisms is also opening new avenues in the study of the theory of evolution. Similarities and

differences between humans and our closest relatives (the primates, and the other mammals) are expected to be illuminated by the genetic variation studies DNA from all organisms is made up of the same chemical and physical components. The DNA sequence is the particular side-by-side arrangement of bases along the DNA strand, like ATTCCGGA. The genome is an organism's complete set of DNA. Genomes vary widely in size: the smallest known genome for a free-living organism (a bacterium) contains about 600,000 DNA base pairs, while human and mouse genomes have some 3 billion. Except for mature red blood cells, all human cells contain a complete genome (**Figure 2.1**).

DNA in the human genome is arranged into 24 distinct chromosomes--physically separate molecules that range in length from about 50 million to 250 million base pairs.

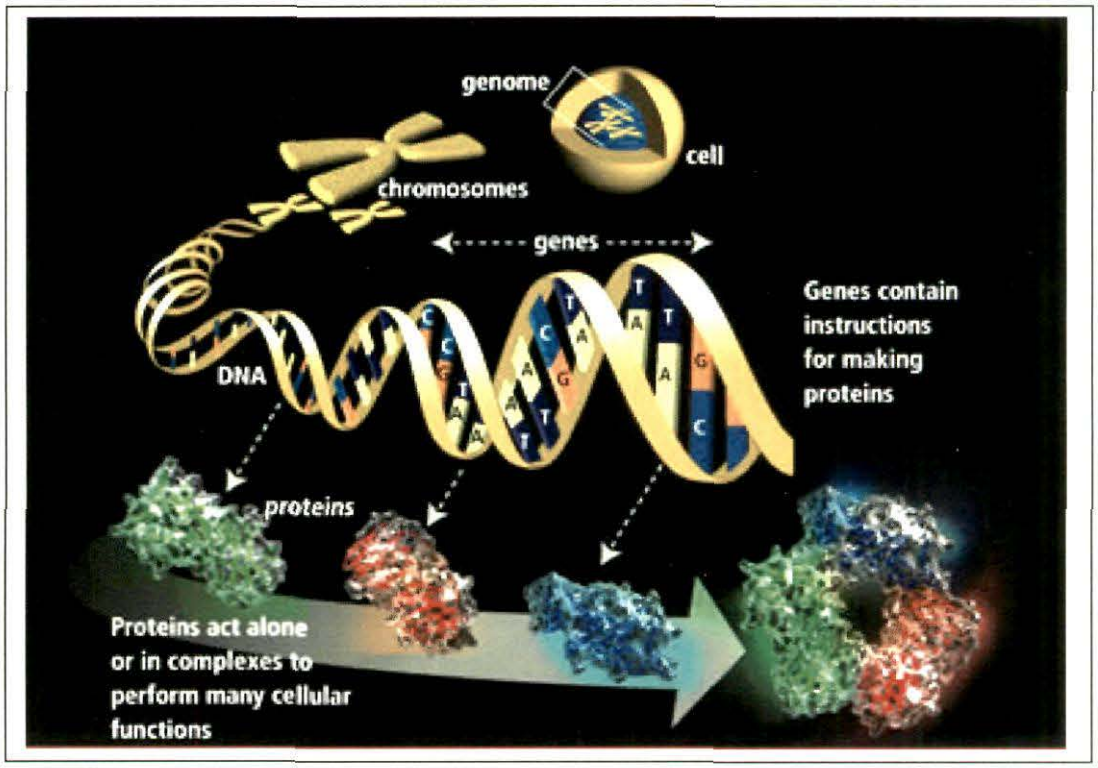


Figure 2.1 Genome organization

Each chromosome contains many genes, the basic physical and functional units of heredity. These genes are encoded in segments of DNA and the information is contained in 23 pairs of chromosomes. DNA transmits its information through four

different nucleotide bases, which are the "letters" of the genetic code A, C, G, and T. The human genome contains approx 3164.7 million chemical nucleotide bases. The average gene consists of 3000 bases, but sizes vary greatly, with the largest known human gene being dystrophin at 2.4 million bases. Almost all (99.9%) nucleotide bases are exactly the same in all people. These bases, when read as strings of three letter word, form the blue prints of molecules responsible for all the functions of the body. But the coding region of these genes takes up only 2-3% of the human genome .The remaining consists of noncoding regions, whose functions may include providing chromosomal structural integrity and regulating where, when, and in what quantity proteins are made. Although genes get a lot of attention, it's the proteins that perform most life functions and even make up the majority of cellular structures. Proteins are large, complex molecules made up of smaller subunits called amino acids. Unlike the relatively unchanging genome, the dynamic proteome changes from minute to minute in response to tens of thousands of intra- and extra cellular environmental signals. A protein's chemistry and behaviour are specified by the gene sequence and by the number and identities of other proteins made in the same cell at the same time and with which it associated.

Human genome varies from individual to individual and therefore no two individuals look alike (**Fischer *et al.*, 2004**). Historically, individual variation was studied on the basis of conventional markers. However, with the advancement of technology various genetic markers were discovered and the gene frequency data for studying the evolution of human races was analyzed using these markers. Initially, the classical serological and biochemical markers have played important roles in various types of human population genetic studies. However, it is important to record population variation because it is helpful to know the various mechanisms involved in causing variation and it further enhances our knowledge about the molecular basis of disease susceptibility. Basic information about the types, frequencies and distribution of common variants are essential not only for the understanding of pathological entities, but also to unravel our evolutionary past and provide guidance about our biological future (**Jorde & Wooding, 2004**).

To study the population variation various genetic markers are being used. Some of the important markers involved for studying ethnic variability are conventional markers like blood groups, protein polymorphisms, restriction fragment length polymorphisms, short tandem repeat polymorphisms, variable tandem repeat polymorphisms, single nucleotide polymorphisms and human leukocyte (HLA) polymorphisms. In **Table 2.1** the characteristic features of various markers is shown

Table 2.1 Genetic markers used to study population variations

Type of marker	Year	No. of loci	Features
Blood groups	1910-1960	~20	May need fresh blood, rare antisera. Genotype cannot always be inferred from phenotype because of dominance. No easy physical localization.
Electrophoretic mobility variants of serum proteins	1960-1975	~30	May need fresh serum, specialized assays, no easy physical localization often limited polymorphisms
Human Leukocyte Antigens (HLA)	1970	1 (multi locus haplotype)	One linked set highly informative. Can only test for linkage to 6p21.3
DNA RFLPs	1975	$>10^5$ (potentially)	Two allele markers, maximum heterozygosity 0.5, initially required Southern blotting, now PCR. Easy physical localization
DNA VNTRs (minisatellites)	1985-	>104 (potentially)	Many alleles, highly informative can be typed by southern blotting easy physical localization. Tend to cluster near ends of chromosomes.
DNA VNTRs (microsatellites) (di-,tri-, and tetranucleotide repeats)	1989-	105 (potentially)	Many alleles, highly informative Can be typed by automated multiplex PCR, easy physical localization. Distributed throughout genome
DNA SNPs	1998-	106 (potentially)	Less informative than microsatellites. Can be typed on a very large scale by automated equipment, without gel electrophoresis, etc.

In present study we have taken into consideration the HLA class II antigens to decipher the social structure of 6 North Indian populations and 3 North Eastern caste and tribal populations. In the following section we have reviewed the human leukocyte antigens in detail.

Human Leukocyte Antigens (HLA)

The major histocompatibility complex (MHC) is a dense complex of genes with immunological and non-immunological functions and is present in all vertebrates. In humans it is known as human leukocyte antigens (HLA) **(Marsh, 2000; Trowsdale, 1995; Gruen & Weissman, 1997)**. Peter Gorer discovered it during transplantation studies in mice (the H-2 complex) in 1937. Jean Dausset described the first human MHC antigen *Mac* (HLA-A2) followed by the discovery of 4a and 4b. MHC is best known with its role in histocompatibility **(Snell, 1981)** and in immune regulation **(Jorde *et al.*, 1999; Benacerraf & Devitt, 1972)**. The main function of the MHC molecules is peptide binding and presentation of these peptides to T lymphocytes. Among the non-immune functions, the noteworthy ones are interactions with receptors on the cell surface **(Jorde *et al.*, 1999; Edidin, 1988; Svejgaard & Ryder, 1976)**, in particular with transferring receptor (TfR), epidermal growth factor **(Schreiber *et al.*, 1984)** and various other hormone receptors and signal transduction **(Abbas *et al* 1991; Schafer *et al.* 1995)**.

In humans, the genes for the HLA antigens are located on the short arm of chromosome 6 in the band 6p 21.3 **(Trachtenberg 1996; Apanius 1997)**. It contains approx 4 million nucleotide pairs and contains approx 200 genes The MHC complex is divided into three subgroups called MHC class I, MHC class II, and MHC class III. Among the genes within the MHC are more than 20 loci encoding proteins involved in binding and presentation of the peptide degradation products of proteins to the T cell antigen receptor **(Figure 2.2a & b)**.

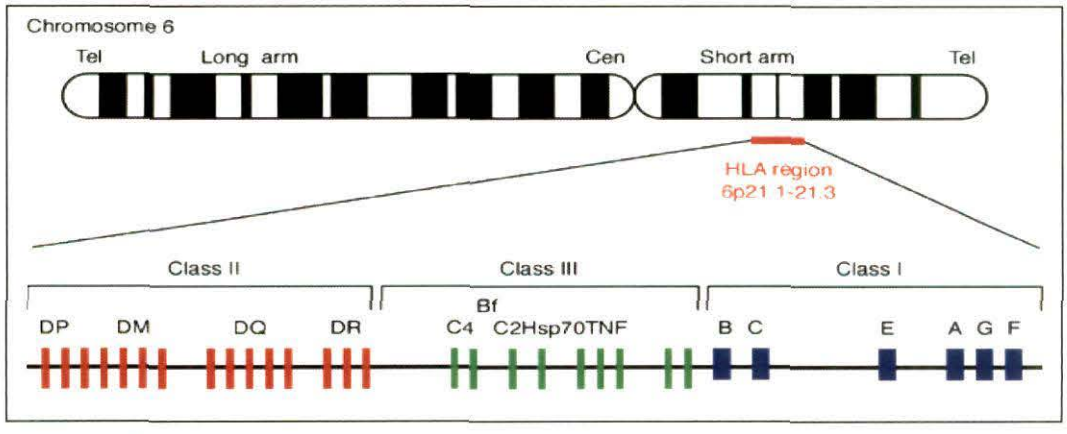
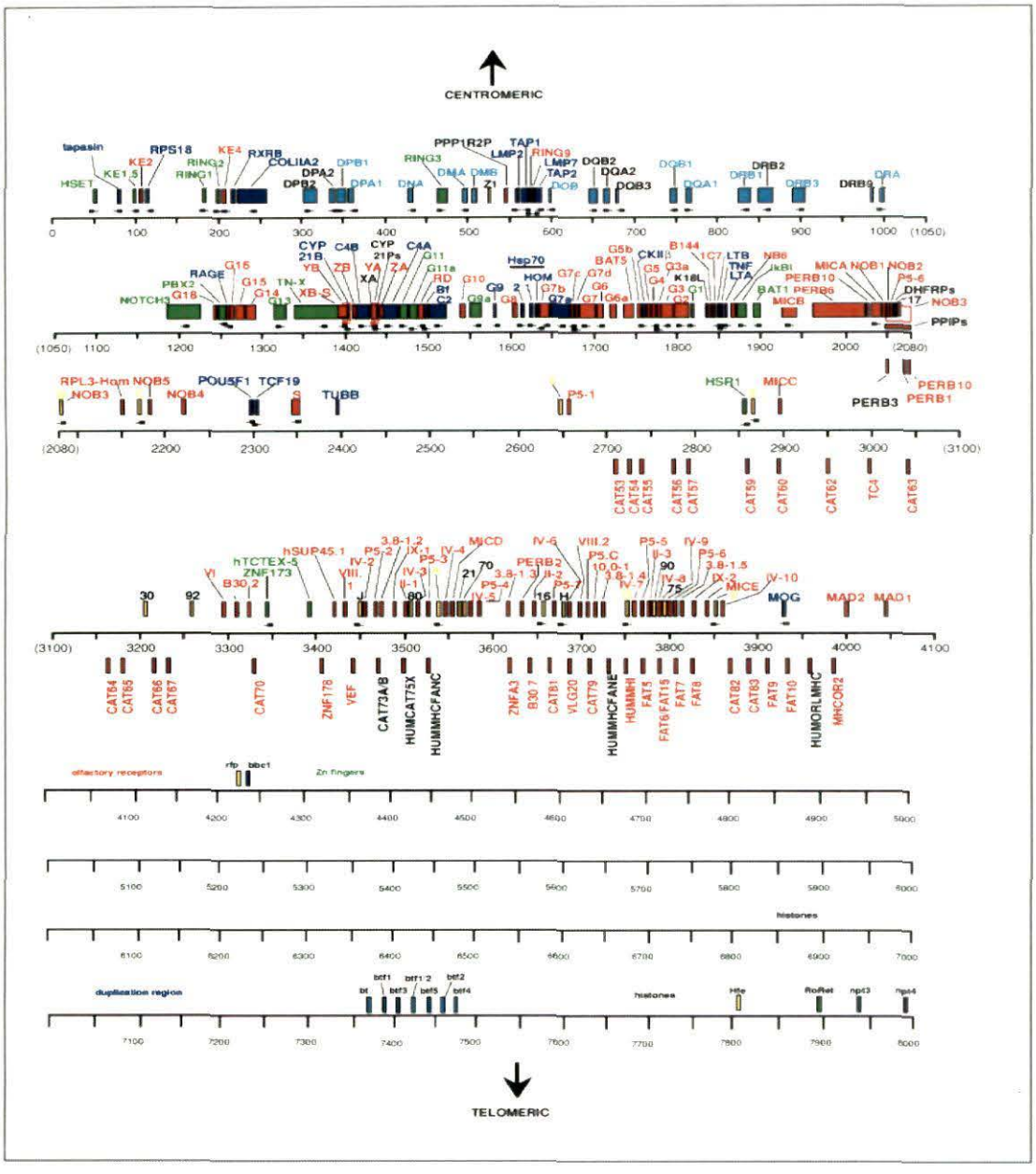


Figure 2.2a Gene map of the human leukocyte antigen (HLA) region



LEGEND

- Class II sequences
- Class I sequences
- Genes of known function
- Genes of putative function
- Genes of unknown function

Figure 2.2b Gene map of the human leukocyte antigen (HLA) region

There is not a definite candidate for the primordial MHC gene. According to one hypothesis the class II MHC evolved first **(Hughes 1993)** whereas another hypothesis holds that the class I MHC originated first as a result of a recombination between an immunoglobulin-like C-domain and the peptide-binding domain of an HSP70 heat-shock protein **(Flajnik,1991)**

A phylogenetic analysis supports a relationship between the class II MHC alpha chain and beta 2-microglobulin and between the class II MHC beta-chain and the class I alpha chain **(Hughes, 1992)**. Most evidence supports the hypothesis that the ancestral MHC molecule had a class II-like structure and it gave rise to the class I molecule **(Hughes,1992; Lawlor 1990; Klein,1990)**

The MHC class I encodes heterodimeric peptide binding proteins, as well as antigen processing molecules such as TAP and Tapasin. The MHC class II encodes heterodimeric peptide binding proteins with help of specialized chaperone invariant chain (Ii chain) The MHC class III region encodes for other immune components, such as complement components (e.g., C2, C4, factor B) and some that encode cytokines (e.g., TNF-alpha). Class I antigens are expressed on all somatic nucleated cells, where as class II is expressed on restricted population of cells, which are usually capable of antigen presentation. The complement system and other soluble molecules constitute the class III component of the MHC molecule (**Figure 2.3**).

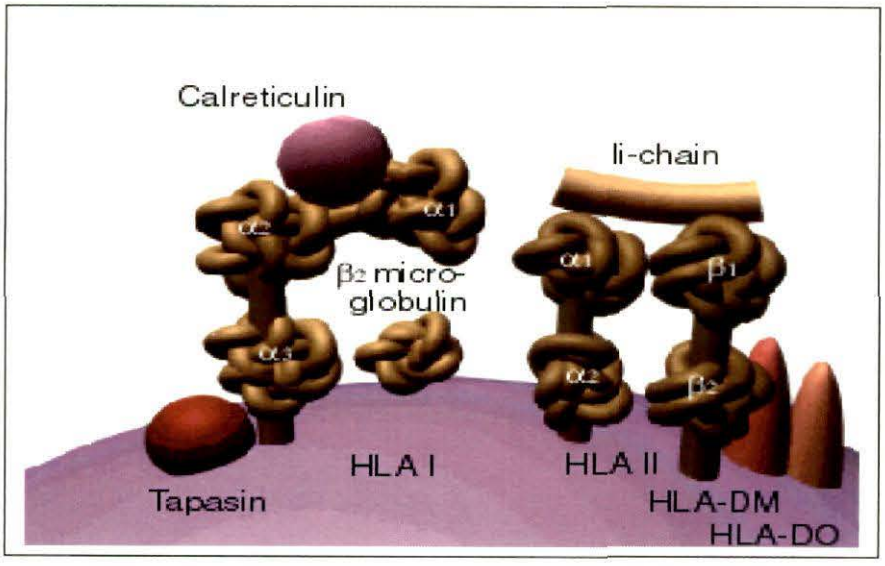


Figure 2.3 Molecular structure of HLA class I and class II antigens

The MHC proteins act as signposts that display fragmented pieces of an antigen on the host cell's surface (**Snell,1981**). They may be self or nonself. If they are nonself, there are two ways by which the host cell may acquire this antigen. If the host is a macrophage or microphage, such as a monocyte or neutrophil, it may engulf the particle (bacterial, viral, or particulate matter), break it apart using lysozymes, and display the fragments on class II MHC molecules. On other hand, if a host cell is infected by a bacteria or virus, or is cancerous, it may display the antigens on its surface with a class I MHC molecule (**Figure 2.4**).

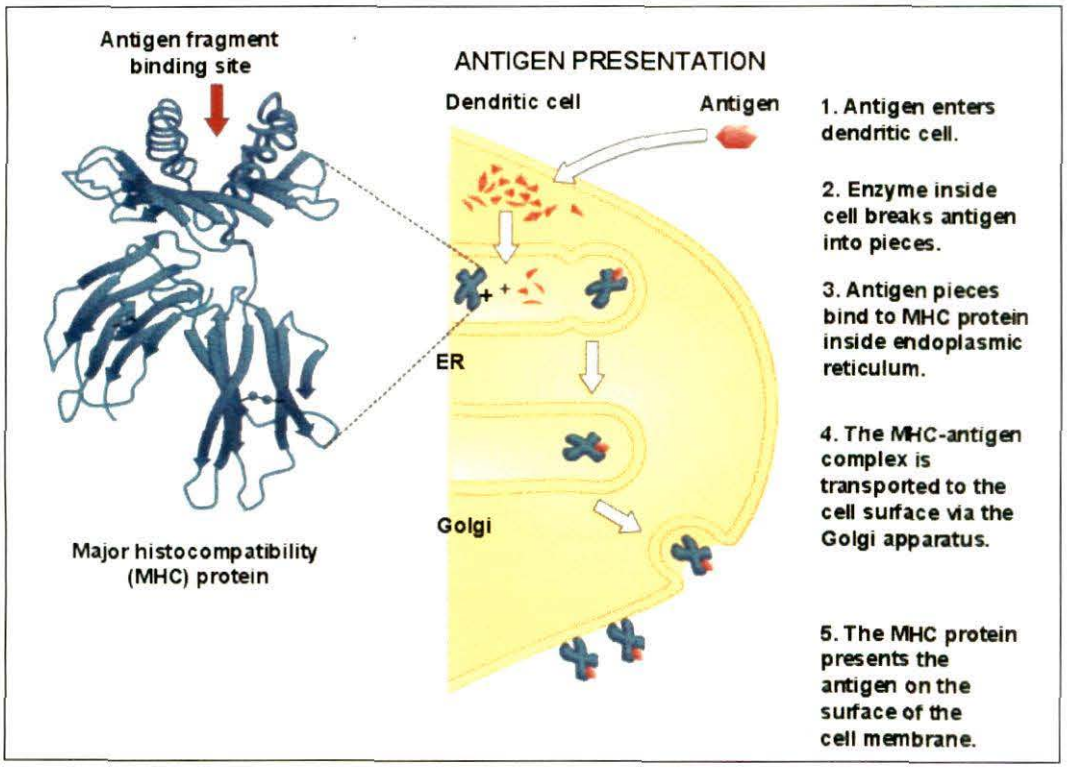


Figure 2.4 Mechanism of MHC processing

Viruses and cancerous cells have a tendency to display unusual, nonself antigens on their surface. These nonself antigens, regardless of which type of MHC molecule they are displayed on, will initiate the specific immunity of the host's body.

The best-known genes in the MHC region are the subset that encodes cell-surface antigen-presenting proteins. In humans, these genes are referred to as human leukocyte antigen (HLA) genes, although people often use the term MHC to refer to

HLA gene products) **(Trowsdale 1995; Gruen & Weissman,1997).** In mouse it is H-2 (Histocompatibility System – 2), in rabbits it is RLA (Rabbit Leukocyte antigens), Guinea Pig it is GLA (Guinea Pig leukocyte antigens) in Chimpanzee ChLA (Chimpanzee Leukocyte Antigen) in Cattle BoLA (Bovine Leukocyte Antigens).

In humans the most intensely studied HLA genes are the nine so-called classical MHC genes: HLA-A, HLA-B, HLA-C, HLA-DPA1, HLA-DPB1, HLA-DQA1, HLA-DQB1, HLA-DRA, and HLA-DRB1. The A, B, and C genes belong to MHC class I, whereas the six D genes belong to class II.

The MHC molecules have a vital role in the complex immunological dialog that must occur between T cells and other cells of the body **(Benacerraf & Devitt, 1972).** At maturity, MHC molecules are anchored in the cell membrane, where they display short polypeptides to T cells, via the T cell receptors (TCRs) **(Janeway, 1997).** The polypeptides may be self, that is, originating from a protein created by the organism itself, or they may be foreign, originating from bacteria, viruses, pollen, etc.

The immune system has another and equally important method to identify antigen: B cells with their membrane-bound antibodies, also known as B cell receptors (BCRs). However, whereas the BCRs of B cells can bind to antigens without much outside help, the TCRs of T cells require "presentation" of the antigen. It is important to realize that, during the vast majority of the time, MHC is kept busy presenting self-peptides, which the T cells should appropriately ignore. A full-force immune response usually requires the activation of B cells via BCRs *and* T cells via the MHC-TCR interaction. All MHC molecules receive polypeptides from inside the cells they are part of and display them on the cell's exterior surface for recognition by T cells. However, there are major differences between MHC class I and II in the method and outcome of peptide presentation.

Besides being scrutinized by immunologists for its pivotal role in the immune system, the MHC has also attracted the attention of many evolutionary biologists, due to the high levels of allelic diversity found within many of its genes

MHC Class I

MHC class I molecules are found on almost every nucleated cell of the body except central nervous system, skeletal and smooth muscle cells, parathyroid cells, pancreatic cells and corneal epithelium. Both male and female germinal cells are also devoid of classical class I antigens. The placenta and extra villous membranes express non-classical class I antigens **(Parham 1996)**. MHC class I molecules are heterodimers, consisting of a single transmembrane polypeptide chain (the α-chain) which is about 44 kiloDaltons (kD) and a β_2 microglobulin is 12 KD protein alpha chain is non-covalently linked with Beta chain **(Figure 2.5)**. The heavy α chain comprises of domains, α1, α2 and α3. β2 micro-globulin is also found as a soluble protein in serum, transmembrane region and a cytoplasm. Close to plasma membrane is located α3 domains.

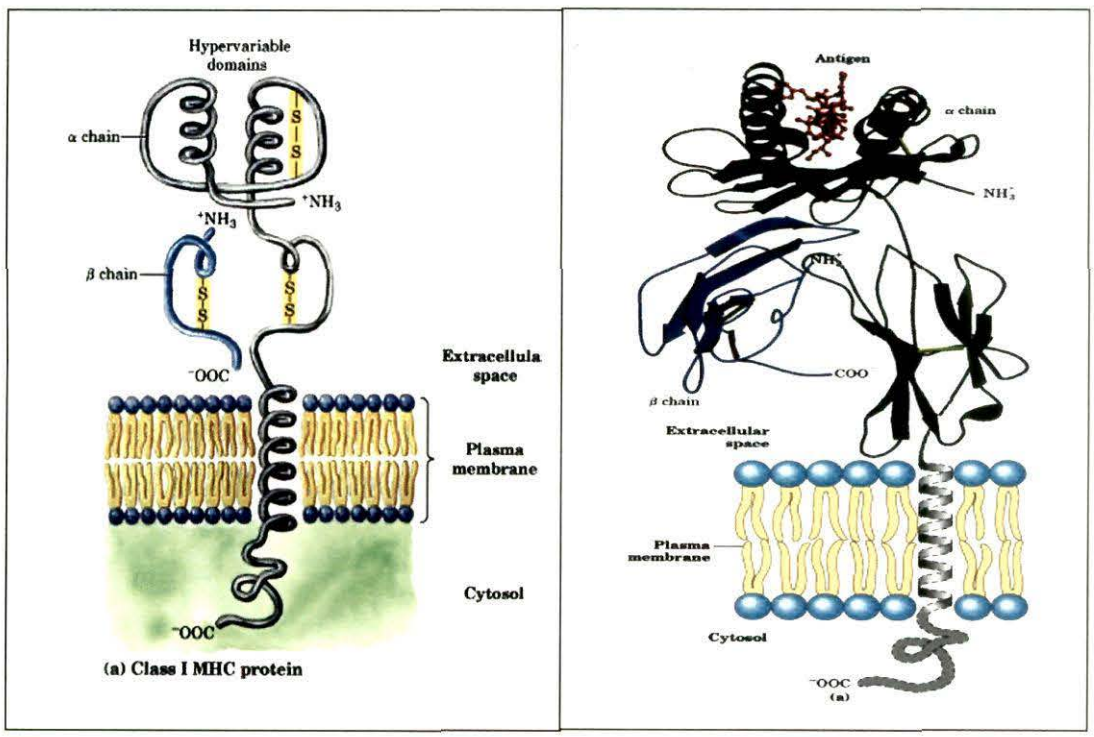


Figure 2.5 Structure of MHC class I antigens

In all the three domains 90 amino acids are present seperately, α2 and α3 domains are linked with inter chain disulphide bonds. In α3 domains at position 86 A glycosylated asparagine residue is situated. In transmembrane region 23 hydrophobic amino acid residues are present due to which α-helical conformation is

achieved. To the C terminal of the membrane arginine and lysine forms a cluster. These then get linked with the polypeptide chain in the membrane by interacting with negatively charged phospholipid groups of the inner membrane. The hydrophobic cytoplasmic domain consists of 30 amino acids; of which 50 are polar amino acids particularly serine and some are phosphorylated by a cAMP dependent protein kinase.

Mechanism of MHC Class 1 Processing

MHC proteins must bind peptide, and Class I must be complexed with β_2-microglobulin in intracellular compartments before MHC can be expressed on the cell surface. Although many MHC alleles have been identified in the human population, each individual has a limited number of MHC proteins with which to present a great many pathogen epitopes to T cells. Peptide binding to MHC is less specific than epitope binding to Ig or TCR; each MHC presents many different epitopes. Peptide must bind MHC with enough affinity to be retained on the plasma membrane and not exchange with soluble peptide. MHC molecules are unstable in the absence of bound peptide and are folded around peptide before transport to the plasma membrane (**Petrovsky, 2004; Terhorst ,1976**).

A virus-infected cell synthesizes virus proteins on ribosomes in its cytoplasm. In order to be presented, these proteins must be broken down into short peptides and transported into the endoplasmic reticulum (ER) to bind to newly synthesized Class I MHC proteins. In the cytosolic processing pathway, cytosolic proteins are degraded to peptides in proteasomes, cylindrical arrays of proteolytic enzymes with their active sites towards the center of the cylinder. Both pathogen proteins and self cell proteins can be complexed with ubiquitin to target them to the proteasome for processing. Two proteases encoded in the MHC Class II region (LMP2 and LMP7) and a third subunit not encoded in MHC is produced in response to interferon, which is synthesized in response to virus infection. These inducible proteases replace constitutive proteases in the proteasome and produce peptides with basic and hydrophobic carboxyl terminal residues preferred as anchor residues in Class I peptide binding sites and for transport from the cytosol into the ER. Two polypeptide

chains. Transporters of Antigen Peptide TAP-1 and TAP-2, are present in the ER membrane with ATP-binding domains on the cytosolic side and hydrophobic transmembrane domains spanning the ER membrane. The TAP-1/TAP-2 complex transports cytosolic peptides into the lumen of the ER with the expenditure of ATP. Both TAP molecules are required for membrane expression of Class I. Newly synthesized, partly folded Class I MHC α chain binds the chaperone calnexin in the lumen of the ER. Chaperone binding prevents mis-folding that otherwise would occur in the absence of antigen peptide. When β_2-microglobulin binds to Class I α chain, calnexin dissociates and Class I α plus β_2-microglobulin form a complex with calreticulin, tapasin, and TAP transporter. When Class I MHC binds peptide, it is released from TAP transporter and the Class I MHC-peptide complex is transported through the Golgi to the plasma membrane. Unbound peptides are thought to be transported back into the cytoplasm for reprocessing and retransport. In uninfected cells, membrane Class I MHC presents self peptides. Viruses which can interfere with TAP function or with transport of Class I MHC to the plasma membrane can evade destruction by cytotoxic T cells (**Petrovsky, 2004; Neumann, 1997**) **(Figure 2.6)**.

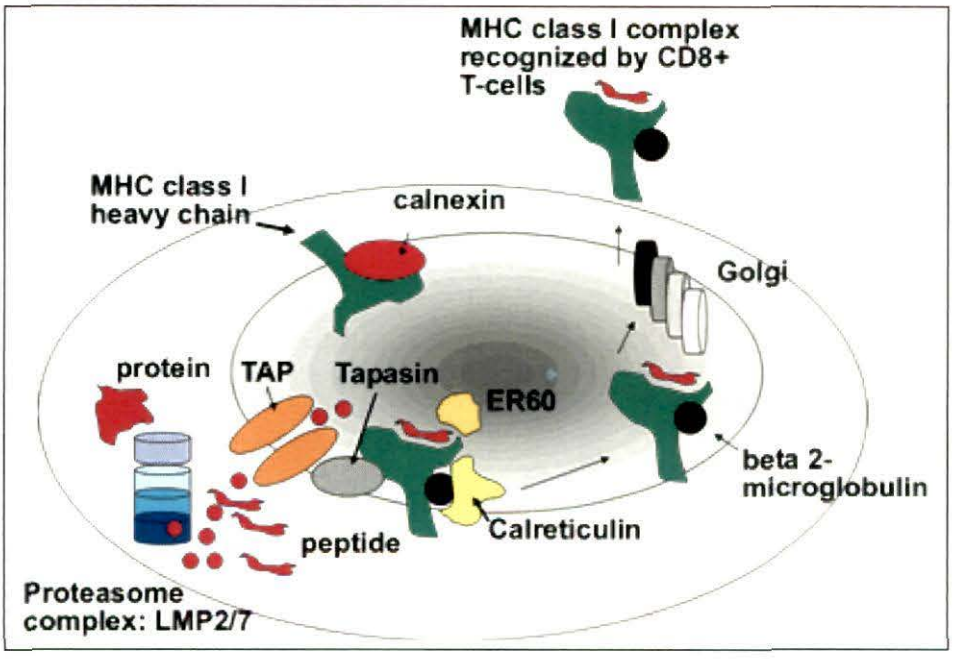


Figure 2.6 Mechanism of MHC Class 1 Processing

Exogenous antigen is processed in the endosomal processing pathway. Bacteria, soluble protein antigens, and antibody-coated viruses which have been taken up by macrophages and B cells, envelope proteins from the plasma membrane of DC, and some bacteria and parasites that live in endosomal vesicles, enter the endosomal processing pathway. The endosomes become increasingly acidic as they move from the plasma membrane farther into the cytoplasm. Increased acidity activates proteases that cut the antigen into peptides.

MHC Class I Gene Expression

Various biological or chemical modulators can modulate MHC gene expression. The promoter region of class I genes can be activated through several pathways. Hence, these genes are not typical domestic genes. Extensive studies on regulation of MHC class I expression, using transfection techniques and transgenic animal models, have resulted in identification of various cis-acting sequences involved in positive and negative regulation of class I genes. Work is in progress to identify the transacting proteins that bind to these sites and to delineate the mechanisms that regulate constitutive and inducible expression of class I genes in normal and diseased cells. Various biological molecules (IFN, GM-CSF, IL-2) and other chemicals up-regulate the MHC expression. If the exact mechanisms are known by which the expression of class I genes is up regulated, the efforts can be made to balance the beneficial and toxic effects of biological molecules with one another, which may facilitate the use of combination of these molecules in subpharmacological doses (to eliminate toxicity) for early and better management of neoplastic diseases, as it is well-known that during malignancy MHC gene expression is down-regulated **(Agrawal & Kishore,2000; Neumann,1997).**

MHC Class II

MHC Class II molecules are found only on a few specialized cell types, including macrophages, dendritic cells, activated T cells, and B cells, all of which are professional antigen-presenting cells (APCs). Like MHC class I molecules, class II molecules are also heterodimers, but in this case consist of two homologous

peptides, an α and β chain. α chain is of 34,000 Daltons and β- chain is of 28000 dalton. These are independently synthesized in the endoplasmic reticulum and become associated with a third type of polypeptide chain, the invariant chain (Ii chain) **(Ceman 1995; Schafer 1995)**. The α-chains are encoded by DRα, DQα, DPα etc. genes whereas β-chains are encoded by DRβ, DQβ, DPβ etc. genes. Each DR, DQ or DP sub-region consists of at least one expressed α and one expressed β gene. Both α and β genes can be polymorphic but most of the polymorphism resides in β genes. All the class II genes lie within the MHC complex **(Parham and Ohta, 1996; Terhorst,1976; Zhong, 1997)**.

The α and β chains consist of 2 extra cellular domains of approximately 90 amino acids each. In the human DR region there is more than one functional β chain gene; the number of genes varies with different class II haplotypes. All these products can associate with the α chain to form class II cell surface molecules, this providing a mechanism for increasing the polymorphism of the DR molecule **(Figure 2.7)**.

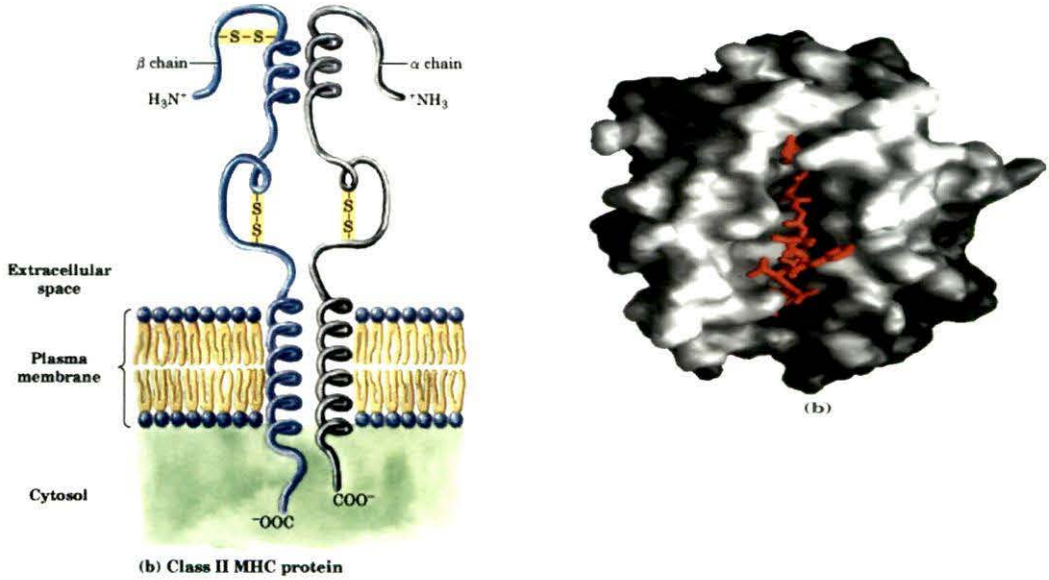


Figure 2.7 Structure of MHC class II antigens

Mechanism of MHC Class II processing

Class II MHC α and β chains are synthesized on the rough endoplasmic reticulum (ER) and transported into the ER lumen, where they assemble with another polypeptide, invariant chain (Ii). Ii and Class II MHC form nine-chain trimeric complexes, with part of each Ii molecule occupying the Class II MHC peptide-binding site. Ii allows Class II MHC to assemble in the absence of foreign peptide and blocks association with normal cell (self) and endogenous peptides present in the ER lumen. Ii also directs the transfer of Class II MHC through the Golgi to a specialized vesicular compartment **(Ceman, 1995)**. MHC (MHC Class II Compartment), where over several hours Ii is degraded by cathepsins and proteases in the low pH. Occupying the peptide –binding groove called CLIP. HLA-DM, a Class II MHC αβ heterodimer resembling other Class II MHC molecules but not expressed on the cell surface, facilitates CLIP removal and peptide binding in the MIIC compartment **(Figure 2.8)**.

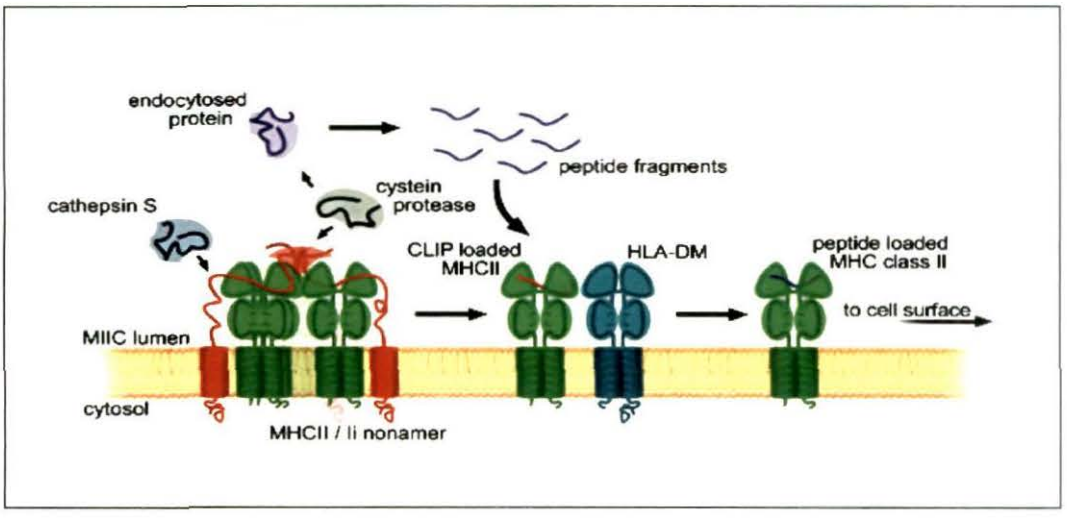


Figure 2.8 Mechanism of MHC Class II Processing

Class II MHC-peptide complex is then transported to the plasma membrane. Class II MHC which does not bind peptide when CLIP dissociates is unstable and is rapidly degraded. In the absence of infection, APC present Class II MHC containing self peptides, including self Class II MHC peptides. Peptide/Class II MHC

complexes are very stable, ensuring that the APC presents its own exogenous peptides and not peptides released from another cell. Like other plasma membrane proteins, Class II MHC proteins shuttle between the membrane and endosomal compartments where they can pick up new exogenous antigen peptide or be degraded (**Zhong,1997**).

MHC Class II Gene expression

The highly complex pattern of expression of major histocompatibility complex class II (MHC-II) molecules determines both the immune repertoire during development and subsequently the triggering and the control of immune responses. These distinct functions result from cell type-restricted expression, developmental control and either constitutive or inducible expression of MHC-II genes , in these various situations, MHC-II gene expression is always under the control of a unique transactivator, CIITA, CIITA gene is controlled by several distinct promoters, two of which direct specific constitutive expression in dendritic cells and B lymphocytes respectively, while another mediates γ interferon-induced expression. Thus the cellular, temporal and functional diversity of MHC-II expression is ultimately controlled by differential activation of different promoters of a single transactivator gene (**Sonntag 2000; Ruemmele,1999; Mehindate,1994**).

Nomenclature

Developing and maintaining a nomenclature for the numerous and rapidly accumulating Class I and II alleles is a challenging problem. The nomenclature for alleles is largely based on earlier serological names since the broad HLA antigen groups were originally defined based on their reaction with antisera in complement mediated microlymphocyotoxicity assays. Antisera were often isolated from women sensitized during pregnancy to the HLA specificities encoded by paternal haplotypes. Serological specificities are referred to as HLA-A1 through A80, HLA-B7 through B81; HLA-Cw1 through Cw10; HLA-DR1 through DR18, and HLA-DQ1 through DQ9 (**Table 2.2**) (**Bodmer,1998**).

Table 2.2 HLA loci and known alleles

Generic locus	Antigen or associated specificity	Number of known alleles
HLA-A	A1 to A80	489
HLA-B	B7 to B81	830
HLA-C	Cw1 to Cw10	266
HLA-E	-	9
HLA-F	-	20
HLA-G	-	23
DRA	DR1 to DR18	3
DRB1	-	463
DQA1	DQ1 to DQ9	34
DQB1	-	78
DPA1	DPw1 to DPw6	23
DPB1	-	125

Source: http://www.ebi.ac.uk/imgt/hla/

Once nucleic acid-based information regarding the sequence of alleles became available, a nomenclature complementary to the serological terms was devised (**Table 2.3**).

The first two digits of an allele name refer to the underlying serological specificity and the third and fourth digits indicate a specific allele sequence. For example, HLA-A*0205 and A*0210 are alleles encoding distinct polypeptides within the A2 serotype. These two alleles both encode the epitope recognized by the anti-A2 antisera but have 5 nucleotide differences elsewhere in exons 2–3 resulting in amino acid variations. Alleles within a serological group may vary from each other by a single or by several nucleotides. For Class II molecules, both the A and the B genes may contribute to antigen variability.

Table 2.3 Nomenclature of HLA alleles

HLA Allele Nomenclature	Examples of alleles Comment
A*24 and A*2404	A*24 refers to any of 33 known alleles with closely related sequences encoding Class I antigens which usually react with A24 anti-sera. A*2404 is a specific allele within this group.
DRB1*0801 and *0805	Differ in exon 2 at codon 74: At this position *0801 has CTG and *0805 has GCG encoding Leu and Ala, respectively
A*01011 and *01012	Differ by a silent polymorphism at codon 142: At this position *01011 has ATC and *01012 has ATT. Both encode Ile
B*1501101 and *1501102N	The B*1501102N (null) allele has a 10-bp deletion near the 3'end of intron 1. The mRNA is improperly spliced with a predicted truncated translated polypeptide.

Thus a DR15 serotype may be found in an individual with one of the DRB1*15 alleles such as DRB1*1501 and DRA*0101. When necessary, a fifth digit is used to identify silent polymorphisms and the sixth and seventh digits are used to denote variation occurring outside of coding regions, such as the promoter and introns. Null allele sequences which result in either no or reduced levels of functional HLA molecules because of transcription changes, aberrant RNA splicing, and frame shift and nonsense mutations, or in frame termination codons are designated by an allele number appropriate to the group and the letter N. Individuals with null alleles may have discrepancies between serological and DNA-based typing. This problem has clinical implications since an inappropriate donor may be sought if more than one laboratory using different techniques are involved in typing a recipient and potential donors.

The origin of diversity of MHC alleles

The major histocompatibility complex (MHC) loci are known to be highly polymorphic in humans, mice and certain other mammals, with heterozygosity as high as 80-90%. Four different hypotheses have been considered to explain this high degree of polymorphism:

(1) A high mutation rate

(2) Gene conversion or inter locus genetic exchange

(3) Over dominant (balancing) selection

(4) Frequency-dependent selection.

The distribution of the pattern of sequence polymorphism in human and mouse class I genes provides evidence for four co-ordinate factors that contribute to the origin and sustenance of abundant allele diversity that characterizes the MHC in the species. These include: (a) a gradual accumulation of spontaneous mutational substitution over evolutionary time but not an unusually high mutation rate; (b) selection against mutational divergence in regions of the class I molecule involved in T cell receptor interaction and also in certain regions that interact with common features of antigens; (c) positive selection pressure in favor of persistence of polymorphism and heterozygosity at the antigen recognition site; and (d) periodic intragenic (interallelic) and more rarely, intergenic, recombination within the class I genes.

Evolutionary interplay between mutation and recombination varies with MHC locus, and even for subregions of the same gene **(Parham & Ohta 1996; Hughes, 1988; Hughes, 1989)**. For example, phylogenetic inferences based on the exon 2 region of HLA-DRB loci are complicated by selection and recombination (gene conversion). Noncoding region analysis may help clarify patterns of allele evolution usually with contrasting results to those obtained from coding region analyzes **(Hickson, 1997)**. The main source for the variability in the HLA gene sequences is point mutation but the mutation rate is by no means higher in the MHC than elsewhere in the genome **(Lawlor, 1988, Parham ,1995)**. Because of trans species polymorphism, accumulation of point mutations over millions of years results in extensive polymorphism. In contrast, gene conversions have produced at

least 80 new class I alleles since the separation of the Homo lineage and the rate of conversion is much higher than that of point mutation **(Little 1999; Marsh 2000)** **(Figure 2.9)**.

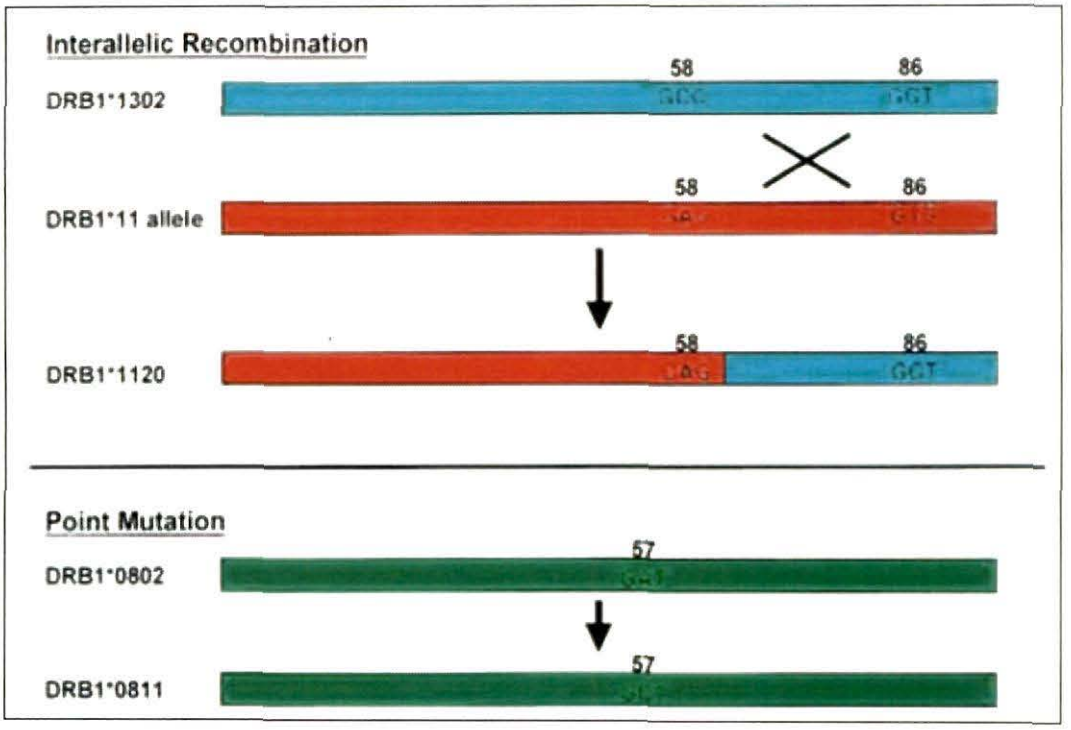


Figure 2.9 Mechanism of generation of new alleles

Alleles arise from existing alleles through several postulated mechanisms **(Marsh & Parham,2000).** DRB1*1120 likely arose via interallelic recombination between DRB1*1302 and a *DRB1**11 allele **(Cizman & While,1996)** DRB1*0811 probably is derived from DRB1*0802 by a point mutation **(Williams,1994)**.

Nature of Class I and II Gene Polymorphism

Multiple alleles are found within most of the known serotypes, although a few serotypes (an example is DR9, *DRB1**09012) are accounted for solely by a single allele. For example the B35 serotype has more than 39 alleles. HLA allele frequencies exhibit ethnic variation, with some alleles found widely distributed among populations and others almost exclusively within a particular ethnic group. The number of different phenotypes that are possible from all combinations of the known HLA alleles is greater than the earth's population. However, the Class I and

II loci reside on a relatively small region of chromosome 6 and specific haplotypes were apparently present at high frequencies in founding populations or were selected for or against by infectious organisms. In this setting, linkage disequilibrium results in a significant overrepresentation of certain haplotypes **(Alper 1992) (Figure 2.10).**

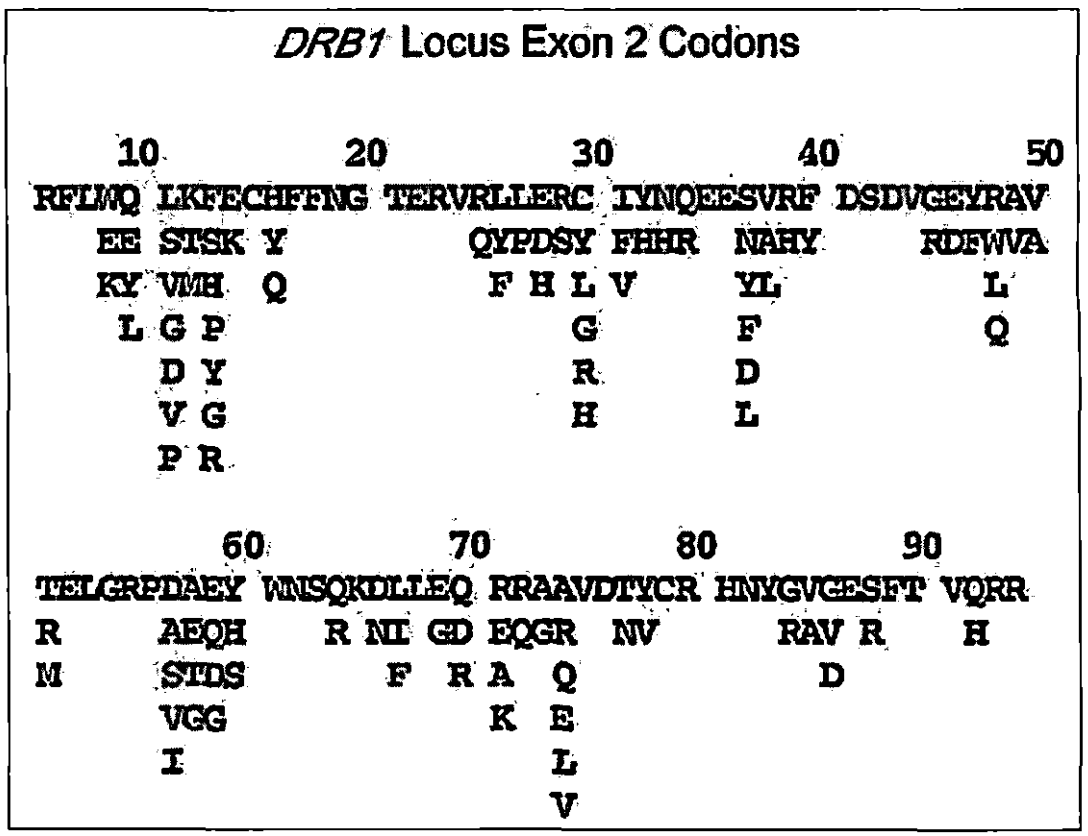


Figure 2.10 DRB1 Gene polymorphism

Single letter amino acid codes are shown for DRB1 exon 2 codons 6–94. About half of the positions are invariant while the remainder displays polymorphism with a few codons encoding as many as seven different amino acids. For example, all *DRB1* alleles have glycine encoded by position 20 while alleles may encode glycine, valine, or aspartic acid at codon 86. *DRB1* alleles arise through the many possible combinations of these polymorphisms.

This fact makes it possible for organizations such as the National Marrow Donor Program to have a fairly high likelihood of finding a well matched donor for a Caucasian recipient with 3–4 million potential volunteer donors in its database **(Kernan,1993).**

Although HLA gene loci are closely linked, meiotic recombination may occur between loci. For example, the crossover rate is approximately 0.8% between the *A* and *B* loci and virtually zero between the *DRB1* and *DQB1* loci. The great majority of the polymorphism found in the Class I and II genes occurs in the exons that encode the α-1 and α-2 (Class I, exons 2–3) and the α-1 and β-1 (Class II, exon 2) domains which bind processed peptides **(Little,1999)**.

Some nucleotide positions in these exons are invariant, others may have two or three or even all four of the possible bases as possibilities. Thus, some codons are constant while others display varying degrees of variability. The polymorphic exons are relatively short in length (about 250 nucleotides), they can easily be amplified in the PCR for molecular diagnostic studies.

Mechanisms maintaining the extreme polymorphism of the MHC

1. Pathogen driven mechanism

Pathogen-driven selection favors genetic diversity of the MHC through both heterozygote advantage (over dominance) and frequency-dependent selection (Potts WK1993). Selection is thought to favor rare MHC genotypes, since pathogens are more likely to have developed mechanisms to evade the MHC-dependent immunity encoded by common MHC genotypes. Six molecular models of pathogen-driven selection have been presented **(Potts 1995)**.

A. Pathogen Evasion Models

Escape of a single T-cell clone recognition

Escape into holes in the T-cell repertoire produced by T-cells energized by pathogen variants

Escape into holes in the T-cell repertoire induced by self-tolerance

Escape of MHC presentation

B. Host-Pathogen Interactions:

Heterozygote advantage

Pathogens bearing allo-MHC antigens

MHC associations with specific infectious diseases have been difficult to demonstrate. The best known one is malaria in humans **(Hill, 1991)**. Since most

infectious agents have multiple epitopes which MHC has to deal **(Fienberg,1970)**, Rather than resistance of specific heterozygous genotypes to specific. In human heterozygote advantage have been reported for a specific genotype in HIV infection **(Carrington,1999)** and in HBV infection **(Thursz,1999; Thio,1999)**

2. Non-pathogen driven mechanisms

MHC is exploited to discriminate against genetic similarity at highly polymorphic loci to avoid inbreeding. MHC-based disassortative mating would produce heterozygous progeny at least at the MHC which would result in increased fitness **(Ober,1997).** Progeny derived from MHC-dissimilar parents would have high fitness because of reduced levels of inbreeding depression and increased resistance to infectious diseases due to high MHC heterozygosity. This selection contributes to the high levels of genetic polymorphism observed at the MHC loci.

Evidence for Selection on MHC alleles

1. One important feature of the MHC genes is that the ratio of non-synonymous (replacement) to synonymous (silent) substitutions (dn /ds ratio) is very high in the codons encoding the antigen recognition site of polymorphic class II molecules compared to other codons **(Hughes, 1994).** This pattern is evidence that the polymorphism at the antigen recognition sites is maintained by over dominant selection of which the most common form is heterozygote advantage. This kind of selection has been noted for all expressed DRB genes including DRB3 and DRB4 **(Klein, 1991).** This feature and the others such as (1) an extremely large number of alleles; (2) ancient allelic lineages that pre-date contemporary species (trans-species evolution) and; (3) extremely high sequence divergence of alleles make the MHC a unique system in the whole genome.
2. The expected number of alleles under neutrality is far lower than the number of MHC alleles observed in natural population which indicates that some form of balancing (diversifying) selection is also acting **(Hedrick, 1983, 1994, 1996).** For a subdivided population over a large range of migration

rates, it appears that the number of self-incompatibility alleles (or MHC-alleles) observed can provide a rough estimate of the total number of individuals in the population but it underestimates the neutral effective size of the subdivided population **(Schierup,1998, 2000, 2002).**

3. The large number of alleles showing a relatively even distribution is against neutrality expectations and indicates that diversifying, and not simply directional, selection operates in contemporary populations.
4. The observed deficiency of homozygotes in humans **(Black, 1997, 1981; Hedrick,1990)** indicates that selection favors heterozygotes, because of high immune response. When the amino acid heterozygosities per site for HLA-A and -B loci were determined, for the 54 amino acid sites thought to have functional importance, the average heterozygosity per site was 0.301. Sixteen positions have heterozygosities greater than 0.5 at one or both loci and the frequencies of amino acids at a given position are very even, resulting in nearly the maximum heterozygosity possible. High heterozygosity is concentrated in the peptide-binding sites, whereas the sites that interact with the T-cell receptor have lower heterozygosity. Overall, these results indicate the importance of some form of balancing selection operating at HLA loci, maybe even at the individual amino acid level **(Hedrick,1996).**
5. The observed linkage disequilibrium among tightly linked MHC genes suggests that the strength of selection is uneven within the MHC **(Apanius, 1997).**
6. Studies in West Africa showed that resistance against malaria is HLA-B53 associated and this is the reason for an increased frequency of B53 in that area. The selection differential for HLA-B*5301 is estimated to be 0.028 **(Hill 1991). Apanius *et al.*, (1997)** suggested that MHC haplotypes can confer resistance to infectious diseases that outweighs the deleterious effects from autoimmunity. Another hypothesis for the maintenance of autoimmune-predisposing MHC haplotypes is that these alleles protect against initial infection, but the pathogen triggers autoimmunity through molecular mimicry.

Allelic diversity and MHC evolution

MHC gene families are found in essentially all vertebrates, though the gene composition and genomic arrangement vary widely. Chickens, for instance, have one of the smallest known MHC regions (19 genes), though most mammals have an MHC structure and composition fairly similar to that of humans. Gene duplication is almost certainly responsible for much of the genic diversity. In humans, the MHC is littered with many pseudogenes.

One of the most striking features of the MHC, particularly in humans, is the astounding allelic diversity found therein, and especially among the nine classical genes. In humans, the most conspicuously-diverse loci, HLA-A, HLA-B, and HLA-DRB1, have 489, 830 and 463 known alleles respectively this allelic diversity is truly exceptional in the human genome. The MHC gene is the most polymorphic in the genome. One remarkable featureof HLA Loci is that many of these alleles are quite ancient. It is often the case that an allele from a particular HLA gene is more closely related to an allele found in chimpanzees than it is to another human allele from the same gene.

Phylogenetically the marsupial MHC lies between eutherian mammals and the minimal essential MHC of birds, although it is closer in organization to non-mammals. It's Class I genes have amplified within the Class II region, resulting in a unique Class I/II region.

The allelic diversity of MHC genes has created fertile grounds for evolutionary biologists (**Arnaiz-Villena *et al.*, 1999**). The most important task for theoreticians is to explain the evolutionary forces that have created and maintained such diversity. Most explanations invoke balancing selection, a broad term that identifies any kind of natural selection in which no single allele is absolutely most fit. Frequency-dependent selection and heterozygote advantage are two types of balancing selection that have been suggested to explain MHC allelic diversity.

Population studies

Molecular HLA-typing has proved to be an invaluable tool in studying the evolutionary origin of human populations (**Arguello *et al.*, 1998; Arnaiz-Villena *et***

***al.*, 1999; Luo *et al.*, 1999; Albis- Camps & Blasczyk 1999**). This information, in turn, contributes to the understanding of cultural and linguistic relationships and practices among and within various ethnic groups. This has become possible due to the continual discovery of new HLA alleles using DNA technology which has increased the power of HLA to distinguish individuals (**Arguello *et al.*, 1998; Arnaiz-Villena *et al.*, 1999; Luo *et al.*, 1999; Albis- Camps & Blasczyk 1999**). It has been postulated that gene conversion events are the main mechanisms for distributing and reshuffling sequences among alleles. In addition reciprocal recombination and point mutations have been suggested to be responsible for the generation of alleles over evolutionary time **(Mason & Parham 1998; Hughes & Nei 1992).**

Some indigenous populations (e.g. groups of Native Americans or from Papua New Guinea) show a very restricted diversity of alleles at DRB1 as well as other HLA loci **(Inman & Rudin 1997).** The extensive polymorphism of the major histocompatibility complex (MHC) genes in humans and the differential allelic distribution in ethnic populations of varied origin has been major focus of immuno genetic research. The presence of certain alleles with high frequency only in specific populations (e.g. A36, A43 African Americans) and the strong linkage disequilibrium between HLA neighboring loci, demonstrates that certain combinations of contiguous alleles (HLA haplotypes) show a characteristic frequency or are distinctive in certain living populations **(Browning & McMichael 1996).** The wide range of allelic diversity and the conserved combinations of different alleles are used as genetic markers and anthropological data is based on the information supplied by population studies **(Direskeneli 2000).** Population studies indicate that there are many alleles and DR-DQ haplotypes that appear to be specific for given ethnic group. The existence of ancestral haplotypes implies conservation of large chromosomal segments.

The extra ordinary power of this small segment of the human genome clusters population in a manner expected from linguistic, anthropological and archaeological evidences. Sequences reveal a dramatic level of diversity. Species specific residues i.e. residues that identify a MHC molecule as belonging to a

particular species, are extremely rare. In contrast, species unique residues i.e. residues that were not characteristic of a species are unique to individuals of that species appear somewhat more frequently. The allele frequency distributions and patterns of variability within the molecule, suggest strong selective forces acting on class II loci and I. The form of selection is unknown and potential selective mechanisms should be examined in the light of classical population genetic theory, which states that it is very difficult to maintain so many alleles even with strong balancing selection. Thus, HLA variability at the level of DNA is useful in unraveling the evolutionary relationships between populations and in investigating the evolutionary forces which shaped the genetic profiles of contemporary populations **(Inman & Rudin,1997).**

The HLA polymorphisms have been created because of balancing selections, which maintain a few allelic lines over very long period **(Harpending,1999).** The extensive variation in HLA markers makes the system highly useful for determining genealogical relationships between populations. **Monsalve.(1999)** have successfully compared the relationship between linguistic and genetic data in Native Americans and Asian populations. They have concluded that gene flow and genetic drift are important factors in shaping the genetic landscape of Native American populations. The results are most congruent with the single migration model. In addition the understanding of the events contributing to MHC class II evolution requires comparison of species that are very closely related. The distribution of alleles in different populations can be used to construct a matrix of genetic distances between populations and a phylogenetic tree or un rooted network in order to examine the historical / evolutionary relationships between these groups.

The high polymorphism, tight linkage among loci, and the random association of alleles make the system of particular interest from the perspective of population genetics. Information on the dynamic evolutionary forces that have acted on a locus can be inferred from the number and distribution of alleles that it carries. The major histocompatibility complex (MHC) is unique in the number of highly polymorphic loci spread over such a small chromosomal region. This creates a context for interpreting HLA region variation in both evolutionary and in clinical

terms. The extensive allelic variation among the HLA class I and class II genes distinguishes these as the most polymorphic coding sequence. The distribution and dispersal of certain HLA Class II alleles in different continents of world is shown below **Figure 2.12**.

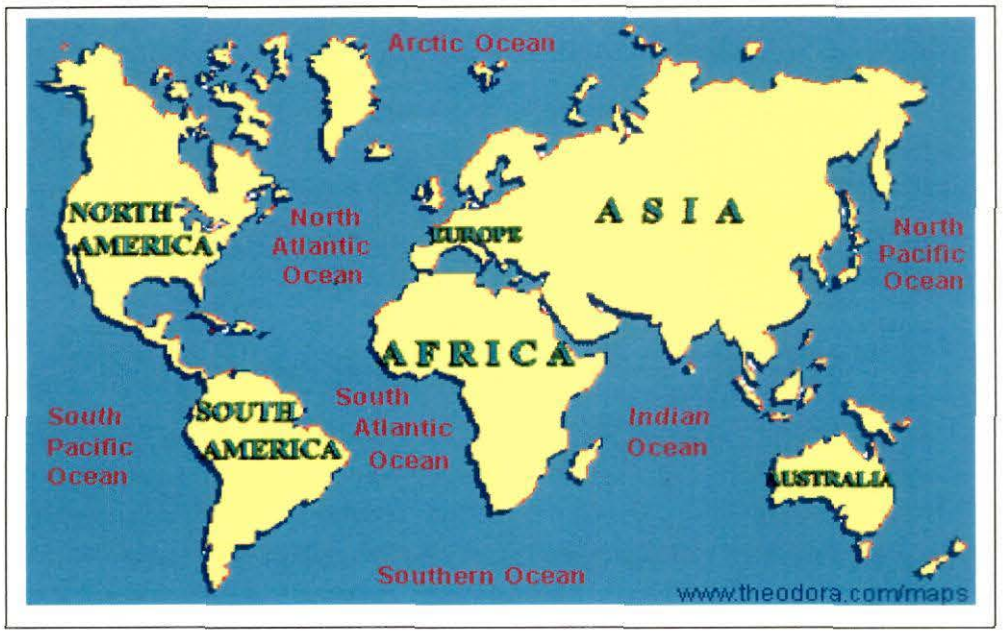


Figure 2.11 Map of Southeast Asia, Australia and the Pacific Ocean, showing the approximate locations of the populations

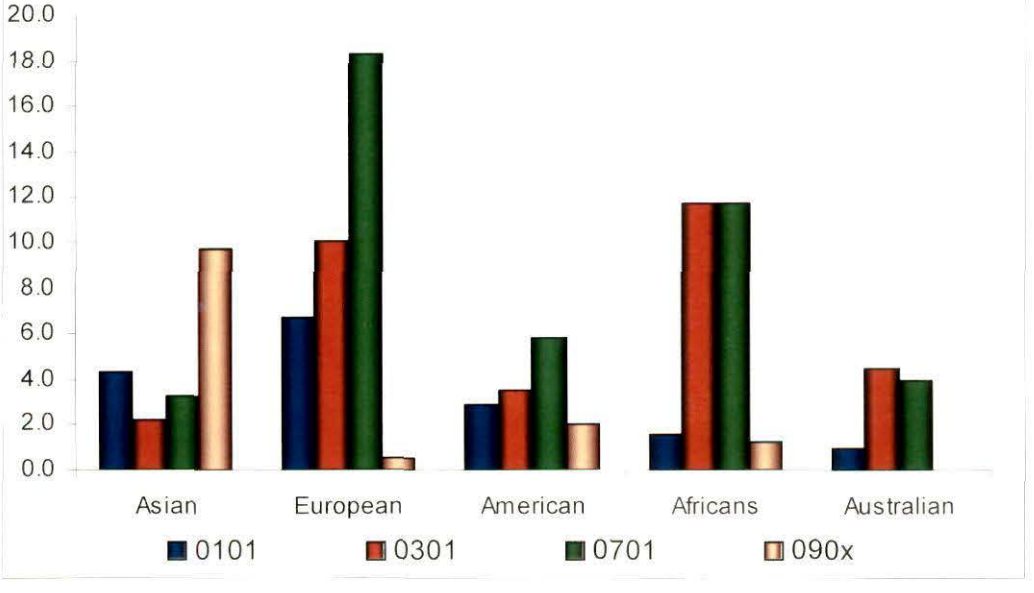


Figure-2.12: Distribution of HLA Class II alleles in different continents of world

It is known that HLA Class II is highly polymorphic as at DRB1 locus 463 alleles, at DQA1 locus 34 alleles and at DQB1 locus 78 alleles have been identified worldwide. The world map as per different continents is shown in **Figure 2.11.** Various alleles of class II are widely distributed in different continents of worlds like Asians, Europeans, Africans, Americans and Australians (**Figure 2.12**).

The present study has been designed to keep the enormous diversity of Indian populations and to decipher the divergence of various populations at HLA class II loci. As there are hundreds of alleles and enormous diversity, many of these alleles are quite ancient, the origin, similarity and the diversity of the Indian subcontinent can be well studied by using this highly polymorphic genetic marker.

The People

The People

The Republic of India is a country in South Asia which comprises the majority of the Indian subcontinent. India has a coastline which stretches over seven thousand kilometers, and shares its borders with Pakistan to the west, the People's Republic of China, Nepal, and Bhutan to the northeast, and Bangladesh and Myanmar (formely known as Burma) on the east. On the Indian Ocean, it is adjacent to the island nations of the Maldives on the southwest, Sri Lanka on the south, and Indonesia on the southeast. India also claims a border with Afghanistan to the northwest. India occupies an area of 32,87,263 sq kms and is located between latitudes 8°4' to 37°6'north and longitudes 68°7'and 97°25'east and measures ~ 3,214 km from north to south and ~ 2,933 km from east to west. Total population of India is 1,065,070,607 where state of Uttar Pradesh is the most populated state with a total population being 100 millions. Politically India is divided into 28 States and 6 unioin territories (**Figure 3.1**).

It is the second most populous country in the world, with a population of over one billion, and is the seventh largest country by geographical area. It is home to some of the most ancient civilizations, and a centre of important historic trade routes. Four major world religions: *Hinduism, Buddhism, Jainism* and *Sikhism* have been originated from India. The official name of the country, India is derived from the Old Persian version of *Sindhu*, the historic local appellation for the *river Indus*. The Constitution of India and general usage also recognizes *Bharat* (Hindi) which is derived from the *Sanskrit* name of an ancient Hindu king. A third name, *Hindustan* (Hindi: *land of the Hindus* in Persian, has been used since *Mughal* times, though its contemporary use is unevenly applied due to domestic disputes over its representative as a national signifier.



Figure 3.1: Geographical map of Republic of India

Brief History of India

Stone Age rock shelters with paintings at *Bhimbetka in Madhya Pradesh* are the earliest known traces of human life in India. The first known permanent settlements appeared 9,000 years ago and developed into the *Indus Valley Civilisation*, which peaked between 2600 BC and 1900 BC. It was followed by the *Vedic Civilisation*. From around 550 BC onwards, many independent kingdoms came into being. In the north, the *Maurya dynasty*, which included the *Buddhist* king Ashoka, contributed greatly to India's cultural landscape. From 180 BC, a series of invasions from Central Asia followed, with the successive establishment in the northern Indian subcontinent of the Indo-Greek, Indo-Scythian and Indo-Parthian kingdoms, and finally the *Kushan Empire*. From the 3rd century onwards the *Gupta dynasty* oversaw the period referred to as ancient India's "Golden Age".In the south, several dynasties including the *Chalukyas, Cheras, Cholas, Kadambas, Pallavas* and

Pandyas prevailed during different periods. Science, art, literature, mathematics, astronomy, engineering, religion, and philosophy flourished under the patronage of these kings.

Following the Islamic invasions in the beginning of the second millennium, much of north and central India came to be ruled by the *Delhi Sultanate*, and later, much of the entire subcontinent by the Mughal dynasty. Nevertheless, several indigenous kingdoms remained or rose to power, especially in the relatively sheltered south.During the middle of the second millennium, several European countries, including the Portuguese, Dutch, French, and British, who were initially interested in trade with India, took advantage of fractured kingdoms fighting each other to establish colonies in the country **(Majumder 2001)**. The English managed to thwart the other colonisers and came to rule much of the country by 1840. After a failed insurrection in 1857 against the British East India Company, popularly known in India as the First War of Indian Independence, most of India came under the direct administrative control of the crown of the British Empire.

India a land of diversity

Survival of human population in Indian sub-continent since last ~60,000 years has been originated from different corners of the world. A large number of migrations from different parts of the world have created an extensaive range of diversity in India. Geography, Language, religion, and caste are major determinants of social and political organisation within the highly diverse Indian population.

Ethnically, Indians belongs to different races with major ethnicity being Indo-Aryan (72%) followed by the Australoids or Dravidians 25% , Mongoloid 2% and traces of Negroid (1%) found in the island regions of Andaman and Nilgiri Hills of Tamil Nadu.

Linguistically, India is home to two major linguistic families, those of the Indo-Aryan and Dravidian-derived languages. Apart from these two linguistic families, Austro-Asiatic languages are spoken by a large number of tribal populations in India, and some groups of the north-eastern India also speak various Sino-Tibetan languaes. The Indian constitution recognises 18 official languages

(Figure 3.2) with Hindi (an Indo-Aryan language) being the official language of Republic of Inida. Hindi is spoken all over north-central and western India. Two classical languages native to the land are *Sanskrit* and *Tamil*. Apart from the the official languages there are more than 875 dialects spoken throughout the country.

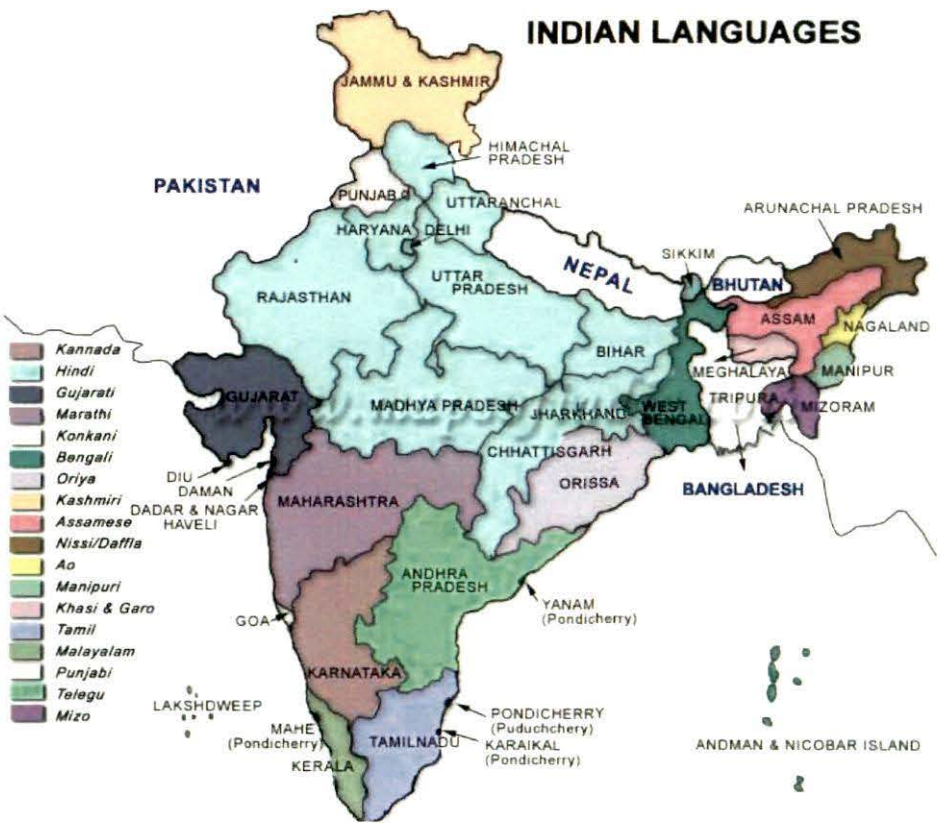


Figure 3.2: Languages spoken in different parts of India

Socially Indian society is fragmented into different non-Tribal and Tribal groups. Tribals are considered to be the original inhabitants of India with majority of them being Austro-Asiatic. They constitute nearly 8% of total Indian Populations rest of the non-Tribal populations follows different religions where Hinduisms is followed by ~75% of Indians. India is also home to the second largest population of *Muslims* in the world (13.4 %) after Indonesia. Other smaller religious minorities include *Christians* (2.33%), *Sikhs* (1.84 %), *Buddhists* (0.76 %), *Jains* (0.40 %), *Ayyavazhi* (0.12 %), *Jews*, *Parsis*, *Ahmadi*, and *Bahá'ís*. Hiundus are found all over India.However, some of the religions have geographical majoies like *Muslim*

constitutes ~70% of Jammu and Kashmir populations and Sikhs forms ~80% of Punjabees **(Figure 3.3)**.



Figure 3.3: Geographical distribution of different Religions in India

Modern India is divided into twenty-eight states and six Union Territories and the National Capital Territory of Delhi. India is the second most populous country in the world, with only China having a larger population. India's literacy rate is 64.8 % with 53.7 % of females and 75.3 % of males being literate. The sex ratio is 933 females for every 1000 males. Work Participation Rate (WPR) (the percentage of workers to total population) stands at 39.1 % with male WPR at 51.7 % and female WPR at 25.6 %. India's median age is 24.66 and has a growth rate of 22.32 births per 1,000.

Although 80.5 % of the people are Hindus, India is also home to the second largest population of *Muslims* in the world (13.4 %) after Indonesia. Other smaller religious minorities include *Christians* (2.33%), *Sikhs* (1.84 %), *Buddhists* (0.76 %), *Jains* (0.40 %), *Ayyavazhi* (0.12 %), *Jews*, *Parsis*, *Ahmadi*, and *Bahá'ís*. India is home to two major linguistic families, those of the Indo-Aryan and Dravidian-

derived languages. The Indian constitution recognises 23 official languages. *Hindi* along with English is the languages used by the Central Government for official purposes. Two classical languages native to the land are *Sanskrit* and *Tamil*. The number of mother tongues in India is as high as 1,652. Language, religion, and caste are major determinants of social and political organisation within the highly diverse Indian population today.

India is home to several major rivers such as the *Ganga (Ganges)*, the *Brahmaputra*, the *Yamuna*, the *Godavari*, and the *Krishna*. The rivers are responsible for the fertile plains in northern India which are conducive to farming. The Indian climate varies from a tropical climate in the south to a more temperate climate in the north. Parts of India which lie in the *Himalaya* have a tundra climate. India gets most of its rains through the monsoons.

India's entire north and northeast states are made up of the *Himalayan Range*. The rest of northern, central and eastern India consists of the fertile *Indo-Gangetic plain*. Towards western India, bordering southeast Pakistan, lies the Thar Desert. The southern Indian peninsula is almost entirely composed of the *Deccan plateau*. The plateau is flanked by two hilly coastal ranges, the *Western Ghats* and *Eastern Ghats*. The *caste system* in India is an important part of ancient Hindu tradition and dates back to 1200 BCE. The term caste was first used by Portuguese travelers who came to India in the 16th century. Caste comes from the Spanish and Portuguese word "casta" which means "race", "breed", or "lineage". Many Indians use the term "*jati*". There are 3,000 *castes* and 25,000 *sub castes* in India, each related to a specific occupation e.g.

Brahmin-Priest, Kshatryas—warriors, Vaishyas—traders, Shudras-- menial jobs

Beyond these four castes there are the *panchamas* (or fifths). These groups were regarded as impure due to their traditional handling of impure substances and therefore were "Untouchable" (*Dalit*) outcastes till a few decades ago. Caste not only dictates one's occupation, but dietary habits and interaction with members of other castes as well. Members of a high caste enjoy more wealth and opportunities while members of a low caste perform menial jobs.

The caste system has existed in India from time immemorial. The word "caste" is supposed to translate the *Sanskrit* word "*Varna*" but sometimes erraneously as "*jati*" which is actually a sub-caste in a local region. The sub-caste (*jati*) belongs to a particular "*varna*". *Jati* designates specialized hereditary functions to an endogamous community locally. Given the long history of humans on the Indian sub-continent and the semi-independent development of numerous socio-cultural groups, the population came to be differentiated by endogamous and occupational groupings, perhaps from pre-historic times. These were and still are subgroups (sub-castes) belonging to the four *Varnas* as shown above. By the time we see the development of a pan-Indian cultural notion of *Varna* essentially that of being a "Hindu", the sub-caste *jatis* had proliferated which led to the complexity of India's social structure.

Hinduism came to be "*Varnashrama Dharma*", the order of (four) castes and (four) stages in life. In pan-indian terms, each local *jatis* were associated with a particular *Varna* in pan-indian sense. The people belong to a *jati* or sub-caste are being identified by their social function/occupation locally and not determined by their economic status i.e, identified by their *jati*-dharma or *jati*-function/occupation. Perhaps the most interesting point about this caste structure is not that it is hierarchical (which it is without a doubt) but rather that the hierarchy disperses only social value but not necessarly economic power. Local *jati* sub-castes have also been largely endogamous except when different *jati* occupational groups merged their jati identification within a Varna. Hence, the proliferation of endogamous groups leading to *jati* diversity within each four Varna, as too the similarities across them.

Much attention has been paid to the limitations on marraiges across *varna/jati* categories. Again it is instructive to note that many influential Brahmins, presumptive superiors in the traditional social ordering, had marriage relations that spanned *Varnas*. The *Varna* and *jati* system do not appear to have been the product of a conquering or dominating socio-economic class. Both systems have survived and indeed identification of *jati* has strengthened, long after the supposed era of "Brahmin dominance". The system is instead an implicit negotiation between India's numerous hereditary occupational (and infrequently ethnic) endogamous groups. The

jati system at any given point in time is a point of equilibrium between these elements of society both at the local and pan-Indian level.

In the beginning the caste system was not a strict system and people could move from one *Varna* to another. Indologists give different dates to this period of change. Some claim the change occurred around 500 B. C. and other claim 500 A. D. Until then, communities and even singular person moved from one Varna to another Varna, because of their desire to adopt different occupations. There were some kings who belonged to *Kshatria* (warrior castes) and changed their status to become religious *Brahmans*. There were also who changed their status to become warriors. And even after the caste system was organized in a strict manner there were many communities who did not always follow their status occupations. There was a case of a Jat that lost its high status because they did not profess the profession worthy of their Varna. The *Kayastha* of east and north east India originally belonged to the *Kshatria Varna* (warrior caste). Some time in the past among warrior's communities, there developed a bureaucratic unit whose job was writing and listing war events and they were called *Kayasthas*. Because these unit members were not warriors, they were excluded from the *Kshatria* status and were given a lower status. But the *Kayasthas* even today claim *Kshatria* status.

Matting pattern among Hindus

As discussed above one of the basic elements of the Indian social structure is the caste system. The institution of caste is so elaborate and pervasive that no aspect of social life in India remains untouched by it. The multiplicity of caste can be illustrated by the fact that **Hutton (1961)** and **Jacob Pandian (1978)** has enumerated about 3000 castes in India. Caste has been defined by **Karve (1961)** as "an endogamous group or an extended family where in the members are related to one another either by blood or by marriage." This definition shows the close interrelationship between the institutions of caste and marriage in India **(Vinay Bahl 2004)**. Marriage is an imperative duty for the Hindu. However, the mate selection for Hindus is a difficult task as there are a number of customary restrictions attached to it, as is implicit in the customary terms like endogamy, exogamy, hyper gamy,

prohibited kin marriage, virgin marriage and other marriage customs and restrictions enumerated by Blunt (1931). These customs have been defined by Blunt as under:

1. The custom of endogamy, which compels him to marry within a certain group. This group may be either the caste or sub caste, to which he / she belong.
2. The custom of exogamy, which forbids him to marry within a certain group. The exogamous group is a subdivision of the endogamous group.
3. The custom of Hindu table of prohibited kin; various castes have various such tables. In some it replaces the restriction of exogamy, in some it reinforces them. The effect is generally to forbid marriage with certain kinds of relatives who are not included in the exogamous groups.
4. The custom of hyper gamy, by which a bride may not marry a man of lower social rank than herself.
5. The custom of virgin marriage, which forbids a man to marry a widow.

Endogamy has been regarded as the most important attribute of the caste system by most of the authors. **Westermarck (1891)** has regarded endogamy as the essence of the caste system. **(Das *et al.*, 1987, 1996)** The principle of endogamy compels a Hindu to marry within his own caste or sub caste. **Blunt (1931)** has indicated that save for certain quite exceptional castes the restriction of endogamy is universal in India. Each caste/ sub caste is further divided into mutually exclusive subgroups known as *gotras* whose members are forbidden to marry within the subgroups (Principle of exogamy).

Until recently breaches of caste endogamy were punished by excommunication from the caste, which was regarded as a social punishment for the violation of this principle. In recent times a trend towards inter caste marriages has been discerned by various authors. **Kapadia (1995)** has shown that willingness on the part of parents to give their children in marriage outside their own caste groups is on the increase. However, he has remarked that the change is in behavior and not in ideology. Some of the recent empirical studies conducted in Punjab have also reported few transformations in this traditional attribute of the caste system. **Bhatnagar (1972)** has shown that although education has not apparently made an

impact on the principle of endogamy but the attitudes of educated persons towards intercaste marriages are quite liberal. It was also reported that willingness of inter caste marriage has tended to increase in rural areas which are in proximity of towns or cities.

Despite such trends which indicate a relaxation in the marriage customs, the statistical estimates of the actual intercaste marriages are not very high. Empirically these deviations from the norm represent just a small fraction of the total number of marriages contracted. From the genetic stand point these departures from the norm can hardly be expected to alter markedly the constitution of gene pools. The magnitude of the genetic disturbances caused by intercaste marriages may attain a greater significance, if and when the trends which have just set in become more pronounced.

Uttar Pradesh

Uttar Pradesh is one of the largest states of India with an area of 294,413 km^2 and a population of about 166,052,859 million with 87466301 males and 78586558 females. The male: female ratio is 916:1000 respectively. Uttar Pradesh is the most populated State of India. Demographic profile of India and Uttar Pradesh shown in **Table 3.1**

Uttar Pradesh has 70 different districts. Lucknow is its capital, which spreads over an area of 2528 sq. kms. It lies between parallels 26°30' and 27°10' north latitude and 80°34' and 81°12' east longitude. Detailed map of Uttar Pradesh with various districts is shown in **Figure 3.4.**

In the present study we have chosen three ethnic groups for studying the genetic makeup of these populations. First group comprised of caste populations i.e. *Kayastha, Mathurs, Vaishyas* and *Rastogies.* Second group is of Muslims i.e. *Shia* and *Sunni* Muslims and in third group we have taken North Eastern populations, which comprises of two tribal populations like *Lachung* and *Mech* while another is caste population i.e. *Rajbanshi.* The brief description of each of these groups is given below.

Table 3.1: Demographic profiles of India and Uttar Pradesh

	India	Uttar Pradesh
Population (2004 est)	1,027,015,247	166,052,859
Percent Population Increase (1991-2001)	23.9	25.8
Density (Population /Km^2)	324	689
Percent Urban	26.1	19.8
Sex Ratio (no. of males per 1000 females)	933	916
Percent 0-14 Years	36.3	39.1
Percent 65+Years Old	3.8	3.8
Percent Schedule Castes	16.7	21.0
Percent Schedule Tribes	8.0	0.2
Percent Literate	65.8	57.36
Exponential Growth Rate	2.14	2.27
Total Fertility Rate	3.6	5.1
Infant Mortality Rate	79	98

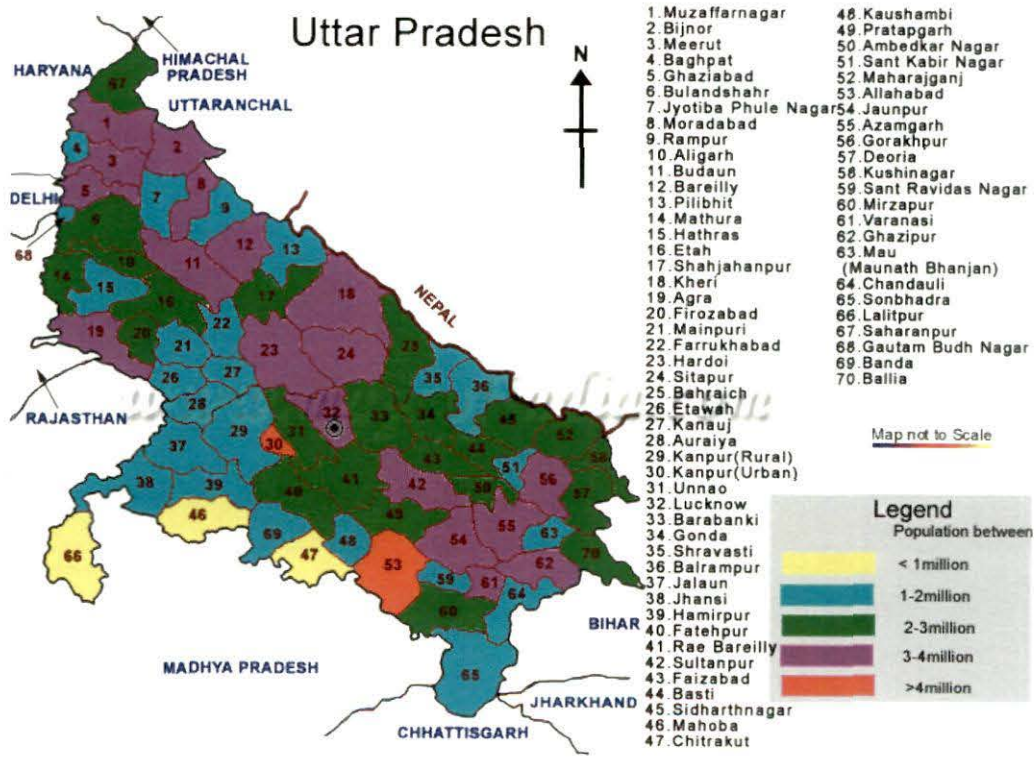


Figure 3.4: Geographical and Population Map of Uttar Pradesh

Kayastha

Origin of *Kayastha*

About the name of this community there is some difference of opinion. Mr.Colebrooke gave as the popular derivation the Sanskrit *kaya-sanstitah*, "staying at home," in reference to their sedentary habits. The caste themselves drive their name from *kaya-stha*, situated in the body; incorporate, with reference to the legend of their descent.

Like all people who are on their promotion, the *Kayastha*, are particularly sensitive as to any imputations on the purity of their descent and it is from every point of view, unless to revive a troublesome controversy. Risley remarked that "the physical characters of the Bihar *Kayastha* afford some grounds for the belief that they may be of tolerably pure Aryan descent, though the group is doubtless a functional one recruited from all grades of the Aryan community."

The *Kayastha* of northern India has recently shown an inclination to admit them to full rights of communion. Risley writes: putting the tradition aside, and looking, on the one hand, to the physical type of the *Kayastha*, and, on the other, to their remarkable intellectual attainments, it would seem that their claim to Aryan descent can not be wholly rejected, though all attempts to lay down their genealogy precisely must necessarily be futile. It appears to be at least a plausible conjecture that they were a functional group, developed within the Aryan community, in response to the demand for an official and literary class, which must in course of time have arisen. This class would naturally have been recruited more largely from the peaceful *Vaisyas* and *Sudras* than from the warlike *Kshatriyas*, while the *Brahmins* would probably have held aloof from it altogether.

In *Padma Puran* (one of the ancient Hindu literature) we read: *Brahma* having remained in mediation for a while, there sprang from his whole body a male of godlike form bearing an inkpot and a pen. This being was named *Chitragupta*, and he was placed by *Brahma* near *Dharmaraja* to register the good and evil actions of all sentient beings. He was possessed of supernatural wisdom, and became the partaker of sacrifices offered to the gods and fire. As he sprang from the body of

Brahma, he is said to be of the *Kayastha* class. His descendants are *Kayastha* of numerous gotras on the face of the earth.

Subgroups of Kayastha

The *Kayasthas* are divided into twelve endogamous sub-castes. These are *Srivastva, Bhatnagar, Saxena, Amisht or Anvasta, Aithan, Asthana, Balmik or Valmiki, Mathurs, Suryadhwaja or Surajdhwaj, Kulsreshta* or *Kulasreshta, Karan* or *Karana, Gauda* or *Gaur* and *Nigam*, with a thirteenth known as Unaya or those of Unao.

Mating Pattern

In the matter of intermarriage, *Kayasthas* follow the standard rule of exogamy based on the text of *Yajnavalkya* as expounded by the author of the *Mitakshara*, which bars marriage between *sapindas*, that is, who are within five degrees of affinity on the side of the mother and seven degrees on the side of the father. Among the sub-castes which still maintain the organization of local groups or sections (al), marriage can not take place between persons belonging to the same al; nor can a man marry a women belonging to the al of his maternal grandfather or great-grandfather. Adult marriage is the rule, and infant marriage the exception, among most of the *Kayasthas*. Sexual license before marriage is neither recognized nor tolerated, and the parents and other members of a girl thus offending would be promptly ex-communicated. Polyandry is utterly prohibited, and polygamy, though allowed, is rarely resorted to unless the first wife be barren, in which case the stringent necessity of begetting male issue makes a second marriage permissible. Re-marriage of widows is absolutely prohibited. There is no legalized divorce, but if a wife commits adultery, she is at once put out of caste, and she can not eat or associate with any member of the community, such women can not, as matter of course, marry again.

Kayasthas follow the highest form of the eight kinds of marriage recognized by *Manu* in his Institutes that known as *Brahma*. The ceremony is performed according to the rules laid down in the *Sanskrit* treatise known as the *Vivaha Paddhati*, with *Vedic* language (*mantra*), as in the case of *Brahmans* and *Kayathas* are Hindus, but belong to various religious sects: some are *Saivas*, some *Saktas*, and

some *Vaishnavas*. Some few are *Nanakshahis*, *Kabirpanthis*, or *Acharis*, or belong to the *Arya Samaj*. The fact of their belonging to different sects does not prevent them from associating freely as members of the same brotherhood. A man of the *Vaishnava* sect may marry a *Sakta* woman, provided the rules of exogamy are observed. Notwithstanding the jealousy with which they are regarded by their less astute neighbors, the social position of the caste is a high one.

Occupation

Some historians hold the view that during the reign of the Mughals, a number of Hindus who were educated and endowed with sharp intellect attained administrative positions through rapid adaptation to the Persian language and culture of India's new rulers. These influential Hindus got together and formed a new caste known as *Kayastha*. Two other regional communities also lay claim to the name *Kayastha*. These are the Prabhu *Kayasthas* of Maharashtra and Bengali *Kayasthas* of West Bengal. They were also the 'writing-castes' in Maharashtra and Bengal, respectively, like the Chitragupta *Kayasthas* of North India

Vaishyas

Origin of *Vaishyas*

Third in the caste system (*Varna*), *Vaishyas* supposedly evolved from *Brahma's* thighs. The *Vaishya's* duty was to ensure the community's prosperity through agriculture, cattle rearing and trade. Initially it was imposed by the society that *Vaishyas* can learn *Vedas* but can not preach *Vedas*. Later, the *Shudras* took over agriculture and cattle rearing while the *Vaishyas* became traders and merchants.

From the end of the 4th century BC, as the country became politically stable, trade routes to previously uncharted areas developed. The merchant community was the first to benefit. Artisans formed guilds and co-operatives in the urban areas and guild leaders became important figures in society. Guilds also provided technical education, though formal education remained the monopoly of the *Brahmins*. As their economic power increased, they were expected to give alms to *Brahmins*, throw feasts for them, and donate generously towards the building of temples and shrines.

However, though they were "twice-born" (*Upanayanam*) and economically strong because they controlled commerce, *Vaishyas* were denied a high social status, for which they resented the upper castes. One expression of this resentment was their support of the anti-Brahminical sects that developed around the 6th century BC, like Buddhism and Jainism. Later, however, many influential *Vaishyas* were honored with titles like 'Nagar Seth' (chief merchant of the town) and 'Tazimi Seth' (privileged to wear an anklet in the king's presence and in royal durbars) Members of trading communities originate mostly from the West Coast and from Sind (in present-day Pakistan). They are commonly known as *banias*, a distortion of the *Sanskrit* word *vanik* meaning "trader". In southern India, the *Chettiars* and *Mudaliars* are prominent *Vaishya* communities, who have contributed to society by building hospitals, endowing universities and developing industries. While the fortunes of the other castes have fluctuated with politics and invasions, the *Vaishyas* alone have preserved their social and financial stability down the ages. Many of them have funded wars and political movements, notably *Seth Amirchand Daga* of Bikaner (Rajasthan), who is believed to have financed the British in the Crimean War. In the last century, some of the most prominent *Vaishya* merchants and philanthropists came from the region of Marwar (Jodhpur), Gujarat and Sindh. Many of them funded Mahatma Gandhi's freedom movement and then evolved into modern India's leading industrialists.

Matting Pattern

Mating pattern among *Vaishya* is strictly regulated in a way which necessitates out breeding and permits marital exchange between and not within, a set of exogamous familial lines. There is persistent and rigid following of the 'sapinda' regulations in this group which regulate marriages within the caste group strictly avoiding consanguinity. The regulations prohibit marriage between individuals related through common male ancestor up to the 7^{th} generation on father's side and fifth generation on mother's side.

Mathurs

Origin of *Mathurs*

According to Indian history *Chitragupta* married *Irawati* and *Nandini*. The 12 *Kayastha* sub-castes are traced to his 12 sons, eight by *Irawati* and four by *Nandini*. *Mathurs, Gaurs, Bhatnagars, Saxenas, Ambhasths, Nigams, Karns* and *Kulshreshths* became the descendants of the eight sons by *Irawati*, whereas *Srivastavas, Suryadhwajs, Asthanas* and *Valmiks* became the descendants of the four sons by *Nandini*. As time passed, the name *Chittagupta* got transformed into *Chitragupta*. Historically *Mathurs* are subgroup of *Kayastha*.

Subgroups of *Mathurs*

As regards the origin of the *Mathur* sub-caste of the *Kayasthas*, there is no manner of doubt that they are so called because their original home was Mathura, much in the same way as *Shrivastavas* are so called because probably their original home was *Shravasti* which was a town famous in *Buddhistic* history and as *Bhatnagars* are do named after *Bhatner*. It is noticeable that there is a class of *Brahmins* too who calls themselves *Mathurs* and trace their ancestory to Mathura. So too there are some Bengalis who are known as *Mathur* Babus. According to Pauranic tradition the eldest son of *Shri Chitragupta* whose name was Charu took up his abode at Mathura ---and so his descendants came to known as *Mathurs*.

The *Mathurs* ruled over Mathura till Qutab-ud-din conquered it. The *Mathurs* are also said to have ruled *Ayodhya*. Later on, their descendants accepted posts of Dewan of the same area under the *Surya Vanshi* family and *Bundra Mathur* and his nineteen generations held the above mentioned posts. The downfall of *Ayodhya* commenced during the Dewanship of *Bal Pratain Mathur*, who took reins of the kingdom in his own hands. His rule stretched through ten generations, after which the reins passed to *Maharaja Daleep*. The kings who followed were *Raghu, Dashrata, Rama, Lakshmana, Bharata and Shatrughana*. The surnames that the *Mathur* frequently adopt are *Dayal, Lall, Chandra, Andley, Berni, Saharia*, and sometimes *Bahadur*. The *Mathurs* are subdivided into *als* and *gotras*. People belonging to the same *al* claim to have descended from the same immediate ancestor and share a common totem. *Gotra* is a larger group being composed of a number of als.

According to tradition, there were only eighty-four big villages in *Brij Bhoomi* - i.e. in the *Ilaka* of Mathura. It is probably on that account that there are eighty-four further sub-sects called *Allas* of the *Mathurs*. Although all these *Allas* no longer represent or indicate those eighty-four villages as their origin and have been considerably changed under various influences such as some peculiarity of the personages belonging to them, nevertheless it is an acknowledged fact that the number of these *Allas* is still reckoned as eighty-four. Out of the 84 *Allas* families belonging to only 25 *Allas* or *Khamps* are to be found in *Rajputana* and of only 13 in Ajmer. The genealogies of quite a large number of *Mathur* families of these parts are given in their history. What is noticeable in the accounts of these families is that several of them had come either from Mathura direct or from Delhi and nearly all of them belonged to the administrative services of the then rulers and held offices such as Diwans, Ministers, Secretaries, or *Kanungoes* etc. receiving shares of the profits of land. That the 84 *Allas* or sub-sections acquired their names after those of the villages or *mauzas* of their origin seems probable enough *Sahariyas* were so named because their original residence was a village *Sahar* which was not very far from Mathura; *Golghotia Alla* is most probably derived by metathesis from *Gokalotia* i.e., residents of *Gokal Narnolias* from *Narnole*, *Mahabani* from *Mahaban*, etc. Other influences, however, operated later on to change them beyond recognition

Mating Pattern

The 12 sub-castes of the *Kayasthas* are not only endogamous subdivisions, but they also have exogamous divisions based on *als*. An *al* would refer to a distinguished ancestor or the place of origin or it could refer to a characteristic acquired during migration. Members of the same *al* are prohibited to inter-marry and hence *als* performs the same function that *gotras* do in other castes. *Mathurs* are divided into different *als* and members of these als are prohibited to marry among themselves. The most common *als* are *Saharia, Kataria, Kakrania, Dewariya, Dilwariya,Tawakaley, Rajauria, Nag, Galgotia, Sarwaria, Andley(Endlay), Ranoria.*

Occupation

During the reign of the Mughals, a number of Hindus who were educated and endowed with sharp intellect attained administrative positions through rapid

adaptation to the Persian language and culture of India's new rulers. Ancient *Sanskrit* texts dating back to the pre-Mughal period, though, have references to *Kayasthas* and *Chitragupta*. The *Smriti of Yajnavalkya* describes the *Kayasthas* as writers, scribes and village accountants. The *Vishnu* and *Parashara Smritis* have also similar references to them, describing them as magistrates, judges and chief executive officers. In the *Garud Puran*, Chitragupta has been described as the giver of the script *(Chitraguptaya namastubhyam veda aksharadatre* — salutation to Chitragupta, the giver of the script). The *Rig Veda* enjoins a salutation to *Chitragupta* before offering sacrifice — *Sri Chitraguptaya vai namah.*

Rastogies

Origin of *Rastogies*

The *Rastogies* form an endogamous caste group concentrated around the indo-gangetic plain in North India. They are fair, handsome, of medium height and stocky build. The community is industrious and has done well in business. Generally they have been identified as *Vaishya* because of their traditional occupation in trade, in particular the money lending. This traditional occupation of *Rastogis* has identified them as of a particular *Varna* in the *Chatur Varna Vyavastha*, but a thorough probe into the community and its history, as also opinions from the established *Vaishya* or *Bania* caste groups reveal a contradictory status for them. While, the majority of the *Rastogies* They considers themselves as definitely of *Vaishya Varna*, a group of enlightened educated elite among them has been claiming a *Rajput* status for the caste. This trend of mobility from within is naturally sought to be substantiated with myths and traditions. What makes the situation more interesting is that certain higher castes consider *Rastogis* as belonging to lower caste group. However, this is contradicted in their behavior as they do not observe any dietary prohibition against the *Rastogies* and freely accept food and water from them. It is also noteworthy that the *Rastogis* wear the sacred thread after due performance of *Yagyopaweet* ceremony just before the marriage, which is a typical *savarna* Hindu caste trait. The point which is important here is not what is thought about them in

caste ranking or in ritual purity scale, but what caste group genetic elements are represented in this group identified as *Rastogis*.

Lastly, the creation of Sudra from the feet symbolizes them as footman, thus their duties are to serve the above three Varnas. The traditional activities of these four Varnas namely *Brahmin, Kshatriya, Vaishya* and *Sudra* confirm the *'Chatur Varna Vyavastha'* of the society. In India, today, there are innumerable castes, all of which associate themselves with one or the other *Varnas* of the above system. There have been inter-mixtures at various levels which must have modified the genetic structure of these castes, sometimes resulting into the formation of other groups of subordinate rank. There are a few *jatis* however, which are not clearly assigned to one or the other *Varna*, in their case their jati status in the *Varna* frame is differently defined at different times by different people, e.g. such is the case with the *Kayasthas* and the *Rastogies*. It is difficult to say that how these groups have come to exist. May be they are the product of fission or fusion within and outside the caste level followed by isolation.

Subgroups of Rastogies

The position of *Rastogis* in the Hindu society of today is not very clear. Conflicting claims regarding their place in the *Chatur Varna Vyavstha* have been made. In view of their anomalous position in the caste hierarchy, it seems essential that a brief survey with respect to their origin and distribution be made. An attempt has therefore, been made here to give an account of the community in terms of its origin, distribution and patterns of mating as prevalent in the society. This survey will enable us to formulate the problem in better perspective.

Matting Pattern

Mating pattern among *Rastogies* is strictly regulated in a way which necessitates out breeding and permits marital exchange between and not within, a set of exogamous familial lines. There is persistent and rigid following of the 'sapinda' regulations in this group, which regulate marriages within the caste group strictly avoiding consanguinity. The regulations prohibit marriage between individuals related through common male ancestor up to the 7^{th} generation on father's side and fifth generation on mother's side.

Indian Muslims

The other two populations selected for the present study from North India are Shia and Sunnis which belong to the Muslim sect. Muslim influx and influence in India started almost at the inception of the religion. The traders from Arabia were frequent visitors to the Indian subcontinent even before Prophet Muhammad revealed the Koran. They brought the word of Muhammad to India in the 7th century and this resulted in some peaceful conversions of Hindus to Islam.

In 711, the Umayyad caliph in Damascus sent an expedition to Baluchistan (an arid region on the Iranian Plateau in Southwest Asia) led by a twenty-year-old Syrian Muslim chieftain named Muhammad bin Qasim, who conquered Sindh (presently a province of Pakistan bordering on Baluchistan, Punjab, and Rajasthan, India) in 712 AD. However, this event in history does not seem to have influenced India as much as expected. Beginning with the invasion of Mohammad of Ghazni in the 11th century, followed by a barrage of invaders from Persia, Turkey, Afghanistan, Northwest and Central Asia in the 11th and 12th centuries, a full force of Islam was thrust upon India. Spread of Muslims in India began with the "Slave Dynasty" of Turkic ruler bur flourished all over India during the reign of Mughal Empire whose first ruler was Babur, King of Farghana (at present in Uzbekistan).

It took several centuries to finally spread Islam in all portions of India. Many Hindus were turned to Islam by laws favoring Muslims or by, others turned to Islam voluntarily. Most Indian Muslims who converted to Islam were Hindu and some of their ancestors embraced Islam under duress, although some did willingly or under the influence of laws favoring Muslims. There were also some converts who belonged to the ruling families of the different kingdoms of the region, many of whom were given little choice in the matter. The Muslim rulers of India also brought businessmen, traders, merchants and slaves from different parts of the world. Many of them married local Indians who converted them to Islam.

Present day Indian Muslims

Today there are more than a quarter of a billion Muslims living in the Indian subcontinent (India, Pakistan and Bangla Desh). This amounts to more than a quarter

of the total Muslim population of the world. Muslims form about fifteen percent of the Indian population.

(1) Categorization of Indian Muslims

The Muslims in India are categorized into two distinct classes based on their origins namely, Ashraf and Ajlaf. The Ashrafs are again grouped as *Sayyeds*, *Sheikhs*, *Mughals* and *Pathans*. The Sayyeds are said to be descendents of the Prophet and regarded in high esteem. The Sheikhs are of Arab descent and are next in line in prestige. The Mughals are descendents of the greatest Muslim rulers of India, the Mughals occupy third place. Pathans including fighters hail from the northwestern regions including Afghanistan and form the last group of Ashraf. The Ajlaf on the other hand are the Indian converts and are considered to be of common ancestry.

(2) Distribution of Muslims in India

Muslims constitute 13.4% of Total Indian population **(Census, 2001).** They are found in all the states of India and are nearly $3/4^{th}$ of the total population of Jammu and Kashmir, more than $1/4^{th}$ of West Bengal and Assam and nearly $1/5^{th}$ of population of Uttar Pradesh. Majority of the Muslim population is of Sunnis (~87%) and they are scattered throughout India. On the other hand, majority of Shiites are found in Uttar Pradesh, Maharashtra, Hyderabad and its adjoining regions.

(3) Distribution of Muslims in Uttar Pradesh

Muslims constitute about 18.5% of the total populations of Uttar Pradesh and are found extensively in areas of Lucknow, Kanpur, Rampur, Hardoi, Saharanpur, Bahraich, Moradabad, Gorakhpur, Ghazipur, Mau, Allahabad, Mirzapur and other eastern districts of Uttar Pradesh. Among these Shiite Muslims are mainly found in Rampur and Lucknow divisions.

(4) Marital Patterns of Muslims in India

Muslims of India practice large scale of consanguinity and except the real sibs, marriages are permissible in between cousins. Marriages between Sunni and Shia sects are very rare, however, marriages in other religions has been frequently seen and has been a major reason of large-scale conversions of Hindus to Muslims. If

one marries a Muslim boy or girl then he or she has to embrace Islam. Overall, it has been seen that migration rates from other religions to Muslims is several fold higher in females. Sunnis being the descendents of the Muslim rulers all over India have been able to marry and admixed with local Indian caste population more profusely while Shiites to a large extent remained confined to their own sects. Another important issue regarding the marital pattern of Muslims is high rate of polygamy.

(5) Possible origin of Indian Muslims

Possible place of origin of Indian Muslims varies from Middle East (Saudi Arabia, Syria, Iraq), to North West Asia (Turkey), Central-west Asia (Afghanistan and Iran) and some part of Eastern Europe (Uzbekistan) as shown in **Figure 3.5.** Although a large number of Indian Muslims are local Hindu converts.

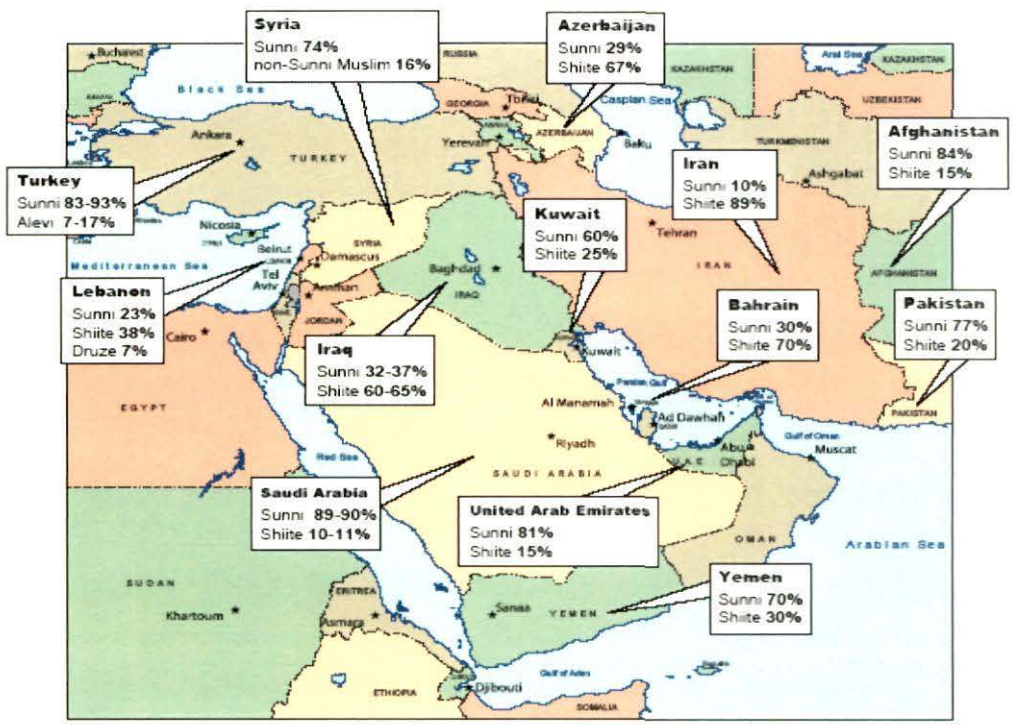


Figure 3.5: Possible place of origin of Indian Muslims

Muslim influx and influence in India started almost at the inception of the religion. The traders from Arabia were frequent visitors to the Indian subcontinent even before Prophet Muhammad revealed the Koran. They brought the word of

Muhammad to India in the 7th century and this resulted in some peaceful conversions of Hindus to Islam.

In 711, the Umayyad caliph in Damascus sent an expedition to Baluchistan (an arid region on the Iranian Plateau in Southwest Asia) led by a twenty-year-old Syrian Muslim chieftain named Muhammad bin Qasim, who conquered Sindh (presently a province of Pakistan bordering on Baluchistan, Punjab, and Rajasthan, India) in 712 AD. However, this event in history does not seem to have influenced India as much as expected. Beginning with the invasion of Mohammad of Ghazni in the 11th century, followed by a barrage of invaders from Persia, Turkey, Afghanistan, Northwest and Central Asia in the 11th and 12th centuries, a full force of Islam was thrust upon India. Spread of Muslims in India began with the "Slave Dynasty" of Turkic ruler bur flourished all over India during the reign of Mughal Empire whose first ruler was Babur, King of Farghana (at present in Uzbekistan).

It took several centuries to finally spread Islam in all portions of India. Many Hindus were turned to Islam by laws favoring Muslims or by, others turned to Islam voluntarily. Most Indian Muslims who converted to Islam were Hindu and some of their ancestors embraced Islam under duress, although some did willingly or under the influence of laws favoring Muslims. There were also some converts who belonged to the ruling families of the different kingdoms of the region, many of whom were given little choice in the matter. The Muslim rulers of India also brought businessmen, traders, merchants and slaves from different parts of the world. Many of them married local Indians who converted them to Islam

North East group

North East India accounts for 7.8 per cent of the total land space of the country, of which again 70 per cent is hilly terrain and about 98% of this region's borders form India's international boundaries.

Eastern and North Eastern Indian region consists of Arunachal Pradesh, Assam, Manipur, Meghalaya, Mizoram, Nagaland, Sikkim, Tripura, and West Bengal. They have about 30 million people (out of a total of about 120 million)

living below (those with incomes of less than $ 1 a day) the poverty line **(Ray & De 2003).**

Table 3.2: Demographic profiles of India and North East India

Particulars b	Unit	NE9a	India
Area	Share in India (%)	9	
Population (2001)	Share in India (%)	12	
Population Density (2001)	Per sq. km.	442	347
PCI in 1970-71[e]	US $ (at current price)	89	320
PCI in 1999-2000[e]	US $ (at current price)	300	363
Percent Literacy		69.22	65.8
Sex Ratio (no. of males per 1000 females)		934	933

a. Arunachal Pradesh, Assam, Manipur, Mizoram, Maghalaya, Nagaland, Sikkim, Tripura and West Bengal.
b. Considers average data of nine NEs.
Sources: Various issues of Economic Survey, Government of India and Profile of States, CMIE. Preliminary Results of 2001 Census, Register of Census Operation, Government of India.

The region shares borders with China in the north, Bangladesh in the southwest, Bhutan in the northwest and Myanmar in the east **(Figure 3.6).** This region has absorbed many waves of migrations. Here Indo-Aryan, Tibeto-Burmese, Chinese, and Mon-Khmer races have mingled with the aborigines to create colorful communities and different political systems amid the fertile Brahmaputra and Surma valleys, the resource-rich Eastern Himalayas, and their foot-hills. It has been hypothe-sized that a plethora of migrations, particularly through the northeast Indian corridor, has contributed to the present day population of northeastern India.

The partition of the country in 1947 broke the natural and age old linkages with rest of the country and this land locked region is connected to mainland India through a narrow corridor in north Bengal, popularly known as the 'Chicken's neck' or the 'Siliguri neck', having an approximate width of 33 km. on the eastern side and 21 km. on the western side. The socio economic and political settings of the northeastern region bear direct linkages and fit in the order where the tribal form of society and economy occupy the vantage positions. The north eastern part of India is inhabited by numerous endogamous tribes and castes that have their own distinct

social, linguistic, and biological identity. Ethnically speaking, most of the tribal groups are Mongoloids, whereas caste groups are either Caucasoid or show a mosaic of features of both the ethnic groups. The Mongoloids/Indo-Mongoloids have come to India from different directions at different times and perhaps earlier than the Caucasoids **(Das *et al.*, 1987).** While the Mongoloids have migrated from eastern, southeastern **(Rapson 1955; Dani 1960),** and central Asian regions, the Caucasoid may have entered from western and northern boundaries of this region. While a majority of the Mongoloids are tribes affiliated with the Tibeto-Chinese linguistic family, excepting Khasi, most of the Caucasoid are caste groups and speak Indo-European languages. Although these groups have been broadly classified on the basis of language and ethnicity, they show considerable variations within these broad categories. Both the Mongoloid and Caucasoid groups show a certain degree of differentiation within themselves in cultural and biological traits such as anthropometry, genetic markers, and dermatoglyphics.

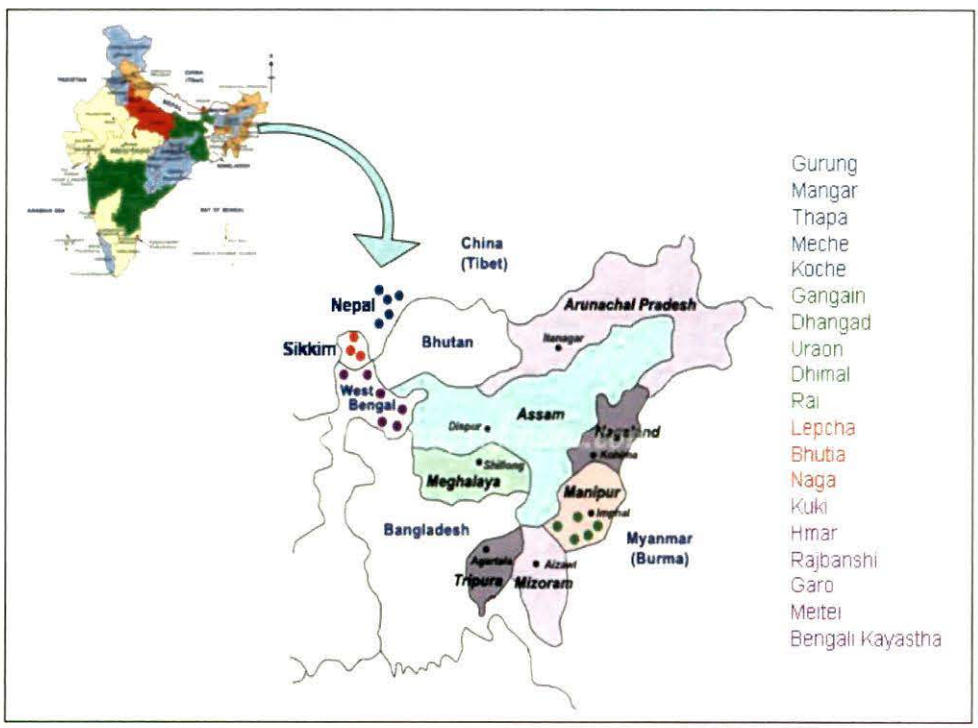


Figure 3.6 Map showing different North Eastern regions

The populations of this region show a mosaic pattern and represent a diverse population from different regions such as Nepali populations of Indo-European and

Tibeto Burman family like Gurung, Mangar, Thapa, Meche, Koche, Gangain, Dhangad, Uraon, Dhimal, Rai etc. Lepcha, Bhutia of Tibeto-Burman family mostly from Sikkim. Naga, Kuki, Hmar, etc. from Manipur, which, are also Tibeto-Burman speaking groups. Rajbanshi, Garo, Meitei, Bengali, Bengali Kayastha etc are few Indo-European populations from West Bengal. On basis of ethnicity, prevalence of endogamy, Linguistic characteristic anthropological significance we selected only three populations from this North East region. These three populations were selected as origin of these populations were from different geographical regions like Lachung from Lachungpa valley (Sikkim), Meches from Jhapa district of Nepal and Rajbanshi from Koch Behar and Jalpaiguri.

Lachung

The Bhutias are people of Tibetan origin, who migrated to Sikkim(India), India and Bhutan some time after the 15th century. They migrated through the different passes ("La" in Tibetan) in the Himalayas. Geographical denotations in the names of Bhutias last names is common. In Northern Sikkim for example, where the Bhutias are the majority inhabitants, they are known as the Lachenpas or Lachungpas, meaning inhabitants of Lachen or Lachung respectively **(Figure 3.7)**.

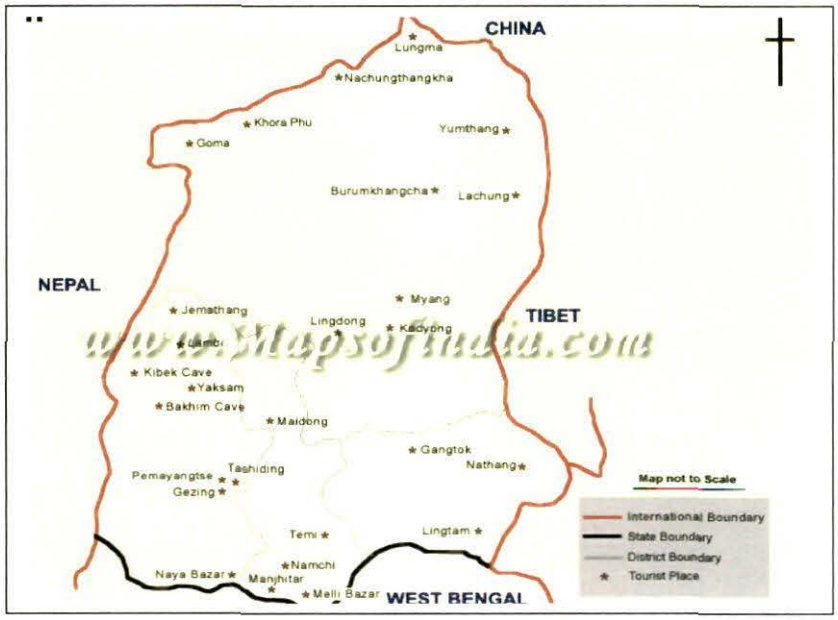


Figure 3.7: Map of Sikkim showing Lachung Valley

The Bhutia aristocrats were called *Kazis* after similar land lord titles in neighboring India especially during the waning days of the Mughal period. This feudal system was an integral part of Chogyal monarchy prior to 1975, when Sikkim was an independent monarchy. Among the Bhutias, the Lachenpas and Lachungpas have their own traditional legal system called the "Dzumsa" which means *the meeting place of the people* The Dzumsa is headed by the village headman known as the *Pipon*. Bhutias are traditionally rice eaters with animal fat fried vegetables or meat usually pork, and beef, and occasionally mutton or chicken. Other well known foods are momo- steamed meat dumplings, and the Thukpa- noodle in broth. The Losar and Losoong are two among many festivals celebrated by the Bhutia community. Almost all Bhutia festivals hold religious Buddhist significance.

Marital Pattern

Bhutias practise intermarriage within their clans and follow a very hierarchical system of bride and groom selection. Clan discrimination is widespread and marriage outside the community is looked down upon.

Mech

Being residents of the Mechi River banks and the neighborhoods in the district of Jhapa, they are called the Meches. Jhapa district, a part of Mechi zone, is one of the seventy-five districts of Nepal, a landlocked country of South Asia **(Figure 3.8)**. The district, with Chandragadhi as its district headquarters, covers an area of 1,606 km² and has a population (2001) of 688,109. Jhapa is the easternmost district of Nepal and lies in the fertile Terai plains. It borders Ilam district in the north, Morang district in the west, the Indian state of Bihar in the South and east, and India's West Bengal state in the east. It has an area of about 1606 km² and a population of about 7,000,000 (according to the census). Jhapa is home to many indigenous tribes such as the Rajbanshi, Satar (Santhal), Meche, Koche, Gangain, Dhangad, Uraon, Dhimal, Rai, Limbu, Magar, Gurung and many others, besides the Brahmnins, Chhettris, and Newars. Jhapa receives 2,500 to 3,000 mm of rainfall a year, and mostly during the monsoon season in the summer and its hilly northern part receives more rainfall than down south. Its major rivers like Mechi, Kankai Mai,

Ratuwa, Biring, Deuniya, Hadiya, and Ninda provide water for irrigation purposes. Due to its alluvial soil best suitable for agriculture. Its name itself is derived from the Rajbanshi word "jhapa" meaning "canopy", which proves that the area was a dense forest in the past.

It was such a dense and dangerous forest that it was called Kaalapaani and prisoners were sent here to die of malaria and other diseases in the jungles.Though the forest area has decreased in recent times due to ignorance, people are waking up to the fact that forests are necessary. Consequently, community forests can be seen in the district these days. Jhapa is diverse and rich in culture and traditions due to its different tribes. All the tribes have their own languages, customs and traditions, and they celebrate their festivals with huge favour and enthusiasm every year.



Figure 3.8: Map of Nepal. Jhapa district is shown at boundry of Sikkim and Bihar

Meches are closer to the Bodos in civilization. According to historians, Meches were nomadic until a few decades ago. Their myth connects them with

Limbus. They became settlers when the land range and forest frontiers of their free roaming became demarcated and restricted. They are animists and Ai Bali Khungri and Batho Barau are their principal deities. They also worship the deities of the forest. Their language derives from the Tibeto-Burman family. Meches are also called Bodos. They are at present engaged in farming. According to Census 2001, their population is 3,763. According to the mythology of Meches, they are brethren of Rais, Limbus and Kirats, who settled in the Terai as they were left behind in course of their journey.

In India Meches are described under the Kachhari community. In the Indian census of 1881, the Kachharis are shown to have 18 groups, including Bodo, Dimasa, Lalung, Madahi, Mech, Rabha, Sarania, Hojai, Garo, Rajbanshi or Koch, Chutiya, Moran, Hajong, Tippera, Mahaliya, Dhimal, Solanimiya and Phulgaria. Their language has been placed in the Tibeto-Burman group. But there is less use of ª (kna), ` (ayan), of (ana), g (na), d (ma) in this language like in between Brahmaputra and Kankai rivers. Boche or Bodo are spread around Duars of North Bengal and Bhutan. Their settlements, in most of the cases, are based in jungles or on the banks of rivers and streams. The way of living, costumes, ornaments, rituals and culture of Meches are unique and resemble those of the Bodo tribe of India. Birth pollution is not over until the umbilical of a newborn falls off its body. Their priest, Raja, makes the family purified. Marriage is of different types. Meches observe Chharkela (the worship of Laxmi). They are very much fond of songs and dances. Drums, pipe and bamboo split canes are their musical instruments. They celebrate both the planting and harvesting of crops. Meches have 13 thars (clans). They have traditional village councils and the councils have a chief. Their Gaunburas (village chiefs) are called Makhal. The vocations of Meches also vary with their thars (Basumatari-Meche, BS2054/55). Mainao or Aibali Khungri and Batho Barau are their prominent gods and goddesses. River worship of Meches is famous. For them siundi (a kind of milk and thorny plant) is a plant where gods and goddesses live. They do not worship idols nor do they have temples to gods or goddesses. Meche women are very industrious. They are seen weaving clothes for their family's use by installing looms in almost every house

Marital Pattern

Meches marry very strictly in their community. They are highly endogamous and they generally do not get admixed with people of different community in particular geographic region.

Rajbanshi

The ethnic status of Rajbanshi is not clear, and different opinions are advanced regarding this **(Dalton 1872; Risley 1915; Waddel 1975; Das *et al.*, 1987).** Rajbanshis are hypothesized to be a mixture of different tribal groups (Rabhas, Tiwas, Kacharis, Garos, Karbis, etc.) that were converted to Hinduism and in the process became admixed with certain Caucasoid caste populations **(Das *et al.*, 1987).**

The Ahom and the Rajbanshi, who were originally tribal groups, held a dominant position in terms of economy as well as power. Both of these groups tried to perpetuate their domination and obtain social sanction for it by becoming an upper caste group. In this process, they might have tried to emulate the cultural and social ideologies of the neighboring upper caste groups and developed marital ties with them. It may be pertinent to note that Das (1981) suggests that Rajbanshi is a fit example of a Tribe–Caste continuum not only in their socio cultural aspects but also in their biological makeup. Therefore, one may infer that what appears to be a purely sociological phenomenon of a Tribe–Caste continuum may well be reflected in their genetic structure. Rajbanshi were originally affiliated with the Tibeto-Chinese linguistic family but now have become semi-Hinduized caste groups, like Chutiya, and Ahom. These groups have currently assumed caste status and speak Indo-European languages **(Vikrant Kumar *et al.*, 2004).** Rajbanshis are mostly the inhabitants of Jalpaiguri and Koch Behar, their geographical position is shown in the **Figure 3.9.**

It is also reported that the Koch-Rajbanshis belong to the Tibeto-Burman group of the Mongoloid origin. Popularly known as the Nomoshudras, this community has always been in the periphery of the Hindu caste hierarchy. In fact in 1496 A.D when Visvasimha the Koch chieftain rose to power, he did so by

converting into Hinduism and adopting Hindu practices through Sanskritization.The Brahmins who were brought in from as far as Kanauj, Mithila, Srihatta legitimized his political power by bringing him into the Brahmanical fold through Hindu rites and rituals.

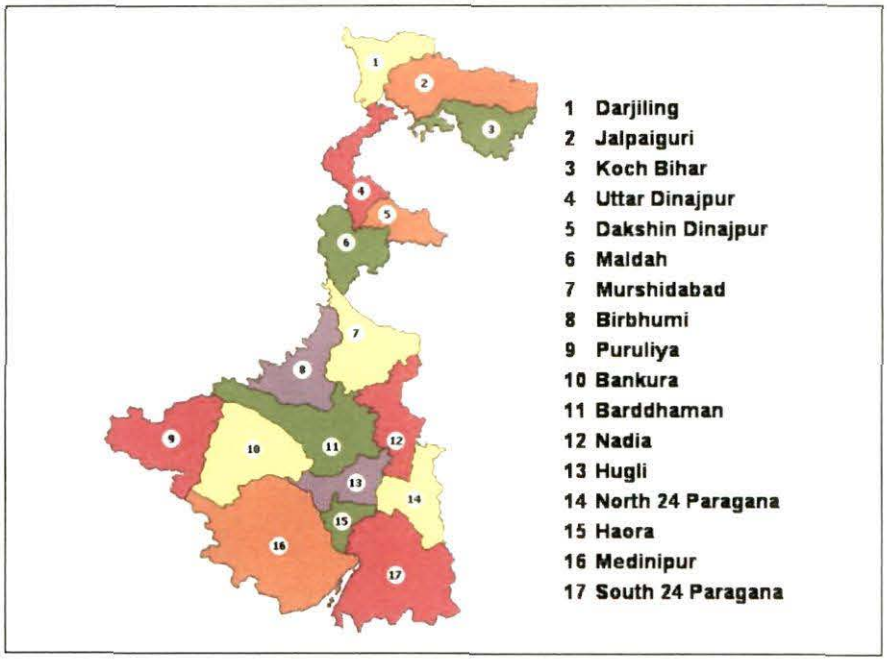


Figure 3.9: Map showing geographical distribution of West Bengal

They also traced the descent of his lineage and tribesmen as Kshtriya's, who had thrown away their sacred thread when fleeing from the wrath of Parashurama. Thus many of Vishwasimha's followers had given up their tribal designation and called themselves Rajbanshis. Thus it is widely held that Koches after undergoing Sanskritisation assumed the name Rajbanshi.

Hinduised, poor and illiterate Rajbanshi's could not really enter the Hindu fold successfully due to the resistance offered by the high caste Hindus, and in fact fell a easy prey to the insatiable greed of the Brahmins. While in 1891 the Rajbanshi's described themselves as Vratya Kshtriya , from 1911 they began to claim pure Kshtriya status legitimized by priests, genealogists and pundits. Since 1912, a number of mass thread wearing ceremonies (Milan Kshetra) were organized in different districts by the "Kshtriya Samiti" where lakhs of Rajbanshi's donned the

sacred thread as a mark of Kshtriya status. The immediate objective of the "Kshtriya samiti" was to regain the lost social status of the Rajbanshi community in the hindu social system.

Marital Pattern

It is considered that Rajbanshis were conglomeration of various tribes that were converted into hindus and in the process they become admixed with certain caste groups after that they started following strict endogamy.

Materials and Methods

Materials and Methods

In the present chapter, we have dealt with the details about the techniques and analytical parameters used for collection, analysis and interpretation of the data. The chapter has been categorized into three parts. At first, we have described the criterion and procedure followed for the sample collection. The second portion deals with the laboratory experimental procedures i.e. HLA typing of DRB1, DQB1 and DQA1 class two antigens. Finally, we have mentioned the details of the statistical tools used for data analysis.

Sample Collection

(a) Populations selected

Present study has been carried out on four endogamous caste groups of North India, two inbreeding Muslim populations and three populations from North East India with the focal rationale of assessing and quantifying the genetic variation existing within and between the populations. The four caste populations and two Muslim populations selected in the present study belong to the state of Uttar Pradesh. They include: (1) *Kayastha* (2) *Mathurs* (3) *Rastogies* (4) *Vaish* (5) *Shia* Muslims and (6) *Sunni* Muslims. Another three populations selected were from North eastern region. They include: (1) *Lachung,* (2) *Mech* and (3) *Rajbanshi.*

Detailed description about the possible origin, socio-cultural practices and the position of these population groups in the social hierarchy network of Indian society have been discussed in detail in the chapter "The People".

(b) Selection criteria of samples

Randomly selected normal healthy individuals from each population have been included in the present study. Three major selection criteria were considered for sample collection which includes:

(i) Un-relatedness of individuals from each other at least for last three generations

(ii) Resident of the state of Uttar Pradesh and North East at least since last three generation

(iii) Exclusion of individuals having inter-caste or inter-religion marriages in last three generations.

(c) Sample collection strategy

Prior to the sample collection, regional addresses and detailed computerized lists of the populations were prepared. Random numbers were generated with the help of computer and adult individuals living in different parts of Uttar Pradesh and North east were questioned about their ethnicity, caste and tribal affiliations and *surnames* and the birthplaces of their parents. Only unrelated subjects were considered eligible to participate. The demographic profile and other ethnical and familial information were filled in a detailed Performa **(Appendix-A)**. Three- generation pedigree charts were prepared to assure un-relatedness in all the samples **(Appendix-B)**. Whole blood was then obtained by venipuncture and about 5 ml blood was collected in EDTA vacutainer tubes after obtaining the informed consent from the subjects.

The study was performed with the approval of the institutional ethical reviewing committee of Sanjay Gandhi Post Graduate Institute of Medical Sciences (SGPGIMS), Lucknow and the government of India.

(d) Demographic profile of the samples

A Total of 1336 samples belonging to four caste populations, two Muslim populations and three North Eastern populations were collected from different regions of Uttar Pradesh and North Eastern parts of India. Male to female ratio was 1.55 as majority of the samples collected were male (n= 786). All samples from North India belong to the Indo-Aryan linguistic family and were residents of Uttar Pradesh since last three generations while form North Eastern region single population belongs to Indo-Aryan family but other two tribals belong to Tibeto-Burman linguistic family. Demographic profile of the samples collected is shown in **Table 4.1.**

Table 4.1: Demographic profile of the studied populations

	Populations								
	Kayastha	***Vaishya***	***Rastogies***	***Mathurs***	***Shia***	***Sunni***	***Lachung***	***Mech***	***Rajbanshi***
Sample Size									
Total (1336)	190	155	196	198	190	188	58	63	98
Male (786)	109	116	118	112	110	121	20	20	60
Female (568)	76	72	82	72	80	67	38	43	38
Mean Age (Years)									
Male	36 ± 5.2	41 ± 2.4	38 ± 2.2	39 ± 4.4	32 ± 4.4	39 ± 2.4	33 ± 3.2	30 ± .4	25 ± 2.2
Female	35 ± 3.1	34 ± 1.5	37 ± 1.4	33 ± 2.2	31 ± 4.1	35 ± 2.4	30 ± 3.1	28 ± 3.4	23 ± 4.4
Linguistic Family									
Major	Indo-Aryan	Indo-Aryan	Indo-Aryan	Indo-Aryan	Indo-Aryan	Indo-Aryan	Tibeto-Burman	Tibeto-Burman	Indo-Aryan
Demographic Profile									
Ethnicity	Caucasians	Caucasians	Caucasians	Caucasians	Caucasians	Caucasians	Mongoloid	Mongoloid	Mongoloid

Laboratory analysis

Laboratory analysis includes extraction of genomic DNA from venous blood samples for HLA class two typing.

(I) Extraction of genomic DNA

(a) Extraction protocol

Laboratory Analysis

DNA extraction

High molecular weight DNA was extracted by using salting out method. The following protocol was used for DNA extraction.

500 µl of EDTA blood was taken in a 1.5 ml eppendorf tube. To each tube 1ml of red cell lysis buffer (**Appendix C**) was added. Contents were mixed gently by inversion for 1min. Tubes were spun at 12,000 rpm for 1 min. at 4 ° C. and the supernatant was discarded. The pellet was resuspended in 200µl of red cell lysis buffer. It was mixed gently by inversion for 1min (washing) and centrifuged at 12,000 rpm for 1 min at 4°C. The supernatant was discarded and step 2 was repeated again (washing) once with red cell lysis buffer and next with water. To the pellet 80µl of Proteinase K buffer **(Appendix C)** and 10µl of 10 % SDS was added. The pellet was resuspended with a wide bore microtip, 400 ul of phenol-chloroform (5:1) **(Appendix C)** was added. The tube was inverted to mix until the contents turned milky. The tubes were centrifuged at 12,000 rpm for 10 min. The aqueous layer was taken out in a fresh tube. Double the volume of chilled ethanol (-70 °C) was added to the aqueous layer. Mixing was done by gently inverting the tubes. The tubes were centrifuged at 12,000 rpm for 10 min at 4°C. The supernatant was decanted off and all the excess fluid was completely drained off. The pellet was washed in 200µl of 70 % ethanol (**Appendix C**). The excess liquid was drained off and pellet was dried at 56 ° C for at least 3-4 hrs. The pellet was finally dissolved in 200 µl of autoclaved HPLC grade water. The tubes were kept at 56 ° C for overnight to dissolve the DNA.

Characterization of DNA

The integrity:

The integrity of high molecular weight DNA is an important factor, which should be considered during extraction steps. Integrity was checked by electrophoresis on 0.8% agarose prepared in 1X TBE buffer (**Appendix C**), containing Ethidium bromide (3µl of 10mg/ml stock for every 50ml of 0.8% agarose). The high molecular weight genomic DNA appeared as a single band near the well. (**Figure 4.1**)

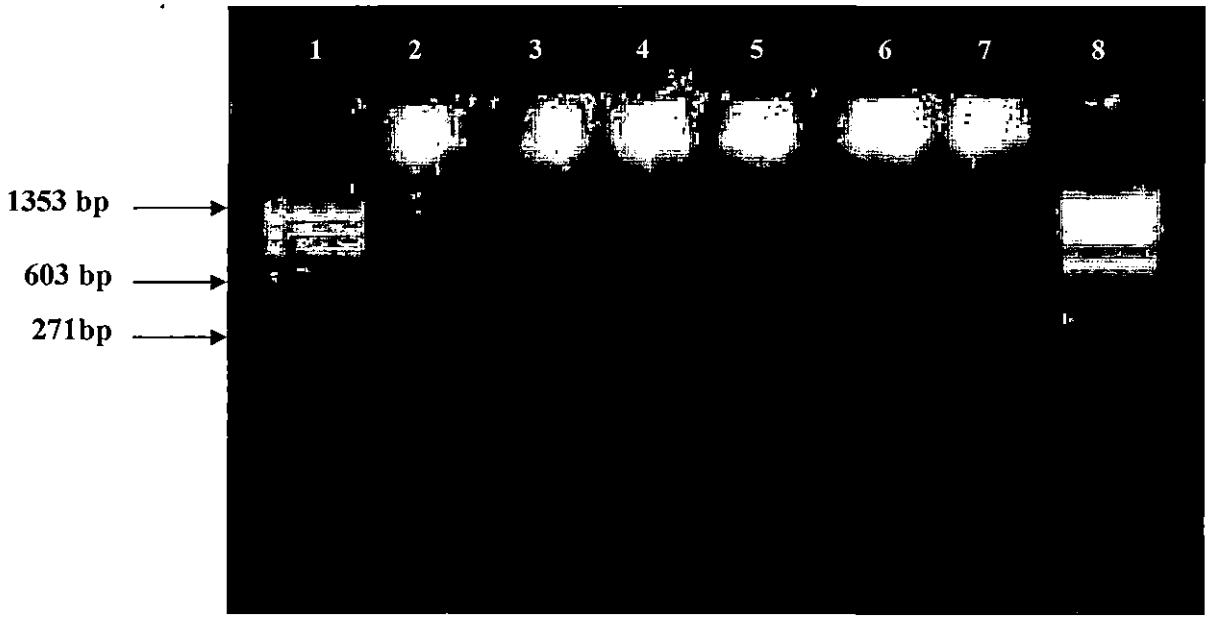


Figure 4.1: High Molecular Weight genomic DNA

Lane 1, 8	Hae III digest of ϕ X 174 DNA
Lane 2-7	DNA samples from Bhargavas (1-6)

The concentration:

DNA was quantified by measuring the optical density at 260nm. 5 µl of stock genomic DNA was taken and 995µl of water was added (Dilution Factor D.F. = 200), mixed well and O.D. was taken at 260 nm in a spectrophotometer (Hitachi). DNA concentration of the sample was calculated as follows:

1 OD = 50μg / ml of ds DNA

XOD = X x 50 μg/ml of ds DNA

The original DNA solution was diluted by a factor, DF = 200

The concentration of DNA in the original stock DNA

= X x 50 x DF μg / ml

= X x 50 x 200 μg / ml

= 10,000 x X μg/ ml

E.g. if OD of the diluted sample = 0.016

DNA concentration = 10,000 x 0.016 μg / ml

= 160 μg / ml

Purity of DNA

Purity of DNA was determined by taking the O.D. of the sample at 280nm for protein concentration and at 260nm for DNA concentration. The ratio OD_{260} / OD_{280} was calculated. DNA samples for which the ratio was 1.7 or above was considered good. If the ratio was less than 1.2, DNA was extracted again.

Each DNA sample was checked on agarose gel stained with ethidium bromide by comparing with known quantity of phage-Lambda DNA as estimation by OD may be inaccurate due to RNA contamination. Samples, where large amount of RNA contamination was seen were treated with RNAase A (Sigma Aldrich, final concentration of 100μg /ml) at 37 ° C for 2 hrs, followed by re-extraction with phenol-chloroform iso-amyl alcohol mixture.

Storage

Samples which were in regular use were kept at 4 °C. Rest of the samples were stored at -80°C.

HLA typing

Various methods are used for DNA typing of MHC class II molecules. These are (i) sequence specific primer based typing (SSP). (ii) Sequence specific

oligonucleotide probe based typing (SSOP). (iii) single stranded conformational polymorphism (SSCP) (iv) sequence based typing (SBT).

In the present study, we have used sequence specific oligonucleotide probe based typing. Polymorphism within class II region is confined mainly to defined hypervariable regions in exon 2, making differentiation between alleles readily achievable through hybridization with the relevant sequence specific oligonucleotide probes (SSOP).

Initial PCR amplification of HLA genes provide a template which can be probed with labelled oligonucleotides which have specificity for particular sequence polymorphisms **(Figure 4.2)**. Panels of these probes can be used to identify which polymorphisms are present in the amplified DNA.

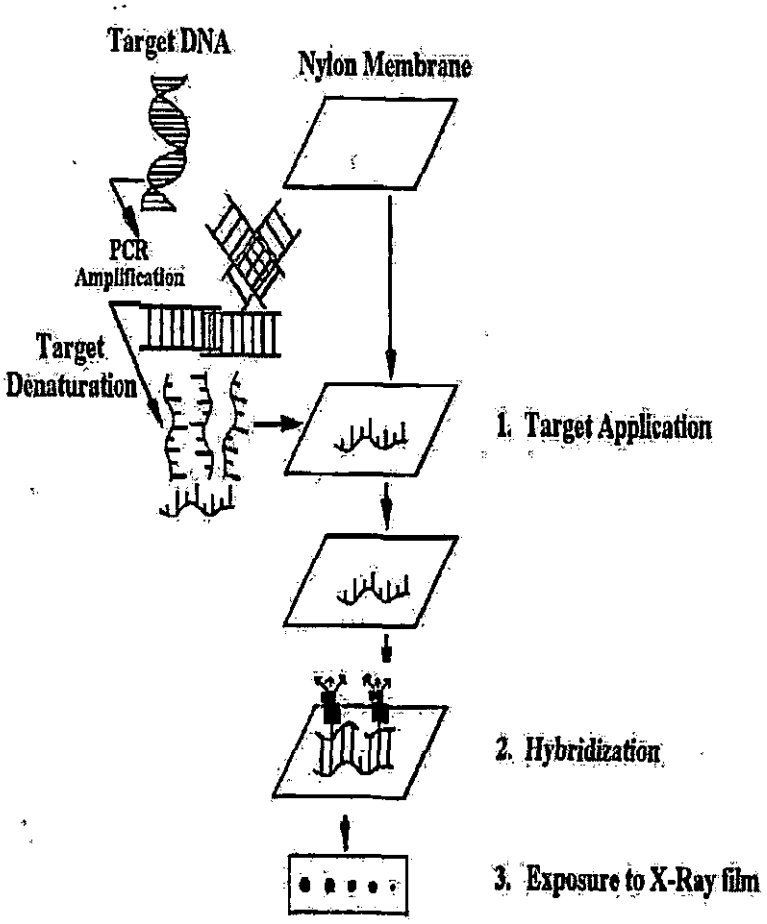


Figure 4.2: Schematic diagram of SSOP based HLA typing

Tissue typing by PCR-SSOP relies on a pattern of hybridization reactivity, which decreases in complexity in relation to increasing PCR specificity. There can be some difficulty in the interpretation of the results, as the majority of the polymorphic differences detected are not unique to a particular allele, but are shared with other alleles, some of which may be found in other loci, depending on the specificity of the PCR. The complex patterns of reactivity detected may require detailed analysis to assign specificities, and certain heterozygous allele combinations may be difficult to assign as hybridization patterns fail to provide a definitive answer.

Dilution of Probes

The probes supplied were unlabelled, HPLC purified, lyophilised (1OD each). We diluted 1 OD probes exactly according to the method given in **Appendix D,** aliquoted the stock solution and stored at -80°C. The stock solutions were diluted to get working solution as per the requirement of each step. Care was taken to avoid repeated freezing and thawing.

SSOP (sequence specific oligonucleotides probes) based typing:

The SSOP technique involves the following steps:

1. Amplification of specific region of the locus to be typed
2. Dot-blotting of the amplified DNA
3. Labelling of oligonucleotide probes
4. Hybridization of blotted product with labelled probe
5. Detection of the hybridized probes
6. Deprobing / dehybridisation of probes.
7. Interpretation of signals.

The detailed procedures of each step are discussed below.

Amplification of specific region

Polymerase Chain reaction is a technique involving primer-directed enzymatic *in vitro* amplification of specific nucleic acid stretches. PCR allows us to

work with an extremely small stretch of HLA gene sequence within the genomic DNA by synthesizing over a million copies within a few hours, significantly facilitating subsequent analysis of HLA polymorphism by sequence specific oligonucleotide probe hybridization.

PCR was carried out in thermostable PCR tubes with 50 µl of PCR mixture that consisted of 5ul of 10X PCR – buffer (**Appendix C**), 0-2ul (depending on loci) of $MgCl_2$ (25mM), 2 ul of each primer (10pm/ul), 1ul each of the deoxynucleotide triphosphates (10mM), and 1ul of Taq *polymerase* (3U/ul) along with 4 - 6 µl (depending on concentration) of the template DNA. PCR mixture was overlaid with 50µl of autoclaved light mineral oil to prevent evaporation during the high temperature cycles. The amplification was done using automated DNA thermocycler (PTC-100, M.J Research) at successive incubation steps for denaturation, annealing and extension. PCR cycling conditions for DRB1, DQA1 and DQB1 loci are shown in **Table 4.2.**

Primer designing:

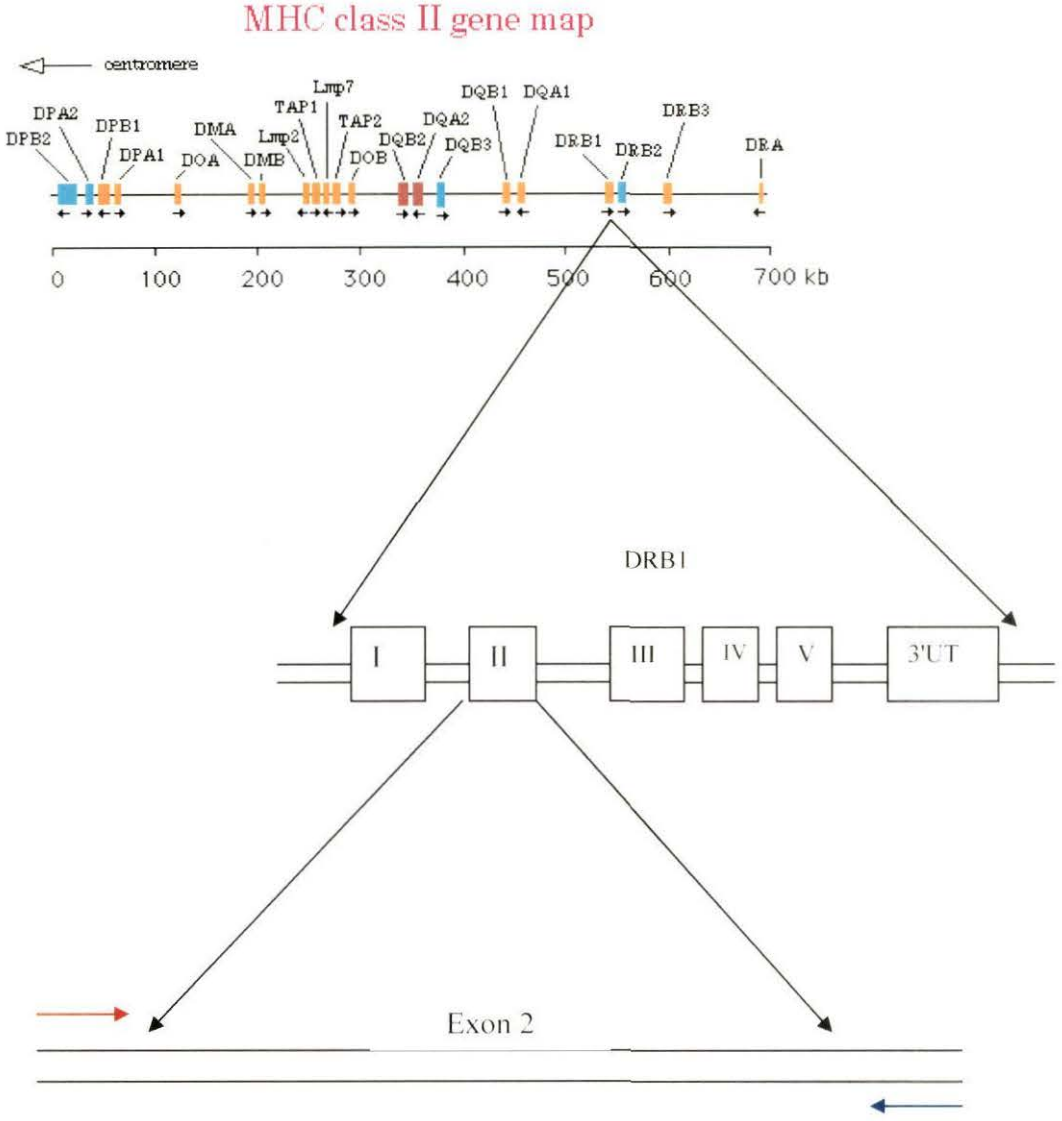
Primers flanking the region of interest were taken in each PCR. For the generic amplification, primers were chosen to amplify all known alleles at the loci and were complementary to sequences shared by all the alleles. An example is shown in **Figure 4.3** for the generic amplification of the exon II of the DRB1 gene. For the specific amplifications for sub-typing, primers having complementarity to a sub set of alleles (e.g. all DR2 alleles) only were chosen. The list of all the primers used for class II typing is given in **Appendix E.**

The details of the samples, their concentrations & the amount of template added for each was noted on experiment sheet ' PCR-Amplification record ' **(Appendix F).**

Table 4.2 PCR conditions for DRB1, DQA1 and DQB1 loci

Name	Primer	Denaturation		Annealing		Extension	
		Temp	**Period**	**Temp**	**Period**	**Temp**	**Period**
Generic							
DRB1	2DRBAMP-A	95^0C	1 min	55^0C	1 min	72^0C	2 min
	2DRBAMP-B						
DQA1	2DQAAMP-A	96^0C	1 min	55^0C	1 min	72^0C	2 min
	2DQAAMP-B						
DQB1	2DQBAMP-A	96^0C	1 min	55^0C	1 min	72^0C	2 min
	2DQBAMP-B						
Specific							
DR1 – DRB1	2DRBAMP-1	96^0C	1 min	60^0C	30 sec	72^0C	2 min
	2DRBAMP-B						
DR2 – DRB1	2DRBAMP-2	96^0C	1 min	55^0C	1 min	72^0C	2 min
	2DRBAMP-B						
DR4 – DRB1	2DRBAMP-4	96^0C	1 min	60^0C	30 sec	72^0C	2 min
	2DRBAMP-B						
DR52 – DRB1	2DRBAMP-3	96^0C	1 min	55^0C	1 min	72^0C	2 min
	2DRBAMP-B						

Figure 4.3: Schematic representation of primer designing for generic amplification of exon 2 of DRB1 gene



Amplification Check

We used mini gel electrophoresis apparatus (Bio-Rad Laboratories, USA) for rapid separation of amplified PCR products. The amplified product was run on 2% Agarose gel with a current strength of 30 mAmps until the tracking dye (**Appendix C**) travelled to 3 cms from the origin. TBE buffer was used as tank buffer, Bromophenol blue as the tracking dye and Hae III digest of ϕX 174 DNA as DNA size marker. The bands were visualised using UV trans-illuminator (Bangalore genei). The band size ranged from 229 – 263 bp according to the primer used (**Table 4.2**). The photographs (**Figure 4.4**, DRB generic amplification product) were taken and permanent records were kept on experiment sheet Ampli-Check (**Appendix F**).

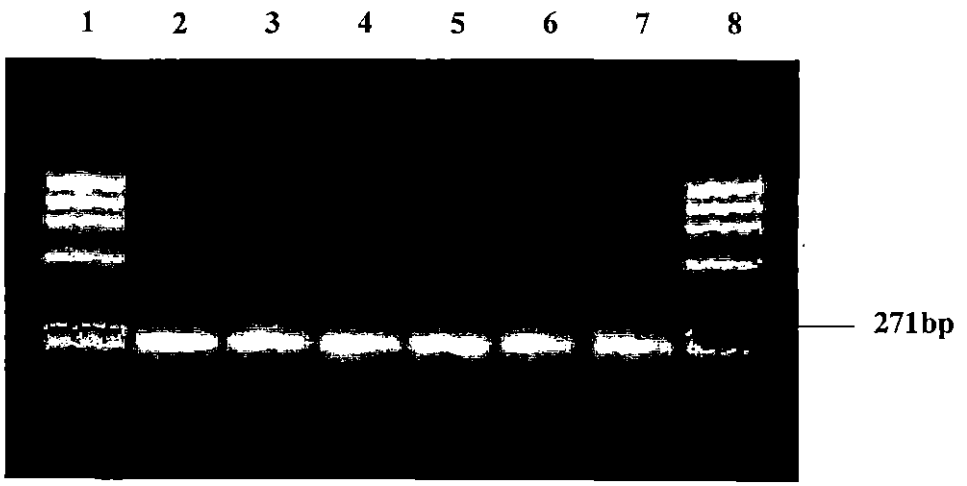


Figure 4.4: Product of DRB1 gene exon 2 amplification using flanking primers

Lane 1	Hae III digest of ϕ X 174 DNA
Lane 2-7	DRB1 PCR product Bhargava 1-6
Lane 8	Hae III digest of ϕ X 174 DNA

Table 4.3 Primer pairs used for the different amplifications and different product sizes

Locus	Primer Pair	Product (bp)
Generic		
DRB1	2DRB AMP-A, 2DRBAMP–B	274
DQA1	2DQAAMP-A, 2DQAAMP–B	229
DQB1	2DQB AMP-A, 2DQBAMP–B	214
Specific		
DR1 – DRB1	DRBAMP-1, 2DRBAMP–B	261
DR2 – DRB1	DRBAMP-2, 2DRBAMP–B	261
DR4 – DRB1	DRBAMP-4, 2DRBAMP–B	263
DR52 – DRB1	DRBAMP-3, 2DRBAMP–B	266

Dot Blotting of PCR products:

A piece of nylon membrane 8cm x 12cm. in size was cut. Membrane was pre-wetted in ddH_2O for 10 min followed by 5 min in 10xSSC. Excess liquid was drained off from the membrane and membrane was placed upon the gasket of the dot-blot apparatus. Membranes were handled with the help of blunt ended forceps. Air bubbles were gently removed using a roller. The sample template was placed on top of the membrane. The vacuum pump was started. 100µl 10xSSC was applied to all 96 sample wells. As soon as the buffer solution drained from all the wells, the pump was turned off.

To 45µl of each PCR product 160 µl 0.4M NaOH/25mM EDTA (denaturing solution, **Appendix C**) was added and incubated for 10 min at room temperature. 50µl of denatured DNA sample was then applied to the membrane corresponding to the order according to experiment sheet 'Blotting details' (**Appendix F**).

Vacuum was then applied. The screws were loosened with the vacuum still on. Finally the sample template was removed. The vacuum pump was turned off and the membrane was removed and allowed to air dry for at least 20 minutes.

Denaturation:

The membranes with DNA face up were placed onto thick Whatmann paper soaked in 0.4M NaOH (**Appendix C**) and left for denaturation for 10 min. Care was taken so that the membranes were not dragged over the denaturation pad. All the air bubbles were removed so that no air bubble was left beneath the membrane.

Neutralization:

The membranes were transferred onto Whatmann paper soaked in 10xSSPE **(Appendix C)** and left for neutralization for 10 min. These were then allowed to air dry for at least 25 min.

UV Crosslinking:

The membranes were wrapped in Saran wrap and were placed on UV trans-illuminator in a manner so that the surface of the membrane containing DNA was facing down. A glass plate was placed on the top of the membranes to hold it flat during the crosslinking procedure. Four minutes exposure was given for crosslinking. The membranes were stored in plastic bags at 4° C.

Labelling of SSO with ^{32}P

Labelling reactions were put up by mixing 1.0 µl of probe (10pm/ µl), 2.5µl 10x Polynucleotide kinase Buffer (Bangalore Genei), 10 µl γ ^{3} ^{2}P ATP (BARC, specific activity 5000 Ci/ mMol), 2.5 µl of T_4 Polynucleotide Kinase (5U/ul, Bangalore Genei) and water to make a final reaction volume of 25µl. This reaction was incubated at 37 ° for 30 min. 1 µl of 0.5 M EDTA (pH 8.0) was added to stop the reaction. To the labelled probe 1.5 µl of salmon sperm DNA (from a 1mg/ml stock solution, **Appendix C**) 60 µl of water and 360 µl of 5 M Ammonium acetate was added and mixed well. 1125 µl of chilled ethanol was then added and the tube was left at -70°C overnight for the probe to precipitate. The tube was then

centrifuged at 12,000 rpm for 15 min at, 4° C, pellet was washed with 70% ethanol, dried and 100 µl double distilled water was added to dissolve the probe. The tube was kept at 56^0C for 4 hrs so that the complete probe was dissolved.

The labelling efficiency was checked by taking the count of 1ul of the labelled probe in liquid scintillation counter (Beckman, USA)

Hybridization

The conditions of pre-hybridisation, hybridization and washing of PCR amplified DNA blots were optimized for SSOPs.

The membrane with denatured amplified DNA fixed on it was pre hybridized with 10ml of Hybridization buffer **(Appendix C)** at 42°C for 4hrs with constant agitation in hybridization oven (Hybaid, England). End labeled SSO (0.5 p moles per ml) was added to the hybridization buffer and incubated for 16 hrs. with constant agitation in the oven at 42°C. The names, numbers and sequences of the appropriate probes for typing different loci are given in **Appendix G.**

The membrane was rinsed in 100ml of 2X SSPE, 0.1 % SDS (Cold wash, **Appendix C**) at room temperature for 10 min, twice.

The filter was then washed in 100ml of 6X SSPE, 1% SDS (critical wash, **Appendix C**) for 10 min at Td twice. The dissociation temperature Td was calculated as

Td = 4 x (no. of G and C in SSO) + 2 x (no. of A and T in SSO) .

The critical washing was necessary to get rid of all nonspecific probe template binding and to let only the specific interaction remain. The Td had to be slightly adjusted from the calculated value from probe to probe to get the optimum results (**Table 4. 4 to 4.6).**

Table 4.4 Critical wash temperature for different Probes used for DRB1 locus.

Probe No.	Td (Critical Temp. ^{0}C)	
	Calculated	Used
2801	58	58
1002	56	56
1008N	52	51.5
7004	62	62
5703	56	55.8
1003	56	55.8
1010	56	55.8
1011	54	53.8
1004	52	51.5
1006	50	49.8
1007	50	49.5
2810	56	56
DRB12	62	61.5
DRB8	64	63.5
DRB6	62	61.8
7031	64	63
DRB14/1	62	61.8
DRB13	62	61.8
7012	66	65.5
2501	56	55.6
2503	56	55
3401	54	53.7

Table 4.5 Critical wash temperature for different Probes used for DQA1 locus.

Probe No.	Td (Critical Temp. ^{0}C)	
	Calculated	Used
3402	54	53.5
3403	54	53.7
4101W	56	55
4102	54	53
4103W	58	57
5501	50	49.5
5502W	56	55.6
5503	50	49.5
6901	54	53.5
6902	50	49
6903	58	57
6904	52	51
7502	48	47
7504	48	47.5
2301	64	63.7
2302	62	61.7
2601	60	59.6
2603	54	53.7
2604	54	53.7
2605W	56	55.6
2606	52	51.7
2607	54	53.6
3701	58	57.6
3702	60	59.6
3703W	52	51.6
3704W	60	59.6

Table 4.6 Critical wash temperature for different Probes used for DQB1 locus.

Probe No.	Td (Critical Temp. ^{0}C)	
	Calculated	Used
3705W	62	61.7
4501	60	59.6
4901	60	59.7
4902W	68	67.7
5701	60	59.7
5702	62	61.7
5703	62	61.7
5704	60	59.7
5705	64	63.7
5706	64	63.7
5707	68	67.6
5708	60	59.7
7001W	66	65.7
7002	60	59.7
7004W	64	63.7
7005	62	61.7
7006	64	63.7
7007	66	65.6
7008	62	61.7
7009	60	59.7
7010	60	59.7

Detection of hybridized probes:

The hybridized probes were detected by Autoradiography. Briefly, the membrane was placed on a Whatman 3 MM filter paper to remove excess liquid. The

membrane was covered in a Saran wrap. It was exposed to X-ray film in a cassette with an intensifying screen. The cassette was kept at -80°C for 5 hrs to overnight depending on the intensity of the signals from the labelled probe as indicated by the count taken on the β counter. The X-ray film was developed and signals were noted (**Figure 4.5** and record sheet 'Probing results'. **Appendix H**).

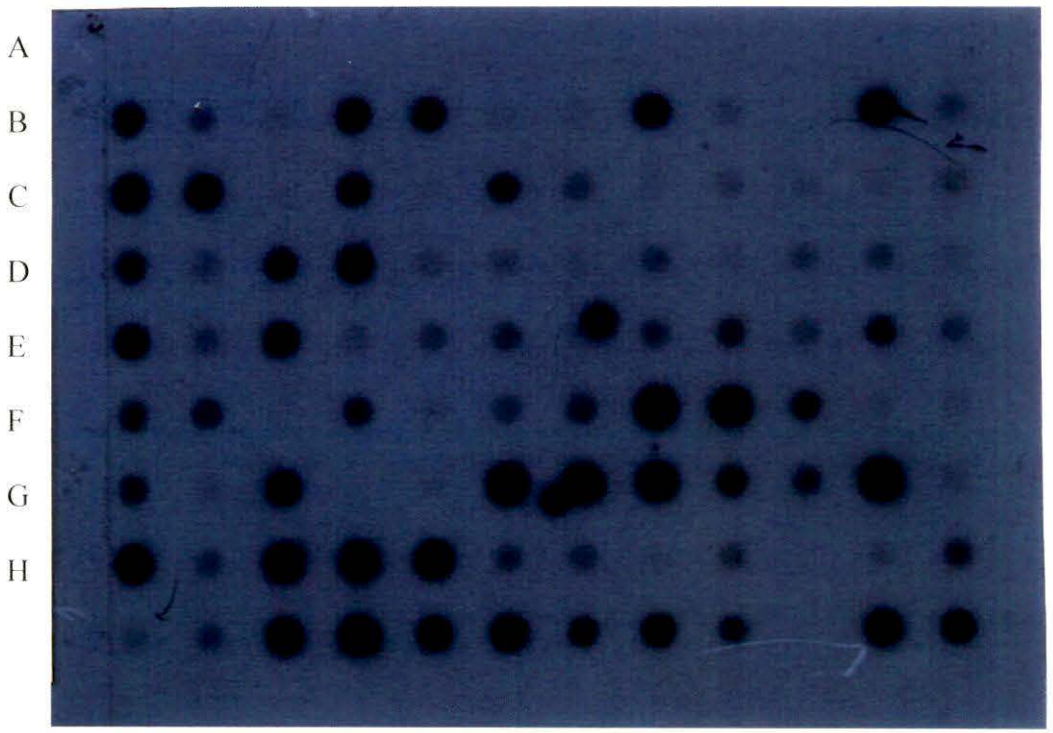


Figure 4.5: A representative autoradiogram showing the positive control and test positive dots

Positive Control	:	H3, H4, H5, H6, H8		
Negative Control	:	H1, H2, H7, H9, H10		
Probe Used	:	3401	**Locus** :	DQA1
Blot No	:	Ky DQA1	**Population** :	Kayastha 1-84

Deprobing / dehybridisation of Probes

As the number of probes ranged from 16 to 35 for the typing of a single locus and we made only 3 copies of each blot, we had to dehybridise the probes from the membranes after hybridisation results were obtained from one probe. The dehybridised membranes were rehybridised with a new probe. For dehybridisation the membranes were rinsed briefly in sterile distilled water followed by washing in 0.2 N NaOH, 0.1% SDS (Deprobing solution, **Appendix C**) at 37 °C for 15 min, twice. The membranes were washed for 30min in 2X SSC **(Appendix C)** at room temperature with shaking. Excess fluid was removed with the help of a filter paper. The membrane was subjected to autoradiography to confirm that the probe was removed completely, so that it was ready for rehybridisation. The membrane was covered with Saran wrap and stored at 4 ° C until use.

Interpretation of Signals

For every SSO, known positive and negative controls were included. Any spot stronger than the negative control was considered positive. When the difference between positive and negative signals were very small, an additional critical wash with a slightly increased temperature was carried out, so as to make the difference between the signals of positive and negative controls more meaningful. In general, for each SSO, the exposure time was such that the negative control on the film shows a very weak spot or no spot at all. The intensity of the spots was represented by grades as follows:

0 Not spotted

1 Negative (definite)

2 Negative (probably)

4 Indefinite

6 Positive (probably)

8 Positive (definite)

9 Positive (definite, more than double intensity).

Individual probes were scored separately (**Appendix H**).

When probing for all essential probes required for the assignment of the particular alleles were completed, the assignment of HLA class II alleles were performed using the interpretation tables shown in **Appendix I.**

Statistical analysis

Allele frequency

Allele frequency was obtained by direct counting method from the observed number of alleles at a locus divided by the number of gametes. The missing values were excluded from such estimation.

Genotypic Frequency

Genotypic Frequency (GF) was obtained from the observed number of a given genotype at each locus. Missing values were excluded from such estimation

Haplotype frequency

The 2 locus and 3 locus haplotype frequencies for the HLA system were computed from genotypic data using the maximum likelihood method.

Heterozygosity

We used heterozygosity as a measure of population diversity. Various estimates were calculated each having its own significance.

Observed Heterozygosity

Observed Heterozygosity was obtained by direct counting method according to **(Levene, 1949)**

$$H_{obs} = \frac{(H_1+H_2+\text{----------}+H_n)/2}{\text{Total number of samples}}$$

where H_1, H_2------- H_n are the number of heterozygotes for alleles

1,2,3----------n at a locus

Expected Heretrozygosities under Hardy-Weinberg Equilibrium

It is defined as the probability that two random alleles taken from the sample are different.

$$H = 1 - \sum_{i=1}^{k} p_i^2$$

where k is the number of alleles

and p_i is the sample frequency of the i th allele at a locus .

Nei's unbiased heterozygosity was calculated as given by

$$H = \frac{n}{n-1}\left(1 - \sum_{i=1}^{k} p_i^2\right)$$

where k = no of alleles, n = no of individuals in the sample

Test of Hardy-Weinberg Equilibrium

We followed several approaches to detect deviation from Hardy-Weinberg equilibrium.

All the tests were performed using both

1. by analyzing all genotypes and
2. by pooling genotypes into 3 classes
   a) homozygotes for the most common allele .
   b) heterozygotes for the most common allele .
   c) all other genotypes, applying the following tests.

Chi-square and G^2 tests

In most of the past studies the Goodness of fit Tests such as Chi-square or G square test were usually employed, when determining the difference between the observed genotypic frequencies and those expected under Hardy-Weinberg Equilibrium. Such tests were often found inaccurate when one or more genotypes had low expected counts and so we used a pooling procedure as discussed above because it increases the expected counts to a point where the bias is removed. But this

pooling is a poor solution as much of the information is lost during the process. For these reasons the use of G^2 or χ^2 (on either pooled or all genotypes) is recommended only for preliminary analysis and we had also used it for the same.

Computer program POPGENE was used to compute the genotypic frequencies under random mating using the algorithm by **Levene (1949).**

Exact Tests

Haldane (1954) described the use of an exact test in 1954 and this forms the basis of our test. The procedure described in **Guo and Thompson (1992)** using a test analogous to Fisher's exact test on a 2x2 contingency table; but extended to a triangular contingency table of arbitrary size was used. The test was done using a modified version of the Markowitz Random walk Algorithm described by **Guo and Thompson (1992).** The modified version is more efficient from a computational point of view.

The test assumes that the allele frequencies are given. A contingency table is first built. The k x k entries of the table are the observed allelic frequencies and k is the number of alleles . The probability of no association (null hypothesis) is given by **Levene (1949)** .

$$Lo = \frac{n! \prod_{k} n!}{(2n)! \prod_{i=1}^{k} \prod_{j=1}^{i} n_{ij}!} 2^{H}$$

where H is the number of heterozygote individuals.

The computer then explores alternative contingency tables having same marginal counts. In order to create a new contingency table from a existing one, 2 distinct lines i_1, i_2 and 2 distinct columns j_1, j_2 at random are selected. The new table is obtained by decreasing the counts of the (i_1,j_1) (i_2,j_2) and increasing the counts of the cells (i_1,j_2) (i_2,j_1) by one unit This leaves the allele counts n_i unchanged . The switch to the new table is accepted with a probability R :

$$R = \frac{L_{n+1}}{L_n} = \frac{n_{i1\,j1}\ n_{i2\,j2}}{(n_{i1\,j2}+1)(n_{i2\,j1}+1)} \cdot \frac{(1+\delta_{i1\,j1})(1+\delta_{i2\,j2})}{(1+\delta_{i1\,j2})(1+\delta_{i2\,j1})}, \text{ if } i_1 \neq j_1 \text{ or } i_2 \neq j_2$$

$$R = \frac{L_{n+1}}{L_n} = \frac{n_{i1\,j1}\ n_{i2\,j2}}{(n_{i1\,j2}+1)(n_{i2\,j1}+2)} \cdot \frac{1}{4}, \text{ if } i_1 = j_1 \text{ and } i_2 = j_2$$

$$R = \frac{L_{n+1}}{L_n} = \frac{n_{i1\,j1}\ (n_{i2\,j2}-1)}{(n_{i1\,j2}+1)(n_{i2\,j1}+1)} \cdot \frac{1}{4}, \text{ if } i_1 = j_2 \text{ or } i_2 = j_1$$

δ denotes the Kronecker function . R is the ratio of the probabilities of the two tables. The switch to the new table is accepted if R>1

The P value of the test is the proportion of the visited tables having a probability smaller or equal to observed (initial) contingency table. The standard error on the P value is estimated using a system of batches.

F statistics

F_{IS} *the inbreeding coefficient of a group of inbred organisms relative to the* subpopulation to which they belong, which measure the reduction in heterozygosity of the inbred organisms and is given by

$$F_{IS} = \frac{H_S - H_I}{\bar{H}_S}$$

where $\bar{H}s$ average observed heterozygosity

H_I average expected heterozygosity in the subpopulations

F_{IT}, the reduction in heterozygosity of the inbred organisms with respect to the total population is given by

$$F_{IT} = \frac{H_T - H_I}{H_T}$$

where H_T : observed heterozygosity of the total population

F_{ST} the reduction in heterozygosity attributable to the mates within subpopulations sharing some common ancestor with respect to the total population is given by

$$F_{ST} = \frac{H_T - H_S}{H_T}$$

The above measures were calculated with the program POPGENE

Gene Diversity Analysis

Nei (1973) defined gene diversity for a single locus as the heterozygosity expected under Hardy Weinberg equilibrium, disregarding the actual genotype frequencies in the population. With this definition, it was shown by Nei that

$H_T = H_S + D_{ST}$.

where D_{ST} is the inter population gene diversity,

H_T is the average expected heterozygosity in the total population and H_S is the average expected heterozygosity in the subpopulations.

$$H_T = H_S + D_{ST}$$

$$\Rightarrow \quad 1 = H_S / H_T + D_{ST} / H_T$$

$$\Rightarrow \quad D_{ST} / H_T = 1 - H_S / H_T$$

$$\Rightarrow \quad G_{ST} = 1 - H_S / H_T$$

G_{ST} is called the co-efficient of gene differentiation. We had calculated H_S, H_T and G_{ST} from average expected heterozygosities at the loci studied and their mean values over loci using the above formula. G_{ST} denotes the extent of gene diversity between populations with respect to the total population and H_S is the gene diversity attributable to that between individuals within subpopulations.

Observed Homozygosity

The proportion of observed homozygotes at a given locus was estimated as

$$\text{Observed Homozygosity} = \frac{\text{Total no of homozygotes}}{\text{No of samples}}$$

Expected Homozygosity

The proportion of expected homozygotes under random mating.

$$F = \sum_{i=1}^{k} p_i^2$$

Where p_i = frequency of the i th allele

k = no. of alleles

Effective allele number

It was measured as the reciprocal of homozygosity.

$$ne = \frac{1}{\sum_{i=1}^{k} p_i^2}$$

n_e or effective no. of alleles is the number of equally frequent alleles that would be required to produce the same homozygosity as observed in an actual population

If a population contains n_e equally frequent alleles

Then $p_1 = p_2 = p_3 \ldots\ldots\ldots\ldots\ldots p_n = 1 / ne$

Expected homozygosity

$$= \sum_{i=1}^{k} p\,i^2$$

$$= \sum_{i=1}^{k} \frac{1}{n_e^2}$$

$$= \frac{n_e}{n_e^2}$$

$$= \frac{1}{n_e}$$

$$\Rightarrow \quad n_e = \frac{1}{\text{Expected Homozygosity}}$$

$$= \frac{1}{\sum_{i=1}^{k} p_i^2}$$

Genetic Distances

With the help of the computer program PHYLIP version 3.5c, subprogram Gendist, Nei's (1972) genetic distance was computed from allele frequency data of our populations and other world populations (wherever available and for as many loci as available) .

The measure assumes that all differences between populations arise from genetic drift. Nei's distance is formulated for an infinite isoalleles model of mutation, in which there is a rate of neutral mutation and each mutant is considered to be a completely new allele. It is assumed that all loci have the same rate of neutral mutation, and that the genetic variability in the population is at equilibrium between mutation and genetic drift, with the effective population size of each population remaining constant.

Nei' s distance is given by

$$D = -\ln\left(\frac{\sum_m \sum_i p_{1mi}\, p_{2mi}}{\left[\sum_m \sum_i p_{1mi}^2\right]^{1/2} \left[\sum_m \sum_I p_{2mi}^2\right]^{1/2}} \right)$$

where m is summation over loci, *i* over alleles at the m th locus and where, p_{1mi} is the frequency of the *i* th allele at the m th locus in population 1 . Subject to the above assumptions, Nei's genetic distance is expected, for a sample of sufficient number of equivalent loci, to rise linearly with time. This measure has been observed to work reasonably well in case of closely related groups within species.

Dendrogram

Phylogenetic trees were constructed based on Nei's measures of genetic distance, using the subprogram Neighbor, of PHYLIP, version 3.5c which generates both Neighbor -Joining and UPGMA trees.

Neighbor Joining trees

The program uses the Neighbor Joining method of **Saitou and Nei (1987)** and constructs a tree by successive clustering of lineages setting branch lengths as the lineages join. The tree does not assume molecular clock, so that in effect it is an unrooted tree.

Unpaired Group Mean Average

The unpaired groups mean average (UPGMA) option constructs a tree by successive (agglomerative) clustering using an average linkage method of clustering.

Gene Flow

To assess the relative amount of gene flow experienced by each population, we have used a regression model originally proposed by **Harpending and Ward (1982).** For this, the heterozygosity of the *i* th population was plotted against the distance of the population from the centroid (r_i) calculated as :

$$r_i = (p_i - P)^2 / P(1 - P)$$

where pi and P are, respectively the frequency of the i th allele in population i and the total population respectively. Under the island model of population structure, Harpending and Ward have shown that there should exist a linear relationship between heterozygosity and distance from the centroid:

$$h_i = H(1 - r_i) ,$$

where h_i and H denote, respectively, the heterozygosities of population i and the total population. Of particular interest in this analysis were the outliers. Populations that have experienced more gene flow than average will have higher heterozygosities than predicted, whilst those that have experienced less gene flow than average will have

lower heterozygosities than predicted. The observed heterozygosities of the three populations were plotted against their distances from the gene frequency Centroid using the program SPSS (ver 8.0).

Ewen- Watterson Neutrality Test

To determine the type of selection (neutrality, balancing selection, directive selection) acting on the loci under study, the homozygosity statistic F (**Manly, 1985**) was calculated as

$$F = \sum_{n=1}^{k} p_i^2$$

where p_i is the frequency of the *i* th allele (i = 1,2,3,.....kth allele) in a panel of n individuals and summation is over k alleles .

The observed frequency distribution of alleles (observed F) was compared to that expected under the neutral marker hypothesis in which mutation and random drift are the only forces altering allele frequencies. The POPGENE program was used to perform the test using the algorithm given by **Manly, 1985.**

The null hypothesis of selective neutrality was tested against selection using a bilateral test at a 5% confidence level.

Observations

Observations

Present study was carried out on different ethnic groups from Uttar Pradesh (North India) and from North Eastern part of India. The populations selected by us from Uttar Pradesh are four endogamous caste groups i.e. *Kayastha, Mathurs, Rastogies* and *Vaish*, two consanguineous groups i.e. *Shia*, and *Sunni* Muslims. From North Eastern part of the country we have selected one caste group i.e. *Rajbanshi* and two tribes of Tibeto-Burman group, these are *Lachung* and *Mech*. High resolution HLA Class II typing i.e. HLA-DRB1, DQA1, and DQB1 was carried out by using Sequence Specific Oligonucleotide probe (SSOP) based typing.

Total number of individuals enrolled for the present study was 1336. The detailed distribution and demographic profile is shown in **Table 4.1**. Under the observation section the data has been presented under four headings i.e. Allele frequency distribution, haplotypes frequency distribution, inter and intra group variation and phylogenetic assessment.

All the caste groups i.e. *Kayastha, Mathurs, Rastogies* and *Vaish* have been clubbed together for showing the allele frequency distribution for different caste groups of Uttar Pradesh. As per the historical records *Kayastha* and *Vaish* are two broader groups and *Mathurs* have been derived from *Kayastha* while *Rastogies* from *Vaish*. In the present study we have made an attempt to prove this hypothesis on the basis of genetic analysis.

Allele Frequency distribution at DRB1 locus in North Indian caste groups

Allele frequency distribution of *Kayastha, Mathurs, Rastogies*, and *Vaish* is shown in **Table 5.1.** Total number of alleles was 36 in North Indian caste groups. Out of 36 alleles 30 alleles were observed in *Kayastha* and *Mathurs*, 27 alleles in *Vaish* and 28 alleles in *Rastogies*.

Table 5.1 Allele Frequency distribution at DRB1 Locus

DRB1	Frequency %							
	n	Kayastha (2n=380)	n	Vaish (2n=310)	n	Mathur (2n=396)	n	Rastogi (2n=392)
DRB1*0101	10	2.60	15	4.80	23	5.80	30	7.65
DRB1*0301	34	8.90	29	9.40	30	7.57	36	9.20
DRB1* 0302	1	0.26	0	0	0	0	6	1.53
DRB1*0401	6	1.57	4	1.30	4	1.00	4	1.00
DRB1*0402	0	0	4	1.30	3	0.76	5	1.30
DRB1*0403	5	1.30	8	2.60	9	2.27	0	0
DRB1*0404	5	1.30	1	0.32	7	1.76	6	1.50
DRB1*0405	2	0.52	0	0	1	0.25	0	0
DRB1*0406	1	0.26	1	0.32	0	0	0	0
DRB1*0407	0	0	2	0.65	1	0.30	0	0
DRB1*0701	80	21.50	74	23.9	102	25.70	77	19.60
DRB1*080X	1	0.26	0	0	1	0.25	4	1.00
DRB1*090X	11	2.80	4	1.30	7	1.76	15	3.80
DRB1*1001	28	7.36	32	10.30	26	6.50	36	9.20
DRB1*1101	24	6.30	15	4.80	18	4.50	20	5.10
DRB1*1102	0	0	2	0.65	2	0.50	1	0.30
DRB1*1103	5	1.31	2	0.65	1	0.25	2	0.50
DRB1*1104	1	0.26	1	0.32	2	0.50	0	0
DRB1*1106	1	0.26	0	0	0	0	0	0
DRB1*1107	0	0	0	0	0	0	1	0.30
DRB1*1108	0	0	0	0	1	0.25	0	0
DRB1*1201	2	0.52	2	0.65	5	1.26	2	0.50
DRB1*1202	11	2.89	1	0.32	8	2.00	9	2.30
DRB1*1203	0	0	0	0	0	0	1	0.30
DRB1*1301	33	8.60	17	5.40	25	6.30	20	5.10
DRB1*1302	9	2.36	6	2.00	12	3.03	7	1.80
DRB1*1303	6	1.57	1	0.30	3	0.80	9	2.30
DRB1*1308	1	0.26	0	0	1	0.25	1	0.30
DRB1*1313	2	0.52	1	0.30	0	0	0	0
DRB1*1401	19	5.00	13	4.30	16	4.00	13	3.30
DRB1*1402	4	1.05	2	0.65	7	1.70	11	2.80
DRB1*1403	4	1.05	0	0	3	0.80	3	0.80
DRB1*1501	37	9.70	26	8.40	34	8.6	39	9.90
DRB1*1502	26	6.80	34	10.90	34	8.6	24	6.10
DRB1*1503	8	2.10	9	3.00	7	1.7	7	1.80
DRB1*1504	3	0.78	4	1.30	3	0.8	3	0.80

Our results revealed that most of the alleles were equally distributed in all the caste groups. However, DRB1*0701 was observed at higher frequencies in all the four North Indian caste populations ranging from 19.6% to 25.7%. At allelic level we have observed that DRB1*0101 was observed at a higher frequency in *Rastogies* (7.6%) while it was found to be less frequent in *Kayastha* (2.6%), *Vaish* (4.8%) and *Mathurs* (5.8%). In *Mathur* and *Vaish* allele DRB1*0302 was absent, however, it was observed with a frequency of 0.26% to 1.53% in *Kayastha* and *Rastogies.* Allele DRB1*0405 was observed only in *Kayastha* (0.52%) and *Mathurs* (0.25%). DRB1*0407 was observed with a frequency of 0.65% in *Vaish* and 0.3% in *Mathurs* which is quite low. However, it was absent in *Kayastha* and *Rastogies.* DRB1*1102 was not observed in *Kayastha,* while DRB1*1313 was observed only in *Kayastha* and *Vaish* but with a low frequency (0.3-0.5%). Some of the alleles were differentially distributed like DRB1*0402 was absent only in *Kayastha,* DRB1*0403 and DRB1*1104 were absent in *Rastogies,* DRB1*080x, DRB1*1308 and DRB1*1403 were absent in *Vaish* but present in other caste groups.

Most of the other the alleles were seen almost with a equal frequency in all the four caste groups. DRB1*0701 and DRB1*1501 were the most frequent alleles observed in these groups. We further compared these Caste groups with one another by applying Chi Square test, which has been discussed separately.

Allele Frequency distribution at DQA1 locus in North Indian caste groups

At DQA1 locus 17 alleles were observed of which 12 alleles were detected in *Kayastha* and *Vaish,* 15 alleles in *Rastogies* and 11 alleles in *Mathurs.* DQA1*0102, DQA*0103, DQA1*0201, DQA1*0301 and DQA1*0501 were observed at a higher frequencies in all the four North Indian caste groups ranging from 9.2% to 27.6%. Some differences were observed like DQA1*0502 was absent in *Kayastha and Mathurs,* DQA1*0503 was absent in *Mathurs.* DQA1*0303, DQA1*0304, DQA1*0604 alleles were seen only in *Rastogies.* DQA1*0402 was a rare allele found only in *Kayastha.* DQA1*0603 was unique for *Mathur.* Some other differences were also observed in between these groups like frequency of allele DQA1*0102 in *Kayastha and Rastogies* was 9.2% and 9.6%, among *Mathurs* and *Vaish* the frequency of DQA1*0102 was relatively higher i.e. in *Vaish* and *Mathurs*

13.6% and 15.1% respectively. *Kayastha* revealed differences at DQA1*0104 as in *Kayastha* its frequency was only 4.6% while among *Vaish, Mathur* and *Rastogies* it was more than double i.e. 13.6%, 15.1% and 10.2% respectively.

Some alleles were commonly observed with a higher frequency these were DQA1*0103, DQA1*0201, DQA1*0301, DQA1*0501. Distribution of allele frequencies is shown in **Table 5.2**

Table 5.2 Allele frequency distribution at DQA1

DQA1	Frequency %							
	n	Kayastha (2n=380)	n	Vaish (2n=310)	n	Mathur (2n=396)	n	Rastogi (2n=392)
DQA1*0101	29	7.50	15	4.70	19	4.79	29	7.39
DQA1*0102	35	9.20	42	13.60	62	15.60	38	9.69
DQA1*0103	61	16.09	47	15.20	44	11.10	58	14.79
DQA1*0104	18	4.60	42	13.60	60	15.15	40	10.20
DQA1*0201	104	27.60	77	24.00	83	20.90	93	23.70
DQA1*0301	59	15.50	40	12.90	40	10.10	58	14.79
DQA1*0302	2	0.53	2	0.65	2	0.50	1	0.25
DQA1*0303	0	0	0	0	0	0	2	0.50
DQA1*0304	0	0	0	0	0	0	1	0.25
DQA1*0401	9	2.36	3	0.90	6	1.50	6	1.50
DQA1*0402	2	0.53	0	0	0	0	0	0
DQA1*0501	53	13.90	35	11.40	66	16.60	48	12.24
DQA1*0502	0	0	1	0.30	0	0	3	0.76
DQA1*0503	2	0.53	1	0.30	0	0	3	0.76
DQA1*0601	6	1.60	5	1.60	13	3.30	11	2.80
DQA1*0603	0	0	0	0	1	0.25	0	0
DQA1*0604	0	0	0	0	0	0	1	0.25

Allele Frequency distribution at DQB1 locus in North Indian caste groups

At DQB1 locus 19 alleles were observed in North Indian populations but when individual populations were considered it was found that among *Kayastha* total number of alleles found were 19, in *Vaish* 15, in *Rastogies* 15 while in *Mathurs* 17 alleles were present. Most common alleles in the four North Indian Caste populations were DQB1*0201, DQB1*0301, DQB1*0303, DQB1*0501, and DQB1*0601. Some major differences were seen for example DQB1*0605 was not represented in *Vaish*, but was found with a low frequency in other three groups i.e. in *Kayastha* 0.53%, in

Rastogies 0.51%, and in *Mathurs* 0.7%. DQB1*0606 was also absent in *Vaish* and *Rastogies*. DQB1*0304 and DQB1*0506 were rare alleles found only in *Kayastha* with a frequency of 0.5%. It is observed that allele DQB1*0301 was found to be raised in *Kayastha* (13.6%) and *Mathurs* (15.1%) while it was comparatively low in *Mathurs* (9.3%) and *Rastogies* (9.6%) Alleles which were commonly observed with high frequency in all the four caste groups were DQB1*0201, DQB1*0303, DQB1*0501, DQB1 *0601. The allele frequency distribution is shown in **Table 5.3.**

Table 5.3 Allele frequency distribution at DQB1 Loci

Allele DQB1	Frequency %							
	n	Kayastha (2n=380)	n	Vaish (2n=310)	n	Mathur (2n=396)	n	Rastogi (392)
DQB1*0201	91	23.9	74	23.8	69	17.4	104	26.5
DQB1*0301	52	13.68	29	9.35	60	15.15	38	9.69
DQB1*0302	13	3.4	17	5.48	11	2.77	12	3.06
DQB1*0303	39	10.26	26	8.38	37	9.3	42	10.7
DQB1*0304	2	0.53	0	0	0	0	0	0
DQB1*0305	2	0.53	2	0.64	3	0.75	1	0.25
DQB1*0401	6	1.57	1	0.32	3	0.75	0	0
DQB1*0402	2	0.53	4	1.29	8	2.02	3	0.8
DQB1*0501	37	9.73	45	14.5	71	17.9	48	12.2
DQB1*0502	1	0.26	3	0.96	6	1.5	6	1.5
DQB1*0503	24	6.3	20	6.45	30	7.6	33	8.3
DQB1*0504	2	0.53	1	0.32	6	1.5	2	0.51
DQB1*0506	2	0.53	0	0	0	0	0	0
DQB1*0601	59	15.5	60	19.3	41	10.35	63	16.07
DQB1*0602	15	3.94	8	2.58	16	4	12	3.06
DQB1*0603	30	7.89	15	4.83	18	4.5	16	4.08
DQB1*0604	2	0.53	5	1.6	8	2	10	2.55
DQB1*0605	2	0.53	0	0	3	0.7	2	0.51
DQB1*0606	3	0.78	0	0	6	1.5	0	0

Distribution of HLA Class II antigens in Shia and Sunni Muslims

Allele Frequency distribution at DRB1 locus in Muslim group

The allele frequency distribution in two *Muslim* populations i.e. *Shia* and *Sunni* are shown in **Table 5.4.**

Table 5.4 Allele frequency distribution at DRB1 Loci

DRB1	Frequency %			
	n	Shia (2n=380)	n	Sunni (2n=376)
DRB1*0101	61	16.00	44	11.70
DRB1*0103	7	1.84	0	0
DRB1*0301	58	15.30	38	10.10
DRB1*0303	0	0	3	0.79
DRB1*0401	5	1.31	3	0.79
DRB1*0402	6	1.57	1	0.27
DRB1*0403	3	0.78	6	1.60
DRB1*0404	3	0.78	3	0.80
DRB1*0405	1	0.26	2	0.53
DRB1*0407	1	0.26	2	0.53
DRB1*0701	56	14.70	60	15.90
DRB1*080X	2	0.50	3	0.79
DRB1*090X	14	3.68	5	1.30
DRB1*1001	30	7.90	34	9.00
DRB1*1101	25	6.60	37	9.80
DRB1*1102	5	1.30	1	0.27
DRB1*1103	6	1.60	2	0.53
DRB1*1104	0	0	1	0.27
DRB1*1105	5	1.30	2	0.53
DRB1*1107	16	4.20	13	3.40
DRB1*1108	1	0.26	0	0
DRB1*1201	1	0.26	14	3.70
DRB1*1202	6	1.60	11	2.90
DRB1*1203	0	0	1	0.27
DRB1*1301	9	2.40	14	3.70
DRB1*1302	6	1.60	4	1.00
DRB1*1303	5	1.30	10	2.60
DRB1*1313	0	0	1	0.27
DRB1*1401	9	2.40	22	5.80
DRB1*1402	7	1.80	3	0.80
DRB1*1403	1	0.26	5	1.30
DRB1*1404	1	0.26	0	0
DRB1*1501	18	4.70	13	3.40
DRB1*1502	8	2.10	13	3.40
DRB1*1503	2	0.50	4	1.00
DRB1*1504	2	0.50	1	0.27

In this group only 36 DRB1 alleles have been identified. Among *Shia's* there were 32 alleles while in *Sunni's* 33 alleles were observed. Some of the DQB1 alleles were observed in both Shia and Sunni Muslims. These were DRB1*0101, DRB1*0301, DRB1*0701 were seen in both the Muslim populations. DRB1*0103 and DRB1*1404 were found only in *Shia's* and alleles DRB1*0303, DRB1*1104, DRB1*1203 and DRB1*1313, were observed only in *Sunni's.* When we closely examined allele frequency distribution in North Indians we found that DRB1*1203 was observed only in *Sunni Muslims* and also in one of the endogamous caste population i.e. *Rastogies.* DRB1*0103 and DRB1*0303 were only seen in *Shia* and *Sunni's.* Alleles which were commonly observed in both the *Muslim* populations were DRB1*0101, DRB1*0301 and DRB1*0701.

Allele Frequency distribution at DQA1 locus in Muslim group

At DQA1 locus 17 alleles were identified, out of 17 alleles 16 were observed in *Shia* while only 12 in *Sunni's.* DQA1*0103, DQA1*0201, DQA1*0301, DQA1*0501 and DQA1*0601 were common alleles which were also found in the other caste groups. *Shia's* and *Sunni's* showed differences at some DQA1 loci like DQA1*0504 was found only in *Sunni's* and DQA1*0602 in *Shia's*, interestingly these alleles were not observed in other studied populations. Some of the other differences were also observed between the two populations like frequency of DQA1*0101 was high in *Shia's* (16%) as compared to *Sunni's* (9.3%). When DQA1*0102 allele was compared it was found that in *Shia* it was 2.9% while *Sunni's* revealed higher frequency of this allele i.e. 7.9%. Differences were also observed at DQA1*0401 and DQA1*0502 allele. The detailed allele frequency distribution is shown in **Table 5.5**

Table 5.5 Allele frequencies at DQA1 Loci

Allele DQA1	Frequency %			
	n	Shia (2n=380)	n	Sunni (2n=376)
DQA1*0101	60	16.00	35	9.30
DQA1*0102	11	2.90	30	7.90
DQA1*0103	56	14.70	47	12.50
DQA1*0104	21	5.50	39	10.30
DQA1*0201	71	18.70	68	18.09
DQA1*0301	36	9.50	42	11.20
DQA1*0302	2	0.53	0	0
DQA1*0303	1	0.26	2	0.53
DQA1*0304	1	0.26	0	0
DQA1*0401	14	3.70	7	1.86
DQA1*0501	49	12.90	63	16.70
DQA1*0502	10	2.60	19	5
DQA1*0503	2	0.53	0	0
DQA1*0504	0	0	1	0.23
DQA1*0601	43	11.30	23	6.10
DQA1*0602	2	0.53	0	0
DQA1*0604	1	0.26	0	0

Allele Frequency distribution at DQB1 locus in Muslim group

At DQB1 locus 20 alleles were found. Twenty alleles were observed in *Shia* while only 18 alleles were seen in *Sunni's*. Some alleles like DQB1*0607 and DQB1*0608 were found to be unique for *Shia's*. These alleles were not found in any of the other caste groups and also in *Sunni's*. DQB1*0604 was seen with less frequency in both *Shia* (0.78%) and *Sunni* (0.8%) while they were common among other caste groups. Allele DQB1*0304 was present in both *Shia's* and *Sunni's* however, with less frequency, this particular allele was absent in *Mathur, Vaish* and *Rastogies* caste groups but present in *Kayastha's*. DQB1*0503 was found with a less frequency (2.3%) among *Shia* while it was high in *Sunni's* (8.2%). Rest of the alleles were almost similar in their frequency distribution. **Table 5.6**

Table 5.6 Allele frequency at DQB1 loci

DQB1	Frequency %			
	n	Shia (2n=380)	n	Sunni (2n=376)
DQB1*0201	81	21.30	74	19.68
DQB1*0301	54	14.20	50	13.30
DQB1*0302	18	4.73	11	2.90
DQB1*0303	36	9.50	29	7.70
DQB1*0304	1	0.26	2	0.53
DQB1*0305	9	2.36	10	2.60
DQB1*0401	8	2.10	3	0.80
DQB1*0402	9	2.37	11	2.90
DQB1*0501	58	15.20	65	17.30
DQB1*0502	3	0.78	5	1.30
DQB1*0503	9	2.37	31	8.20
DQB1*0504	17	4.47	8	2.10
DQB1*0601	44	11.57	51	13.50
DQB1*0602	9	2.36	5	1.30
DQB1*0603	13	3.40	13	3.45
DQB1*0604	3	0.78	3	0.80
DQB1*0605	1	0.26	3	0.80
DQB1*0606	5	1.30	2	0.53
DQB1*0607	1	0.26	0	0
DQB1*0608	1	0.26	0	0

Distribution of HLA Class 11 antigen allele frequency among North Eastern caste and tribal groups

Three North Eastern caste and tribal populations from different parts of North Eastern Regions of India were studied. The allele frequency distribution at DRB1, DQA1, DQB1is shown below.

Allele Frequency distribution at DRB1 locus in North Eastern group

At DRB1 locus, allele frequency distribution revealed some interesting results. In three populations of North Eastern region one was Indo Aryan speaking caste population i.e. *Rajbanshi's* while other two were Tibeto-Burman speaking groups. These were *Lachung* and *Mech.* There were total 21 DRB1 alleles. **(Table 5.7)** DRB1*0701, DRB1*1101, DRB1*1501 and DRB1*080x were frequently observed

in all the three populations. DRB1*0402 was absent in Lachung but present in *Rajbanshi's* and *Mech*. DRB1*1104 was present in both *Lachung* and *Mech* while not represented in *Rajbanshi's*. DRB1*1202 was not observed in *Lachung* and *Mech* but it was found in *Rajbanshi's* (1.0%). DRB1*0408 was found only in *Lachung* with low frequency.

Table 5.7 Allele frequency of DRB1

Allele DRB1	Frequency %					
	n	Lachung (2n=116)	n	Mech (2n=126)	n	Rajbanshi (2n=196)
DRB1*0101	2	1.70	2	1.60	6	3.06
DRB1*0301	5	4.30	7	5.50	11	5.60
DRB1*0401	4	3.40	4	3.20	6	3.06
DRB1*0402	0	0	2	1.60	1	0.50
DRB1*0408	2	1.70	0	0	0	0
DRB1*0701	16	13.80	15	11.90	27	13.90
DRB1*080X	9	7.70	14	11.10	18	9.20
DRB1*090x	12	10.30	11	8.70	6	3.06
DRB1*1001	6	5.20	6	4.76	11	5.70
DRB1*1101	13	11.20	12	9.50	19	9.70
DRB1*1103	1	0.80	1	0.80	2	1.03
DRB1*1104	2	1.70	1	0.80	0	0
DRB1*1201	9	7.70	11	8.70	20	10.20
DRB1*1202	0	0	0	0	2	1.00
DRB1*1301	2	1.70	2	1.60	4	2.04
DRB1*1302	3	2.60	2	1.60	7	3.60
DRB1*1401	6	5.20	7	5.50	14	7.14
DRB1*1404	5	4.30	4	3.20	4	2.04
DRB1*1405	2	1.70	2	1.60	4	2.04
DRB1*1501	11	9.50	17	13.50	27	13.70
DRB1*1502	6	5.20	6	4.70	7	3.60

It was found that allele 090x was found with high frequency in *Lachung* (10.3%) and *Mech* (8.7%), however, in *Rajbanshi's* it was comparatively less (3%). Some of the Common alleles were DRB1*0701, DRB1*1101, DRB1*1201, and DRB1*1501.

Allele Frequency distribution at DQA1 locus in North Eastern group

At DQA1 loci eleven alleles were observed in this group while in all the three populations i.e. *Lachung, Mech* and *Rajbanshi* 10 alleles were observed. DQA1*0102, DQA1*0201, DQA1*0301 were commonly observed in all the three populations. It was observed that DQA1*0402 was absent in *Rajbanshi* and at very low frequency in *Lachung* (0.8%) while high frequency of this allele was seen in *Mech* (13.5%). DQA1*0601 was totally absent in Mech but the other two populations were having very high frequency of this alleles (5.2% in *Lachung* and 8.7% in *Rajbanshi*). DQA1*0302 was missing in *Lachung*. In *Rajbanshi's* it was observed that at allele DQA1*0104 was found with a frequency of 7.7% while in *Lachung* it was 1.7% and in *Mech* it was 2.4%. DQA1*0501 was observed with a high frequency in *Lachung* and *Rajbanshi* (11.2% and 14.2%) **Table 5.8.**

Table 5.8 Allele frequency of DQA1

DQA1	Frequency %					
	n	Lachung (2n=116)	n	Mech (2n=126)	n	Rajbanshi (2n=196)
DQA1*0101	9	7.70	11	8.70	12	6.12
DQA1*0102	18	15.50	17	13.50	26	13.30
DQA1*0103	11	9.50	15	11.90	16	8.16
DQA1*0104	2	1.70	3	2.40	15	7.70
DQA1*0201	26	22.40	36	28.50	52	26.50
DQA1*0301	28	24.10	19	15.08	24	12.20
DQA1*0302	0	0	3	2.40	3	1.53
DQA1*0401	2	1.70	2	1.60	3	1.53
DQA1*0402	1	0.80	17	13.50	0	0
DQA1*0501	13	11.20	3	2.40	28	14.20
DQA1*0601	6	5.20	0	0	17	8.70

Allele Frequency distribution at DQB1 locus in North Eastern group

At DQB1 locus total 12 alleles were observed. There were 10 alleles among *Lachung* while in *Mech* there were only 9 alleles and in *Rajbanshi's* 11 alleles were observed. DQB1*0303 was the only allele which was observed with frequency of 2% to 5.1%. When compared to North Indian caste populations and Muslims. DQB1*0302 and DQB1*0401 were found in *Rajbanshi's* only. Allele DQB1*0603 was seen in *Lachung*, Frequency distribution at DQB1 loci for these tribal groups is given in **Table 5.9**

Table 5.9 Allele frequency of DQB1

DQB1	Frequency %					
	n	Lachung (2n=116)	n	Mech (2n=126)	n	Rajbanshi (2n=196)
DQB1*0201	22	19.00	27	21.40	32	16.30
DQB1*0301	27	23.30	33	26.20	46	23.40
DQB1*0302	0	0	0	0	3	1.50
DQB1*0303	6	5.17	4	3.17	4	2.04
DQB1*0401	0	0	0	0	2	1.03
DQB1*0402	1	0.90	2	1.60	2	1.03
DQB1*0501	20	17.20	21	16.67	57	29.08
DQB1*0502	8	6.90	9	7.10	5	2.55
DQB1*0503	3	2.60	1	0.80	3	1.50
DQB1*0601	16	13.80	19	15.08	37	19.00
DQB1*0602	11	9.50	10	7.90	5	2.55
DQB1*0603	2	1.70	0	0	0	0

Comparison of HLA class II allele frequencies among three groups:

Different alleles at HLA Class 11 loci were not equally distributed in all the studied populations i.e. caste groups from North India, Inbreeding Muslim populations from North India and one caste and two tribal populations from North Eastern parts of India. Chi -square analysis was done and only those alleles, which revealed significant differences among the studied populations, are shown in **Table 5.10**, Maximum differences were observed between different caste populations and Muslims. When the North Eastern caste and tribal groups were compared it was found that these groups were significantly different from Caste and Muslim populations of North India. It was observed that the frequencies of DRB1* 0101, DRB1*0301, DRB1*1201, DRB1*1301 and DRB1*1502 were significantly different in caste populations, and Muslim populations. However, the frequency of DRB1*0401, DRB1*080X, DRB1*090x, DRB1*1101, DRB1*1201, DRB1*1301, were significantly different in *Lachung, Mech, Rajbanshi's* of North Eastern region and caste populations of North India. Differences were also observed between *Shia* and *Sunni Muslims* at DRB1*0301, DRB1*0401 and DRB1*1201. When we compared the three North Eastern populations it was observed that *Rajbanshi's* differ with other two tribals i.e *Lachung* and *Mech* at 090x allele.

Table: -5.10 Chi Square analysis at DRB1 Loci

	DRB1	0101	0301	0401	080x	090x	1101	1201	1202	1301	1303	1502
Kayastha	Vaish								4.67*			
	Mathur	4.05*										
	Rastogi	5.42*								3.87*		
	Shia	37.7***	6.67**					24.38***		14.51**		8.89**
	Sunni	22.096***						9.32**		7.97**		
	Lachung			11.95**	12.068**	9.53**		21.43***		6.56*		
	Mech			12.78**	35.03***	6.41*		25.44***		7.40**		
	Rajbanshi			20.36***	24.52***			35.009***		9.49**		
Vaish	Rastogi											5.36*
	Shia	22.61***	4.78*									23.47***
	Sunni	11.34**						7.66**			4.49*	15.01**
	Lachung			14.11**				17.25***				9.33*
	Mech			14.90**		16.72***		20.63***				
	Rajbanshi			22.32***		12.77**		28.73***				8.82**
Mathur	Shia	20.23***	11.39**	6.12*						7.20**		15.91**
	Sunni	8.45**					8.17**	5.10*			8.87*	8.87**
	Lachung			19.54***	12.66**	18.47***	6.94**	13.89**		3.89*		6.46*
	Mech			20.41***	36.58***	13.91**	4.37*	17.73***		4.35*		
	Rajbanshi			28.49***	30.85***		4.94*	28.17***		5.14*		5.11*
Rastogi	Shia	16.49***	7.36**	4.1*								
	Sunni	6.02*				4.72*	5.61*	6.45*				
	Lachung			14.39**	9.99**	6.32**	4.53*	16.07***				
	Mech			15.29***	32.15***	3.86*		20.07***				
	Rajbanshi			23.49***	26.89***			30.59***				
Shia	Sunni		4.08*	5.91*				9.92**				
	Lachung	14.81**	8.65*	3.9*	12.08**	6.65**		21.62***				
	Mech	16.45***	7.12**	4.11*	34.62***	4.11*		25.75***				
	Rajbanshi	19.41***	10.52**	8.75**	28.91***			35.65**				
Sunni	Lachung	9.27**		18.15***	9.44**							
	Mech	10.41*		19.07***	30.68***			3.99*				
	Rajbanshi	11**		27.15***	25.61***			9.77**				
Lachung	Rajbanshi					8.69**						
Mech	Rajbanshi					6.29*						

P value<0.05*, P value<0.01* , P value<0.001*

At DQA1* locus χ^2 analysis (**Table 5.11**) revealed that Caste populations of North India did not reveal any significant differences between themselves for the common alleles like DQA1* 0101, DQA1*102, DQA1*0103, DQA1*0201, DQA1*0301 and DQA1*0501. However, some of the alleles were not present in different caste groups like DQA1*0303, DQA1*0304, DQA1*0502, DQA1*0603, DQA1*0604. A closer examination revealed that these populations were significantly different from Muslims and North Eastern populations. It was also observed that Muslims deviated from North Eastern Indian populations. Caste populations showed significant difference with Muslims at DQA1* 0101, DQA1*102, DQA1*0104, DQA1*0301, DQA1*0401 DQA1* 0502 and DQA1* 0601 alleles. Both *Shia* and *Sunni Muslims* showed differences at DQA1*0102, DQA1*0104 and DQA1*0601. North Eastern populations i.e. *Lachung, Mech* and *Rajbanshi's* differ from caste populations of North India and Muslims at DQA1*0102, DQA1*0104 and DQA1*0301 alleles while *Rajbanshis* were significantly different at DQA1*0101and DQA1*0601 alleles. Among North Eastern group *Lachung* were different from *Mech* at DQA1*0301 alleles, while they showed differences with *Rajbanshis* at DQA1*0104 and DQA1*0301 allele. *Mech* differed from *Rajbanshis* at DQA1*0104 allele.

At DQB1* locus significant differences were observed between caste groups and Muslims at different alleles like DQB1*0301, DQB1*0305, DQB1*0401, DQB1*0402, DQB1*0501, DQB1*0503, DQB1*0504, DQB1*0601, and DQB1*0603. Alleles DQB1*0301, DQB1*0303, DQB1*0501, DQB1*0502, and DQB1*0503 were significantly different in *Lachung, Mech* and *Rajbanshi* when compared to caste groups while *Rajbanshi's* differed at DQB1*0302 allele. Between Muslims and North Eastern group differences were found at DQB1*0301, DQB1*0302, DQB1*0303, DQB1*0501, DQB1*0502, DQB1*0503, DQB1*0601, and DQB1*0602 alleles. *Lachung* and *Mech* revealed differences when compared to *Rajbanshi's* at DQB1*0501, DQB1*0601 and DQB1*0602 alleles **Table 5.12.**

Table 5.11 Chi Square analysis at DQA1 Loci

	DQA1	0101	0102	0104	0301	0401	0502	0503	0601
Kayastha	Shia	3.84***	4.50***	2.21**					
	Sunni								
	Lachung			7.08**	4.22**				
	Mech			5.09**					
Vaish	Shia	20.77***	27.01**	14.12**	4.44*	4.17*	4.42*		23.35***
	Sunni	4.38*	5.6*				11.81**		7.69**
	Lachung			11.92**	5.78*				4.95*
	Mech				11.34**				
	Rajbanshi				3.98*				12.74**
Mathur	Shia	14.53***	14.72***	5.60*	5.31*	3.634*	4.448*	4.44*	20.21***
	Sunni						13.40**		4.24*
	Lachung		4.22*	8.36**	5.42*				
	Mech				7.53**				
	Rajbanshi								8.96**
Rastogi	Shia	26.08***	36.60***	18.84***		3.59*			18.35***
	Sunni	5.84*	10.56**	3.77*					3.36**
	Lachung			15.04**	14.99**				
	Mech			14.52**					
	Rajbanshi			6.50*					7.74**
Shia	Sunni		29.45***	6.05*					6.41*
	Lachung				15.77***				3.76*
	Mech								8.09**
	Rajbanshi	7.48**							
Sunni	Lachung		5.72*	8.68**	12.21**				
	Mech		3.38*	7.86**					
	Rajbanshi		4.07*						
Lachung	Mech				8.62**				
	Rajbanshi			3.88*	7.31**				
Mech	Rajbanshi			4.03*					5.21*

P value<0.05*, P value<0.01*, P value<0.001*

Table 5.12 Chi Square analysis at DQB1 loci

	DQB1	0301	0302	0303	0305	0401	0402	0501	0502	0503	0504	0601	0602	0603
Kayastha	Mathur							4.69*						
	Vaish													4.01*
	Rastogi							8.19**	4.36*					
	Shia				3.32*			5.98*		7.66**	10.58**	3.92*		5.8*
	Sunni				5.09*			10.74**						5.63*
	Lachung	8.33**						5.81*	5.29*				8.04**	4.98*
	Mech	13.58**		5.15*				5.34*	19.70***	5.70*			5.52*	
	Rajbanshi	10.80**		11.05**				36.34***	4.53*	6.55*				
Vaish	Shia	3.80*			3.9*	4.21*				6.94**	10.0**	8.6**		
	Sunni				4.75*						4.27*	5.09*		
	Lachung	14.3**							9.55**				9.43**	
	Mech	20.81***		4.15*					10.56**	5.49*			6.49*	
	Rajbanshi	17.8***	4.94*	9.28**				14.07**		6.19*				
Mathur	Shia	3.95*			6.82**					13.47**	12.78**	3.39*		
	Sunni						9.0**	4.12*			3.77*			
	Lachung	15.14***		3.35*					9.76**	4.53**			8.70**	
	Mech	22.39***		6.88**					10.84**	8.92**			5.67*	
	Rajbanshi	19.42***		13.94**				24.5***		10.62**				
Rastogi	Shia									7.36**	10.51**	3.61*		
	Sunni				5.0*		3.63*							
	Lachung	11.25**							8.15**				8.53**	
	Mech	17.43***		5.43*					9.14**	6.23*			5.57*	
	Rajbanshi	14.55***		11.53**				12.64**		6.76**				
Shia	Sunni									13.02**				
	Lachung	5.34*							15.28***				11.62**	
	Mech	9.53**		5.15**					16.49***				8.11**	
	Rajbanshi	6.95**	3.78*	11.05**				15.10**	4.23*			5.77*		
Sunni	Lachung	6.68**							8.62**	3.57*			16.22**	
	Mech	11.36**							9.71**	7.57**			12.02**	
	Rajbanshi	8.68**						9.17**		9.22**				
Lachung	Mech													
	Rajbanshi							4.48*					5.84*	
Mech	Rajbanshi							5.33*				3.86*		

P value<0.05*, P value<0.01*, P value<0.001*

Haplotype Analysis

We have calculated the two locus haplotypes of DRB1-DQB1 **(Table 5.13)** and three locus extended haplotypes DRB1- DQA1- DQB1 **(Table 5.14)**. Our results revealed that there were total 25 haplotypes observed in our study. In various caste groups there were 17 haplotypes, among Muslims 12 haplotypes and in North Eastern populations 13 haplotypes were observed. Most frequent haplotypes were DRB1*0301-DQB1*0201, DRB1*0701-DQB1*0201, DRB1* 1501-DQB1*0601 which were found in all the nine studied population. It was observed that among *Kayastha* and *Mathurs* haplotypes DRB1*0403-DQB1*0302 was not seen but was observed in *Rastogies* (2.2%) and *Vaish* (2.2%). In *Kayastha* and *Rastogies* haplotype DRB1*1202-DQB1*0301 was observed with similar frequencies but was not detected in *Mathurs* and *Vaish.* Haplotype DRB1*1301-DQB1*0601 was absent in *Rastogies* while this particular haplotypes was present in other three caste populations. In the present study we have observed certain unique haplotypes like DRB1*1301-DQB1*0201 and DRB1*0701-DQB1*0301 were unique for *Kayastha,* while DRB1*1503- DQB1*0601 was unique for all the four North Indian caste groups.

Among Muslims DRB1* 0101- DQB1*0601 was unique haplotype which was not observed in any other populations. Haplotype differences were seen in *Shia* and *Sunni Muslims* as haplotypes DRB1*1301-DQB1*0201, DRB1*1401-DQB1*0503, and DRB1*1501- DQB1*0602 were observed in *Sunni* only and not among *Shia's.*

In North Eastern populations haplotype DRB1*1101- DQB1*0501was commonly observed and was unique for this group only, haplotype DRB1* 0701-DQB1*0501 was unique for *Lachung* (2.5%) and *Rajbanshis* (3.6%) as this haplotypes was not observed in caste populations and Muslims, while DRB1*0301-DQB1*0501 was unique for *Mech* population. Overall, the haplotype analysis has shown that HLA-DRB1-DQB1 in nine populations revealed some common and some unique haplotypes in different populations.

Table 5.13 Two locus haplotypes in North Indian Caste populations

DRB-DQB	Kayastha	Mathur	Rastogi	Vaish	Shia	Sunni	Lachung	Mech	Rajbanshi
1001*0501	4.6	9.0	7.6	9.0		7.1	3.4		2.0
0101*0501	1.4	3.1	2.8	3.1	6.6	3.6			
0101*0601					1.6	4.4			
0301*0201	7.3	7.0	6.0	7.0	5.9	7.4	3.4	1.5	4.6
0301*0501								3.9	
0403*0302			2.2	2.2					
0701*0201	14.3	16.4	11.1	16.4	8.6	8.5	10.3	10.2	7.1
0701*0301	1.3								
0701*0303	4.8	5.8	5.0	5.8	2.4	3.3			
0701*0501							2.5		3.6
080x*0301							3.4	7.9	6.7
090x*0301							5.1	3.2	
1101*0301	4.4	3.8	3.8	4.0	3.5	6.3	4.3		
1101*0501							4.9	5.5	6.9
1201*0301							7.7	8.7	8.7
1202*0301	1.5		1.7		1.5	1.5			
1301*0201	1.1					2. 3			
1301*0601	1.5	1.2		1.2					
1301*0603	4.2	3.5	3.3	3.5					
1401*0501							2.5	2.3	6.1
1401*0503	3.4	3.5	1.7	3.5		3.4			
1501*0601	6.0	6.0	7.0	6.0	1.1	2.0	3.4	7.9	8.4
1501*0602	1.0	1.2	1.2	1.2		2.9	4.3	3.1	
1502*0601	4.4		4.5	7.0					
1503*0601	1.5	2.5	1.2	2.5					

Three Locus Extended haplotype

The three locus extended haplotype results have shown that there were total 31 HLA-DRB1-DQB1-DQA1 haplotypes. 19 haplotypes were observed in North Indian caste groups, 13 in Muslims and 16 in North Eastern populations. Most frequent haplotypes where the frequency ranged between 4.9% to 14.8% these were found in all the caste groups. These were DRB1*0701-DQA1*0201-DQB1*0201, DRB1*0701-DQA1*0201-DQB1*0303 and DRB1*1001-DQA1*0104-DQB1*0501. DRB1*0101-DQA1*0101-DQB1*0501, DRB1*0301-DQA1*0301-DQB1*0201, DRB1*1101-DQA1*0501-DQB1*0301, DRB1*1401-DQA1*0104-DQB1*0503, DRB1*1501*0103 -DQB1*0601, DRB1*1502-DQA1*0103-DQB1*0501 were common haplotypes for all the four North Indian caste populations. Some unique haplotypes were also observed like DRB1*1501-DQA1*0103-DQB1*0503 was unique for *Mathurs* (1.5%) while it was entirely absent in all the studied populations

When Muslim Group was explored it was found that haplotype DRB1*0301-DQA1*0301-DQB1*0201 was observed in *Sunni* Muslims which was frequent in all the studied caste populations also, some unique haplotypes were in this group. These were DRB1*1501*0101*0301 unique for *Shia* Muslim, DRB1*0101-DQA1*0501-DQB1*0601 was unique for both the Muslim populations (**Table 5.14**).

In North Eastern populations DRB1*1502-DQA1*0102-DQB1*0601 was unique haplotype for *Lachung* with a frequency of 1.7%, DRB1*0701-DQA1*0201-DQB1*0501 were found only in *Lachung* and *Rajbanshi's*, DRB1*1501-DQA1*0601-DQB1*0201 was found in *Mech* and *Rajbanshi's* while DRB1*1101-DQA1*0301-DQB1*0501 was observed in both *Lachung* and *Rajbanshi's*.

Test of Hardy-Weinberg equilibrium (HWE)

Multi locus genotype frequencies were calculated for each population and were utilized for testing the conformity with the assumption of HWE. It was interesting to note that all the studied populations were in hardy Weinberg equilibrium at all the three loci.

Intra population genetic variation analysis

In the present study, different parameters were used to quantify intra population diversity namely, average gene diversity and average observed heterozygosity. The average values of the different parameters depicting intra population diversity at three HLA Class II loci namely DRB1, DQA1 and DQB1 in four North Indian caste populations, two Muslim groups, and three North Eastern populations groups is shown in **Table 5.15**.

Table 5.14 Three locus haplotypes in North Indian Caste populations

DRB-DQA-DQB	Kayastha	Mathur	Vaish	Rastogies	Shia	Sunni	Lachung	Mech	Rajbanshi
0101*0101*0501	0.3	2.9	2.0	1.7	2.8	1.6	0.8		3.1
0101*0501*0601					1.5	1.5			
0301*0102*0201							2.6	2.4	
0301*0301*0201	1.8	1.9	2.2	1.7		1.2			
0301*0501*0201	4.9	4.5	3.3	3.8	3.3	3.3			
0403*0301*0302		2.2		1.0					
0701*0201*0201	12.0	14.8	12.6	9.4	6.2	6.2	7.0	4.2	8.6
0701*0201*0303	4.2	5.1	6.3	4.0	1.8				
0701*0201*0501							2.6		4.1
080x*0101*0601							2.5	2.3	
080x*0301*0301								3.9	2.5
0901*0301*0301							5.1	1.5	
1001*0104*0501	3.4	8.1	3.5	6.8	2.6	4.2			
1001*0401*0402								1.6	
1101*0201*0301							3.4	1.5	1.0
1101*0301*0501							1.7		4.0
1101*0501*0301	3.6	3.2	3.7	3.3	2.8	4.0	2.6	3.2	
1201*0501*0301							3.4	2.4	2.1
1202*0601*0301	1.5			1.7	1.5				
1301*0103*0603	2.6	1.3		2.1	1.8	1.2			
1401*0104*0503	2.2	2.5	1.7	1.0			3.1	2.5	2.4
1501*0101*0301					1.3				
1501*0102*0601	2.0		2.0	3.2					
1501*0103*0503		1.5							
1501*0103*0601	2.8	2.1	2.8	2	0.8	1.3	1.7	1.5	2.0
1501*0601*0201								1.5	2.5
1502*0101*0501		2.4							
1502*0102*0601				1.4			1.7		
1502*0103*0501		3.3	1.0						
1502*0103*0501	3.1	3.3	2.8	2.3					
1502*0103*0601	2.8		2.8	2.0		1.4			

Table 5.15 Average genomic diversity and observed heterozygosity

	Average Gene Diversity	Average observed Heterozygosity
Kayastha	0.7724	0.7765
Mathur	0.7783	0.7607
Vaish	0.7693	0.7628
Rastogi	0.7717	0.7526
Shia	0.7420	0.7024
Sunni	0.7223	0.7060
Lachung	0.7592	0.7213
Mech	0.6286	0.5747
Rajbanshi	0.7360	0.7090

(i) Average gene diversity

Average gene diversity at diploid loci is indicative of average expected heterozygosity and high value of gene diversity corresponds to high intra population diversity. All the populations revealed high degree of gene diversity. It varied from 0.6286 to 0.7783. Minimum average gene diversity was found among Mech was 0.6286.

(ii)Average observed heterozygosity

The average observed heterozygosity at three HLA class II loci among four caste populations, two Muslim groups and three North Eastern groups was calculated to determine intra population diversity. Locus wise average observed heterozygosity in studied nine populations is presented in **Table 5.16**. The observed heterozygosity values value ranged between 0.5747 in *Meches* to 0.7765 in *Kayastha*. The results of hetrozygosity estimation further strengthen our observation of high allelic diversity among all the Populations investigated for the present studied.

Table 5.16 Average Observed Heterozygosity

	Kayastha	Mathur	Vaish	Rastogi	Shia	Sunni	Lachung	Mech	Rajbans
DRB1	0.7855	0.7832	0.7553	0.7727	0.7031	0.7140	0.7377	0.5862	0.7143
DQB1	0.7790	0.7452	0.7422	0.7531	0.6940	0.7111	0.7191	0.6034	0.7143
DQA1	0.7701	0.7539	0.7910	0.7321	0.7101	0.6930	0.7072	0.5345	0.6984
Mean	0.7765	0.7607	0.7628	0.7526	0.7024	0.7060	0.7213	0.5747	0.7090

Inter population genetic variation

Inter-population genetic variation corresponds to the analysis of population differentiation. In the present study, two different strategies have been used for the intra level population genetic variation analysis: (i) The measure of portioning genetic diversity-Fst, (ii) The estimation of the proportion of genetic variance due to subdivision- Gst and (ii) Analysis of Molecular variance –AMOVA.

(i) Analysis of F-statistics

F-statistic was applied to study the population sub-division and inter-population variation. The measure for portioning genetic diversity within population relative to between populations is Co ancestry coefficient or F_{st}. Its value ranges from a value of zero that corresponds to no differentiation to the value of one indicating no shared genetic variation. Locus wise value of Fst is shown in **Table 5.17.** Average Fst value over all loci in caste groups was found to be 0.0358, in *Muslims* it was 0.0108 while in *North Eastern group* it was 0.0193.

Table: 5.17 Level of inter and intra population variation at HLA loci in the 9 study populations (Fst Analysis)

	Caste groups	Muslims	North Eastern Group
DRB1	0.03	0.0115	0.0105
DQB1	0.0587	0.0018	0.0416
DQA1	0.0181	0.019	0.0064
Mean	0.0358	0.0108	0.0193

Combined Fst analysis revealed that the level of differentiation (Fst) increased as we included the populations belonging to different socio cultural strata **(Table 5.18).** The Fst value increased to 0.0413 indicating that the level of differentiation within the groups is very low as compared to between groups.

Table 5.18: Fst analysis in different groups of population

Groups	Fst
Studied four caste populations	**0.0358**
4 Caste populations, 2 Muslim groups and 3 North Eastern groups	**0.0413**
Only Muslims (Shia and Sunni)	**0.0108**
Only North Eastern group(Lachung, Mech and Rajbanshi)	**0.0193**

Gene-diversity based Population subdivision analysis

There are three values estimated at gene diversity level which is based on population sub-structuring analysis these are total genomic diversity (Ht), genomic diversity of subpopulations (Hs) and average gene diversity between sub populations (Gst). Gst analysis revealed that the total genomic diversity (H_T) among populations was quite high ranging from 0.7191 for DQB1 to 0.799 for DRB1in *North Indian caste group* average being 0.7502. In *Muslim groups* it ranged from 0.719 for DQB1 to 0.7521 for DRB1 average being 0.7345. In *North Eastern group* it ranged from 0.722 for DQB1 to 0.7365 for DRB1 loci and average being 0.73. However, most of the genomic diversity is because of diversity between individuals within subpopulations (H_S). Genomic diversity between populations (G_{ST}) varied from 0.002 (DQB1) to 0.008 (DRB1) in North Indian caste group. However, average Gst value between populations was 0.0043.

Muslims/consanguineous group revealed Gst value (0.0053), which was higher than caste populations, and North Eastern Group revealed Gst value (0.004). Locus wise Ht,Hs and Gst values are given in **Table 5.19**

Table 5.19 Genomic diversity in the studied populations

N Indian Caste group	Ht	Hs	Gst
DRB1	0.799	0.7925	0.008
DQB1	0.7191	0.718	0.002
DQA	0.7325	0.7295	0.003
Mean	0.7502	0.746	0.004
Muslim group			
DRB1	0.7521	0.7428	0.011
DQB1	0.7191	0.718	0.002
DQA	0.7325	0.7295	0.003
Mean	0.7345	0.7301	0.005
N East group			
DRB1	0.7365	0.7259	0.004
DQB1	0.7222	0.7185	0.004
DQA	0.7315	0.7281	0.004
Mean	0.73	0.7241	0.004

Analysis of Molecular Variance (AMOVA)

AMOVA was carried out on HLA loci with Arlequin software package. Three parameters are calculated using AMOVA these are: variance within individual populations, variance between populations of a group, and variance between different groups. AMOVA was carried out in three groups comprising the *Muslims*, *four middle caste populations* of North India and three *North Eastern populations* studied in the present study. AMOVA analysis shows that 98.13% of the genetic variation was found to exist within each population **(Table 5.20)**. Most of the variation of the remaining fraction was due to the differences between the groups i.e1.66% while only 0.21% variance was the difference observed between the populations within the group. This suggests that the populations in each of these groups are genetically more similar to each other than the populations of different groups.

Table 5.20 Analysis of Molecular variance (AMOVA) based on Three HLA Class11 loci

Groups	Number of populations (n)	Variance Components		
		Within populations	Among populations within groups	Among Groups
Four Middle Caste populations, two Muslim population and three North Eastern populations	9	98.13%	0.21%	1.66%

Estimation of genetic distances

To assess the genetic relationship between the studied populations, pair wise genetic distances were calculated based on the multi locus genotypic data of HLA Class II loci. However, Nei's D_A based distance approach is formulated for an infinite iso alleles model of mutation, in which there is a rate of neutral mutation and each mutant results into completely new allele. Therefore, Nei's D_A includes both mutation and drift as possible cause of genetic differences. The distance matrices generated from this method are given in **Table 5.21**. *Kayastha* and *Mathur*; *Rastogies* and *Vaish* were genetically more similar to each other. *Kayastha* and *Mathurs* were more distant from rest of the other two caste populations.

Shia's from consanguineous group were very far away from Caste group while *Sunni's* show some closeness with Caste groups.

From North Eastern group *Lachung* and *Meche* were genetically more similar to each other but *Rajbanshi's* were more distant with both *Lachung* and *Meche's* as these two are tribals of Tibeto-Burman group while *Rajbanshi's* are believed to be hindu caste of Indo Aryan family.

Table 5.21 Genetic distance between the populations

	Kayastha	Mathur	Rastogi	Vaish	Shia	Sunni	Lachung	Mech	Rajbanshi
Kayastha	0								
Mathur	0.0234	0							
Rastogi	0.085	0.0436	0						
Vaish	0.0507	0.0541	0.0166	0					
Shia	0.1484	0.1305	0.1188	0.1182	0				
Sunni	0.0958	0.0668	0.0498	0.0677	0.0495	0			
Lachung	0.1307	0.1709	0.1435	0.1529	0.2119	0.1575	0		
Mech	0.1284	0.1692	0.1445	0.1573	0.1999	0.1604	0.0286	0	
Rajbanshi	0.1283	0.1308	0.0957	0.1315	0.169	0.1093	0.0678	0.0485	0

Discussion

Phylogenetic assessment of studied populations

Phylogenetic reconstruction was carried out to cluster together populations who are genetically similar, identify distinct clusters and perhaps see how these clusters relate to geographical, linguistic or socio-cultural information available about these individuals (**Figure 5.1**).

Phylogenetic tree shows three clusters corresponding to *Muslims, middle caste groups* and *North Eastern group. Muslims* were clustered together while the middle caste population cluster has been further bifurcated into two branches: one carrying *Kayastha* and *Mathurs* and other carrying *Rastogies* and *Vaish*. All the three North Eastern populations clustered in a separate group.

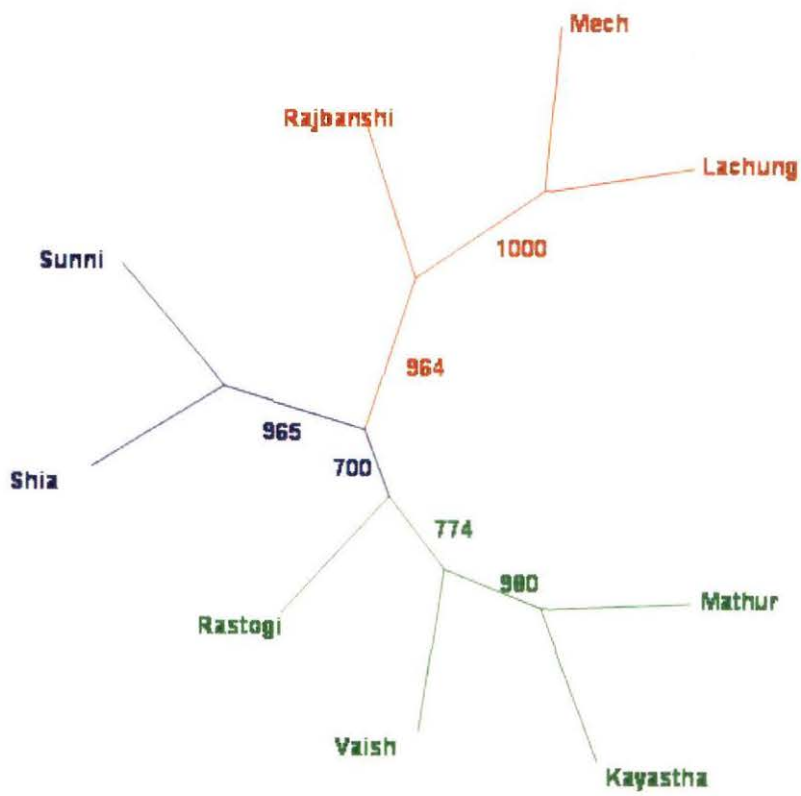


Figure: 5.1 Phylogenetic trees depicting clustering of the nine studied populations

Phylogenetic analysis is in accordance with the allele frequency and haplotype analysis revealing distinct caste, inbreeding and tribal groups and show distinct and unique features corresponding to particular groups.

Discussion

Regarding the evolution and divergence of world populations various hypothesis have been put forwarded. It is thought that human populations arose first from a small pool of breeding individuals 100,000-200,000 years ago **(Cavalli-Sforza, 2003; Maynard V, 1999; Harpending *et al.*, 1998)**, followed by population expansions as a result of which these populations adopted different occupations and geographic territories **(Cavalli Sfroza, 1994a)**. Most founder populations have been diluted or lost as a result of later admixtures following large-scale immigrations. However, a small number of groups remained, particularly in geographically isolated regions. These populations can be labeled as ethnic outliers using measures of genetic diversity. **(Cavalli Sfroza, 1994b)**. Efficient farming methods 10,000-20,000 years ago following the last Glacial Maxima, resulted in a major wave of demographic expansion **(Cavalli Sfroza, 1998; Harpending *et al.*, 1998; Rogers and Comuzzie, 1995)** resulting into various ethnic groups. These ethnic groups may resemble or differ from one another depending upon the nature of evolution they have undergone.

Several measures to evaluate genetic variation are being used. These are useful for reconstructing the phylogenetic relationships of such groups. For calculating the genetic distances there is a need to know about the distribution of allelic frequencies which can be compared with that of other populations. Human species is genetically polytypic, in the sense that allele frequencies at a larger number of loci differ from one population to another. Each breeding population is polymorphic, that is at many loci two or more alleles are found which usually remain in Hardy-Weinberg equilibrium under the balancing forces like mutation, drift or selection. Very few alleles are entirely restricted to the populations of any one region. Such a pattern of allele frequency would be

expected only if the populations have remained undisturbed for many generations **(Cavalli Sfroza,1967)**. Neighboring populations, however, are more likely to resemble one another in allele frequencies than populations living at a distance.

Although it is quite difficult to get reliable frequency estimates for less common or rare alleles, some of the alleles are by no means uniformly distributed and present some interesting problems in human genetics. It will, therefore, prove to be of additional information, if in addition to more frequent traits, a systematic attempt is made to determine the frequencies of rare conditions in the general populations. A study of the distribution of these and many other genetic traits in various populations can be helpful in mapping geographical clines, which, in turn are useful in reconstructing ethnic relationships and help in the understanding of selective pressures which determine frequencies of these genes. Several authors have proposed statistical methods to evaluate the influence of various forces acting on populations. These methods are in the form of measures of genetic distance, which take into account the variation accumulated through the mutational processes along with the effects of other evolutionary forces.

Various genetic markers used to evaluate genetic diversity among populations are blood groups, electrophoretic mobility variants of serum proteins, human leukocyte antigens, DNA restriction fragment length polymorphisms (RFLP), variable number of tandem repeats (VNTR's), single nucleotide polymorphisms (SNP's) etc. When these markers are appropriately analyzed, they can reveal useful information about the genetic structure, divergence, and microevolution of human populations.

The Indian subcontinent offers excellent opportunities in this respect. Its population is an assemblage of small clusters of human groups, separated from each other by geographic space and social distance. This has led to the formation of numerous closed gene pools, which have remained virtually undisturbed for many generations as a result of breeding in isolation.

In the present study an attempt has been made to describe the genetic structure of four North Indian caste populations, two Muslim populations and three North Eastern populations. To evaluate the genetic structure of populations in the present study we have used highly polymorphic HLA class II antigens marker.

Human Leukocyte Class II Antigens:

Human past can be read through a history written in HLA. The HLA polymorphism has proven to be useful for singling out individuals and populations. The discoveries of new loci and presently available DNA typing and sequencing of new alleles have dramatically increased the variety of HLA alleles. Certain alleles are frequent only in specific populations e.g. A36 and A43 in African – Americans. Strong linkage disequilibrium between HLA neighboring loci demonstrates that certain combinations of contiguous alleles (HLA haplotypes) also occur more frequently in a population compared to others. **(Gottellii *et al.*, 1994; Paetkau and Strobeck 1994; Taylor *et al.*, 1994).** They can be used to estimate effective population size **(Allen *et al.*, 1995)** and to gain insight into the degree of population substructure including both the amount of migration between subpopulations **(Allen *et al.*, 1995 ; Gottelli et. al. 1994)** and genetic relationships among the various subpopulations **(Bowcock *et al.*, 1994; Estoup *et al.*, 1996; Bamshad *et al.*,2001).**

In the present study we have analyzed the immunogentic profile of different caste groups i.e *Kayastha, Mathur, Vaish,* and *Rastogies* and two inbreeding populations i.e. *Shia* and *Sunni* Muslims have been selected from North India. Populations from North Eastern region have also been included in this study and these are *Lachung, Mech* and *Rajbanshi.* These populations were compared with one another to elucidate the anthropological relationship between various groups. Multi locus haplotypes were estimated and the genetic relation between populations was evaluated using HLA class II data. The main objective was to understand the population structure and their genetic profile etc.

Combined analysis of HLA class II alleles revealed interesting observations on the basis of allele frequency and haplotype frequencies at HLA- DRB1, DQA1and DQB1 loci.

HLA-DRB1 allele frequency distribution

In the present study we have observed that at DRB1 locus, alleles where frequency distribution ranged between 2.6% to 25.7% were present in all the

populations from North India and also among North Eastern India. There were only five such alleles i.e. DRB1*0101, DRB1*0301, DRB1*0701, DRB1*1001 and DRB1*1101. Although these alleles were present in all the populations but frequency gradation was seen which is described below.

The allele DRB1*0701 which was found to be highest among *Mathurs* (25.7%) was seen with a low frequency in *Mech* (11.9%) from North Eastern populations. Among other caste populations i.e. four caste populations the frequency of this allele was 19.6% to 25.7%. While among Muslims and North Eastern populations i.e. *Lachung, Mech* and *Rajbanshi* it ranged from 11.9%to 15.9%.

It was interesting to note that although there was no significant difference at this locus among the studied populations. However, when Muslims were compared with one another there was greater similarity among them. *Lachung* and *Rajbanshi* from North Eastern region of India were showing almost similar allele frequency at DRB1*0701. *Mech* showed little variation from that of *Lachung* and *Rajbanshi* the frequency distribution at this locus was 11.9%.

At DRB1*1001 the allele frequency varied between 6.5% to 10.3% in North Indian populations. Among North Eastern populations the frequency of this allele was low i.e 4.7% to 5.7%.

On comparing allele DRB1*1101 in all the populations it was evident that in caste populations and *Shia Muslims* frequency of this allele ranged between 4.5% to 6.6%. While in *Sunni* Muslims and North Eastern populations its frequency was high i.e 9.5% to 11.2%. Further analysis revealed that *Sunni* muslims differed significantly with Vaish and *Rastogies* . Differences between *Mathurs* from caste populations and all the three North Eastern populations were statistically significant. *Rastogies* revealed significant difference from that of *Lachung*.

Another allele which occurred with a high frequency was DRB1*0301. Among caste populations it ranged between 7.5% to 9.4%. However, among Muslims it ranged between 10.1to 15.3% which was significantly different from caste population. In North Eastern population frequency of this allele was between 4.3% to 5.6%. However, it was found that North eastern populations differ significantly only with *Shia* Muslims and not from any of the other studied caste groups. It has been found that some of the alleles

where the frequency was as high as 16% was restricted only in some of the populations but their frequency decreased to the extent of 1.7% in other populations for e.g. the allele DRB1*0101 is represented in all the populations, highest in Muslims (16% and 11.7%), while in caste groups its frequency ranged between 2.6% in *Kayastha* to 7.65% in *Rastogies*.

The gradation was much clear when we analyzed the frequency of this allele in North eastern populations. It ranged from 1.7% to 3.7%. At this allele it was interesting to find that there was significant difference between caste populations and Muslims. Differences were also seen among Muslims and North eastern populations.

We have further categorized the allele frequency distribution into moderately high frequency category (2.2% to 11.1%). Under this category following alleles were seen DRB1*0403, DRB1*1201, DRB1*1202, DRB1*1301, DRB1*1302, DRB1*1401, DRB1*1503, DRB1*080x and DRB1*090x. It has been found that at moderately high frequency distribution there was only one allele i.e. DRB1*1202 at which there was a significant difference between *Kayastha* and *Vaish.* Inspite of having moderate frequency at DRB1 loci, some of the alleles were altogether absent in the North Eastern populations like DRB1*0403, DRB1*1202 and DRB1*1503 this demonstrates that North Eastern populations are different from North Indian caste population and the other two highly inbreeding Muslim populations. There were certain alleles which were found with a very low frequency and we labelled these as less frequent alleles. The allele frequency distribution ranged between 0.26% to 1.5%. This category revealed some of the interesting findings i.e. there were certain alleles which were specific only to the caste groups for example among caste groups DRB1*1308 and DRB1*1403. The low frequency of some of the HLA-DRB1 locus may be due to genetic isolation caused by caste endogamy. However, between *Shia* and *Sunni* Muslims there were certain unique alleles like, DRB1*1107 which are suggestive of the fact that *Shia* and *Sunni* are different from caste groups of North India which has also been proved on the basis of Mitochondrial, Y chromosomes and STR studies from our laboratory.

It is interesting to note that at high and moderate high allele frequency distribution there was not much difference between North Indian and North eastern

populations but at the less frequent alleles there were certain alleles like DRB1*1405 which was restricted to *Lachung* (1.7%), *Mech* (1.6%) and *Rajbanshi* (2.04%) only.

Comparison of HLA-DRB1 locus with the world populations

Further we have compared the allele frequency distribution of our population groups with other world populations. For comparing the allele frequency data of our populations we have taken different world populations **(Table 6.1).**

Among the high frequency alleles when DRB1*0101 was compared with world populations having different ethnicity it was found that this allele was mostly present in high frequency in Asians, Middle East and Europeans. Among North Indian caste populations frequency ranged between 2.6% to 7.65%, which was comparable to most of the Caucasian populations like Czech (7.5%), Germans (10.5%), Italy (6.8%), England Cauc (10.3%), Poland (9.5%) and Turkey (5.8%) **(Cerna *et al.*, 1992; Ferencik *et al.*, 1997; Rendine, Borelli *et al.*, 1998; Doherty *et.al*, 1992; Kapustin *et.al*, 1999; Saruhan-Direskeneli *et al.*, 2000)** Muslim group showed much similarity with Middle East populations like Iran Baloch (15%) and Iran Yazd Parsi (6.9%) in *Shia* and *Sunni* Muslims it was 16% and 11.7% respectively **(Farjadian *et al.*, 2006)**. while frequency of this allele in North Eastern population was very low i.e 1.6% to 3.06% and frequency of this allele was found to be comparable to populations like China Naxi (1.3%), China Yunnan (3.1%), China Wuhan (1.8%) and Khalkha Mongols (2.5%). **(Fu Y, Liu Z *et al.*, 2003; Shi, Xu *et al.*, 2006; Ferencik *et al.*, 1998; Munkhbat,1997; Machulla *et al.*, 2003)**.

In the present study DRB1*0301 allele was found in high frequency in all the populations with frequency range of 4.3% to 15.3%. Among caste populations the average frequency of this allele was 7.5% to 9.2%, which was comparable to Caucasion populations like Czech (7.1%), Germans (11.4%), Italy (7.0%) and Turkey (9.2%). **(Cerna *et al.*, 1992; Ferencik *et al.*, 1997; Rendine & Borelli *et al.*, 1998; Saruhan-Direskeneli *et al.*, 2000).** *Shia* and *Sunni* Muslims revealed its frequency in the range of 15.2% and 10.1% respectively, which was comparable to Middle Eastern populations like Iran Baloch (29%) and Iran Yazd Parsees (5.4%) **(Farjadian *et al.*, 2006).** North Eastern populations were different from North Indian caste and inbreeding groups and

were also different from Caucasians and populations from Middle East. Frequency of this allele in North Eastern group was 4.3% to 5.6% only. However, these populations showed similarity with oriental populations like China Naxi (2.5%), China Wuhan (4.1%) and North Chinese (2.8%) etc. **(Fu Y, Liu Z *et al.*, 2003; Ferencik *et al.*, 1998)**

It was also seen among DRB1*0701 was predominantly present in almost all the world populations like Basque (31.3%), Czechs (16.1%), Italy (16.8%), France South (22%), Germans (12.3%) etc from Europe. In Asians like Uygur (16%), SD Han (11.3%), North Chinese (13.7%), Pomors (12.3%), Khalkha Mongols (12%) etc. **(Sanchez-Velasco *et al.*, 2003; Charron D,1997; Cerna M *et al.*, 1992; Ferencik *et al.*, 1997; Rendine & Borelli *et al.*, 1998; Doherty *et al.*, 1992; Kapustin *et al.*, Saruhan-Direskeneli *et al.*, 2000; Mizuki & Ohno *et al.*, 1998; Zhou *et al.*, 2005; Evseeva *et al.*, 2002; Munkhbat 1997; Machulla *et al.*, 2003)** This allele was also found in high frequency among Africans populations like Algerians (13.8%), Oromo (22.1%), Amhara (26.3%) **(Arnaiz-Villena *et al.*, 1995; Fort M *et al.*, 1998)**. This allele was also frequently seen among some American populations like Argentina, Brazil. **(Cerna *et al.*, 1993)** with frequency between 11.8% to 17.8%. Among studied populations also high frequency of this allele was seen like in *Kayastha* (25.8%), *Vaish* (23.9%), *Mathurs* (25.7%), *Rastogies* (19.6%), *Shia* (14.7), *Sunni* (15.9%), *Lachung* (13.8%), *Mech* (11.9%) and *Rajbanshis* (13.9%). The allele frequencies were comparable to other populations mentioned above. However, it was completely absent in population like Brazilian Kaiowa, South American Indians (Sikuanis, Koguis Inganos, Emberas, Waunanas, Guahibo Nukakas). **(Tsuneto *et al.*, 2003)**

Allele DRB1*1001 was found in most of the populations. In the study group the frequency of this allele among North Indian populations ranged between 6.5% to 10.1%. However, among Caucasians like Turkey, Romania and Chuetas its frequency was between 2.0% to 3.4%. **(Borelli *et al.*, 1998, Crespí *et al.*, 2002)**. The North Eastern belt of our study revealed the frequency from 4.7 to 5.6% which was similar to Minman (5.3%) and Hakka (4.9%) from China. North Eastern populations also revealed similarity with Khotons (3.5%) and Khalkha (4.0%). **(Munkhbat 1997; Machulla *et al.*, 2003)**

DRB1*1101 was observed in all the populations in this study. Caste populations show frequency in the range of 4.5% to 6.3% which was found to be similar frequencies to

Caucasian populations like Czechs (6.7%), Germans (6.0%) and Belgium (7.2%) **(Cerna *et al.*, 1992; Ferencik *et al.*, 1997; Andrien & Dupont *et al.*,1998).** Among North East populations the frequency was between 9.5% to 11.2% and its frequencies were found to be comparable to China Naxi (6.8%), China Beizing and Xian (9.4%), North Chinese (6.6%), and Khalkha Mongols (6.5%). **(Fu Y, Liu Z *et al.*, 2003; GaoX *et al.*, 1997; Munkhbat 1997; Machulla *et al.*, 2003)**

Muslim populations showed high frequency of this allele i.e. 6.5% to 9.8%and were also found to be comparable to Middle Eastern populations like Iranian Baloch (8.5%) and Iran Yazd parsees (14.6%) **(Farjadian *et al.*, 2006).**

DRB1*0403 allele was found in North Indian caste populations and Muslims with very low frequency i.e 0.78% to 2.27%. It was totally absent in all the North Eastern populations. Frequency of this allele in different ethnic populations was also low except in South and Central American like Ingano, Waunana, Embera, Tule, Kogui, Ijka, Mexican and Argentina Toba Rosario **(Tsuneto *et al.*, 2003; Carmen *et al.*, 2001; Cera *et al.*, 1993)** in which the frequency was very high i.e. 5% to 22%.

We have listed DRB1*08 as moderately high frequency allele. This allele was found in high frequency in different world populations like in American population's i.e Bolivian Aymara (29.9%), Brazil (26.3%), Mexican Mazatecans (15%), Ingano (39%), Ijka (64%), Sikuani (18%) and Argentina (20.3%). **(Donadi *et al.*, 2002; Tsuneto *et al.*, 2003; Carmen Alaez *et al.*, 2001; Cera M *et al.*, 1993)** and in Asian populations i.e Tsaatan Mongols (7.0%), Japan Hyogo (14.3%), SD Han (7.1%) **(Munkhbat 1997; Machulla *et al.*, 2003; Nobuo Araki *et al.*, 1998 Zhou & Lin B *et al.*, 2005)** in comparison to that this allele was predominant in only Tibeto-Burman group i.e. *Mech* (11.1%), *Rajbanshi* (9.3%), and *Lachung* (7.7%).

DRB1*09 was found to be present at a high frequency in populations of Mongoloid origin i.e. Thais, Japanese and Chinese (9.5 – 14.1%). **(Chandanayingyong D *et al.*, 1997, Saito S *et al.*, 2000)** this allele was observed in two North eastern populations *lachung* (10.3%) and *Mech* (8.7%) however, it was present with a very low frequency in most of the Mediterranean populations (0.1 – 1%) This allele was totally absent in many other world populations e.g. in Caucasoids like Greeks, Bulgarians, Croatians and in American Indians like Tules, Koguis, Inganos, Ijakas, Emberas,

Guahibos and Nukakas and also Algerians, Canadians, Oromos and Amharas (Ethiopia). **(Reveille JD *et al.*, 1995, Ivanova *et al.*, 2002, Tsuneto LT *et al.*, 2003, Arnaiz-Villena *et al.*, 1995;Fort M *et al.*, 1998)**

Among all the studied populations allele DRB1*1201 was highest in *Rajbanshi* (10.2%), *Mech* (8.7%), and *Lachung* (7.7%) which was also observed in most of the Oriental populations like China Beizing and Xian (7.4%), China Naxi (8.4%), China Urumqui Kazaq (8.2%), South Korea (7.2%) and Russian Nentsy (16.4%). **(Gao *et al.*, 1997; Fu & Liu Z *et al.*, 2003; Mizuki *et al.*, 1997; Lee KW *et al.*, 2005; Evseeva *et al.*, 2002)**

In North Indian Caste populations high frequency of DRB1*1301 was found with frequency range between 5.1% to 8.6%, whereas in all other populations studied the frequency was too low. The high frequency of this allele was comparable to populations like Africans i.e Algeria (10%), Cameroon (13%) and Gabonese (11.3%). **(Arnaiz-Villena *et al.*, 1995, Pimtanothal *et al.*, 2001).** This allele was also seen among Asians like Uygur (6.1%), Russia (6.4%), Pomors (11.1%), Nentsy (11,8%). **(Mizuki & Ohno *et al.*,1998; Evseeva *et al.*, 2002)** while in Iran Parsi frequency was 15% **(Farjadian *et al.*, 2006).** Among Europeans it was found in Czech, Germans, Minorca, South France and Poland with frequency of 7% to 13%. **(Cerna M *et al.*, 1992; Ferencik *et al.*, 1997; Crespí & Milà *et al.*, 2002; Charron 1997; Jungerman).**

Among caste populations DRB1*1501 was seen in frequency range between 8.5% to 9.9% and similar frequencies were also seen among Caucasian populations e.g. Czech (10.8%), Italy (5.5%), South France (11.9%) and Croatians (9.0%). **(Cerna *et al.*, 1992; Rendine & Borelli *et al.*, 1998; Charron 1997; Grubic & Zunec *et al.*, 1995).** Muslims show deviation at this allele (3.4% to 4.7%), which was seen to be comparable to Iran Baloch (4.5%) **(Farjadian *et al.*, 2006).** On comparing our North Eastern with Mongolian populations it revealed some similarity. In North Eastern population frequency was 9.5% to 13.7% while it was comparable to Mongoloids like China Yunnan (12.2%), Wuhan (13.7%), SD Han (16.5%) and Japanese (15%). **(Shi &Xu *et al.*, 2006; Ferencik *et al.*, 1998; Zhou & Lin B *et al.*, 2005; Saito *et al.*, 2000)**

Table 6.1 DRB1 frequency distribution in world population

	1	2	3	4	5	6	7	8	9	10	11	12
Alleles	**0101**	**0103**	**0301**	**0302**	**0303**	**0401**	**0402**	**0403**	**0404**	**0405**	**0406**	**0407**
Asians												
Kayastha	0.0260	0.0000	0.0895	0.0026	0.0000	0.0157	0.0000	0.0130	0.0113	0.0052	0.0026	0.0000
Mathur	0.0480	0.0000	0.0949	0.0000	0.0000	0.0130	0.0130	0.0260	0.0032	0.0000	0.0032	0.0065
Vaish	0.0581	0.0000	0.0758	0.0000	0.0000	0.0101	0.0076	0.0227	0.0176	0.0025	0.0000	0.0030
Rastogi	0.0765	0.0000	0.0920	0.0153	0.0000	0.0101	0.0227	0.0128	0.0150	0.0000	0.0000	0.0000
Shia	0.1605	0.0184	0.1526	0.0000	0.0000	0.0131	0.0157	0.0078	0.0078	0.0026	0.0000	0.0026
Sunni	0.1170	0.0000	0.1011	0.0000	0.0080	0.0079	0.0027	0.0160	0.0080	0.0053	0.0000	0.0053
Lachung	0.0170	0.0000	0.0430	0.0000	0.0000	0.0340	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Mech	0.0160	0.0000	0.0550	0.0000	0.0000	0.0320	0.0160	0.0000	0.0000	0.0000	0.0000	0.0000
Rajbanshi	0.0306	0.0000	0.0560	0.0000	0.0000	0.0306	0.0050	0.0000	0.0000	0.0000	0.0000	0.0000
North Indian	0.0136	0.0000	0.0955	0.0000	0.0000	0.0000	0.0091	0.0454	0.0181	0.0000	0.0091	0.0000
China Naxi	0.0130	0.0000	0.0250	0.0000	0.0040	0.0000	0.0000	0.0090	0.0000	0.0760	0.0170	0.0000
China Yunnan	0.0310	0.0000	0.0310	0.0000	0.0000	0.0080	0.0000	0.0080	0.0120	0.0160	0.0040	0.0000
China Wuhan	0.0080	0.0000	0.0410	0.0040	0.0000	0.0080	0.0000	0.0120	0.0080	0.0250	0.0290	0.0000
Uygur	0.0180	0.0000	0.1400	0.0000	0.0000	0.0090	0.0180	0.0000	0.0260	0.0530	0.0000	0.0000
Sd Han	0.0050	0.0000	0.0360	0.0000	0.0000	0.0150	0.0000	0.0150	0.0000	0.0460	0.0150	0.0000
Ch Bei Xian	0.0260	0.0000	0.0260	0.0000	0.0000	0.0220	0.0090	0.0180	0.0000	0.0630	0.0370	0.0000
Urumqi Kaz	0.0370	0.0000	0.1310	0.0000	0.0000	0.0140	0.0000	0.0240	0.0240	0.0350	0.0000	0.0000
South Korea	0.0650	0.0000	0.0280	0.0020	0.0000	0.0060	0.0000	0.0220	0.0170	0.0870	0.0600	0.0030
Thailand	0.0040	0.0000	0.0700	0.0000	0.0000	0.0040	0.0000	0.0210	0.0000	0.0620	0.0390	0.0000
Japanese	0.0660	0.0000	0.0030	0.0000	0.0000	0.0110	0.0000	0.0390	0.0030	0.1690	0.0410	0.0000
Ainu Japa	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0900	0.0000	0.0000
Japan Hyogo	0.0940	0.0000	0.0000	0.0000	0.0000	0.0160	0.0000	0.0000	0.0000	0.1510	0.0310	0.0000
Gypsy	0.0590	0.0000	0.0590	0.0000	0.0000	0.0880	0.0000	0.0000	0.0590	0.0000	0.0000	0.0000
Russia	0.0880	0.0100	0.0940	0.0000	0.0000	0.0600	0.0050	0.0450	0.0230	0.0000	0.0000	0.0050
Pomors	0.1220	0.0000	0.0000	0.0000	0.0000	0.2020	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Saami	0.1360	0.0000	0.0000	0.0000	0.0000	0.3270	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Nentsy	0.0250	0.0000	0.0000	0.0000	0.0000	0.2180	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
N Chinese	0.0280	0.0000	0.0280	0.0000	0.0000	0.0170	0.0060	0.0110	0.0000	0.0440	0.0500	0.0000
N E Russia	0.1312	0.0000	0.0000	0.0000	0.0000	0.1677	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Nivkhi Rus	0.0000	0.0000	0.0190	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0280	0.0000	0.0000
Khoton	0.0780	0.0000	0.1420	0.0000	0.0060	0.0810	0.0000	0.0000	0.0290	0.0550	0.0000	0.0000
Khalka	0.0250	0.0000	0.0900	0.0000	0.0000	0.0600	0.0000	0.0250	0.0000	0.0500	0.0000	0.0000
Vietnam	0.0000	0.0000	0.1260	0.0000	0.0000	0.0000	0.0000	0.0180	0.0000	0.0230	0.0000	0.0000
Tsaatan	0.0350	0.0000	0.1000	0.0000	0.0000	0.0800	0.0000	0.0140	0.0000	0.0600	0.0100	0.0000
M Main China	0.1890	0.0000	0.2320	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Min Nan	0.3400	0.0000	0.0837	0.0000	0.0400	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Hakka	0.3900	0.0000	0.0750	0.0000	0.0700	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Pacific												
Macedonia	0.0570	0.0000	0.0690	0.0000	0.0000	0.0190	0.0310	0.0190	0.0190	0.0060	0.0000	0.0060
Phillipines	0.0000	0.0000	0.0000	0.0000	0.0000	0.0550	0.0000	0.0370	0.0000	0.0190	0.0000	0.0000
Cooks Isl	0.0100	0.0000	0.0100	0.0000	0.0000	0.0000	0.0000	0.2600	0.0000	0.0300	0.0000	0.0000
Samoa Isl	0.0000	0.0000	0.0200	0.0000	0.0000	0.0000	0.0000	0.1600	0.0000	0.0900	0.0000	0.0000
Tokelau Isl	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.3800	0.0000	0.0700	0.0000	0.0100
Tonga Isl	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.1700	0.0000	0.0400	0.0000	0.0000
Nauru	0.0000	0.0000	0.0000	0.0000	0.0000	0.0370	0.0000	0.1230	0.0000	0.0080	0.0000	0.0000
N. Caledonia	0.0080	0.0080	0.0000	0.0000	0.0000	0.0310	0.0000	0.0620	0.0000	0.0230	0.0000	0.0000

Table 6.1 DRB1 Continued.....

	13	14	15	16	17	18	19	20	21	22	23	24
Alleles	**0701**	**0801**	**0802**	**0803**	**080x**	**090x**	**1001**	**1101**	**1102**	**1103**	**1104**	**1105**
Kayastha	0.2150	0.0000	0.0000	0.0000	0.0026	0.0280	0.0736	0.0630	0.0000	0.0131	0.0026	0.0000
Vaish	0.2390	0.0000	0.0000	0.0000	0.0000	0.013	0.1013	0.0475	0.0063	0.0065	0.0032	0.0000
Mathur	0.2570	0.0000	0.0000	0.0000	0.0025	0.0176	0.0650	0.0455	0.0051	0.0025	0.0050	0.0000
Rastogi	0.1960	0.0000	0.0000	0.0000	0.0102	0.0380	0.0918	0.0510	0.0026	0.0050	0.0000	0.0000
Shia	0.1474	0.0000	0.0000	0.0000	0.0050	0.0368	0.0789	0.0660	0.0130	0.0160	0.0000	0.0130
Sunni	0.1596	0.0000	0.0000	0.0000	0.0080	0.0130	0.0904	0.0980	0.0027	0.0053	0.0027	0.0053
Lachung	0.1379	0.0000	0.0000	0.0000	0.0770	0.1030	0.0520	0.1120	0.0000	0.0080	0.0107	0.0000
Mech	0.1190	0.0000	0.0000	0.0000	0.1110	0.0870	0.0476	0.09.50	0.0000	0.0080	0.0080	0.0000
Rajbanshi	0.1390	0.0000	0.0000	0.0000	0.0920	0.0306	0.0570	0.0970	0.0000	0.0103	0.0000	0.0000
North Indian	0.1955	0.0000	0.0000	0.0000	0.0045	0.0091	0.0955	0.0682	0.0000	0.0091	0.0091	0.0000
China Naxi	0.0040	0.0170	0.0170	0.1100	0.0000	0.0640	0.0130	0.0680	0.0000	0.0000	0.0040	0.0000
Yunnan	0.0620	0.0080	0.0000	0.0930	0.0000	0.0970	0.0120	0.0190	0.0000	0.0040	0.0120	0.0000
China Wuhan	0.0710	0.0040	0.0080	0.0668	0.0000	0.1590	0.0080	0.0920	0.0040	0.0000	0.0000	0.0000
Uygur	0.1640	0.0180	0.0000	0.0000	0.0000	0.0350	0.0000	0.0260	0.0000	0.0530	0.0530	0.0000
Sd Han	0.1130	0.0000	0.0150	0.0560	0.0000	0.1030	0.0100	0.0560	0.0000	0.0000	0.0150	0.0000
Ch Bei Xian	0.1170	0.0000	0.0090	0.0560	0.0000	0.1460	0.0180	0.0940	0.0000	0.0000	0.0230	0.0000
Urumqi Kaz	*0.1150*	*0.0000*	*0.0120*	*0.0480*	*0.0000*	*0.0710*	*0.0120*	*0.0480*	*0.0000*	*0.0000*	*0.0000*	*0.0000*
South Korea	0.0600	0.0000	0.0200	0.0750	0.0000	0.0860	0.0140	0.0460	0.0000	0.0000	0.0020	0.0000
Thailand	0.0910	0.0000	0.0040	0.0210	0.0000	0.0940	0.0140	0.0550	0.0000	0.0000	0.0000	0.0000
Japanese	0.0040	0.0000	0.0420	0.0760	0.0000	0.1420	0.0070	0.0310	0.0000	0.0000	0.0000	0.0000
Ainu Japa	0.0000	0.0000	0.1200	0.0200	0.0000	0.1600	0.0000	0.0320	0.0000	0.0000	0.0000	0.0000
Japan Hyogo	0.0000	0.0000	0.0580	0.0850	0.0000	0.1100	0.0310	0.0000	0.0000	0.0000	0.0000	0.0000
Gypsy	0.0294	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.1580	0.0000	0.0290	0.0000	0.0000
Russia	0.1190	0.0230	0.0000	0.0120	0.0000	0.0090	0.0080	0.0790	0.0150	0.0000	0.0150	0.0000
Pomors	0.1230	0.0480	0.0000	0.0000	0.0000	0.0340	0.0000	0.0620	0.0000	0.0000	0.0000	0.0000
Saami	0.0370	0.0740	0.0000	0.0000	0.0000	0.0440	0.0000	0.0440	0.0000	0.0000	0.0000	0.0000
Nentsy	0.1180	0.0550	0.0000	0.0000	0.0000	0.1820	0.0000	0.0180	0.0000	0.0000	0.0000	0.0000
N Chinese	0.1370	0.0000	0.0110	0.0330	0.0000	0.1440	0.0220	0.0660	0.0000	0.0000	0.0060	0.0000
N E Russia	0.1270	0.0450	0.0000	0.0000	0.0000	0.0520	0.0000	0.1470	0.0000	0.0000	0.0000	0.0000
Nivkhi Rus	0.0860	0.0000	0.0280	0.0090	0.0000	0.1610	0.0090	0.0090	0.0000	0.0000	0.0000	0.0000
Khoton	0.0730	0.0150	0.0100	0.0120	0.0000	0.0640	0.0350	0.0320	0.0000	0.0000	0.0350	0.0000
Khalka	0.1200	0.0200	0.0300	0.0250	0.0000	0.0950	0.0400	0.0650	0.0000	0.0100	0.0050	0.0000
Vietnam	0.0060	0.0000	0.0150	0.0150	0.0000	0.0770	0.0300	0.0120	0.0000	0.0000	0.0000	0.0000
Tsaatan	0.1180	0.0700	0.0000	0.0210	0.0000	0.0280	0.0560	0.0760	0.0000	0.0000	0.0000	0.0000
M Main China	0.0890	0.0000	0.0000	0.0000	0.0000	0.0000	0.1420	0.0000	0.0000	0.0000	0.0000	0.0000
Min Nan	0.0219	0.0000	0.0000	0.0000	0.0000	0.0000	0.0534	0.0000	0.0000	0.0000	0.0000	0.0000
Hakka	0.0210	0.0000	0.0000	0.0000	0.0000	0.0000	0.0490	0.0000	0.0000	0.0000	0.0000	0.0000
Pacific												
Macedonia	0.0250	0.0120	0.0000	0.0000	0.0000	0.0000	0.0060	0.0950	0.0000	0.0190	0.1830	0.0000
Phillipines	0.0000	0.0000	0.0000	0.0400	0.0000	0.0100	0.0000	0.0550	0.0000	0.0000	0.0000	0.0000
Cooks Isl	0.0100	0.0000	0.0200	0.0800	0.0000	0.0800	0.0000	0.1200	0.0000	0.0000	0.0000	0.0000
Samoa Isl	0.0100	0.0000	0.0200	0.1000	0.0000	0.2500	0.0000	0.0800	0.0000	0.0000	0.0000	0.0000
Tokelau Isl	0.0000	0.0000	0.0000	0.1200	0.0000	0.0400	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Tonga Isl	0.0000	0.0100	0.0000	0.1200	0.0000	0.1600	0.0000	0.1600	0.0000	0.0000	0.0000	0.0000
Nauru	0.0000	0.0150	0.0000	0.0550	0.0000	0.0000	0.0000	0.0680	0.0000	0.0000	0.0000	0.0000
N. Caledonia	0.0150	0.0080	0.0000	0.0770	0.0000	0.0230	0.0000	0.3290	0.0000	0.0000	0.0000	0.0000

Table 6.1 DRB1Continued.....

	25	26	27	28	29	30	31	32	33	34	35	36
Alleles	**1106**	**1107**	**1108**	**1201**	**1202**	**1203**	**1301**	**1302**	**1303**	**1308**	**1401**	**1402**
Asians												
Kayastha	0.0026	0.0000	0.0000	0.0052	0.0289	0.0000	0.0863	0.0236	0.0157	0.0026	0.0500	0.0105
Vaish	0.0000	0.0000	0.0000	0.0065	0.0032	0.0000	0.0538	0.0190	0.0030	0.0000	0.0430	0.0065
Mathur	0.0000	0.0000	0.0025	0.0126	0.0202	0.0000	0.0631	0.0303	0.0080	0.0025	0.0404	0.0177
Rastogi	0.0000	0.0030	0.0000	0.0050	0.0230	0.0030	0.0510	0.0179	0.0230	0.0030	0.0332	0.0280
Shia	0.0000	0.0420	0.0026	0.0026	0.0158	0.0000	0.0237	0.0158	0.0132	0.0030	0.0240	0.0180
Sunni	0.0000	0.0340	0.0000	0.0370	0.0290	0.0027	0.0372	0.0100	0.0260	0.0000	0.0580	0.080
Lachung	0.0000	0.0000	0.0000	0.0770	0.0000	0.0000	0.0170	0.0260	0.0000	0.0000	0.0520	0.0000
Mech	0.0000	0.0000	0.0000	0.0870	0.0000	0.0000	0.0159	0.0160	0.0000	0.0000	0.0550	0.0000
Rajbanshi	0.0000	0.0000	0.0000	0.1020	0.0100	0.0000	0.0204	0.0360	0.0000	0.0000	0.0714	0.0000
North Indian	0.0000	0.0000	0.0045	0.0045	0.0181	0.0045	0.0773	0.0273	0.0000	0.0045	0.0436	0.0200
China Naxi	0.0000	0.0000	0.0000	0.0810	0.1340	0.0250	0.0170	0.0040	0.0040	0.0000	0.0550	0.0000
China Yunnan	0.0040	0.0000	0.0000	0.0810	0.0890	0.0040	0.0310	0.0310	0.0040	0.0000	0.0760	0.0000
China Wuhan	0.0000	0.0000	0.0000	0.0080	0.0920	0.0380	0.0220	0.0120	0.0000	0.0000	0.0590	0.0000
Uygur	0.0000	0.0000	0.0000	0.0530	0.0180	0.0000	0.0610	0.0440	0.0000	0.0000	0.0180	0.0000
Sd Han	0.0000	0.0000	0.0000	0.0200	0.0820	0.0000	0.0260	0.0510	0.0000	0.0000	0.0410	0.0000
Ch Bei Xian	0.0000	0.0000	0.0000	0.0410	0.0330	0.0000	0.0180	0.0000	0.0000	0.0000	0.0410	0.0000
Urumqi Kaz	0.0000	0.0000	0.0000	0.0600	0.0220	0.0000	0.0480	0.0120	0.0000	0.0000	0.0710	0.0000
South Korea	0.0020	0.0000	0.0000	0.0450	0.0270	0.0020	0.0110	0.1040	0.0000	0.0000	0.0410	0.0000
Thailand	0.0210	0.0000	0.0000	0.0110	0.1590	0.0000	0.0180	0.0115	0.0000	0.0000	0.0355	0.0040
Japanese	0.0000	0.0000	0.0000	0.0390	0.0120	0.0000	0.0000	0.0890	0.0000	0.0000	0.0390	0.0160
Ainu Japa	0.0520	0.0000	0.0000	0.0420	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.2220	0.0000
Japan Hyogo	0.0000	0.0000	0.0000	0.0310	0.0160	0.0000	0.0160	0.0470	0.0000	0.0000	0.0160	0.0000
Gypsy	0.0000	0.0000	0.0000	0.0000	0.0290	0.0000	0.0206	0.0590	0.0000	0.0000	0.0290	0.0000
Russia	0.0000	0.0000	0.0000	0.0270	0.0000	0.0000	0.0640	0.0370	0.0300	0.0000	0.0130	0.0000
Pomors	0.0000	0.0000	0.0000	0.0210	0.0000	0.0000	0.1110	0.0000	0.0000	0.0000	0.0000	0.0000
Saami	0.0000	0.0000	0.0000	0.0370	0.0000	0.0000	0.0940	0.0000	0.0000	0.0000	0.0000	0.0000
Nentsy	0.0000	0.0000	0.0000	0.1640	0.0000	0.0000	0.1180	0.0000	0.0000	0.0000	0.0000	0.0000
N Chinese	0.0000	0.0000	0.0000	0.0220	0.0440	0.0000	0.0170	0.0060	0.0000	0.0000	0.0440	0.0060
N E Russia	0.0000	0.0000	0.0000	0.0270	0.0000	0.0000	0.0840	0.0000	0.0000	0.0000	0.0000	0.0000
Nivkhi Rus	0.0940	0.0000	0.0000	0.1040	0.0000	0.0000	0.0090	0.0000	0.0000	0.0000	0.1230	0.0660
Khoton	0.0000	0.0060	0.0060	0.0120	0.0120	0.0000	0.0150	0.0280	0.0000	0.0000	0.0240	0.0000
Khalka	0.0050	0.0000	0.0000	0.0250	0.0450	0.0000	0.0350	0.0350	0.0000	0.0050	0.0450	0.0100
Vietnam	0.0000	0.0000	0.0000	0.0000	0.1010	0.0000	0.0060	0.0060	0.0130	0.0000	0.1500	0.0000
Tsaatan	0.0000	0.0000	0.0000	0.0560	0.0000	0.0000	0.0390	0.0460	0.0000	0.0000	0.0660	0.0000
M Main China	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Min Nan	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Hakka	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Pacific												
Macedonia	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0620	0.0440	0.0060	0.0000	0.0680	0.0000
Phillipines	0.0000	0.0000	0.0000	0.0000	0.2800	0.0000	0.0000	0.0000	0.0000	0.0000	0.0730	0.0000
Cooks Isl	0.0000	0.0000	0.0000	0.1600	0.0100	0.0000	0.0000	0.0000	0.0000	0.0000	0.0600	0.0000
Samoa Isl	0.0000	0.0000	0.0000	0.0800	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0100	0.0000
Tokelau Isl	0.0000	0.0000	0.0000	0.2200	0.0000	0.0000	0.0300	0.0000	0.0000	0.0000	0.0000	0.0000
Tonga Isl	0.0000	0.0000	0.0000	0.0800	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0500	0.0000
Nauru	0.0000	0.0000	0.0000	0.0080	0.2900	0.0000	0.0000	0.0150	0.0000	0.0000	0.0520	0.0000
New Caledonia	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0230	0.0000

Table 6.1 DRB1 Continued.....

	37	38	39	40	41	42	43	44	45
Alleles	**1403**	**1404**	**1405**	**1501**	**1502**	**1503**	**1504**	**1601**	**1602**
Asians									
Kayastha	0.0105	0.0000	0.0000	0.0970	0.0680	0.0210	0.0078	0.0000	0.0000
Vaish	0.0000	0.0000	0.0000	0.0840	0.1090	0.0300	0.013	0.0000	0.0000
Mathur	0.0080	0.0000	0.0000	0.0859	0.0859	0.0177	0.0080	0.0000	0.0000
Rastogi	0.0080	0.0000	0.0000	0.0990	0.0612	0.0179	0.0080	0.0000	0.0000
Shia	0.0026	0.0026	0.0000	0.0470	0.0210	0.0050	0.0050	0.0000	0.0000
Sunni	*0.0130*	*0.0000*	*0.0000*	*0.0340*	*0.0340*	*0.0100*	*0.0027*	*0.0000*	*0.0000*
Lachung	0.0000	0.0430	0.0170	0.0950	0.0520	0.0000	0.0000	0.0000	0.0000
Mech	0.0000	0.0320	0.0160	0.1350	0.0470	0.0000	0.0000	0.0000	0.0000
Rajbanshi	0.0000	0.02.04	0.0204	0.1370	0.0360	0.0000	0.0000	0.0000	0.0000
North Indian	0.0000	0.0000	0.0000	0.1171	0.0757	0.0000	0.0166	0.0045	0.0000
China Naxi	0.0090	0.0470	0.0090	0.0890	0.0170	0.0000	0.0000	0.0000	0.0170
China Yunnan	0.0230	0.0270	0.0040	0.1220	0.0270	0.0000	0.0000	0.0000	0.0200
China Wuhan	0.0000	0.0000	0.0000	0.1370	0.0210	0.0000	0.0000	0.0000	0.0552
Uygur	0.0090	0.0090	0.0000	0.0960	0.0350	0.0000	0.0000	0.0000	0.0000
Sd Han	0.0050	0.0100	0.0150	0.1650	0.0600	0.0000	0.0050	0.0000	0.0150
Ch Bei Xian	0.0060	0.0060	0.0120	0.0670	0.0530	0.0000	0.0000	0.0000	0.0230
Urumqi Kaz	0.0360	0.0060	0.0120	0.0560	0.0360	0.0000	0.0000	0.0000	0.0220
South Korea	0.0060	0.0000	0.0300	0.0840	0.0380	0.0000	0.0000	0.0000	0.0000
Thailand	0.0000	0.0355	0.0040	0.0780	0.1240	0.0000	0.0040	0.0000	0.0040
Japanese	0.0000	0.0000	0.0000	0.0690	0.0950	0.0000	0.0000	0.0000	0.0070
Ainu Japa	0.0000	0.0000	0.0000	0.0300	0.0150	0.0000	0.0000	0.0000	0.0000
Japan Hyogo	0.0000	0.0000	0.0310	0.1500	0.0630	0.0000	0.0000	0.0000	0.0310
Gypsy	0.0000	0.0294	0.0000	0.0290	0.1048	0.0000	0.0000	0.0290	0.1180
Russia	0.0000	0.0030	0.0030	0.1140	0.0180	0.0000	0.0000	0.0280	0.0160
Pomors	0.0000	0.0000	0.0000	0.1720	0.0000	0.0000	0.0000	0.0000	0.0000
Saami	0.0000	0.0000	0.0000	0.1270	0.0000	0.0000	0.0000	0.0000	0.0000
Nentsy	0.0000	0.0000	0.0000	0.0470	0.0000	0.0000	0.0000	0.0000	0.0000
N Chinese	0.0110	0.0000	0.0000	0.1240	0.0780	0.0000	0.0000	0.0000	0.0280
N E Russia	0.0000	0.0000	0.0000	0.1340	0.0000	0.0000	0.0000	0.0000	0.0000
Nivkhi Rus	0.0850	0.0000	0.0380	0.0280	0.0090	0.0000	0.0000	0.0000	0.0000
Khoton	0.0120	0.0000	0.0000	0.0370	0.0800	0.0000	0.0000	0.0120	0.0060
Khalka	0.0050	0.0000	0.0000	0.0600	0.0350	0.0000	0.0000	0.0000	0.0050
Vietnam	0.0000	0.0000	0.0000	0.0540	0.1310	0.0000	0.0060	0.0000	0.1990
Tsaatan	0.0000	0.0000	0.0000	0.0530	0.0420	0.0000	0.0000	0.0000	0.0000
M Main China	0.0000	0.0000	0.0000	0.1323	0.0000	0.0000	0.0000	0.0267	0.0000
Min Nan	0.0000	0.0000	0.0000	0.0920	0.0000	0.0000	0.0000	0.0390	0.0000
Hakka	0.0000	0.0000	0.0000	0.0170	0.0000	0.0000	0.0000	0.0380	0.0000
Pacific									
Macedonia	0.0000	0.0120	0.0000	0.1080	0.0120	0.0000	0.0000	0.0980	0.0120
Phillipines	0.0000	0.0200	0.0100	0.0100	0.3410	0.0000	0.0000	0.0000	0.0500
Cooks Isl	0.0000	0.0000	0.0000	0.0100	0.0600	0.0000	0.0000	0.0000	0.0000
Samoa Isl	0.0000	0.0000	0.0000	0.0300	0.0700	0.0000	0.0000	0.0000	0.0000
Tokelau Isl	0.0000	0.0000	0.0000	0.0000	0.0600	0.0000	0.0000	0.0000	0.0000
Tonga Isl	0.0000	0.0000	0.0000	0.0300	0.1000	0.0000	0.0000	0.0000	0.0000
Nauru	0.0000	0.0000	0.0080	0.0820	0.2390	0.0000	0.0000	0.0000	0.0000
New Caledonia	0.0000	0.0000	0.0000	0.1390	0.1150	0.0000	0.0000	0.0000	0.0080

Table 6.1 DRB1Continued.....

	1	2	3	4	5	6	7	8	9	10	11	12
Alleles	**0101**	**0103**	**0301**	**0302**	**0303**	**0401**	**0402**	**0403**	**0404**	**0405**	**0406**	**0407**
Europeans												
Europeans	0.0770	0.0080	0.0970	0.0000	0.0000	0.0970	0.0040	0.0160	0.0810	0.0040	0.0000	0.0080
Basque	0.0200	0.0070	0.1730	0.0000	0.0000	0.0200	0.0200	0.0470	0.0200	0.0000	0.0000	0.0070
Czech Republic	0.0750	0.0050	0.0710	0.0050	0.0000	0.0670	0.0100	0.0050	0.0140	0.0100	0.0000	0.0000
German	0.1050	0.0030	0.1140	0.0000	0.0000	0.0750	0.0000	0.0030	0.0230	0.0030	0.0000	0.0030
Italy	0.0680	0.0030	0.0700	0.0010	0.0000	0.0200	0.0170	0.0150	0.0060	0.0110	0.0050	0.0080
Majorca	0.1509	0.0000	0.1321	0.0000	0.0000	0.0991	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Minorca	0.1304	0.0000	0.1232	0.0000	0.0000	0.1087	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Ibiza	0.1154	0.0077	0.1154	0.0000	0.0000	0.0846	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Chuetas	0.0576	0.0000	0.0632	0.0000	0.0000	0.1494	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Belgium	0.0870	0.0050	0.1580	0.0000	0.0000	0.0630	0.0050	0.0050	0.0350	0.0150	0.0000	0.0100
Azores	0.0820	0.0170	0.0900	0.0020	0.0000	0.0300	0.0300	0.0170	0.0220	0.0130	0.0000	0.0130
Eng Caucas	0.1030	0.0250	0.1280	0.0000	0.0000	0.1240	0.0080	0.0680	0.0000	0.0230	0.0000	0.0000
France South	0.0820	0.0000	0.1100	0.0000	0.0000	0.0510	0.0140	0.0090	0.0000	0.0170	0.0000	0.0090
Greece	0.0890	0.0000	0.1000	0.0000	0.0000	0.0420	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Bulgaria	0.0580	0.0000	0.0820	0.0000	0.0000	0.0090	0.0360	0.0090	0.0470	0.0090	0.0000	0.0090
Croatians	0.0820	0.0000	0.0638	0.0036	0.0000	0.0390	0.0245	0.0245	0.0036	0.0036	0.0000	0.0071
Polish	0.0959	0.0101	0.0757	0.0000	0.0000	0.0354	0.0051	0.0101	0.0000	0.0000	0.0151	0.0101
Romania	0.0720	0.0000	0.1330	0.0000	0.0000	0.0360	0.0240	0.0120	0.0120	0.0060	0.0000	0.0000
Mexi Mazatecan	0.0070	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0980	0.0000	0.0191	0.0000	0.3210
C Athabaskan	0.0080	0.0000	0.0310	0.0000	0.0000	0.0080	0.0000	0.0970	0.0000	0.0080	0.0000	0.0240
Italy Bergamo	0.0600	0.0050	0.0990	0.0000	0.0000	0.0410	0.0050	0.0050	0.0150	0.0000	0.0000	0.0050
Cantabr	0.0420	0.0120	0.0180	0.0000	0.0000	0.0360	0.0180	0.0540	0.0420	0.0000	0.0000	0.0060
Cabuern	0.0040	0.0050	0.0100	0.0000	0.0000	0.0210	0.0160	0.0630	0.0210	0.0000	0.0000	0.0050
Pasiegos	0.0480	0.0050	0.0160	0.0000	0.0000	0.0380	0.0160	0.0430	0.0380	0.0000	0.0000	0.0050
Turkey	0.0580	0.0000	0.0920	0.0000	0.0000	0.0200	0.0260	0.0380	0.0200	0.0180	0.0000	0.0100
Aka Pygmies	0.0050	0.0000	0.1180	0.0000	0.0000	0.0000	0.0000	0.0050	0.0000	0.0110	0.0000	0.0000
Bantu Congolese	0.0180	0.0000	0.0650	0.0290	0.0000	0.0120	0.0000	0.0000	0.0000	0.0120	0.0060	0.0000
Americans												
Argentina Aires	0.0580	0.0120	0.1090	0.0000	0.0000	0.0310	0.0170	0.0140	0.0170	0.0050	0.0000	0.0190
Archirigu	0.0090	0.0000	0.0000	0.0000	0.0000	0.0090	0.0000	0.0180	0.0090	0.0000	0.0000	0.0540
Ar Toba	0.0350	0.0230	0.0060	0.0000	0.0000	0.0060	0.0000	0.0520	0.0580	0.0060	0.0000	0.0810
Brazil	0.0520	0.0040	0.0840	0.0160	0.0000	0.0260	0.0240	0.0160	0.0300	0.0220	0.0060	0.0120
Guar Mbya	0.0000	0.0000	0.0160	0.0000	0.0000	0.0000	0.0000	0.0000	0.0050	0.0000	0.0000	0.0000
Gua Kaiowa	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0030	0.0000	0.0000	0.0810
Guanandeva	0.0110	0.0110	0.0060	0.0060	0.0000	0.0180	0.0000	0.0000	0.0000	0.0000	0.0000	0.0110
Ache	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0060	0.0000	0.0000	0.0000	0.0000
Kaingang	0.0000	0.0000	0.0000	0.0000	0.0000	0.0080	0.0000	0.0000	0.2590	0.0000	0.0000	0.0000
Ticuna	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0410	0.0200	0.0000	0.0000	0.0000
Mexi Mazatec	0.0070	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0980	0.0000	0.0191	0.0000	0.3210
Kogui	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0950	0.0000	0.0000	0.0000	0.5850
Ijka	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0700	0.0000	0.0000	0.0000	0.1200
Guahibo	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0200	0.0200	0.0000	0.0000	0.0900
Nukak	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.1500	0.0000	0.0000	0.0000
Sikuani	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.3200	0.0000	0.0000	0.0000
Ingano	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0900	0.1400	0.0000	0.0000	0.1800
Waunana	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0500	0.0200	0.0000	0.0000	0.1450
Embera	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0600	0.0350	0.0350	0.0000	0.0250

Table 6.1 DRB1Continued.....

	13	14	15	16	17	18	19	20	21	22	23	24
Alleles	**0701**	**0801**	**0802**	**0803**	**080x**	**090x**	**1001**	**1101**	**1102**	**1103**	**1104**	**1105**
Europeans												
Europeans	0.1450	0.0320	0.0000	0.0000	0.0000	0.0000	0.0120	0.0770	0.0040	0.0120	0.0000	0.0000
Basque	0.3130	0.0000	0.0000	0.0000	0.0000	0.0000	0.0130	0.0270	0.0000	0.0000	0.0530	0.0000
Czech Republic	0.1610	0.0330	0.0000	0.0100	0.0000	0.0050	0.0100	0.0670	0.0100	0.0050	0.0430	0.0000
German	0.0920	0.0260	0.0030	0.0030	0.0000	0.0140	0.0030	0.0600	0.0000	0.0060	0.0350	0.0000
Italy	0.1680	0.0120	0.0000	0.0050	0.0000	0.0010	0.0060	0.1380	0.0150	0.0230	0.1030	0.0000
Majorca	0.1415	0.0519	0.0000	0.0000	0.0000	0.0047	0.0094	0.1509	0.0000	0.0000	0.0000	0.0000
Minorca	0.1957	0.0217	0.0000	0.0000	0.0000	0.0072	0.0145	0.1449	0.0000	0.0000	0.0000	0.0000
Ibiza	0.2462	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.2077	0.0000	0.0000	0.0000	0.0000
Chuetas	0.2069	0.0172	0.0000	0.0000	0.0000	0.0000	0.0345	0.2356	0.0000	0.0000	0.0000	0.0000
Belgium	0.1120	0.0310	0.0000	0.0100	0.0000	0.0100	0.0150	0.0720	0.0100	0.0100	0.0150	0.0000
Azores	0.1920	0.0370	0.0000	0.0070	0.0000	0.0280	0.0090	0.0048	0.0200	0.0130	0.0350	0.0000
Eng Caucas	0.1330	0.0200	0.0030	0.0000	0.0000	0.0140	0.0030	0.0420	0.0140	0.0000	0.0000	0.0000
France South	0.2200	0.0580	0.0000	0.0000	0.0000	0.0080	0.0080	0.0950	0.0000	0.0000	0.0000	0.0000
Greece	0.0320	0.0100	0.0000	0.0000	0.0000	0.0000	0.0100	0.2920	0.0000	0.0000	0.0000	0.0000
Bulgaria	0.0650	0.0270	0.0000	0.0000	0.0000	0.0000	0.0000	0.0820	0.0000	0.0000	0.1550	0.0000
Croatians	0.1000	0.0248	0.0000	0.0036	0.0000	0.0036	0.0177	0.0606	0.0213	0.0000	0.0770	0.0000
Polish	0.1212	0.0151	0.0000	0.0000	0.0000	0.0101	0.0101	0.0657	0.0051	0.0101	0.0151	0.0000
Romania	0.0960	0.0060	0.0060	0.0000	0.0000	0.0060	0.0240	0.0720	0.0060	0.0000	0.0540	0.0000
Mexi Mazatecan	0.0070	0.0000	0.1519	0.0000	0.0000	0.0000	0.0070	0.0070	0.0000	0.0000	0.0000	0.0000
Canada Athabaskan	0.0160	0.0000	0.0480	0.0000	0.0000	0.1040	0.0000	0.0080	0.0000	0.0000	0.0000	0.0000
Italy Bergamo	0.1290	0.0250	0.0000	0.0100	0.0000	0.0000	0.0050	0.1580	0.0200	0.0410	0.0410	0.0000
Cantabr	0.2050	0.0060	0.0000	0.0000	0.0000	0.0060	0.0120	0.0540	0.0180	0.0000	0.0720	0.0000
Cabuern	0.1680	0.0000	0.0000	0.0000	0.0000	0.0000	0.0100	0.1050	0.0050	0.0000	0.1420	0.0000
Pasiegos	0.1790	0.0110	0.0000	0.0000	0.0000	0.0050	0.0160	0.0430	0.0160	0.0000	0.0650	0.0000
Turkey	0.0900	0.0140	0.0100	0.0080	0	0.0100	0.0200	0.1040	0.0140	0.0160	0.0620	0.0000
Aka Pygmies	0.2410	0.0000	0.0000	0.0000	0.0000	0.0110	0.0110	0.0000	0.0000	0.0000	0.0000	0.0000
Bantu Congolese	0.0650	0.0530	0.0000	0.0000	0.0000	0.0530	0.0180	0.0940	0.0590	0.0000	0.0290	0.0000
Americans												
Argentina Aires	0.1360	0.0140	0.0270	0.0050	0.0000	0.0190	0.0120	0.0650	0.0070	0.0100	0.0310	0.0000
Archirigu	0.0180	0.0000	0.1070	0.0000	0.0000	0.0180	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Ar Toba	0.0520	0.0060	0.2030	0.0000	0.0000	0.0170	0.0000	0.0120	0.0000	0.0000	0.0000	0.0000
Brazil	0.1180	0.0170	0.0170	0.0060	0.0000	0.0170	0.0220	0.0750	0.0140	0.0040	0.0190	0.0000
Guar Mbya	0.0050	0.0000	0.0270	0.0000	0.0000	0.0270	0.0050	0.0110	0.0000	0.0000	0.0000	0.0000
GUA Kaiowa	0.0000	0.0000	0.1000	0.0000	0.0000	0.1060	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Guanandeva	0.0170	0.0000	0.0990	0.0000	0.0000	0.0470	0.0000	0.0390	0.0060	0.0000	0.0220	0.0000
Ache	0.0000	0.0000	0.0340	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Kaingang	0.0020	0.0000	0.5100	0.0000	0.0000	0.0040	0.0200	0.0160	0.0270	0.0000	0.0000	0.0000
Ticuna	0.0000	0.0100	0.0310	0.0000	0.0000	0.0810	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Mexi Mazatec	0.0070	0.0000	0.1519	0.0000	0.0000	0.0000	0.0070	0.0070	0.0000	0.0000	0.0000	0.0000
Kogui	0.0000	0.0200	0.1300	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Ijka	0.0700	0.0200	0.6200	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0200	0.0000
Guahibo	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Nukak	0.0000	0.0000	0.0300	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Sikuani	0.0000	0.0000	0.1800	0.0000	0.0000	0.0200	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Ingano	0.0000	0.0500	0.3600	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Waunana	0.0000	0.0000	0.0300	0.0000	0.0000	0.0200	0.0000	0.0000	0.0200	0.0000	0.0000	0.0000
Embera	0.0000	0.0000	0.0950	0.0000	0.0000	0.0000	0.0000	0.0000	0.0550	0.0000	0.0600	0.0000

Table 6.1 DRB1Continued.....

	25	26	27	28	29	30	31	32	33	34	35	36
Alleles	**1106**	**1107**	**1108**	**1201**	**1202**	**1203**	**1301**	**1302**	**1303**	**1308**	**1401**	**1402**
Europeans												
Europeans	0.0000	0.0000	0.0000	0.0080	0.0000	0.0000	0.0440	0.0280	0.0120	0.0000	0.0400	0.0000
Basque	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0600	0.0200	0.0070	0.0000	0.0000	0.0000
Czech Republic	0.0000	0.0000	0.0000	0.0430	0.0000	0.0000	0.0850	0.0240	0.0330	0.0000	0.0380	0.0000
German	0.0000	0.0000	0.0000	0.0320	0.0000	0.0000	0.0910	0.0260	0.0120	0.0000	0.0320	0.0000
Italy	0.0000	0.0000	0.0000	0.0170	0.0000	0.0000	0.0480	0.0290	0.0260	0.0000	0.0680	0.0010
Majorca	0.0000	0.0000	0.0000	0.0094	0.0000	0.0000	0.0991	0.0000	0.0000	0.0000	0.0472	0.0000
Minorca	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.1304	0.0000	0.0000	0.0000	0.0000	0.0000
Ibiza	0.0000	0.0000	0.0000	0.0077	0.0000	0.0000	0.0692	0.0000	0.0000	0.0000	0.0231	0.0000
Chuetas	0.0000	0.0000	0.0000	0.0057	0.0000	0.0000	0.0977	0.0000	0.0000	0.0000	0.0345	0.0000
Belgium	0.0000	0.0000	0.0000	0.0100	0.0000	0.0000	0.0770	0.0360	0.0000	0.0000	0.0160	0.0000
Azores	*0.0000*	*0.0000*	*0.0000*	*0.0240*	*0.0000*	*0.0000*	*0.0820*	*0.0250*	*0.0100*	*0.0000*	*0.0090*	*0.0000*
Eng Caucas	0.0000	0.0000	0.0000	0.0140	0.0000	0.0000	0.0420	0.0250	0.0030	0.0000	0.0280	0.0000
France South	0.0000	0.0000	0.0000	0.0290	0.0000	0.0000	0.1050	0.0000	0.0000	0.0000	0.0630	0.0000
Greece	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.1150	0.0000	0.0000	0.0000	0.0420	0.0000
Bulgaria	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.1070	0.0270	0.0000	0.0000	0.0090	0.0000
Croatians	0.0000	0.0000	0.0000	0.0177	0.0000	0.0000	0.0625	0.0450	0.0177	0.0000	0.0450	0.0000
Polish	0.0101	0.0051	0.0000	0.0101	0.0051	0.0051	0.0707	0.0000	0.0252	0.0000	0.0051	0.0000
Romania	0.0000	0.0000	0.0000	0.0300	0.0060	0.0000	0.0540	0.0180	0.0120	0.0000	0.0300	0.0000
Mexi Mazatecan	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0350
Canada Athabaskan	0.0000	0.0000	0.0000	0.0880	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.1670	0.3450
Italy Bergamo	0.0000	0.0000	0.0000	0.0250	0.0050	0.0000	0.0540	0.0300	0.0100	0.0000	0.0540	0.0000
Cantabr	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0840	0.0180	0.0060	0.0000	0.0000	0.0000
Cabuern	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0580	0.0160	0.0000	0.0000	0.0000	0.0000
Pasiegos	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0760	0.0160	0.0050	0.0000	0.0000	0.0000
Turkey	0.0000	0.0000	0.0000	0.0180	0.0020	0.0000	0.0420	0.0500	0.0200	0.0000	0.0480	0.0020
Aka Pygmies	0.0000	0.0000	0.0000	0.1290	0.0160	0.0000	0.1610	0.0480	0.0000	0.0000	0.0380	0.0000
Bantu Congolese	0.0000	0.0000	0.0000	0.0290	0.0000	0.0000	0.0650	0.1110	0.0060	0.0060	0.0060	0.0000
Americans												
Argentina Aires	0.0020	0.0000	0.0020	0.0100	0.0000	0.0000	0.0560	0.0590	0.0100	0.0000	0.0230	0.0140
Archirigu	0.0000	0.0000	0.0000	0.0450	0.0000	0.0000	0.0000	0.0090	0.0000	0.0000	0.0000	0.2140
Ar Toba	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0120	0.0170	0.0000	0.0000	0.0060	0.1450
Brazil	0.0000	0.0000	0.0000	0.0140	0.0000	0.0000	0.0590	0.0460	0.0080	0.0000	0.0240	0.0150
Guar Mbya	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0050
GUA Kaiowa	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.2080
Guanandeva	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.2050
Ache	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0530
Kaingang	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0020	0.0000	0.0000	0.0000	0.0000	0.0160
Ticuna	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Mexi Mazatec	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0350
Kogui	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.1700
Ijka	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0300
Guahibo	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.4000
Nukak	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.6400
Sikuani	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Ingano	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Waunana	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.2670
Embera	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.2400

Table 6.1 DRB1 Continued.....

	37	38	39	40	41	42	43	44	45
Alleles	**1403**	**1404**	**1405**	**1501**	**1502**	**1503**	**1504**	**1601**	**1602**
Europeans									
Europeans	0.0000	0.0040	0.0000	0.1820	0.0040	0.0000	0.0000	0.0000	0.0000
Basque	0.0000	0.0000	0.0000	0.1660	0.0000	0.0000	0.0000	0.0000	0.0000
Czech Republic	0.0000	0.0000	0.0000	0.1080	0.0190	0.0000	0.0000	0.0240	0.0050
German	0.0000	0.0000	0.0000	0.1720	0.0090	0.0000	0.0000	0.0260	0.0000
Italy	0.0000	0.0060	0.0000	0.0550	0.0120	0.0000	0.0000	0.0220	0.0010
Majorca	0.0000	0.0000	0.0000	0.1038	0.0000	0.0000	0.0000	0.0000	0.0000
Minorca	0.0000	0.0000	0.0000	0.1233	0.0000	0.0000	0.0000	0.0000	0.0000
Ibiza	0.0000	0.0000	0.0000	0.1230	0.0000	0.0000	0.0000	0.0000	0.0000
Chuetas	0.0000	0.0000	0.0000	0.0977	0.0000	0.0000	0.0000	0.0000	0.0000
Belgium	0.0000	0.0000	0.0000	0.1430	0.0000	0.0000	0.0000	0.0250	0.0050
Azores	0.0000	0.0020	0.0000	0.0840	0.0070	0.0000	0.0000	0.0180	0.0020
Eng Caucas	0.0000	0.0000	0.0000	0.1470	0.0080	0.0000	0.0000	0.0080	0.0000
France South	0.0000	0.0000	0.0000	0.1190	0.0000	0.0000	0.0000	0.0000	0.0000
Greece	0.0000	0.0000	0.0000	0.0420	0.0420	0.0000	0.0000	0.1420	0.0420
Bulgaria	0.0000	0.0090	0.0000	0.0180	0.0370	0.0000	0.0000	0.1510	0.0180
Croatians	0.0000	0.0000	0.0000	0.0900	0.0142	0.0071	0.0000	0.1050	0.0177
Polish	0.0000	0.0000	0.0000	0.1616	0.0151	0.0101	0.0101	0.0252	0.0000
Romania	0.0000	0.0000	0.0180	0.0740	0.0420	0.0000	0.0000	0.1090	0.0180
Mexi Mazatecan	0.0000	0.0000	0.0000	0.0000	0.0420	0.0070	0.0000	0.0000	0.2210
Canada Athabaskan	0.0000	0.0000	0.0000	0.0080	0.0000	0.0000	0.0000	0.0000	0.0000
Italy Bergamo	0.0000	0.0000	0.0000	0.0640	0.0000	0.0000	0.0000	0.0540	0.0000
Cantabr	0.0000	0.0000	0.0000	0.2550	0.0000	0.0000	0.0000	0.0000	0.0000
Cabuern	0.0000	0.0000	0.0000	0.3360	0.0000	0.0000	0.0000	0.0000	0.0000
Pasiegos	0.0000	0.0000	0.0000	0.3270	0.0000	0.0000	0.0000	0.0000	0.0000
Turkey	0.0000	0.0060	0.0000	0.0640	0.0400	0.0040	0.0000	0.0000	0.0600
Aka Pygmies	0.0000	0.0000	0.0000	0.0130	0.0070	0.1160	0.0000	0.0050	0.0000
Bantu Congolese	0.0000	0.0000	0.0000	0.0999	0.0000	0.0000	0.0000	0.0000	0.0000
Americans									
Argentina Aires	0.0000	0.0020	0.0000	0.0700	0.0170	0.0070	0.0000	0.0360	0.0090
Archirigu	0.0000	0.0000	0.0000	0.0090	0.0180	0.0000	0.0000	0.0000	0.1860
Ar Toba	0.0000	0.0000	0.0000	0.0060	0.0000	0.0000	0.0000	0.0000	0.0060
Brazil	0.0000	0.0000	0.0000	0.0380	0.0020	0.0290	0.0000	0.0140	0.0340
Guar Mbya	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.3730
GUA Kaiowa	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.3390
Guanandeva	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.3100
Ache	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Kaingang	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.1300
Ticuna	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.2450
Mexi Mazatec	0.0000	0.0000	0.0000	0.0000	0.0420	0.0070	0.0000	0.0000	0.2210
Kogui	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Ijka	0.0000	0.0000	0.0000	0.0200	0.0000	0.0100	0.0000	0.0000	0.0000
Guahibo	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.3900
Nukak	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0800
Sikuani	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0200
Ingano	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Waunana	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.3300
Embera	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.3400

Table 6.1 DRB1 Continued.....

Alleles	0101	0103	0301	0302	0303	0401	0402	0403	0404	0405	0406	0407
Tule	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.2200	0.0000	0.0300	0.0000	0.1200
Brazil Cent	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0700	0.0000	0.0000	0.0770
Ny Af Am	0.0170	0.0000	0.0680	0.0660	0.0000	0.0190	0.0000	0.0000	0.0080	0.0150	0.0000	0.0020
N W Colombia	0.0650	0.0000	0.0650	0.0050	0.0000	0.0000	0.0050	0.0000	0.0000	0.0000	0.0000	0.0150
Aymara	0.0000	0.0000	0.0150	0.0000	0.0000	0.0000	0.0000	0.0340	0.0830	0.0100	0.0000	0.0930
African												
Cameroon	0.0040	0.0000	0.0710	0.0670	0.0000	0.0000	0.0000	0.0000	0.0000	0.0080	0.0000	0.0000
Oromo	0.0000	0.0000	0.1080	0.0000	0.0000	0.0240	0.0000	0.0300	0.0300	0.0540	0.0180	0.0000
Amhara	0.0000	0.0050	0.0760	0.0000	0.0000	0.0100	0.0000	0.0200	0.0100	0.0510	0.0150	0.0000
Algerian	0.0210	0.0320	0.1380	0.0000	0.0000	0.0000	0.0960	0.0110	0.0000	0.0320	0.0320	0.0000
Arab Moracc	0.0200	0.0000	0.1480	0.0100	0.0000	0.0100	0.0460	0.0200	0.0150	0.0510	0.0200	0.0000
Gabonese	0.0090	0.0000	0.0690	0.0090	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Jerba	0.0484	0.0455	0.0273	0.0000	0.0000	0.0000	0.0293	0.0364	0.0000	0.0000	0.0000	0.0000
Matmat	0.0185	0.0000	0.1111	0.0062	0.0000	0.0123	0.0617	0.0370	0.0062	0.0000	0.0000	0.0000
Gabes	0.0000	0.0000	0.1700	0.0052	0.0000	0.0270	0.0288	0.0365	0.0104	0.0156	0.0000	0.0000
Providencia	0.0700	0.0300	0.1500	0.0700	0.0000	0.0800	0.0000	0.0200	0.0000	0.0000	0.0000	0.0300
Cauca	0.0500	0.0000	0.1100	0.2400	0.0000	0.0000	0.0000	0.0000	0.0000	0.0250	0.0800	0.0050
Choco	0.0250	0.0000	0.1000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0500
Australasia												
Ausl Cape Pen	0.0100	0.0050	0.0450	0.0000	0.0000	0.0050	0.0000	0.0100	0.0150	0.0860	0.0000	0.0000
Aul Ab Kimberly	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0490	0.0000	0.0000
Middle East												
Iran Baloch	0.1500	0.0200	0.2900	0.0850	0.0000	0.0000	0.0000	0.0100	0.0000	0.0150	0.0000	0.0000
Iran Yazd Parsi	0.0690	0.0230	0.0540	0.0000	0.0000	0.0000	0.0150	0.0770	0.0000	0.0000	0.0000	0.0540

Alleles	0701	0801	0802	0803	080x	090x	1001	1101	1102	1103	1104	1105
Tule	0.0200	0.0000	0.0300	0.0000	0.0000	0.0000	0.0000	0.0200	0.0000	0.0000	0.0000	0.0000
Brazil Cent	0.0000	0.0000	0.2630	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Ny Af Am	0.1220	0.0040	0.0040	0.0230	0.0000	0.0060	0.0190	0.0810	0.0290	0.0020	0.0120	0.0000
N W Colombia	0.1785	0.1245	0.0000	0.0000	0.0000	0.0205	0.0155	0.0150	0.0000	0.0000	0.0000	0.0000
Aymara	0.0050	0.0000	0.2990	0.0000	0.0000	0.1950	0.0050	0.0000	0.0000	0.0000	0.0050	0.0000
African												
Cameroon	0.0550	0.0000	0.0000	0.0000	0.0000	0.0040	0.0110	0.0480	0.0310	0.0000	0.0000	0.0000
Oromo	0.2210	0.0000	0.0000	0.0000	0.0000	0.0000	0.0240	0.0180	0.0060	0.0000	0.0000	0.0000
Amhara	0.2030	0.0000	0.0000	0.0000	0.0000	0.0000	0.0306	0.0250	0.0000	0.0000	0.0000	0.0000
Algerian	0.1380	0.0210	0.0000	0.0110	0.0000	0.0000	0.0210	0.1600	0.0430	0.0000	0.0110	0.0000
Arab Moracc	0.1490	0.0260	0.0000	0.0000	0.0000	0.0100	0.0310	0.0400	0.0260	0.0000	0.0020	0.0000
Gabonese	0.0660	0.0030	0.0030	0.0000	0.0000	0.0270	0.0030	0.1350	0.0780	0.0000	0.0000	0.0000
Jerba	0.1282	0.0364	0.0000	0.0182	0.0000	0.0455	0.0316	0.1095	0.0818	0.0000	0.0000	0.0000
Matmat	0.1792	0.0185	0.0123	0.0062	0.0000	0.0123	0.0185	0.0432	0.0185	0.0000	0.0370	0.0000
Gabes	0.1912	0.0000	0.0052	0.0000	0.0000	0.0052	0.0313	0.1117	0.0156	0.0000	0.0052	0.0000
Providencia	0.0300	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0200	0.0000	0.0300	0.0200	0.0000
Cauca	0.0500	0.0000	0.0500	0.0000	0.0000	0.0250	0.0250	0.0250	0.0250	0.0000	0.0250	0.0000
Choco	0.2200	0.0000	0.0000	0.0000	0.0000	0.0500	0.0250	0.0250	0.0000	0.0000	0.0000	0.0000
Australasia												
Ausl Cape Pen	0.0400	0.0000	0.0000	0.3560	0.0000	0.0000	0.0050	0.0000	0.0000	0.0000	0.0000	0.0000
Aul Ab Kimberly	0.0000	0.0000	0.0000	0.2560	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Middle East												
Iran Baloch	0.0300	0.0050	0.0000	0.0000	0.0000	0.0000	0.0000	0.0850	0.0000	0.0000	0.0000	0.0000
Iran Yazd Parsi	0.2230	0.0540	0.0000	0.0000	0.0000	0.0080	0.0080	0.1460	0.0150	0.0000	0.0000	0.0000

Table 6.1 DRB1 Continued.....

Alleles	1106	1107	1108	1201	1202	1203	1301	1302	1303	1308	1401	1402
Tule	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.2600
Brazil Cent	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.2950
Ny Af Am	0.0000	0.0000	0.0000	0.0390	0.0040	0.0000	0.0370	0.0850	0.0120	0.0000	0.0230	0.0000
N W Colombia	0.0000	0.0000	0.0000	0.0055	0.0055	0.0050	0.1050	0.0000	0.0000	0.0000	0.0990	0.0000
Aymara	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0050	0.0050	0.0000	0.0000	0.0000	0.1670
African												
Cameroon	0.0000	0.0000	0.0000	0.0590	0.0080	0.0000	0.1300	0.0430	0.0350	0.0000	0.0150	0.0000
Oromo	0.0000	0.0000	0.0000	0.0060	0.0000	0.0000	0.0240	0.1730	0.0300	0.0000	0.0000	0.0000
Amhara	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0200	0.1778	0.0240	0.0000	0.0100	0.0000
Algerian	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0110	0.0210	0.0320	0.0040	0.0210	0.0210
Arab Moracc	0.0000	0.0000	0.0000	0.0050	0.0000	0.0000	0.0360	0.0920	0.0410	0.0000	0.0150	0.0000
Gabonese	0.0000	0.0000	0.0000	0.0660	0.0000	0.0000	0.1230	0.0570	0.0000	0.0000	0.0120	0.0000
Jerba	0.0000	0.0000	0.0000	0.0040	0.0000	0.0000	0.0828	0.0364	0.0545	0.0000	0.0000	0.0000
Matmat	0.0000	0.0000	0.0062	0.0000	0.0000	0.0000	0.0802	0.0309	0.0185	0.0000	0.0000	0.0000
Gabes	0.0052	0.0000	0.0000	0.0156	0.0000	0.0000	0.0280	0.0320	0.0593	0.0000	0.0054	0.0000
Providencia	0.0000	0.0000	0.0000	0.0700	0.0000	0.0000	0.0300	0.0800	0.0000	0.0000	0.0300	0.0200
Cauca	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0250	0.0000	0.0000	0.0000	0.0000	0.0000
Choco	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0250	0.0500	0.1000	0.0000	0.0000	0.0000
Australasia												
Ausl Cape Pen	0.0000	0.0000	0.0000	0.0050	0.0050	0.0000	0.0150	0.0000	0.0000	0.0000	0.0450	0.0710
Aul Ab Kimberly	0.0000	0.0000	0.0000	0.0370	0.0000	0.0000	0.0120	0.0000	0.0000	0.0000	0.0490	0.0000
Middle East												
Iran Baloch	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0450	0.0000	0.0000	0.0000	0.0000	0.0000
Iran Yazd Parsi	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.1540	0.0000	0.0000	0.0000	0.0080	0.0000

Alleles	1403	1404	1405	1501	1502	1503	1504	1601	1602
Tule	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.1600
Brazil Cent	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.2950
Ny Af Am	0.0000	0.0000	0.0000	0.1620	0.0000	0.0060	0.0000	0.0190	0.0000
N W Colombia	0.0000	0.0000	0.0000	0.1300	0.0000	0.0000	0.0000	0.0100	0.0360
Aymara	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0690
African									
Cameroon	0.0000	0.0000	0.0000	0.0000	0.0000	0.2960	0.0000	0.0000	0.0000
Oromo	0.0000	0.0000	0.0000	0.0300	0.0000	0.0900	0.0000	0.0000	0.0000
Amhara	0.0000	0.0000	0.0000	0.0150	0.0000	0.0920	0.0000	0.0000	0.0000
Algerian	0.0000	0.0000	0.0000	0.0490	0.0210	0.0000	0.0000	0.0000	0.0210
Arab Moracc	0.0000	0.0000	0.0000	0.0760	0.0150	0.0100	0.0000	0.0100	0.0000
Gabonese	0.0000	0.0000	0.0000	0.3220	0.0000	0.0000	0.0000	0.0000	0.0030
Jerba	0.0000	0.0000	0.0000	0.0751	0.0040	0.0000	0.0000	0.0273	0.0000
Matmat	0.0000	0.0000	0.0000	0.1173	0.0062	0.0000	0.0000	0.0000	0.0000
Gabes	0.0000	0.0000	0.0000	0.0754	0.0212	0.0000	0.0000	0.0105	0.0000
Providencia	0.0000	0.0000	0.0000	0.0200	0.0000	0.1800	0.0000	0.0000	0.0000
Cauca	0.0000	0.0000	0.0000	0.0000	0.0000	0.1600	0.0000	0.0000	0.0800
Choco	0.0000	0.0000	0.0000	0.0000	0.0000	0.1250	0.0000	0.0000	0.0250
Australasia									
Ausl Cape Pen	0.0000	0.0000	0.0000	0.0450	0.0510	0.0000	0.0000	0.0000	0.0100
Aul Ab Kimberly	0.0000	0.0000	0.0000	0.0000	0.1340	0.0000	0.0000	0.0000	0.0120
Middle East									
Iran Baloch	0.0000	0.0100	0.0000	0.0450	0.0800	0.0000	0.0000	0.0450	0.0600
Iran Yazd Parsi	0.0000	0.0000	0.0000	0.0150	0.0380	0.0000	0.0000	0.0000	0.0080

HLA- DQA1 allele frequency distribution

Total six most frequent alleles were found at DQA1 loci in the studied populations. These alleles were DQA1*0101, DQA1*0102 DQA1*0103, DQA1*0201, DQA1*0301 and DQA1*0501. The Frequency ranged between 7.9% to 28.5%. Allele 0104 and 0601 were listed under moderately high frequency alleles its frequency ranged between 1.6% to 15%. While some alleles were identified as low frequency alleles, these were DQA1*0302, DQA1*0303, DQA1*0304, DQA1*0401, DQA1*0502 and DQA1*0604 At DQA1 locus there was no difference between different caste groups. However, when a comparison was made between endogamous caste populations and two inbreeding groups we observed a significant deviation between these groups, which was up to, the extent of 70%. The differences were seen at DQA1*0101, DQA1*0102, DQA1*0104, DQA1*0301, DQA1*0401, DQA1*, DQA1*0502 and DQA1*0601. We further compared North Indian groups with North Eastern populations and found that *Rajbanshi's* were similar to the caste populations but different from *Shia* at DQA1*0101allele and from *Sunni's* at DQA1*0102 allele. The other populations i.e. *Lachung* and *Mech* deviated significantly from the caste and other two inbreeding populations. **(Table 5.11)**

Comparison of HLA-DQA1 locus with the world populations

Further we compared our results with different world populations **(Table 6.2)** The DQA1*0101 was found in almost all the world populations. However, there was gradiant in the distribution of this allele i.e 4.8% to 16.0%. Highest frequency of this allele was seen in Asian populations like China Hong Kong singaporese (9.6%), China Urumqi Kazaq (15.5%), Khotons (11.2%) **(Gao X *et al.*,1997; Mizuki & Ohno *et al.*, 1997 Munkhbat 1997; Machulla *et al.*, 2003)** In Caucasians it was found in Romania (19.1%) and Croatia (15.7%) **(Reed Ho & Lupu *et al.*, 1992; Grubic & Zunec *et al.*, 1995)**. Among Middle Eastern populations it was highest among Iran Baloch (42.5%). Frequency of this allele in *Shia* Muslims was as high as16% and was comparable to Iranian Yazd Parsis (20.0%) **(Farjadian *et al.*,2006)**.

Very interesting result was found at allele DQA1*0102. Among North Indian caste populations and North Eastern populations frequency was almost similar (9.6% to 15.5%). It was found that this allele was also higher among Orientals populations like

North Chinese (14.8%), South Korea (20.2%), Pomors (21%) **(GaoX & Sun Y *et al.*, 1991; Lee & Oh *et al.*, 2005; Evseeva *et al.*, 2002)**. In Caucasians it was occurring with a frequency of 23% in Romania, 15.7%in Croatians and 22.7% in Germans **(Reed & Ho Lupu *et al.*, 1992; Grubic, Zunec *et al.*, 1995; Ferencik, Grosse-Wilde 1997)**. Among Africans it was also seen to be high in Cameroon (38.0%) and Oromo (28.9%) **(Pimtanothai *et al.*, 2001; Fort M *et al.*, 1998)**. Among Middle Eastern populations like Iran Parsees **(Farjadian *et al.*,2006)** and Iran Baloch this allele was absent; interestingly *Shia* Muslim also had low frequency of this allele i.e. 2.8% as compared to other studied populations.

At DQA1*0103 allele it was found that among caste populations frequency of this allele was 11.2% to 16.09% which was found to be comparable to Caucasians like Czech (14.5%), Germans (8.1%), Poland (10.5%) and Azores (9.7%) **(Cerna M *et al.*,1992; Ferencik *et al.*, 1997; Jungerman; Spinola *et al.*, 2005)**. Frequency of this allele was 14.7% in *Shia* and 12.5% in *Sunni*, which was comparable to Iran Yazd Parsees (14.6%). **(Farjadian *et al.*, 2006)**. Our North Eastern populations show frequency in range of 8.2% to 11.9% which was comparable to Mongolian populations like China Beizing and Xian (12.0%), Khotons (13.1%) and North Chinese (11.5%) respectively **(GaoX *et al.*,,1991; 1999; Munkhbat 1997; Machulla *et al.*, 2003)**

Allele DQA1*0201 was the most frequent allele in our populations and also among other world populations like among Orientals i.e. Pomors, North Chinese, China Beizing and Xian (11.7% to 13.7%) **(GaoX *et al.*, 1991; 1999; Evseeva *et al.*, 2002)**. In Caucasians i.e. Croatians, Italy and Czechs it ranged between 10.2% to 19.3% **(Grubic Zunec *et al.*, 1995; Rendine & Borelli *et al.*,1998; Cerna M *et al.*, 1992)** this allele was also seen in high frequency among Africans like Choco (20.0%) and Oromo (22.0%). **(Fort *et al.*,1998)**

When comparison was drawn on basis of allele DQA1*0301 it was found that all the caste populations and Muslims show similarity with Caucasians like Germans (7.5%), France (15.0%), Basque (11.0%) and Middle east populations like Iranian Parsis (18.5%) **(Ferencik *et al.*, 1997; Charron 1997; Sanchez-Velasco *et al.*, 2003; Farjadian *et al.*, 2006)** Frequency of this allele in caste population was 10.2% to 15.5% while among *Muslims* it was 9.4% to 11.1%. But when we compared frequency of North

Eastern populations (11.8% to 24.1%) it was found that these were comparable to Mongolian populations like China Hong Kong Singaporese (28.6%), China Beizing and Xian (27.8%), Khalkha Mongols (23.1%) and Khotons (25.4%). **(Gao X *et al.*,1991;1997; Munkhbat 1997 and Machulla *et al.*, 2003)**

DQA1*0501 was another high frequency allele found in all the studied populatons (11.2% to 16.7%) and other world populations. High frequency of this allele is seen in Oriental populations such as China Hong Kong Singaporese (15.2%), Khotons (18.9%) **(Gao 1997; Munkhbat 1997; Machulla *et al.*, 2003)**, In Africans like Gabonese and Amhara its frequency was 18.5% to 20.9%) **(Migot-Nabias *et al.*, 1999; Fort & De Stefano *et al.*, 1998)**. In Caucasians it was found in Romania, Turkey, Bulgaria, Czech etc with frequency range between 29.0% to 34.0% **(Reed, Ho *et al.*, 1992; Saruhan-Direskeneli *et al.*, 2000; Ivanova, Rozemuller *et al.*, 2002; Cerna *et al.*, 1992)**. In Middle East populations i.e Iran Baloch and Iranian Parsis also this allele was seen in high frequency i.e. 21.5% to 35.5%. **(Farjadian *et al.*, 2006)**

Table 6.2 DQA1 frequency distribution in world populations

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Allele	**0101**	**0102**	**0103**	**0104**	**0201**	**0301**	**0302**	**0303**	**0401**	**0501**	**0502**	**0503**	**0504**	**0601**
Asians														
Kayastha	0.0750	0.0920	0.1609	0.0460	0.2759	0.1552	0.0053	0.0000	0.0236	0.1390	0.0000	0.0053	0.0000	0.0160
Vaish	0.0470	0.1361	0.1519	0.1360	0.2400	0.1290	0.0065	0.0000	0.0090	0.1139	0.0030	0.0030	0.0000	0.0160
Rastogi	0.0739	0.0969	0.1479	0.1020	0.2370	0.1479	0.0025	0.0050	0.0150	0.1224	0.0076	0.0076	0.0000	0.0280
Mathur	0.0479	0.1556	0.1110	0.1515	0.2092	0.1010	0.0050	0.0000	0.0150	0.1658	0.0000	0.0000	0.0000	0.0330
Shia	0.1600	0.0289	0.1470	0.0550	0.1870	0.0950	0.0053	0.0026	0.0368	0.1290	0.0260	0.0053	0.0000	0.1130
Sunni	0.0930	0.0790	0.1250	0.1030	0.1809	0.1120	0.0000	0.0053	0.0186	0.1670	0.0505	0.0000	0.0023	0.0610
Lachung	0.0770	0.1550	0.0948	0.0170	0.2240	0.2410	0.0000	0.0000	0.0170	0.1120	0.0000	0.0000	0.0000	0.0520
Mech	0.0870	0.1349	0.1190	0.0238	0.2850	0.1508	0.0240	0.0000	0.0160	0.0240	0.0000	0.0000	0.0000	0.0000
Rajbanshi	0.0612	0.1330	0.0816	0.0770	0.2650	0.1220	0.0153	0.0000	0.0155	0.1420	0.0000	0.0000	0.0000	0.0870
Chi Hk Singap	0.0960	0.2300	0.0590	0.0000	0.0480	0.2860	0.0000	0.0000	0.0070	0.1520	0.0000	0.0000	0.0000	0.1220
Chi Bei Xian	0.1170	0.1430	0.1200	0.0000	0.1170	0.2780	0.0000	0.0000	0.0090	0.1810	0.0000	0.0000	0.0000	0.0350
Uru Kazak	0.1550	0.0480	0.0830	0.0000	0.1070	0.2380	0.0000	0.0000	0.0480	0.2970	0.0000	0.0000	0.0000	0.0240
Pomors	0.1350	0.2140	0.0620	0.0000	0.1300	0.2260	0.0000	0.0000	0.0340	0.1990	0.0000	0.0000	0.0000	0.0000
Saami	0.1540	0.1290	0.0490	0.0000	0.0440	0.3830	0.0000	0.0000	0.0620	0.1790	0.0000	0.0000	0.0000	0.0000
Nentsy	*0.0270*	*0.0555*	*0.0920*	*0.0000*	*0.1090*	*0.4130*	*0.0000*	*0.0000*	*0.0455*	*0.2580*	*0.0000*	*0.0000*	*0.0000*	*0.0000*
N E Russian	0.1330	0.1950	0.0550	0.0000	0.1370	0.1500	0.0000	0.0000	0.0150	0.3150	0.0000	0.0000	0.0000	0.0000
Khoton	0.1120	0.0600	0.1310	0.0000	0.0920	0.2540	0.0000	0.0000	0.0450	0.1890	0.1170	0.0000	0.0000	0.0000
Khalkha	0.1100	0.0410	0.2070	0.0000	0.0290	0.2310	0.0000	0.0000	0.0670	0.1690	0.1460	0.0000	0.0000	0.0000
North Indian	0.0318	0.1273	0.1455	0.1727	0.2000	0.0773	0.0182	0.0000	0.0091	0.1909	0.0000	0.0000	0.0000	0.0272
Russia NW	0.1550	0.1850	0.0850	0.0000	0.1050	0.1600	0.0000	0.0000	0.0450	0.2650	0.0000	0.0000	0.0000	0.0000
South Korea	0.0760	0.2020	0.1230	0.0630	0.0560	0.0970	0.0970	0.1130	0.0120	0.0310	0.0000	0.0280	0.0000	0.0340
North Chinese	0.1210	0.1480	0.1150	0.0000	0.1370	0.2860	0.0000	0.0000	0.0110	0.1380	0.0000	0.0000	0.0000	0.0440
Russia Tuva	0.4090	0.0000	0.0000	0.0000	0.0990	0.1920	0.0000	0.0000	0.0410	0.2590	0.0000	0.0000	0.0000	0.0000
Japan Hyogo	0.2200	0.1830	0.1520	0.0000	0.0000	0.3650	0.0000	0.0000	0.0800	0.0000	0.0000	0.0000	0.0000	0.0000
Uygur	0.0530	0.1400	0.0790	0.0000	0.2110	0.2190	0.0000	0.0000	0.0090	0.2630	0.0000	0.0000	0.0000	0.0260
Africa														
Providencia	0.2700	0.2800	0.0300	0.0000	0.0300	0.1300	0.0000	0.0000	0.0800	0.1800	0.0000	0.0000	0.0000	0.0000
Cauca	0.0750	0.2750	0.0250	0.0000	0.0500	0.1750	0.0000	0.0000	0.1750	0.2250	0.0000	0.0000	0.0000	0.0000
Choco	0.1500	0.2000	0.0250	0.0000	0.2000	0.1250	0.0000	0.0000	0.1000	0.2000	0.0000	0.0000	0.0000	0.0000
Equador	0.0520	0.3800	0.1290	0.0000	0.1030	0.0340	0.0000	0.0000	0.0950	0.2070	0.0000	0.0000	0.0000	0.0000
Africa														
Gabonese	0.1500	0.5000	0.0390	0.0000	0.0500	0.0750	0.0000	0.0000	0.0090	0.1770	0.0000	0.0000	0.0000	0.0000
Oromo	*0.0904*	*0.2892*	*0.0241*	*0.0000*	*0.2229*	*0.1746*	*0.0000*	*0.0000*	*0.0181*	*0.1807*	*0.0000*	*0.0000*	*0.0000*	*0.0000*
Amhara	0.1684	0.2806	0.0204	0.0000	0.2041	0.1122	0.0000	0.0000	0.0051	0.2092	0.0000	0.0000	0.0000	0.0000
Cameroon	0.0510	0.3800	0.0430	0.0710	0.0640	0.0040	0.1120	0.0000	0.1120	0.0000	0.0000	0.0000	0.0000	0.0040
Algerian	0.1060	0.1170	0.0640	0.0000	0.1490	0.1490	0.0000	0.0000	0.0110	0.3830	0.0000	0.0000	0.0000	0.0210
Ticuna	0.0000	0.0000	0.4500	0.0000	0.0000	0.0000	0.0000	0.0000	0.2700	0.2800	0.0000	0.0000	0.0000	0.0000
Moraccan	0.1210	0.2120	0.0560	0.0000	0.1260	0.1720	0.0000	0.0000	0.0400	0.2730	0.0000	0.0000	0.0000	0.0000
S Americans														
NY Afameric	0.1350	0.3440	0.0270	0.0000	0.0910	0.1040	0.0000	0.0000	0.1080	0.1890	0.0000	0.0000	0.0000	0.0020
Arg Bu Aeres	0.1310	0.1610	0.0720	0.0000	0.1390	0.1760	0.0000	0.0000	0.0500	0.2710	0.0000	0.0000	0.0000	0.0000
Arg Chiriguan	0.0450	0.0270	0.0000	0.0000	0.0180	0.1790	0.0000	0.0000	0.1340	0.5970	0.0000	0.0000	0.0000	0.0000
Brazil N East	0.1740	0.1610	0.0600	0.0000	0.1200	0.1920	0.0000	0.0000	0.0740	0.2070	0.0040	0.0000	0.0000	0.0080
Guarani Mbya	0.0050	0.0000	0.0110	0.0000	0.0050	0.3010	0.0000	0.0000	0.1300	0.5480	0.0000	0.0000	0.0000	0.0000
Guarani Kai	0.0000	0.0000	0.0000	0.0000	0.0000	0.2360	0.0000	0.0000	0.2210	0.5430	0.0000	0.0000	0.0000	0.0000
Guarani Nand	0.0340	0.0000	0.0000	0.0000	0.0170	0.1460	0.0000	0.0000	0.2150	0.5880	0.0000	0.0000	0.0000	0.0000
Ache	0.0000	0.0000	0.0000	0.0000	0.0000	0.7470	0.0000	0.0000	0.1150	0.1380	0.0000	0.0000	0.0000	0.0000

Table 6.2 DQA1 Continued.....

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Allele	**0101**	**0102**	**0103**	**0104**	**0201**	**0301**	**0302**	**0303**	**0401**	**0501**	**0502**	**0503**	**0504**	**0601**
S Americans														
Kaingang	0.0280	0.0000	0.0020	0.0000	0.0020	0.2720	0.0000	0.0000	0.5140	0.1820	0.0000	0.0000	0.0000	0.0000
Brazil Plateau	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.1540	0.0000	0.0000	0.2650	0.5810	0.0000	0.0000	0.0000
Argentinatoba	0.0640	0.0290	0.0120	0.0000	0.0520	0.3080	0.0000	0.0000	0.1750	0.3600	0.0000	0.0000	0.0000	0.0000
Kogui	0.0000	0.0000	0.0000	0.0000	0.0000	0.7200	0.0000	0.0000	0.1300	0.1500	0.0000	0.0000	0.0000	0.0000
Ijka	0.0200	0.0500	0.0000	0.0000	0.0700	0.1800	0.0000	0.0000	0.6300	0.0500	0.0000	0.0000	0.0000	0.0000
Guahibo	0.0200	0.0000	0.0000	0.0000	0.0000	0.1500	0.0000	0.0000	0.0600	0.7700	0.0000	0.0000	0.0000	0.0000
Nukak	0.0000	0.0000	0.0000	0.0000	0.0000	0.2500	0.0000	0.0000	0.0300	0.7200	0.0000	0.0000	0.0000	0.0000
Sikuani	0.0000	0.0000	0.0000	0.0000	0.0000	0.7500	0.0000	0.0000	0.1800	0.0700	0.0000	0.0000	0.0000	0.0000
Ingano	0.0000	0.0000	0.0000	0.0000	0.0000	0.5900	0.0000	0.0000	0.3600	0.0500	0.0000	0.0000	0.0000	0.0000
Waunana	0.0000	0.0000	0.0000	0.0000	0.0000	0.3500	0.0000	0.0000	0.0300	0.6200	0.0000	0.0000	0.0000	0.0000
Embera	0.0000	0.0000	0.0000	0.0000	0.0000	0.3500	0.0000	0.0000	0.1000	0.5500	0.0000	0.0000	0.0000	0.0000
Tule	0.0000	0.0000	0.0000	0.0000	0.0200	0.6500	0.0000	0.0000	0.0300	0.3000	0.0000	0.0000	0.0000	0.0000
European														
Romania	0.1910	0.2300	0.0530	0.0000	0.0920	0.0920	0.0000	0.0000	0.0130	0.3290	0.0000	0.0000	0.0000	0.0000
Turkey Ankara	0.3600	0.0000	0.0000	0.0000	0.0600	0.2400	0.0000	0.0000	0.0100	0.3200	0.0000	0.0000	0.0000	0.0100
Croatians	0.1570	0.2685	0.0730	0.0000	0.1020	0.1020	0.0000	0.0000	0.0213	0.2687	0.0000	0.0000	0.0000	0.0075
Bulgaria	0.1750	0.2200	0.0950	0.0000	0.0600	0.0900	0.0000	0.0000	0.0200	0.3400	0.0000	0.0000	0.0000	0.0000
France CEPH	0.1410	0.2050	0.0440	0.0000	0.1410	0.2130	0.0000	0.0000	0.0360	0.2200	0.0000	0.0000	0.0000	0.0000
Italy Central	0.1000	0.1400	0.0670	0.0400	0.1560	0.0470	0.0060	0.0110	0.0190	0.4080	0.0000	0.0000	0.0000	0.0060
Canada	0.1760	0.0000	0.0000	0.0000	0.0160	0.2820	0.0000	0.0000	0.0480	0.4780	0.0000	0.0000	0.0000	0.0000
Czech Republic	0.0100	0.2010	0.1450	0.0540	0.1930	0.0200	0.0000	0.0190	0.0570	0.2940	0.0000	0.0000	0.0000	0.0070
German	0.1290	0.2270	0.0810	0.0400	0.0920	0.0750	0.0580	0.0000	0.0350	0.2630	0.0000	0.0000	0.0000	0.0000
Italy Bergamo	0.1000	0.1380	0.0590	0.0600	0.1280	0.0500	0.0250	0.0000	0.0250	0.1000	0.0000	0.0000	0.0000	0.0050
Denmark	0.1640	0.2180	0.0820	0.0000	0.1450	0.2100	0.0000	0.0000	0.0270	0.1450	0.0000	0.0000	0.0000	0.0090
France South	0.1450	0.2100	0.0540	0.0000	0.1290	0.1500	0.0000	0.0000	0.0290	0.2200	0.0000	0.0000	0.0000	0.0630
Greece	0.1510	0.2700	0.0790	0.0000	0.0100	0.0520	0.0000	0.0000	0.0100	0.4280	0.0000	0.0000	0.0000	0.0000
Mazatecans	0.0140	0.0140	0.0350	0.0000	0.0070	0.4850	0.0000	0.0000	0.1360	0.3090	0.0000	0.0000	0.0000	0.0000
Cantabrians	0.0780	0.2710	0.0660	0.0120	0.2050	0.1930	0.0000	0.0000	0.0120	0.1630	0.0000	0.0000	0.0000	0.0000
Cabuernigos	0.0470	0.3160	0.0530	0.0110	0.1680	0.1420	0.0000	0.0000	0.0050	0.2580	0.0000	0.0000	0.0000	0.0000
Pasiegos	0.0760	0.3370	0.0610	0.0160	0.1790	0.1740	0.0000	0.0000	0.0160	0.1410	0.0000	0.0000	0.0000	0.0000
Polish	0.1807	0.1977	0.1050	0.0000	0.1273	0.1206	0.0000	0.0000	0.0454	0.2182	0.0051	0.0000	0.0000	0.0000
Azores	0.1120	0.1090	0.0970	0.0160	0.1740	0.0890	0.0190	0.0580	0.0500	0.1280	0.0000	0.0000	0.0000	0.0040
Basques	0.1370	0.1320	0.0710	0.0000	0.2600	0.1100	0.0000	0.0000	0.0200	0.2700	0.0000	0.0000	0.0000	0.0000
Canadian	0.1530	0.1753	0.0448	0.0000	0.1541	0.1976	0.0000	0.0000	0.0485	0.2230	0.0037	0.0000	0.0000	0.0000
Australasia														
Austral Abcape	0.1670	0.0960	0.3590	0.0000	0.0400	0.2020	0.0000	0.0000	0.0000	0.1310	0.0000	0.0000	0.0000	0.0050
Aust Abkimberly	0.0850	0.1460	0.2800	0.0000	0.0000	0.1830	0.0000	0.0000	0.0000	0.3060	0.0000	0.0000	0.0000	0.0000
Middle East														
Iran Baloch	0.4250	0.0000	0.0800	0.0000	0.0350	0.0600	0.0000	0.0000	0.0450	0.3550	0.0000	0.0000	0.0000	0.0000
Iran Yazd	0.2000	0.0000	0.1460	0.0000	0.1850	0.2080	0.0000	0.0000	0.0460	0.2150	0.0000	0.0000	0.0000	0.0000
Pacific Island														
PNG	0.2280	0.4650	0.0970	0.0000	0.0000	0.1490	0.0000	0.0000	0.0000	0.0610	0.0000	0.0000	0.0000	0.0000
New Caledonia	0.1460	0.2460	0.0920	0.0000	0.0150	0.1620	0.0000	0.0000	0.0080	0.3310	0.0000	0.0000	0.0000	0.0000
Peru Quecha	0.0200	0.0500	0.0000	0.0120	0.0500	0.3850	0.0000	0.0000	0.1660	0.3170	0.0000	0.0000	0.0000	0.0000
Trinidad	0.4600	0.0000	0.0000	0.0000	0.0950	0.1500	0.0000	0.0000	0.0650	0.2300	0.0000	0.0000	0.0000	0.0000

HLA- DQB1 allele frequency distribution

At DQB1 locus high frequency alleles which were common in all the populations were in the range of 9.7%to 29.08%. These alleles were DQB1*0201, DQB1*0301, DQB1*0303, DQB1*0501 and DQB1*0601.

The frequency of DQB1*0201 in all the populations studied were very high i.e. 17.4% in *Mathurs* to 26.5% in *Rastogies*. Frequency of DQB1*0301 was also very high in all the populations but highest frequency was only seen in North Eastern populations. Range was between 23.3% to 26.2%. Further analysis revealed that *Shia* Muslims differed with caste populations at this allele. It has been found that there was significant difference between Muslims and North eastern populations. High frequency of allele DQB1*0501 was found in all the nine populations i.e. 9.7% to 29.08%. It was interesting to see that there was significant difference between caste populations at this locus. Further it was found that *Kayastha* differed with Muslims and North Eastern populations. *Rajbanshi* from North Eastern group also revealed significant difference with all the caste populations and Muslims.

It was found that high frequency of allele DQB1*0601 was seen among all the nine populations (10.3% to 19.3%). *Lachung* and *Mech* differed significantly with caste and Muslim populations *Rajbanshi* differed with *Lachung* and *Mech* significantly. We have identified some alleles which ranged between 1.5% to 10.7% these are DQB1*0302, DQB1*0303, DQB1* 0503, DQB1*0602 and DQB1*0603. It was interesting to see that *Rajbanshi's* differed with *Vaish* and *Shia* populations at DQB1*0302 locus. North Eastern populations showed significant differences with caste and Muslim populations at allele DQB1*0303 and DQB1*0503 allele. Certain alleles could be listed as less frequent alleles. These were DQB1*0304, DQB1*0305, DQB1*0401, DQB1*0402, DQB1*0504, DQB1*0506, DQB1*0604, DQB1*0605 and DQB1*0606 and the frequency ranged between 0.26% to 2.6%. At alleles DQB1*0305, DQB1*0401, DQB1*0402, DQB1*0504 there was significant difference between caste populations and *Muslims*.

Comparison of HLA-DQB1 locus with the world populations

We have compared our data with the different world population. Frequency distribution at this locus is given in **Table 6.3**.

Table 6.3 DQB1 frequency distribution in world populations

	1	2	3	4	5	6	7	8	9	10	11
	0201	0202	0301	0302	0303	0304	0305	0401	0402	0501	0502
Asian											
Japanese	0.0000	0.0030	0.1140	0.1100	0.1350	0.0000	0.0000	0.1150	0.0390	0.0750	0.0350
China Be Xian	0.1260	0.0000	0.2050	0.0610	0.1640	0.0000	0.0000	0.0470	0.0180	0.0580	0.0290
Uru Kazak	0.2380	0.0000	0.2130	0.0710	0.0710	0.0000	0.0000	0.0120	0.0710	0.0480	0.0480
Uygur	0.3750	0.0000	0.1510	0.0000	0.0440	0.0000	0.0000	0.0440	0.0180	0.0530	0.0000
Gypsy	0.2350	0.0000	0.3240	0.0590	0.1180	0.0000	0.0000	0.0000	0.0000	0.1202	0.0176
Russia N W	0.1850	0.0000	0.2110	0.1110	0.0500	0.0000	0.0000	0.0000	0.0400	0.1310	0.0650
South Korea	0.0290	0.0540	0.1280	0.1020	0.1050	0.0020	0.0000	0.0770	0.0390	0.0880	0.0520
China Singap	0.1000	0.0000	0.2190	0.0700	0.1630	0.0000	0.0000	0.0560	0.0070	0.0370	0.1040
Nivkhi Russian	0.1050	0.0000	0.4160	0.0090	0.1600	0.0000	0.0000	0.0280	0.0570	0.0090	0.0090
Ainu Japan	0.0000	0.0000	0.3700	0.0300	0.1400	0.0000	0.0000	0.0800	0.1200	0.0100	0.0000
North Chinese	0.1430	0.0000	0.1580	0.0770	0.1700	0.0000	0.0000	0.0500	0.0110	0.0770	0.0280
Sd Han	0.1610	0.0000	0.2120	0.0410	0.1220	0.0000	0.0000	0.0460	0.0150	0.0150	0.0150
China Naxi	0.0580	0.0000	0.2260	0.0270	0.0790	0.0000	0.0070	0.0680	0.0480	0.0550	0.1630
Khoton	0.2830	0.0000	0.2460	0.0120	0.0800	0.0000	0.0000	0.0550	0.0350	0.1030	0.0310
Khalka	0.1150	0.0600	0.2550	0.0400	0.1350	0.0000	0.0000	0.0500	0.0450	0.0750	0.0350
Oold	0.0870	0.0670	0.4420	0.0190	0.0770	0.0000	0.0000	0.0190	0.0390	0.0480	0.0480
Tsaatan	0.1250	0.0700	0.2640	0.0350	0.0280	0.0100	0.0000	0.0680	0.0550	0.0850	0.0550
Pomors	0.1640	0.0000	0.2050	0.1710	0.1100	0.0000	0.0000	0.0000	0.0000	0.1030	0.0000
Saami	0.1480	0.0000	0.2350	0.1850	0.1120	0.0000	0.0000	0.0000	0.0000	0.1480	0.0000
Nentsy	0.1610	0.0000	0.2270	0.2090	0.2180	0.0000	0.0000	0.0000	0.0000	0.0180	0.0000
N E Russia	0.1850	0.0000	0.2570	0.1350	0.0720	0.0000	0.0000	0.0000	0.0000	0.0960	0.0000
Japan Hyogo	0.0000	0.0000	0.0470	0.0630	0.1130	0.0000	0.0000	0.1490	0.0780	0.1420	0.0310
Russia Tuva	0.1270	0.0000	0.2840	0.0280	0.0710	0.0100	0.0000	0.0360	0.0380	0.0660	0.0180
Shia	0.2130	0.0000	0.1420	0.0470	0.0947	0.0026	0.0236	0.0210	0.0237	0.1520	0.0078
Sunni	0.1968	0.0000	0.1330	0.0290	0.0770	0.0053	0.0260	0.0080	0.0293	0.1729	0.0130
Mathur	0.1740	0.0000	0.1515	0.0277	0.0930	0.0000	0.0075	0.0075	0.0202	0.1790	0.0152
Vaish	0.2380	0.0000	0.0935	0.0548	0.0838	0.0000	0.0064	0.0032	0.0129	0.1450	0.0096
Rastogi	0.2650	0.0000	0.0969	0.0306	0.1070	0.0000	0.0025	0.0000	0.0080	0.1220	0.0150
Kayastha	0.2390	0.0000	0.1368	0.0340	0.1026	0.0053	0.0053	0.0157	0.0053	0.0973	0.0026
Lachung	0.1900	0.0000	0.2330	0.0000	0.0517	0.0000	0.0000	0.0000	0.0086	0.1720	0.0690
Mech	0.2140	0.0000	0.2619	0.0000	0.0317	0.0000	0.0000	0.0000	0.0159	0.1667	0.0710
Rajbanshi	0.1630	0.0000	0.2340	0.0150	0.0204	0.0000	0.0000	0.0103	0.0103	0.2908	0.0255
N Indian	0.2364	0.0000	0.1227	0.0727	0.0636	0.0000	0.0000	0.0000	0.0091	0.1227	0.0230
Gypsy	0.2350	0.0000	0.3240	0.0590	0.1180	0.0000	0.0000	0.0000	0.0000	0.1202	0.0176
America											
Brazil Xava	0.0000	0.0000	0.5810	0.1540	0.0000	0.0000	0.0000	0.0000	0.2650	0.0000	0.0000
Kogui	0.0000	0.0000	0.1500	0.7200	0.0000	0.0000	0.0000	0.0000	0.1300	0.0000	0.0000
Ijka	0.0700	0.0000	0.0500	0.1800	0.0000	0.0000	0.0000	0.0000	0.6300	0.0200	0.0000
Guahibo	0.0000	0.0000	0.7700	0.1500	0.0000	0.0000	0.0000	0.0200	0.0400	0.0200	0.0000
Nukak	0.0000	0.0000	0.7200	0.2500	0.0000	0.0000	0.0000	0.0000	0.0300	0.0000	0.0000
Sikuani	0.0000	0.0000	0.0200	0.7800	0.0200	0.0000	0.0000	0.0000	0.1800	0.0000	0.0000
Ingano	0.0000	0.0000	0.1800	0.5450	0.0000	0.0000	0.0000	0.0000	0.2250	0.0500	0.0000
Waunana	0.0000	0.0000	0.6200	0.3300	0.0200	0.0000	0.0000	0.0000	0.0300	0.0000	0.0000
Embera	0.0000	0.0000	0.5750	0.3250	0.0000	0.0000	0.0000	0.0000	0.1000	0.0000	0.0000
Tule	0.0200	0.0000	0.4100	0.4100	0.0000	0.0000	0.0000	0.0000	0.1600	0.0000	0.0000
Bulgaria	0.1370	0.0000	0.2760	0.0520	0.0000	0.0000	0.0000	0.0390	0.0000	0.1030	0.1600
Ny Af Amer	0.2070	0.0000	0.1640	0.0440	0.0120	0.0000	0.0000	0.0000	0.0660	0.1540	0.0230
Arg Aeres	0.2030	0.0000	0.2170	0.1030	0.0510	0.0000	0.0000	0.0000	0.0450	0.1080	0.0310
Ar Chirigu	0.0180	0.0000	0.6120	0.1300	0.0180	0.0000	0.0000	0.0000	0.1490	0.0460	0.0000
Arg Toba	0.0530	0.0000	0.4010	0.2440	0.0230	0.0000	0.0000	0.0000	0.1740	0.0580	0.0000
Brazil NE	0.1830	0.0000	0.1670	0.1200	0.0600	0.0000	0.0000	0.0000	0.0750	0.1600	0.0260
Guarani Mbya	0.0210	0.0000	0.5350	0.2660	0.0350	0.0000	0.0000	0.0000	0.1370	0.0050	0.0000
Guar Kai	0.0000	0.0000	0.5410	0.1690	0.1090	0.0000	0.0000	0.0000	0.1810	0.0000	0.0000

Table 6.3 DQB1 Continued.....

	12	13	14	15	16	17	18	19	20	21
	0503	0504	0601	0602	0603	0604	0605	0606	0607	0608
Asian										
Japanese	0.0270	0.0000	0.1850	0.0820	0.0070	0.0730	0.0000	0.0000	0.0000	0.0000
China Be Xian	0.0530	0.0000	0.1040	0.1020	0.0180	0.0060	0.0090	0.0000	0.0000	0.0000
Uru Kazak	0.0480	0.0240	0.0360	0.0360	0.0600	0.0240	0.0000	0.0000	0.0000	0.0000
Uygur	0.0440	0.0000	0.0440	0.1560	0.0350	0.0180	0.0180	0.0000	0.0000	0.0000
Gypsy	0.0412	0.0000	0.0000	0.0000	0.0260	0.0590	0.0000	0.0000	0.0000	0.0000
Russia N W	0.0000	0.0000	0.0150	0.0960	0.0600	0.0360	0.0000	0.0000	0.0000	0.0000
South Korea	0.0310	0.0000	0.1030	0.0740	0.0110	0.0630	0.0400	0.0000	0.0000	0.0000
China Singap	*0.0330*	*0.0000*	*0.1110*	*0.0590*	*0.0220*	*0.0040*	*0.0150*	*0.0000*	*0.0000*	*0.0000*
Nivkhi Russian	0.1510	0.0000	0.0190	0.0280	0.0000	0.0090	0.0000	0.0000	0.0000	0.0000
Ainu Japan	0.2000	0.0000	0.0200	0.0200	0.0000	0.0100	0.0000	0.0000	0.0000	0.0000
North Chinese	0.0440	0.0000	0.0990	0.1200	0.0170	0.0000	0.0060	0.0000	0.0000	0.0000
Sd Han	0.0460	0.0000	0.1100	0.1460	0.0200	0.0460	0.0000	0.0000	0.0000	0.0000
China Naxi	0.1170	0.0000	0.0750	0.0210	0.0070	0.0000	0.0140	0.0000	0.0000	0.0000
Khoton	0.0000	0.0000	0.0940	0.0220	0.0150	0.0240	0.0000	0.0000	0.0000	0.0000
Khalka	0.0100	0.0000	0.0550	0.0500	0.0350	0.0250	0.0000	0.0000	0.0000	0.0000
Oold	0.0000	0.0000	0.0480	0.0480	0.0390	0.0190	0.0000	0.0000	0.0000	0.0000
Tsaatan	0.0000	0.0000	0.0630	0.0630	0.0490	0.0000	0.0000	0.0000	0.0000	0.0000
Pomors	0.0000	0.0000	0.0000	0.2470	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Saami	0.0000	0.0000	0.0000	0.1720	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Nentsy	0.0000	0.0000	0.0000	0.1670	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
N E Russia	0.0000	0.0000	0.0000	0.2550	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Japan Hyogo	0.0310	0.0000	0.1570	0.1420	0.0000	0.0310	0.0000	0.0000	0.0160	0.0000
Russia Tuva	0.0480	0.0000	0.0380	0.1140	0.0510	0.0330	0.0000	0.0050	0.0150	0.0000
Shia	0.0237	0.0447	0.1157	0.0236	0.0340	0.0078	0.0026	0.0130	0.0026	0.0026
Sunni	0.0820	0.0210	0.1350	0.0130	0.0345	0.0080	0.0080	0.0053	0.0000	0.0000
Mathur	0.0830	0.0051	0.1607	0.0306	0.0408	0.0255	0.0051	0.0000	0.0000	0.0000
Vaish	0.0645	0.0032	0.1930	0.0258	0.0483	0.0158	0.0000	0.0000	0.0000	0.0000
Rastogi	0.0764	0.0139	0.1042	0.0417	0.0486	0.0208	0.0069	0.0139	0.0000	0.0000
Kayastha	0.0630	0.0053	0.1550	0.0394	0.0789	0.0053	0.0053	0.0078	0.0000	0.0000
Lachung	0.0259	0.0000	0.1379	0.0948	0.0170	0.0000	0.0000	0.0000	0.0000	0.0000
Mech	0.0079	0.0000	0.1508	0.0790	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Rajbanshi	0.0150	0.0000	0.1900	0.0255	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
N Indian	0.0681	0.0045	0.1362	0.0364	0.0864	0.0182	0.0000	0.0000	0.0000	0.0000
Gypsy	0.0412	0.0000	0.0000	0.0000	0.0260	0.0590	0.0000	0.0000	0.0000	0.0000
America										
Brazil Xava	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Kogui	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Ijka	0.0000	0.0000	0.0000	0.0500	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Guahibo	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Nukak	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Sikuani	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Ingano	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Waunana	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Embera	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Tule	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Bulgaria	0.0650	0.0000	0.0130	0.0520	0.0810	0.0090	0.0130	0.0000	0.0000	0.0000
Ny Af Amer	0.0190	0.0000	0.0000	0.2320	0.0150	0.0230	0.0410	0.0000	0.0000	0.0000
Arg Aeres	0.0400	0.0000	0.0140	0.0890	0.0520	0.0470	0.0000	0.0000	0.0000	0.0000
Ar Chiriguano	0.0000	0.0000	0.0000	0.0180	0.0000	0.0090	0.0000	0.0000	0.0000	0.0000
Arg Toba	0.0060	0.0000	0.0000	0.0120	0.0120	0.0170	0.0000	0.0000	0.0000	0.0000
Brazil NE	0.0460	0.0000	0.0000	0.0830	0.0490	0.0310	0.0000	0.0000	0.0000	0.0000
Guarani Mbya	0.0000	0.0000	0.0000	0.0010	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Guar Kai	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

Table 6.3 DQB1 Continued.....

	1	2	3	4	5	6	7	8	9	10	11
	0201	0202	0301	0302	0303	0304	0305	0401	0402	0501	0502
Guar Nandeva	0.0240	0.0000	0.5770	0.1050	0.0460	0.0000	0.0000	0.0000	0.2080	0.0340	0.0000
Ache	0.0000	0.0000	0.1380	0.7470	0.0000	0.0000	0.0000	0.0000	0.1150	0.0000	0.0000
Kaingang	0.0050	0.0000	0.1880	0.2620	0.0040	0.0000	0.0000	0.0000	0.5100	0.0290	0.0000
N W Colombia	0.2150	0.0000	0.1120	0.1730	0.1140	0.0100	0.0050	0.0960	0.0000	0.0150	0.0000
Aymara	0.1700	0.0000	0.1980	0.2150	0.1350	0.0120	0.0000	0.0050	0.2500	0.0050	0.0000
African											
Choco	0.3300	0.0000	0.2000	0.0500	0.0000	0.0000	0.0000	0.0000	0.0500	0.1800	0.0000
Jerba	0.1900	0.0000	0.2736	0.0000	0.0182	0.0000	0.0000	0.0000	0.0364	0.2273	0.0909
Matmata	0.3086	0.0000	0.1667	0.0864	0.0000	0.0000	0.0062	0.0000	0.0494	0.1420	0.0123
Gabes	0.3161	0.0000	0.2484	0.0750	0.0052	0.0000	0.0000	0.0000	0.0360	0.0900	0.0156
Cameroon	0.0715	0.0715	0.1320	0.0150	0.0870	0.0000	0.0000	0.0040	0.0710	0.1210	0.0110
Moraccan	0.2970	0.0000	0.1410	0.0890	0.0100	0.0000	0.0000	0.0000	0.0680	0.1209	0.0160
Algerians	0.2450	0.0000	0.3510	0.0530	0.0210	0.0000	0.0000	0.0210	0.0110	0.0640	0.0530
Gabonese	0.1680	0.0000	0.1020	0.0000	0.0300	0.0030	0.0000	0.0000	0.0090	0.1470	0.0060
Oromo	0.3370	0.0000	0.0900	0.0900	0.0300	0.0000	0.0000	0.0000	0.0480	0.0840	0.0000
Amhara	0.3010	0.0000	0.1370	0.0560	0.0100	0.0000	0.0000	0.0000	0.0250	0.1610	0.0000
Equador Africa	0.1640	0.0000	0.1720	0.0090	0.0090	0.0000	0.0000	0.0000	0.0860	0.0520	0.0690
European											
Polish	0.1850	0.0000	0.1350	0.1080	0.0303	0.0202	0.0051	0.0000	0.0454	0.1526	0.0151
Majorca	0.3084	0.0000	0.2583	0.0000	0.0000	0.0000	0.0000	0.0417	0.0000	0.2083	0.0000
Minorca	0.2348	0.0000	0.2803	0.0000	0.0000	0.0000	0.0000	0.0152	0.0000	0.2045	0.0000
Ibiza	0.3559	0.0000	0.3051	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.1949	0.0000
Chuetas	0.2969	0.0000	0.3828	0.0000	0.0000	0.0000	0.0000	0.0156	0.0000	0.1250	0.0000
Denmark	0.2130	0.0000	0.1210	0.1310	0.0550	0.0000	0.0000	0.0000	0.0280	0.1390	0.0000
England Cauc	0.2420	0.0000	0.1710	0.1300	0.0370	0.0000	0.0000	0.0030	0.0170	0.1430	0.0080
France South	0.2720	0.0000	0.2080	0.0640	0.0300	0.0000	0.0000	0.0000	0.0520	0.1330	0.0250
Greece	0.1690	0.0000	0.3380	0.0310	0.0150	0.0000	0.0000	0.0100	0.0000	0.1300	0.2050
Italy Bergamo	0.0910	0.0950	0.3340	0.0550	0.0300	0.0060	0.0000	0.0000	0.0250	0.1040	0.0590
Mazatecan	0.0000	0.0000	0.3160	0.4848	0.0000	0.0000	0.0000	0.0000	0.1330	0.0140	0.0000
Basque	0.0000	0.4800	0.1130	0.0670	0.0400	0.0000	0.0000	0.0000	0.0000	0.0730	0.0000
Cantabria	0.0000	0.2230	0.2050	0.0840	0.0480	0.0000	0.0000	0.0060	0.0060	0.0960	0.0060
Cabuerni	0.0000	0.2000	0.2890	0.0530	0.0260	0.0000	0.0000	0.0000	0.0050	0.0630	0.0110
Pasiegos	0.0000	0.1850	0.1900	0.0760	0.0430	0.0000	0.0000	0.0110	0.0050	0.0980	0.0050
Turkey Ankara	0.1600	0.0000	0.4600	0.0000	0.0000	0.0000	0.0000	0.0200	0.0000	0.2900	0.0700
German	0.1870	0.0060	0.2130	0.0520	0.0320	0.0000	0.0000	0.0000	0.0290	0.1310	0.0260
Romania	0.2310	0.0000	0.2160	0.0820	0.0070	0.0000	0.0000	0.0000	0.0150	0.1570	0.0900
Croatians	0.1480	0.0000	0.2520	0.0378	0.0126	0.0000	0.0000	0.0084	0.0084	0.1413	0.1210
France CEPH	0.1810	0.0000	0.1850	0.1450	0.0600	0.0000	0.0040	0.0000	0.0360	0.0960	0.0160
Italy Central	0.1730	0.0190	0.3170	0.0510	0.0460	0.0080	0.0000	0.0000	0.0290	0.1170	0.0460
Canada	0.0480	0.0000	0.4520	0.1285	0.1055	0.0000	0.0000	0.0000	0.0880	0.0080	0.0000
Czech Republic	0.2010	0.0000	0.2010	0.0960	0.0470	0.0000	0.0050	0.0000	0.0380	0.0910	0.0290
Azores	0.0580	0.1730	0.1940	0.1010	0.0500	0.0120	0.0000	0.0000	0.0470	0.1120	0.0190
Middle East											
Iran Baloch	0.3200	0.0000	0.1050	0.0150	0.0250	0.0000	0.0000	0.0000	0.0300	0.1050	0.1900
Iran Yazd	0.2390	0.0000	0.2230	0.0080	0.1380	0.0000	0.0000	0.0000	0.0620	0.0920	0.0150
Pacific											
PNG	0.0130	0.0000	0.0710	0.0000	0.0060	0.0000	0.0000	0.1530	0.0000	0.0000	0.0260
Cooks	0.0200	0.0000	0.3000	0.2500	0.0900	0.0000	0.0000	0.0000	0.0300	0.0100	0.0800
Samoa	0.0000	0.0000	0.1600	0.1600	0.2600	0.0000	0.0000	0.0100	0.1000	0.0000	0.0300
Tokelau	0.0000	0.0000	0.2200	0.3800	0.0600	0.0000	0.0000	0.0000	0.0800	0.0000	0.0000
Tonga	0.0000	0.0000	0.2300	0.1700	0.1600	0.0000	0.0000	0.0000	0.0500	0.0000	0.0100
New Caledonia	0.0150	0.0000	0.3540	0.0770	0.0230	0.0000	0.0000	0.0000	0.0460	0.0150	0.0310
Peru Quecha	0.0790	0.0000	0.3280	0.1920	0.1920	0.0000	0.0000	0.0000	0.1400	0.0120	0.0000
Ticuna	0.0100	0.0000	0.2800	0.3789	0.0710	0.0000	0.0000	0.0000	0.2601	0.0000	0.0000

Table 6.3 DQB1 Continued.....

	12 0503	13 0504	14 0601	15 0602	16 0603	17 0604	18 0605	19 0606	20 0607	21 0608
Guar Nande	0.0000	0.0000	0.0060	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Ache	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Kaingang	0.0000	0.0000	0.0000	0.0000	0.0020	0.0000	0.0000	0.0000	0.0000	0.0000
N Wcolombi	0.0100	0.0000	0.0500	0.2000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Aymara	0.0000	0.0000	0.0000	0.0000	0.0050	0.0050	0.0000	0.0000	0.0000	0.0000
African										
Choco	0.0000	0.0000	0.0000	0.1500	0.0200	0.0200	0.0000	0.0000	0.0000	0.0000
Jerba	0.0000	0.0000	0.0000	0.0401	0.0780	0.0091	0.0000	0.0000	0.0364	0.0000
Matmata	0.0000	0.0000	0.0000	0.1173	0.0679	0.0247	0.0000	0.0000	0.0000	0.0000
Gabes	0.0000	0.0000	0.0208	0.0896	0.0308	0.0266	0.0000	0.0000	0.0000	0.0000
Cameroon	0.0110	0.0000	0.0080	0.3090	0.0190	0.0230	0.0080	0.0000	0.0000	0.0080
Moraccan	0.0160	0.0000	0.0100	0.1140	0.0501	0.0680	0.0000	0.0000	0.0000	0.0000
Algerians	0.0210	0.0000	0.0210	0.0530	0.0430	0.0430	0.0000	0.0000	0.0000	0.0000
Gabonese	0.0030	0.0000	0.0000	0.4200	0.0210	0.0910	0.0000	0.0000	0.0000	0.0000
Oromo	0.0000	0.0000	0.0000	0.0900	0.0480	0.1050	0.0780	0.0000	0.0000	0.0000
Amhara	0.0100	0.0000	0.0000	0.0760	0.0510	0.1070	0.0660	0.0000	0.0000	0.0000
Equador Africa	0.0000	0.0000	0.0000	0.2330	0.1030	0.0000	0.0000	0.0000	0.0000	0.0000
European										
Polish	0.0051	0.0000	0.0404	0.1730	0.0797	0.0051	0.0000	0.0000	0.0000	0.0000
Majorca	0.0000	0.0000	0.1833	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Minorca	0.0000	0.0000	0.2652	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Ibiza	0.0000	0.0000	0.1441	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Chuetas	0.0000	0.0000	0.1797	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Denmark	0.0090	0.0000	0.0090	0.1660	0.0830	0.0460	0.0000	0.0000	0.0000	0.0000
England Cauc	0.0250	0.0030	0.0110	0.1440	0.0350	0.0310	0.0000	0.0000	0.0000	0.0000
France South	0.0280	0.0000	0.0100	0.0860	0.0700	0.0190	0.0030	0.0000	0.0000	0.0000
Greece	0.0310	0.0000	0.0000	0.0200	0.0310	0.0200	0.0000	0.0000	0.0000	0.0000
Italy Bergamo	0.0550	0.0000	0.0000	0.0590	0.0550	0.0260	0.0000	0.0000	0.0000	0.0000
Mazatecan	0.0000	0.0000	0.0350	0.0070	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Basque	0.0000	0.0000	0.0000	0.1670	0.0530	0.0070	0.0000	0.0000	0.0000	0.0000
Cantabria	0.0000	0.0000	0.0000	0.2480	0.0720	0.006	0.0000	0.0000	0.0000	0.0000
Cabuerni	0.0000	0.0000	0.0000	0.2900	0.0580	0.005	0.0000	0.0000	0.0000	0.0000
Pasiegos	0.0000	0.0000	0.0000	0.3170	0.0650	0.005	0.0000	0.0000	0.0000	0.0000
Turkey Ankara	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
German	0.0320	0.0000	0.0090	0.1620	0.0920	0.0260	0.0000	0.0000	0.0030	0.0000
Romania	0.0520	0.0000	0.0300	0.0450	0.0600	0.0150	0.0000	0.0000	0.0000	0.0000
Croatians	0.0550	0.0000	0.0126	0.0950	0.0700	0.0379	0.0000	0.0000	0.0000	0.0000
France CEPH	0.0440	0.0000	0.0040	0.1520	0.0420	0.0000	0.0000	0.0000	0.0000	0.0000
Italy Central	0.0480	0.0030	0.0190	0.0570	0.0350	0.0320	0.0000	0.0000	0.0000	0.0000
Canada	0.1700	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Czech Republic	0.0380	0.0000	0.0240	0.1150	0.0910	0.0240	0.0000	0.0000	0.0000	0.0000
Azores	0.0160	0.0000	0.0040	0.0540	0.1010	0.0120	0.0000	0.0000	0.0000	0.0000
Middle East										
Iran Baloch	0.0850	0.0000	0.0800	0.0350	0.0000	0.0100	0.0000	0.0000	0.0000	0.0000
Iran Yazd	0.0080	0.0000	0.0460	0.1000	0.0000	0.0690	0.0000	0.0000	0.0000	0.0000
Pacific										
PNG	0.3590	0.0000	0.1090	0.2630	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Cooks	0.1300	0.0000	0.0800	0.0100	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Samoa	0.1100	0.0000	0.1400	0.0300	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Tokelau	0.1000	0.0000	0.1400	0.0000	0.0200	0.0000	0.0000	0.0000	0.0000	0.0000
Tonga	0.1700	0.0000	0.1900	0.0200	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
New Caledonia	0.1230	0.0000	0.1770	0.1390	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Peru Quecha	0.0000	0.0000	0.0570	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Ticuna	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

The frequency of DQB1*0201 in all the caste populations ranged between 17.4% to 26.5%, in Muslims it was 19.6% and 21.3% while among North Eastern group frequency ranged between 16.3% to 21.4%. High frequency of this allele is seen among most of the ethnic populations of world. Among Asians i.e. China Beizing and Xian (12.6%), China Urumqi Kazaq (23.8%) and North Chinese (14.3%). **(GaoX *et al.*, 1991: 1997; Mizuki *et al.*, 1997)**.Similar trend was seen in Caucasians like Poland (18.5%), Denmark (21.3%) and South France (27.2%) **(Jungerman; Lindblom;Charrod 1997)** While among Africans and Middle Eastern population it was very high in Choco (33.0%), Oromo (33.7%), Iranian Baloch (32%) and Iranian Parsis (23.9%) **(Fort M, De Stefano *et al.*, 1998; Farjadian *et al.*, 2006).**

At DQB1*0301 locus it was found that in caste populations frequency was 9.3% to 15.1% which was found to be comparable to Caucasian populations i.e. Poland (13.5%), Denmark (12.1%) and Basque (11.3%) etc **(Jungerman & Lindblom; Sanchez-Velasco *et al.*, 2003)** Frequency of North Eastern populations (23.2% to 26.19%) was almost similar to Mongolian populations like China Hong Kong Singaporese (21.9%), Khoton (24.6%) and Khalkha (23.5%) **(Gao 1997; Munkhbat 1997; Machulla *et al.*, 2003)**. Frequency of this allele in Muslims sects (13.3% to 14.2%) was found to be comparable to Iran Baloch (10.5%) and Iran Parsi (22.3%) **(Farjadian *et al.*, 2006)**

When we compared our populations at DQB1*0501 locus it was seen that this allele was also present in most of the world populations of different ethnic origins. Frequency of this allele in our populations ranged between of 9.7% to 29.08%. Among Asians it was found in Japan Hyogo (14.2%), Saami (14.8%), Russian North West (13.1%) and Gypsies (12.02%) **(Nobuo Araki *et al.*,**1998; **Evseeva *et al.*,2002)**,while among Caucasians it was 15.2% in Polish, 20.8% in Majorcans 13.9% in Denmark, 13.3%in France and 13.1%in Germans **(Jungerman; Lindblom Charrod 1997; Ferencik *et al.*,1997)**.This allele was also seen to be high in African populations like Jerba (22.7%) and Amhara (16.0%) **(Fort *et al.*, 1998; Abdennaji Guenounou *et al.*, 2006)**. Our comparisons have revealed that this allele is one of the frequent allele in all the subcontinents of world.

DQB1*0601 locus revealed that frequencies of Caste and Muslim populations (10.3% to 16.07%) were comparable to Caucasian and Middle Eastern populations like Majroca (18.3%), Ibiza (14.4%), Minorca (26.5%) and Iran Baloch (8.0%) **(Crespí *et al.*,2002)**. While frequencies of North Eastern population (13.7% to 19.07%) were comparable to Oriental populations like Japanese (18.5%), China Beizing and Xian (10.4%) and Japan Hyogo (15.7%) respectively **(Saito *et al.*, 2000; Gao 1997; Nobuo Araki 1998)**

Among moderate and low frequency allele some interesting findings were seen for example frequency of DQB1*0302 in caste and Muslims populations ranged between (2.7% to 5.4%) was almost similar to Caucasian populations i.e. Italy Bergamo (5.5%), Germans (5.2%) and Greece (3.1%) etc. **(Ferrara, Delfino *et al.*, 1998; Ferencik *et al.*, 1997; Reveille *et al.*, 1995)** While frequency in Middle East populations was low i.e. Iranian Baloch (1.5%). In North Eastern populations this allele was seen only in *Rajbanshis* but with very low frequency (1.5%), it is further seen that this allele was absent in most of the mongoloid populations like China Uygur but relatively low frequency was also seen in Nikhvi (0.9%) and Khoton Mongols (1.2%) etc. **(Mizuki N *et al.*, 1998; Munkhbat, 1997; Machulla *et al.*, 2003)**

The frequency of DQB1*0602 allele in caste populations was 4.0% to 7.8%which was almost similar to European populations i.e. Italy Bergamo (5.9%), Romania (4.5%) etc. **(Ferrara; Reed & Ho *et al.*, 1992)** while frequency of this allele in Muslims was low as compared than Caste populations (1.3% to 2.3%) and was comparable to Iran Baloch (3.5%) and Caucasians.

In North Eastern populations it was found that allele DQB1*0603 was absent in Mech and Rajbanshi from North Eastern belt but was seen only in *Lachung* with frequency of only 1.7%. It was found that this allele was absent in populations having Mongoloid origin like Nikhvi, Ainu, Pomors, Saami etc but was seen in Japanese with 0.7% frequency only. **(Evseeva *et al.*, 2002; Saito *et al.*, 2000)** Frequency of this allele in North Indian caste and Muslim populations was in the range of 3.4% to 7.8%, which was comparable to Polish (7.9%), Danish (8.3%), Greece (3.1%), Italy Bergamo (3.5%), and Basque (5.3%) etc **(Jungerman; Lindblom; Reveille *et al.*, 1995; Ferrara;**

Sanchez-Velasco *et al.*, 2003) In Middle East populations this allele was altogether absent.

Haplotype distribution

At HLA Class II locus two and three locus extended haplotype were constructed. In the present study we have included only those haplotypes, which occurred with a frequency of more than 1% **(Table 5.13 and 5.14).**

Two locus haplotypes

In the present section we have discussed two locus haplotypes in all the nine populations studied by us and compared all the populations on the basis of haplotypic distribution and further comparison was made with the other world populations. Our results revealed that there were some common haplotypes shared by all the populations, these were DRB1*0301-DQB1*0201, DRB1*0701-DQB1*0201. Among caste and inbreeding Muslim populations two haplotypes i.e. DRB1*0101-DQB1*0501 and DRB1*0701-DQB1*0303 were commonly seen. It has been revealed that *Shia* Muslims were not having the haplotype DRB1*1001-DQB*0501 and DRB1*1401-DQB1*0503 while this haplotype was common in caste populations and *Sunni* Muslims. The haplotypes designated as DRB1*080x-DQB1*0301, DRB1*090x-DQB1*0303, DRB1*1101-DQB1*0501, DRB1*1201-DQB1*0301 and DRB1*1401-DQB1*0501 were seen only in North Eastern populations.

Our analysis revealed the presence of five unique haplotypes. These haplotypes were DRB1*0101-DQB1*0601 (1.6%-4.4%) in Muslim populations, DRB1*1301-DQB1*0601 (1.2%-1.5%) among caste populations, DRB1*1301-DQB1*0201(1.1%-2.3%) in *Kayastha* and *Sunni* Muslims, DRB1*0301-DQB1*0501 (3.9%) in *Mech* and DRB1*1101-DQB1*0501(4.9%-6.9%) in the entire North Eastern group.

Further comparison with world populations showed that haplotype DRB1*0101-DQB1*0501 which was seen in Caste and Muslim populations with frequency between 1.4%-6.6% was distributed in different world population like among Caucasian it was seen in Pasiegos, Azores, Germans, French, etc with frequency ranged between 5.0% to 9.3, **(Sanchez-Velasco *et al.*, 1999; Spinola *et al.*, 2005; Ferencik *et al.*, 1997; Charron 1997)** Middle Eastern populations like Parsees (17.2%) **(Farjadian *et al.*,**

2006), in Mongolian like Koreans and Japanese frequency was 6.5%-6.8% **(Lee & Oh *et al.*, 2005; Saito *et al.*, 2000)** while in SD Han, Hakka and South China this haplotype was absent **(Zhou & Lin *et al.*, 2005; Shaw & Chen *et al.*, 1999).**

High frequency of DRB1*0301-DQB1*0201 was found in caste and Muslim populations (5.9-7.4%) which was comparable to Caucasoid i.e. Pasiegos, Bulgaria, Germans and Romania etc (7.0-14.9%)%**(Sanchez-Velasco *et al.*, 1999; Ivanova 2002; Ferencik *et al.*, 1997; Reed, Ho *et al.*, 1992)** and Parsees from Middle East (6.3%) **(Farjadian *et al.*, 2006),** while frequency of this haplotype was considerably low in North Eastern populations (1.5% to 4.6%).and was similar to that of Mongoloid populations like Northern Chinese, SD Han, Koreans, Japanese etc (1.7%-3.6%) **(Gao *et al.*, 1991; Zhou *et al.*, 2005; Lee *et al.*, 2005; Saito *et al.*, 2000)**

In our study we found DRB1*0701-DQB1*0201 to be the most frequent haplotypes seen in high frequency (7.1% to 16.4%) DRB1*0701-DQB1*0201 was significantly represented in all the studied nine populations (7.1%-16.4%) and was almost universally present with high frequency in Caucasians, Mongoloids, Arabians and Middle Eastern populations. **(Sanchez-Velasco *et al.*, 1999; Ivanova *et al.*, 2002; Ferencik *et al.*, 1997; Reed, Ho *et al.*, 1992; Gao *et al.*, 1991;. Zhou *et al.*, 2005; Lee *et al.*, 2005; Saito *et al.*, 2000; Abdennaji Guenounou 2006; Farjadian *et al.*, 2006)**

The haplotype DRB1*080x-DQB1*0301 was seen only in North Eastern populations (3.4-7.9%) Interestingly this was present only in Japanese populations with a frequency of 1.8% **(Saito *et al.*, 2000)** showing Mongoloid element in North Eastern part of India.

Another interesting finding was seen at DRB1*090x-DQB1*0303. This haplotype was found in high frequency in *Lanchung* and *Mech* (3.2-5.1%) and was comparable to Oriental populations like Northern Chinese, Hans, Koreans and Japanese etc (6.8-14.2%). **(Gao *et al.*, 1991; Zhou *et al.*, 2005; Lee *et al.*, 2005; Saito *et al.*, 2000; Lee *et al.*, 2005)** This haplotype was also seen to be frequent among Pacific Islanders (9-25%).

Among caste and Sunni populations DRB1*1001-DQB1*0501 was seen with frequency of 4.6-9.0% and was comparable to Caucasians i.e. Pasiegos, Romania and Croatians (2.2-12.7%). **(Sanchez-Velasco *et al.*, 1999; Reed, Ho *et al.*, 1992;**

Grubic *et al.*, 1995). This haplotype was absent in Shia Muslims and it was interesting to find that this haplotype was also not seen in Parsees (**Farjadian *et al.*, 2006**) and Pakistani populations like Pathans and Kalash etc (**Mohyuddin *et al.*, 2002**) showing resembelence of different Muslim populations with our *Shia* populations. Between *Lachung* and *Rajbanshis* this haplotype was seen but with much less frequeny (2.0-3.4%), which was similar to that of Northern Chinese and Koreans (1.7-3.2%) (**GaoX *et al.*, 1991; Lee KW *et al.*, 2005**). DRB1*1301-DQB1*0603 was seen only in caste populatons only (3.3-4.2%), which was comparable to Caucasians like Pasiegos, Azores, Bulgaria etc (7.6-8.9%). (**Sanchez-Velasco *et al.*, 1999; Spinola *et al.*, 2005; Ivanova *et al.*, 2002**) **(Table 6.4)**

Extended haplotype

When the three locus haplotype comparison of our populations was done, it was seen that. DRB1*0701–DQA1*0201-DQB1*0201 and DRB1*1501-DQA1*0103-DQB1*0601 were common in all the studied populations. It was interesting to see that all the Caste and Muslim populations have some shared haplotypes, these were DRB1*0301–DQA1*0501-DQB1*0201, DRB1*1001–DQA1*0101-DQB1*0501 and DRB1*1301-DQA1*0103-DQB1*0603. Further it was seen that *Sunni* Muslims also share haplotype DRB1*0301-DQA1*0301-DQB1*0201 with caste populations while *Shia* share DRB1*0701-DQA1*0201-DQB1*0301 with Caste populations. Haplotype DRB1*1401-DQA1*0101-DQB1*0503 was seen to be common in North Eastern and caste populations.

Extended haplotypes analysis also revealed the presence of some unique haplotypes among all the studied populations. In caste populations DRB1*0301-DQA1*0301-DQB1*0201 was seen as unique haplotype while among Muslims DRB1*0101-DQA1* 0501-DQB1*0601 was unique and not observed in any of the other studied populations and different world populations. While maximum number of unique haplotypes were seen in North Eastern populations these were DRB1*0301-DQA1*0102-DQB*0201, DRB1*0701-DQA1*0201-DQB*0501, DRB1*1101-DQA1*0201-DQB*0301, DRB1*1101 -DQA1*0301 -DQB*0501 and DRB1*1501-DQA*0601-DQB*0201. DRB1*0101- DQA1*0101 -DQB1*0501 (0.3-2.9%) was

common haplotypes present in the studied populations and was also seen in different world populations of varying ethnicity like in Bulgaria, Czech, Romania and Croatia (7.0-20%) of Caucasian origin **(Ivanova *et al.*, 2002; Cerna , Fernandez-Vina *et al.*, 1992; Reed, Ho *et al.*, 1992; Grubic *et al.*, 1995)** Africans like Gabonese, Moraccan Arabs and Algerians (0.8-2.1%) **(Migot-Nabias *et al.*, 1995; Djoulah *et al.*,1994)** and Mongols like Khalkha, Khotons, Pomors, Saami etc (2.4-13.0%) **(Munkhbat *et al.*, 1997; Evseeva *et al.*, 2002)** further indicates this haplotype to be one of the ancient haplotype.

The haplotype DRB1*0701-DQA1*0201-DQB1*0201 was the most common haplotype found in our study (4.2-14.8%) and was comparable to different ethnic population of world like Koton Mongols, Korean, China Urumqi Kazaq, French, Germans, Italians etc. **(Munkhbat *et al.*, 1997; Lee *et al.*, 2005; Mizuki & Ohno 1997; Ferencik *et al.*, 1997; Petrone *et al.*, 2001)** It was found in almost all the world populations, indicating that this might be an founder haplotype.

Interestingly the haplotype DRB1*0701-DQA1*0201-DQB1*0303 found in North Indian caste populations was also present in Muslim populations with frequency ranged between 1.8-6.3%. The caste populations are dominantly of Caucasoid origin and hence show similarity with populations like Czech (7.0%), Gypsies (4.0%) and Italians (5.5%) **(Cerna *et al.*, 1992; Petrone *et al.*, 2001)**, In Africans like Oromo, Amhara and Cameroon its frequency was very low (0.4-1.2%) **(Fort 1998; Pimtanothai 2001)** while it was entirely absent in Mongoloid populations like Khotons, Khalkha, Japanese, Koreans etc. **(Munkhbat *et al.*, 1997; Saito *et al.*, 2000; Lee *et al.*, 2005)**.

Another important finding was seen in North Eastern populations that haplotype DRB1*090x-DQA1*0301-DQB1*0301 present only in tribal populations of North East i.e. *Lachung* and *Mech* with frequency of 5.1% and 1.5% respectively which was comparable to only Mongolian populations like Khotons (3.0%), Pomors (3.4%), Saami (4.3%) and Japanese (4.5%) **(Munkhbat *et al.*, 1997; Evseeva I *et al.*, 2002; Saito *et al.*, 2000)** suggesting the presence of Mongoloid element.

The most frequent DRB1*1501 haplotype in our population was DRB1*1501-DQA1*0103-DQB1*0601 (0.8-2.8%). This association was also common in Caucasian populations like Czech (3.0%), Turkish (4.0%), Italians (2.3%) **(Cerna *et al.*, 1992;**

Saruhan-Direskeneli 2000; Petrone *et al.*, 2001) This haplotype was also seen in Moraccan Arabs of African origin **(Gomez-Casado *et al.*, 2000)**

The allele frequency and haplotype analysis of our data revealed that our North Indian Caste populations and Muslim Populations show admixture of mainly Caucasoid component this is supported by number of European scholars which states that the observed west Eurasian haplogroups in India are due to the contribution of the near East west Eurasian gene pool. In general, their interpretation is based on the haplogroup frequency **(Kivisild *et al.*, 2003, Bamshad *et al.*, 2001,)**. The Indian genetic landscapes do not occur gradually, but are structured as a multitude of endogamous pockets **(Cavalli-Sforza *et al.*, 1994)**. While in our study it revealed that Sub-Himalayan North Eastern populations showed Mongoloid/Orientals admixture. It has already been reported that genetic profile of India and East Asia suggest that the expansion of the super-haplogroup M in Indians and Mongoloids, including the genesis of the region-specific sub-clusters which was clearly separated in space since then, there has been only very limited gene flow between India and eastern Asia. **(Chen *et al.*, 1995) (Table 6.5)**

Table 6.4 Two locus haplotypes of DRB*DQB

Haplotype	N.India	Kayastha	Mathur	Vaish	Rastogi	Shia	Sunni	Lachung	Mech	Rajbanshi	Pasiegos	Cantabrian	Azores	Bulgaria
0101-0501	1.00	1.40	3.10	3.10	2.80	6.60	3.60	0.00	0.00	0.00	5.00	1.80	5.40	2.00
0101-0601	0.00	0.00	0.00	0.00	0.00	1.60	4.40	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0301-0201	8.60	7.30	7.00	7.00	6.00	5.90	7.40	3.40	1.50	4.60	9.00	2.05	3.50	8.00
0301-0501	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	3.90	0.00	0.00	0.00	0.00	0.00
0403-0302	2.70	0.00	0.00	0.00	2.20	2.20	0.00	0.00	0.00	0.00	4.00	2.30	0.00	1.00
0701-0201	14.50	14.30	16.40	16.40	11.10	8.60	8.50	10.30	10.20	7.10	4.60	3.08	15.00	0.00
0701-0301	0.00	1.30	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0701-0303	4.00	4.80	5.80	5.80	5.00	2.40	3.30	0.00	0.00	0.00	0.00	0.00	2.80	0.00
080x-0301	0.00	0.00	0.00	0.00	0.00	0.00	0.00	3.40	7.90	6.70	0.00	0.00	0.00	0.00
090x-0303	1.00	0.00	0.00	0.00	0.00	0.00	0.00	5.10	3.20	0.00	0.17	0.20	1.10	0.00
1001-0501	8.60	4.60	9.00	9.00	7.60	0.00	7.10	3.40	0.00	2.00	12.70	0.00	0.00	1.00
1101-0301	5.90	4.40	3.80	4.00	3.80	3.50	6.30	4.30	0.00	0.00	3.80	1.00	5.00	16.00
1101-0501	0.00	0.00	0.00	0.00	0.00	0.00	0.00	4.90	5.50	6.90	0.00	0.00	0.00	0.00
1201-0301	0.00	0.00	0.00	0.00	0.00	0.00	0.00	7.70	8.70	8.70	0.00	0.00	1.90	0.00
1202-0301	1.40	1.50	0.00	0.00	1.70	1.50	1.50	0.00	0.00	0.00	0.00	0.00	0.00	0.00
1301-0201	0.00	1.10	0.00	0.00	0.00	0.00	2.30	0.00	0.00	0.00	0.00	0.00	0.00	0.00
1301-0601	0.00	1.50	1.20	1.20	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
1301-0603	5.40	4.20	3.50	3.50	3.30	0.00	0.00	0.00	0.00	0.00	7.70	1.00	8.90	7.00
1401-0501	0.00	0.00	0.00	0.00	0.00	0.00	0.00	2.50	2.30	6.10	0.00	0.00	0.00	0.00
1401-0503	0.00	3.40	3.50	3.50	1.70	0.00	3.40	0.00	0.00	0.00	0.00	0.00	0.00	0.00
1501-0601	0.00	6.00	6.00	6.00	7.00	1.10	2.00	3.40	0.00	0.00	0.00	0.00	4.30	0.00
1501-0602	0.00	1.00	1.20	1.20	1.20	0.00	2.90	4.30	3.10	0.00	14.07	2.05	0.00	1.00
1502-0602	0.00	4.40	0.00	7.00	4.50	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00
1502-0601	0.00	1.50	2.50	2.50	1.20	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Table 6.4 Two locus haplotypes of DRBDQB

Haplotype	Italy	German	French	PakSindh	Pathan	Kalash	Burushu	Brahui	Baloch	Parsi	Macedonia	Croatia	Romania	Canada
0101-0501	1.80	6.30	9.30	0.00	0.00	0.00	11.30	13.50	11.20	17.20	0.00	8.90	7.40	0.00
0101-0601	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0301-0201	49.20	7.70	11.00	22.50	12.40	8.60	21.00	32.00	20.90	6.30	0.00	0.00	14.90	3.20
0301-0501	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0403-0302	1.10	6.30	2.10	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.80	1.70	2.90	9.60
0701-0201	1.20	8.50	11.20	6.10	7.90	12.30	0.00	0.00	0.00	12.60	0.00	7.20	7.46	1.60
0701-0301	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.85	0.00	0.00
0701-0303	0.00	2.80	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.20	0.00	0.00
080x-0301	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
090x-0303	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.75	0.00
1001-0501	1.10	0.00	1.50	0.00	0.00	0.00	5.40	0.00	5.30	0.00	0.00	2.12	2.24	0.00
1101-0301	1.30	9.50	8.60	11.50	10.90	7.80	0.00	0.00	0.00	18.80	9.30	6.78	6.70	0.00
1101-0501	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
1201-0301	0.00	2.30	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.20	3.73	0.00
1202-0301	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
1301-0201	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
1301-0601	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
1301-0603	0.00	4.10	5.70	11.50	0.00	12.30	11.60	0.00	11.20	0.00	0.00	5.50	1.49	0.00
1401-0501	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
1401-0503	0.00	0.00	1.50	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	3.40	0.00	0.00
1501-0601	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	8.60	0.00	0.00	0.00	0.00
1501-0602	0.00	7.40	7.20	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	8.90	4.48	0.00
1502-0602	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
1502-0601	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.85	0.75	0.00

Table 6.4 Two locus haplotypes of DRB*DQB

Haplotype	Russia	N China	SD Han	S Korea	Japan	Hakka	S China	M China	N China	Abo-rigens	Jerba	Mat-mata	Tunisia	Cook	Samoa	To-kelau	Tonga
0101-0501	8.50	2.80	0.00	6.30	6.50	0.00	0.00	1.90	1.32	0.00	4.90	1.90	3.10	0.00	0.00	0.00	0.00
0101-0601	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0301-0201	9.00	2.80	3.60	3.20	1.70	0.00	5.41	3.08	2.63	0.00	0.00	11.00	15.00	0.00	0.00	0.00	0.00
0301-0501	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0403-0302	0.00	0.00	0.00	4.20	3.80	0.00	4.50	4.01	3.95	0.00	0.00	5.60	2.90	24.00	15.00	36.00	17.00
0701-0201	8.00	11.00	8.70	8.90	1.90	0.98	2.14	6.88	11.94	0.00	11.00	17.00	15.00	0.00	0.00	0.00	0.00
0701-0301	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0701-0303	2.50	0.00	2.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.80	1.90	1.90	0.00	0.00	0.00	0.00
080x-0301	0.00	0.00	0.00	0.00	1.80	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
090x-0303	0.00	14.20	9.70	6.80	12.30	12.78	4.56	7.62	4.08	2.78	0.00	0.00	4.80	9.00	25.00	4.00	15.00
1001-0501	0.00	1.70	0.00	3.20	1.00	0.00	0.00	0.00	0.00	0.00	3.30	1.90	4.80	0.00	0.00	0.00	0.00
1101-0301	5.50	6.80	4.60	2.60	2.70	7.38	1.45	3.08	1.75	8.01	3.20	4.30	4.37	11.00	7.00	0.00	16.00
1101-0501	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
1201-0301	0.00	2.20	0.00	2.10	3.00	9.86	3.63	6.76	4.12	7.10	0.00	0.00	1.60	0.00	0.00	0.00	0.00
1202-0301	0.00	4.50	7.70	0.00	1.10	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.60	16.00	8.00	22.00	6.00
1301-0201	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
1301-0601	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
1301-0603	5.50	0.00	2.00	0.00	0.70	2.19	0.00	3.32	4.97	0.00	0.00	6.80	1.90	0.00	0.00	0.00	0.00
1401-0501	0.00	0.00	2.00	0.00	2.20	4.00	8.93	4.60	4.93	16.30	0.00	0.00	0.00	0.00	0.00	0.00	0.00
1401-0503	0.00	0.00	0.00	0.00	2.00	0.00	0.00	0.00	0.00	0.00	1.80	0.00	0.97	0.00	0.00	0.00	0.00
1501-0601	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.80	12.00	2.10	0.00	0.00	0.00	0.00
1501-0602	0.00	11.70	13.20	8.40	8.20	4.76	7.39	9.86	14.27	7.36	2.00	0.00	6.80	0.00	0.00	0.00	0.00
1502-0602	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
1502-0601	0.00	6.20	4.60	3.20	10.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	2.10	0.00	0.00	0.00	0.00

Table 6.5 Three locus haplotypes of DRB*DQA*DQB

DRB-DQA-DQB	Kayasth	Mathur	Vaish	Rastogi	Shia	Sunni	Lachung	Mech	Rajbanshi	Ashk-jew	NonAshk	Russia	Bulgaria	Czech	Gypsy
0101*0101*0501	0.30	2.90	2.00	1.70	2.80	1.60	0.80	0.00	3.10	4.70	0.00	8.20	7.00	20.00	2.00
0101*0501*0601	0.00	0.00	0.00	0.00	1.50	1.50	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0301*0102*0201	0.00	0.00	0.00	0.00	0.00	0.00	2.60	2.40	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0301*0301*0201	1.80	1.90	2.20	1.70	0.00	1.20	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0301*0501*0201	4.90	4.50	3.30	3.80	3.30	3.30	0.00	0.00	0.00	7.00	4.40	0.00	7.00	18.00	1.00
0403*0301*0302	0.00	2.20	0.00	1.00	0.00	0.00	0.00	0.00	0.00	2.30	2.90	7.20	2.00	0.00	0.00
0701*0201*0201	12.00	14.80	12.60	9.40	6.20	6.20	7.00	4.20	8.60	9.40	13.90	9.60	0.00	34.00	6.00
0701*0201*0303	4.20	5.10	6.30	4.00	1.80	0.00	0.00	0.00	0.00	0.00	0.00	3.10	0.00	7.00	4.00
0701*0201*0501	0.00	0.00	0.00	0.00	0.00	0.00	2.60	0.00	4.10	0.00	0.00	0.00	0.00	0.00	0.00
080x*0101*0601	0.00	0.00	0.00	0.00	0.00	0.00	2.50	2.30	0.00	0.00	0.00	0.00	0.00	0.00	0.00
080x*0301*0301	0.00	0.00	0.00	0.00	0.00	0.00	0.00	3.90	2.50	0.00	0.00	0.00	0.00	0.00	0.00
0901*0301*0301	0.00	0.00	0.00	0.00	0.00	0.00	5.10	1.50	0.00	0.00	0.70	0.30	0.00	1.00	0.00
1001*0101*0501	3.40	8.10	3.50	6.80	2.60	4.20	0.00	0.00	0.00	0.00	0.00	0.00	0.00	2.00	0.00
1001*0401*0402	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.60	0.00	0.80	5.90	0.00	0.00	0.00	0.00
1101*0201*0301	0.00	0.00	0.00	0.00	0.00	0.00	3.40	1.50	1.00	0.00	0.00	0.00	0.00	0.00	0.00
1101*0301*0501	0.00	0.00	0.00	0.00	0.00	0.00	1.70	0.00	4.00	0.80	0.00	0.00	0.00	0.00	0.00
1101*0501*0301	3.60	3.20	3.70	3.30	2.80	4.00	2.60	3.20	0.00	0.00	0.00	0.00	18.00	10.00	6.00
1201*0501*0301	0.00	0.00	0.00	0.00	0.00	0.00	3.40	2.40	2.10	0.00	0.00	2.70	0.00	10.00	0.00
1202*0601*0301	1.50	0.00	0.00	1.70	1.50	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00
1301*0103*0603	2.60	1.30	0.00	2.10	1.80	1.20	0.00	0.00	0.00	3.10	0.70	5.10		10.00	6.00
1401*0101*0503	2.20	2.50	1.70	1.00	0.00	0.00	3.10	2.50	2.40	1.60	2.20	0.00	7.00	1.00	0.00
1501*0101*0301	0.00	0.00	0.00	0.00	1.30	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
1501*0102*0601	2.00	0.00	2.00	3.20	0.00	0.00	0.00	0.00	0.00	0.80	1.50	10.00	5.00	20.00	0.00
1501*0103*0503	0.00	1.50	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
1501*0103*0601	2.80	2.10	2.80	2.00	0.80	1.30	1.70	1.50	2.00	0.00	0.00	0.00	1.00	3.00	0.00
1501*0601*0201	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.50	2.50	0.00	0.00	0.00	0.00	0.00	0.00
1502*0101*0501	0.00	2.40	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
1502*0102*0601	0.00	0.00	0.00	1.40	0.00	0.00	1.70	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
1502*0103*0501	0.00	3.30	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
1502*0103*0501	3.10	3.30	2.80	2.30	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
1502*0103*0601	2.80	0.00	2.80	2.00	0.00	1.40	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Table 6.5 Three locus haplotypes of DRB*DQA*DQB

DRB-DQA-DQB	Turkey	Polish	Romania	Croatia	Dane	French	German	Italian	Sardinian	Spanish	Gypsy	USA	Canada	Gabon	Oromo	Amhara	Moroccan
0101*0101*0501	0.00	8.80	7.40	8.90	12.30	9.40	6.30	10.20	4.90	5.00	2.30	4.90	5.40	0.80	0.00	0.00	2.00
0101*0501*0601	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0301*0102*0201	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0	0.00	0.00	0.00
0301*0301*0201	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.60	0.00	0.00
0301*0501*0201	9.20	7.00	14.90	7.60	11.30	11.20	7.70	14.10	24.10	6.80	7.00	4.60	11.70	6.70	9.60	7.60	14.10
0403*0301*0302	3.60	0.00	2.20	1.27	0.00	0.00	0.00	3.10	0.00	0.00	0.00	0.00	0.00	0.00	3.01	2.04	1.50
0701*0201*0201	7.20	10.10	6.70	7.20	9.40	10.90	8.50	24.20	4.40	14.30	0.00	9.80	8.50	4.80	20.48	19.39	12.10
0701*0201*0303	0.60	0.00	0.00	1.27	0.00	0.00	0.00	5.50	0.00	0.00	0.00	0.00	0.00	0.00	1.20	1.00	0.00
0701*0201*0501	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
080x*0101*0601	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
080x*0301*0301	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0901*0301*0301	1.00	0.00	0.75	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.20	0.00	0.00	0.00	1.00
1001*0101*0501	2.00	0.00	2.24	2.12	0.00	1.50	0.00	1.60	2.20	1.30	0.00	0.00	0.00	0.00	2.40	3.06	3.00
1001*0401*0402	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
1101*0201*0301	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
1101*0301*0501	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.80	2.50	0.00
1101*0501*0301	10.00	6.50	6.70	6.78	0.00	0.00	0.00	25.80	0.00	0.00	0.00	0.00	0.00	2.40	0.00	0.00	2.00
1201*0501*0301	1.60	0.00	3.73	1.27	0.00	0.00	0.00	3.10	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.50
1202*0601*0301	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
1301*0103*0603	3.70	5.50	0.00	5.50	0.00	0.00	0.00	5.50	0.00	0.00	0.00	0.00	0.00	3.50	1.20	1.50	1.00
1401*0101*0503	4.60	0.00	2.20	3.40	0.00	0.00	0.00	10.90	0.00	0.00	0.00	0.00	0.00	0.30	0.00	1.00	1.50
1501*0101*0301	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
1501*0102*0601	5.80	13.00	4.48	8.90	0.00	0.00	0.00	3.90	0.00	0.00	0.00	0.00	0.00	31.00	0.00	0.00	0.00
1501*0103*0503	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
1501*0103*0601	4.00	0.00	0.00	0.85	0.00	0.00	0.00	2.30	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	6.50
1501*0601*0201	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
1502*0101*0501	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.60	0.00	0.00	0.00	0.00	0.00	0.30	0.00	0.00	0.00
1502*0102*0601	0.00	0.00	0.75	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	3.01	0.00
1502*0103*0501	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
1502*0103*0501	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
1502*0103*0601	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.80	0.00	1.30	4.50	0.00	0.00	0.00	0.00	0.00	1.00

Table 6.5 Three locus haplotypes of DRB*DQA*DQB

DRB-DQA-DQB	Algeria	Cameroon	Khoton	Pomor	Saami	Nentsy	Khalkh	Japan	Korea	SingCh	Af Am	Ijka	Sikua	Nukak	Ingano	Waunana	Embera	Tule
0101*0101*0501	2.10	0.40	8.10	8.90	13.00	1.80	2.40	4.90	6.30	0.00	1.70	2.00	2.00	0.00	0.00	0.00	0.00	0.00
0101*0501*0601	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0301*0102*0201	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0301*0301*0201	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0301*0501*0201	11.70	6.30	6.40	0.00	0.00	0.00	6.10	0.00	3.20	4.80	6.80	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0403*0301*0302	2.10	0.80	0.00	16.40	17.20	20.00	0.00	0.00	0.00	0.00	1.50	0.00	2.00	15.00	9.00	2.00	3.00	0.00
0701*0201*0201	12.80	4.40	5.90	6.80	3.10	7.30	0.00	0.00	8.40	0.00	7.10	7.00	0.00	0.00	0.00	0.00	0.00	2.00
0701*0201*0303	1.10	0.40	0.00	5.50	0.60	1.80	0.00	0.00	0.00	0.00	0.60	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0701*0201*0501	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.20	0.00	0.00	0.00	0.00	0.00	0.00	0.00
*080x*0101*0601*	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
080x*0301*0301	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	9.00	0.00	0.00	0.00
0901*0301*0301	0.00	0.00	3.00	3.40	4.30	18.10	0.00	4.50	0.00	11.20	0.00	0.00	0.00	0.00	0.00	2.00	0.00	0.00
1001*0101*0501	1.10	1.20	2.40	0.00	0.00	0.00	2.40	0.00	3.20	0.00	1.90	0.00	0.00	0.00	0.00	0.00	0.00	0.00
1001*0401*0402	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
1101*0201*0301	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
1101*0301*0501	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
1101*0501*0301	13.80	0.40	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	2.30	0.00	0.00	0.00	0.00	0.00	0.00	2.00
1201*0501*0301	0.00	0.00	0.00	2.10	2.50	15.40	0.00	0.00	0.00	0.00	0.20	0.00	0.00	0.00	0.00	0.00	0.00	0.00
1202*0601*0301	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
1301*0103*0603	2.10	2.00	0.00	4.80	3.70	5.40	0.00	0.00	0.00	0.00	0.20	0.00	0.00	0.00	0.00	0.00	0.00	0.00
1401*0101*0503	0.00	1.20	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
1501*0101*0301	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
1501*0102*0601	4.30	0.00	0.00	16.40	6.80	2.70	0.00	0.00	0.00	0.00	15.00	2.00	0.00	0.00	0.00	0.00	0.00	0.00
1501*0103*0503	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
1501*0103*0601	1.10	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
*1501*0601*0201*	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
1502*0101*0501	1.10	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
1502*0102*0601	0.00	0.00	0.00	0.00	0.00	0.00	1.50	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
1502*0103*0501	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
1502*0103*0501	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
1502*0103*0601	0.00	0.00	6.50	0.00	0.00	0.00	4.90	9.50	3.20	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Distribution of genetic variation

The informativity of a locus, also known as the degree of diversity, is measured by an index called heterozygosity that is based on both the number and proportion of individual alleles. Heterozygosity ranges from 0 to 1, where "0" is ***no variation*** and "1"is ***infinite variation***. Most of the studies carried out in last one decade based on different numbers of autosomal STR and HLA loci have showed very high level of genetic diversity (74-85% average observed heterozygosity) among African populations **(Calafell *et al.*, 1998; Destro-bisol 1999)**. Furthermore, it was also figured out that the non-Africans carry only a fraction of the diversity found in Africans **(Calafell *et al.*, 1998; Underhill *et al.*, 2000)**, with the notable exception of Indian populations, which are reported to harbor more genetic diversity than any contemporary population other than Africans **(Khan *et al.*, 2003; 2004; Agrawal and Khan 2005)**.

The use of HLA in population studies has an advantage because of their high heterozygosity, this feature makes HLA as a markers of choice for many genetic and forensic applications. It is important that for population variation studies one need, not only highly variable loci, but loci in which the allele frequencies differ across the study populations. The mean observed heterozygosity over all the HLA loci was 0.7765 in *Kayastha*, 0.7607 in *Mathur,* 0.7628 in *Vaish*, 0.7526 in *Rastogies* , 0.7024 in *Shia Muslims*, 0.7060 in *Sunni Muslims,* 0.7213 in *Lachung*, 0.5747 in *Mech* and 0.7090 in *Rajbanshi*. The mean over all the nine populations was 0.7184 which was comparable to the mean heterozygosity in the Asian populations like Buyi (0.8880), Koryacs (0.8921) **(Imanishi, 1992, 1998)**, Pakistan Brahui (0.76%) **(Mohyuddin *et al.*, 2003)**. The existence of high genetic diversity either infer that population concern is an ancestral population that have maintained a larger effective population size and had a long existence that allowed mutations and recombination to increase the level of heterozygosity **(Gabriel *et al.*, 2002; Kidd *et al.*, 2004)**. The second possibility is of the colossal gene flow from different corners of the world having created immense genetic diversity in a population **(Majumder 1997)**. However, a long and in depth period of human survival and constant gene flow from different parts of Asia and Europe in recent past have been a possible reason for the high level of genetic diversity among the studied nine populations. Intra and inter population variation analysis revealed a high range of intra population diversity deduced from the analysis of average gene diversity and average observed heterozygosity within each

population. However, values of gene diversity estimates (0.6286 in *Mech* to 0.7783 in *Mathurs*) indicated high intra-population variation but more inter-population similarity. It has also been reported that the intra population diversity is not only present at the national level, but also within smaller geographical regions of the country. On comparing various parameters assessing genetic diversity in nine studied populations, it was clearly evident that each of these parameters runs in parallel between the nine studied groups suggesting very less inter-population genetic variation.

Apportionment of genetic variation

Various genetic studies ranging from blood groups, protein polymorphism, RFLPs, SNPs and STR loci have shown that most of the genetic variation (~85-90%) is distributed among individuals rather than among populations **(Jorde *et al.*, 2000; Tishkoff and Kidd, 2004; Kidd *et al.*, 2004)**. The variation reported to occur between populations or between population groups together varies between 9-15%. However, it has been reported recently **(Rosenberg *et al.*, 2002)** that the component of between population variance decreases when analysis is carried out within a geographical region. Three different analyses have been carried out in the present data set to structure the genetic variation into different hierarchical status of within and between populations.

At first, Nei's gene diversity based analysis was done, which established that most of the genetic variation existing in the studied populations is due to the variation between individuals as the Gst value in North Indian Caste population was 0.43%, Muslim population 0.53% and was 0.40% in North Eastern populations. This indicated that the studied populations of different groups have not yet diverged from one another to a great extent and might have originated from a common ancestor. Fst analysis was performed to deduce the genetic differentiation between different population groups. This analysis corresponds to the findings of gene diversity analysis. Low value of Fst, 0.0358 in North Indian Caste populations, 0.0108 in Muslim populations and 0.0193 in North Eastern populations shows less degree of genetic differentiation between the populations in particular groups owing to the presence of same recent common ancestor or heavy gene flow between the population groups in the past.

Finally we carried out analysis of variation (AMOVA) and found that results of AMOVA were similar with the Gst and Fst results taking three groups comprising the two Muslims, four caste populations of North India and three North Eastern populations in the present study. It was found that 98.13% of the genetic variation was within each population. Most of the variation was due to the differences between the groups i.e 1.66% while only 0.21% variance was the difference observed between the populations within the group. This suggests that the populations in each of these groups are genetically more similar to each other than the populations of different groups. But the differences between the population groups were only 1.66%, which indicates that there might be some gene flow between the studied populations. A synthesis of similar hierarchical distribution of variance between 19 global populations based on HLA Class I and Class II loci was performed by **Diogo Meyer *et al.*, 2006.** Their analysis is based on groups of four Sub- Saharan, three European, four South East Asian, three North East Asians, two Oceania, and three North and South American populations. Their results revealed that the most of the variation was within population (87.7-92.8%). While proportion of variance within groups decreases like only 2.98-7.47%. Still the populations inhabiting these regions differ from each other at genetic level **(Table 6.6)**. Our findings are in concordance with **Diogo Meyer *et al.*, 2006** analysis and suggest that the amount of genetic variation between populations in a socio-religious group is minimal as compared to variance between groups. Our results show that caste populations, two inbreeding Muslim population and North Eastern populations reveal significant difference at various HLA loci and also at different haplotypes. We further observed the significant differences when performed analysis of variance which revealed that these differences may be due to socio cultural structuring.

Table: 6.6 AMOVA analysis of 19 global populations based on HLA Class I and Class II Loci (Diogo Meyer *et al.*, 2006)

Variance Components %	Variance Componenets and 95% confidence interval%				
	HLA-A	HLA-B	HLA-C	HLA-DRB1	HLA-DQB1
Within Population	88.71	92.8	91.9	88.11	87.79
Among population within region	3.87	4.22	4.55	4.42	5.21
Among Regions (Groups)	7.41	2.98	3.55	7.47	7

Diogo Meyer *et al.*, 2006

An insight into the effect of socio-cultural barriers on Indian gene pool

An elite historical, demographical and socio-cultural contour makes Indian populations a melting pot for study of genetic variation and differentiation. On one hand, copious migratory events have created an extensive range of genetic diversity while inflexible and stern socio-cultural barriers have structured this diversity into different endogamous groups identified by the name of "castes". This term "Caste" is an assemblage of various socio-cultural customs; traditions and barriers that have created abundant number of hierarchically arranged 'endogamous' groups. This social hierarchy system is unique and exquisite because in caste system, birth of an individual governs and decides most of the proceedings of his life including the choice of mating partner. Marriage between partners of equal status is preferred, and reproduction in the caste system is largely endogamous **(Heinz 1999).**

In order to testify the hypothesis of social cleavage resulting into genetic structuring even in a confined geographical area, we have carried out various statistical analyses to determine the level of genetic differentiation between the studied populations. As mentioned earlier, all the studied populations belong to different religious strata and practice highly restricted marital patterns. The level of populations structuring triggered by the caste system further gets enhanced due to an additional level of endogamy called 'surname endogamy' practiced mainly by the most stringent higher caste group **(Agrawal *et al.*, 2004).** As four of the studied

populations belong to middle caste group, two are the Muslim populations who practices high level of consanguinity and from rest three caste and tribal group from North eastern belt of India therefore probability of genetic differentiation increases in these groups.

At first, we structured data set into three groups namely, only Muslims (2 populations), only middle caste populations (4 populations) and North Eastern group (e populations) studied for the same set of HLA markers were included. When we repeated Gst, Fst and AMOVA assessment with the structured data set, it became clearer that the populations in each of the three groups, middle caste and Muslims and North Eastern group are genetically more similar to populations of the same group.

The findings suggested that the amount of genetic variation attributable to between populations in a socio-religious group is minimal as compared to variance between groups. Therefore, the three groups differ from each other at genetic level owing to the socio-cultural structuring. However, the genetic profile of all the nine populations included in the analysis exhibit extensive genetic overlap either due to sharing of same common recent ancestor or due to the fact that caste system is ~3000-4000 years old **(Balakrishnan *et al.*, 1978; Roychaudhary *et al.*, 2000)** and the time period is significantly small to create the genetic differentiation. It has been suggested that there has been a considerable admixture between Islamic traveler who introduced their culture and language into other Indian groups; it might have brought certain changes which at present are not visible. However, when *Shia* and *Sunni* Muslims were analysed at Mitochondrial and Y chromosome DNA in one of our earlier studies we observed that there were no difference at Mitochondrial DNA level but *Shia* Muslims showed presence of significant YAP element showing a substantial greater similarity with the Middle East Asian populations. At HLA loci the results are similar to that of Mitochondrial DNA rather than Y chromosome.

Our earlier studies on Mitochondrial DNA have shown similarity with the Andaman Islanders. We have proposed that North Eastern populations might have been the most ancestral populations of India. However, this has not been emerged through the HLA data.

Phylogenetic assessment

Finally, we carried out the phylogenetic analysis to infer the genetic similarity and differences between the nine studied populations based on two different approaches- genetic distance and maximum likelihood model. The population pair wise matrices created from genetic distance calculations, **Nei's D_A**.

Most distinct observation that emerged out from the phylogenetic analysis was differential genetic relationship of two *Muslim* population groups with that of four caste populations despite of having same origin. *Shia Muslims* showed nearly double genetic distances from caste populations as compared to *Sunni Muslims* in distance matrices. Sunni Muslims constitute the major sects of Muslims in India and are ~87% of the total Muslim population. They have ruled different parts of India for around 800 years and during their rein, they expanded within the sub-continent. Despite practicing consanguinity, they married outside their religion. This is supported by various studies **(Mukherjee *et al.*, 2001)** that Indian Muslims, specifically North Indian Muslims, have high genetic similarity with other Hindu caste populations. On the contrary, Shia populations have lesser admixture with the caste populations. In the North Indian caste group, *Kayastha*, *Mathurs*, Vaish and *Rastogies* were genetically more similar. It was also observed that *Kayastha* and *Mathurs* were close to each other as the distance between these two was very low while Vaish and *Rastogies* were close to each other. As it has been earlier postulated that *Mathurs* have been derived from *Kayastha* while *Rastogies* from Vaish. Here we would like to mention that the historical, cultural and social evidences have been strengthened by a genetic study based upon HLA class II loci.

From North Eastern group it was found that this group was very far from *Muslims* and North Indian Caste populations, *Lachung* and *Mech* were found to be genetically more similar to each other but Rajbanshi's show high genetic distance than *Lachung* and *Mech* as these two are tribals of Tibeto-Burman group while *Rajbanshis* are believed to be admixtured population.

When four middle caste populations were compared with Muslim population and North Eastern populations, a patristic separation of Muslims, middle caste populations and North Eastern population was visible. In phylogram Muslims and North Indian Caste populations clustered on a same branch revealing more genetic similarity than from the four middle caste populations. While North Eastern populations clustered separately.

Comparison with world populations

Comparison of the nine populations with the world populations was carried out by using both ML phylogram and PC-plot analysis. It was observed that basal cluster pattern of ML phylogram carries three geo-ethnic groups indicating that role of genetic drift as a major force of evolution **(Figure 6.1)**. Both the trees have longer African branch. Such a patristic separation was also visible in the PC-plot analysis **(Figure 6.2)**. The African populations have been clustered into groups. Such clustering was also reported by **Cavalli-Sforza and Feldman, (2003) Underhill *et al.*, (2001)** on the basis of polymorphisms of 120 protein-coding genes and Y-chromosome binary haplogroup respectively.

This sub clustering further strengthens the utility of HLA loci in deciphering the accurate phylogenies even within the same geographical region. Middle Eastern Arabs displayed a branch nearer to Caucasians suggesting strong Caucasian element suggestive of the Demic expansion of the Middle East genes. Recently, Y-chromosome SNP analysis by **Al-Zahery *et al.*, 2003** also revealed similar pattern in other Middle Eastern populations. European branching pattern have resolved Bulgaria, Greece, Italy, Czech, Turkey and Spaniards etc to a separate cluster from that of Amenrican branch under which populations like Sikuanis, Tules, Koguis, Inganos, African-Americans and Ijkas etc fall (**Trachtenberg *et al.*, 1996**). It has been suggested by **Trachtenberg *et al.*, 1996** that these populations underwent little or no admixture with other neighboring populations in the last 300 years and hence are distinct and form a separate cluster. Muslims and Caste populations clustered on a separate branch sharing the same node with the Caucasians like Azores, Basque and South France in ML phylogram. While North Eastern populations clustered with Mongoloid populations like Pomors, Saami, Nentsy, South Korea, China Beizing and Xian, North Chinese and Japan.

When PC plot analysis was performed similar results were obtained as that of ML phylogram. Four separate clusters were seen in PC plot. It was found that all the Africans were seen on separate axis, our North Indian caste populations and Muslim populations were near to European populations like France, Romania, Polish and Basque while our North Eastern populations were found near to Oriental populations like China Beizing and Xian. South and Central Americans were found separately.

Clinal distribution of Caucasian and Asian genes in Indian populations

Indian sub-continent has experienced a massive gene flow from at least two Neolithic episodes of migrations. Firstly, about 10-15 KYA, when agriculture developed in the Fertile Crescent region, a part of an eastward wave of human migration **(Cavalli-Sforza *et al.*, 1994; Renfrew 1987)** has entered India. This wave brought Dravidian languages into India **(Renfrew 1987)** mainly, Elamo-Dravidian languages **(Ruhlen 1991)**, which are now confined to southeastern India. A later episode, the arrival of pastoral nomads from the central Asian steppes to the Iranian plateau, ~4,000 YBP, brought with it the Indo-European language family, which eventually replaced Dravidian languages from most of Pakistan and northern India, perhaps by an elite dominance process **(Renfrew 1987; 1996; Quintana-merci *et al.*, 2001)**. These nomadic migrants may have consolidated their power by admixing with native Dravidic-speaking (e.g., Telugu) proto-Asian populations **(Cavalli-Sforza *et al.*, 1994)**, and subsequently established the Hindu caste hierarchy to legitimize and maintain this power **(Poliakov 1974; Cavalli- Sforza *et al.*, 1994)**.

Therefore, the genetic makeup of Indian populations is stated to be an admixture of Caucasian and Asian gene pool. **Bamshad *et al.*, 2001** studied 8 south Indian populations and deduced that south Indian populations are genetically closer to Asian populations. The data generated from present study has shown that the six North Indian (Caste and Muslim populations) populations are genetically closer to the Caucasian than from Asian populations while North Eastern populations have affinity towards Mongoloids.

Figure 6.1: Neighbour-joining tree depicting genomic affinities among different world populations based on HLA class II loci polymorphisms (Unrooted)

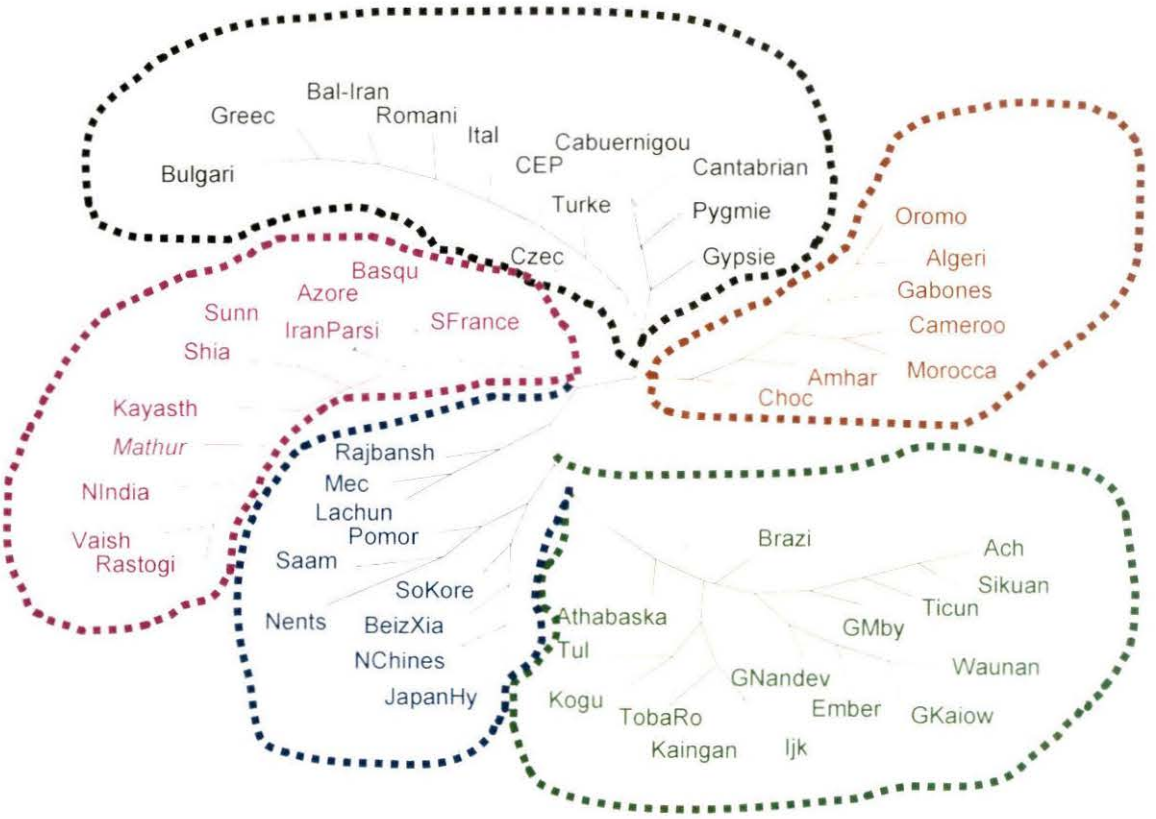
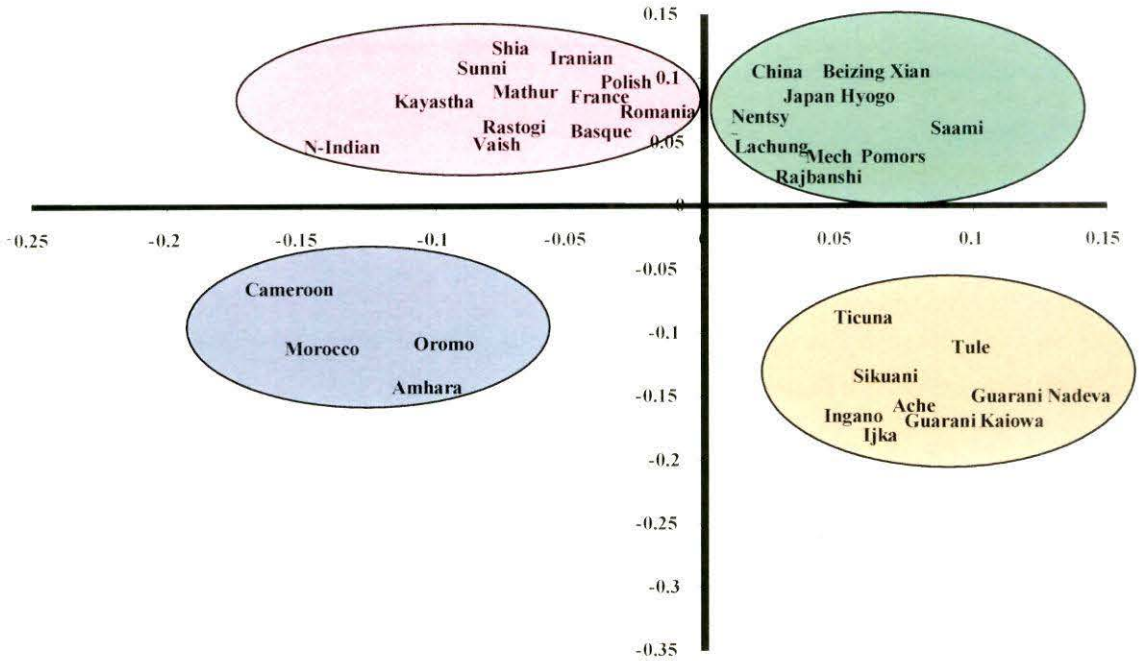


Figure 6.2: PC Plot analysis Based on HLA Class II Loci



Conclusively, the analysis of HLA loci in nine populations of North India have propvided a plethora of information about the genetic profile of studied populations. At first, the impending role of extensive gene flow through a series of migrations and invasions have shaped and distributed the contemporary genetic variation across all population groups of north India and has created enormous amount of genetic diversity that marks high level of intra-population genetic variation observed in all the five populations. Furthermore, the meta-analysis of the diversity estimates revealed that the north Indians (Indo-Europeans) exhibits maximum genetic diversity among contemporary Indian populations. Secondly, although intra-population differences were marginal but there occurs a definite pattern of genetic variation distribution in which different populations structured into socio-religious groups have more genetic similarity with the populations of the same group and are genetically more distant from populations of other groups. Thirdly, there have been a massive gene flow between the population groups or they have same recent common ancestor due to which there occurs a genetic overlap among these groups. Finally, the genetic affinity of Indians and that of different caste groups

towards Europeans is distributed in a cline where geographically north Indians and socially both caste and Muslim populations are genetically closer to the Europeans.

However, it should be emphasized that the DNA variation studied here is thought to be selectively neutral and thus represents only the effects of population history. These results permit no inferences about the phenotypic differences between populations. In addition, different caste/religious populations, reflecting a shared history, share alleles and haplotypes needs to be analyzed for more genetic markers to reach a conclusion. Indeed, the findings of the present study emphasized the longstanding appreciation that the distribution of genetic polymorphisms in India is highly complex. Further investigation of the spread of anatomically modern humans throughout South Asia will need to consider that such complex patterns are normal rather than the exception.

Summary

Amalgamation of the genotypic and haplotypic organization of the nine populations based on the analysis of HLA ClassII loci have deciphered vital information about the amount, pattern and distribution of genetic variation in different studied populations of India. The perspective correlation of the genetic profile of studied populations with the past human movements and other historical records facilitated determination of their genetic ancestry/origin. Furthermore, the data generated from the present study provided valuable information to comprehend the effect of socio-cultural barriers on the genetic makeup of Indians population. The results also provided an inference about the pattern and causes of genetic similarity and differences between different Indian and world populations.

The analyses of HLA Class II loci in all the populations have divulged surfeit facts about the genetic sketch of studied populations. The long range of observed allelic states, high allelic diversity depicted by gene diversity, haplotype distribution, heterozygosity and several other parameters including analysis of variance have demonstrated that the contemporary studied populations are highly diverse population groups. The reason of such high diversity include both accumulation of large number of mutation during long term survival and the colossus gene flow from all over the world that have introduced a plethora of different allelic states to India. Furthermore, distribution of the genetic variation within and between populations revealed that the intra-population differences were marginal but there occurs a definite pattern of diversity distribution where populations structured into socio-religious groups have more genetic similarity within their own group and are genetically more distant from populations of other groups.

Overall, synthesis of the genetic data at HLA loci indicated a strong genetic overlap among the studied groups owing either to the recent massive gene flow between the populations or possible identical recent common ancestor. Genetic distance based study revealed that in north Indian caste group genetic similarity was seen between *Kayastha* and *Mathurs* while the genetic distance between *Vaish* and *Rastogies* was low. It has been postulated earlier

that *Mathurs* have been derived from *Kayastha* while *Rastogies* from *Vaish* so our data gives some clue about the possible origin of *Mathurs* and *Rastogies*.

Sunni Muslims shared more genetic similarity with the caste populations. This may be because of the fact that although Muslims practiced consanguinity, they married outside their religion. This was further supported by various studies that Indian Muslims, specifically North Indian Muslims, have high genetic similarity with other Hindu caste populations. On the contrary, *Shia* populations have lesser admixture with the caste populations. Populations from North Eastern region were distant from Muslims and North Indian Caste populations, while the two tribal populations i.e. *Lachung* and *Mech* were genetically similar but *Rajbanshi's* were distant from the tribal populations. This is because of the fact that *Rajbanshi's* adopted the process of transformation of a tribe into a caste resulting in a Tribe-Caste continuum and in between this process populations like *Rajbanshi's* who were once tribal but gradually adopted the attributes of caste system.

Finally, the phylogenetic assessment based on HLA Loci reveals that the genetic affinity of Indians towards other world populations is distributed in a cline where geographically north Indians and socially both middle caste and Muslim populations are genetically more closer to the Eurasians than East Asians. Muslims and caste populations clustered on a separate branch with the Caucasians like Azores, Basque and South France in ML phylogram. Some Middle Eastern populations also displayed a branch nearer to Caucasians suggesting strong Caucasian element suggestive of the Demic expansion of the Middle East genes. North Eastern populations clustered with Mongoloid populations like Pomors, Saami, Nentsy, South Korea, China Beizing and Xian, North Chinese and Japan. Similar results were also seen on PC Plot analysis.

The phylogeny of HLA polymorphisms studied in various studied nine Indian populations generated a molecular narration of the population demography, relatedness, and genetic substrata reflecting possible signatures of ancestors and novel micro-evolutionary genetic differentiations. Additionally, the haplotypic profile of HLA deciphered the underlying variation in the genealogy of studied populations. The complete synthesis of two and three locus haplotypes also demonstrated that north Indian populations (caste and Muslim populations) showed the presence of mostly Eurasian and Middle East (specific for Muslims) element while Sub-Himalyan North Eastern

populations showed presence of haplotypes which were more frequent in Mongoloids suggesting admixturing with most of the Mongoloid /Oriental. The micro-evaluation of some of these haplotypes revealed that the perspective view of the weak Garden of Eden hypothesis that Eurasians have received some waves of migration from India could not be ruled out.

A vital outcome of the HLA analysis was the robust signals of each of the movements in north India starting from earliest hominids that reached India through southern coastal route, to the farmers of Neolithic agriculture expansion, Indo Aryan speaking nomads from central Asian steppes and most recent Muslim influx from central Asia and Middle East while in north eastern part there was also limited gene flow between India and Eastern Asia. The haplotype distribution magnified the conclusions of HLA analysis that there occurs a disproportion in rate of admixture between the two Muslim sects with four caste populations of north India. High sharing of two and three locus haplotypes were suggestive of genetic homogeneity of Muslim and caste groups while presence of unique haplotype DRB1*1501-DQA1*0101-DQB1*0301 in *Shia* corresponds to the preservation of original gene pool. Finally, the geographic cline of genetic affinities of north Indians with Europeans and North Eastern populations with Mongoloids was more realistic as clearly the native Indian haplotypes were more pronounced in Caucasian haplotypes and North Eastern haplotypes have similarity with Mongoloids.

Conclusively, the combined picture represented by the empirical results of the high-resolution analysis of HLA Loci infers that (1) the studied nine populations are genetically highly diverse people with most of the variation scattered between individuals. (2) The genetic differentiation does exist between numerous of endogamous groups across Indian mainland but the differentiation is mainly configured geographically and linguistically, and the stringent socio-cultural norms have only a trivial contribution. (3) The advents of migrations and invasions in northwestern and north eastern corridor of India have more genetic evidences towards native Indian origin. (4) The legacy of Indo-Aryans is discernible in the gene pool of each of the studied north Indian population (5) There is differential admixture rate of *Sunnis* and *Shiites* with other Indian populations.(6) North Eastern populations showed affinity towards Mongoloid populations.

Overall, the genetic configuration of Indians is as complex as the history of Indian mainland, interwoven in numerous threads of unknown facts. More genetic data from this part of the world and other critically important regions like Afghanistan and Iraq is highly necessitated in tracing the missing blocks of the causes and consequences of human genetic variation. Present study has provided some clues that might help in unwinding the complex interwoven threads of Indian genetic composition.

Bibliography

Bibliography

Abbas A.K, et al. *Cellular and Molecular Immunology.* Philadelphia: W.B. Saunders, 1991.

Abdennaji Guenounou B, Yacoubi , LoueslatiB, Buhler S, Hmida S, Ennafaa, Khodjet-Elkhil, Moojat, Dridi, Boukef, Ben Ammar Elgaaied& A. Sanchez-Mazas HLA class II genetic diversity in southern Tunisia and the Mediterranean area International Journal of Immunogenetics 2006; 33 Issue 2 Page 93

Agrawal S, Arundhati K, Bhardwaj U et al.. HLA antigens and haplotype frequencies in Bhargavas and Chaturvedies of UP (India). *Ind J Hum Genet,* 1999 5(1): 25-30.

Agrawal S, Khan F. Reconstructing recent human phylogenies with forensic STR loci: a statistical approach.BMC Genet. 2005 Sep 28;6:47.

Agrawal S, Kishore MC MHC class I gene expression and regulation. J Hematother Stem Cell Res. 2000 Dec;9(6):795-812

Agrawal S, Muller B, Bharadwaj U,Bhatnagar S, Sharma A, Khan F, Agarwal SS. Microsatellite variation at 24 STR loci in three endogamous groups of Uttar Pradesh, India. Human Biology 2003 Feb;75(1):97-104

Albis-Camps,M. and Blasczyk,R. Fluorotyping of HLA-DRB by sequence specific priming and fluorogenic probing. *Tissue Antigens* 1999: 53: 301-307.

Allen,P.J., Amos,W., Pomeroy,P.P. and Twiss,S.D : Microsatellite variation in gray seals (*Halichoerus grypus*) shows evidence of genetic differentiation between two British breeding colonies. *Mol. Ecol* 1995: 4: 653-662

Almawi WY, et al. Distribution of HLA class II (DRB1/DQB1) alleles and haplotypes among Bahraini and Lebanese Arabs *Transplant Proc* 2004 Jul-2004 Aug; 36(6):1844-6

Alper CA, Awdeh Z, Yunis EJ: Conserved, extended MHC haplotypes.Exp Clin Immunogenet 1992, 9:58–72

Andrien M, Dupont E HLA 1998,Pg119

ApaniusV, PennD, Slev_P.R, Ruff, and Potts "The nature of selection on the major histocompatibility complex.". *Critical Reviews in Immunology*,1997. 17: 179-224.

Arguellolo,J.R, Little,A.M, Bohan,E, Goldman,J.M, Marsh, S.G.E., Madrigal ,J.A. High resolution HLA class I typing by Reference strand mediated conformation analysis (RSCA). *Tissue Antigens.* 1998: 52: 57-66.

Arnaiz-Villena A, Benmamar D, Alvarez M, Diaz-Campos N, Varela P, Gomez-Casado E, Martinez-Laso HLA allele and haplotype frequencies in Algerians. Relatedness to Spaniards and Basques.Human Immunology 1995: 43: 259

Arnaiz-Villena,A. Iliakis,P, Gonzalez-Hevilla,M, Longas,J, Gomez-Casado,E. Sfyridaliv,K., Trapaga,J., Silvera-Redondo,C. Matsouka,C. and Martinez-laso,J. The origin of Cretan populations as determined by characterisation of HLA alleles. *Tissue Antigens* 1999; 53: 213-226.

Balakrishnan K, Pitchappan RM, Suzuki K, Kumar US, Shanthakumari R, Tokunaga K. HLA affinities of Iyers: A Brahmin population of Tamil Nadu, South India. *Human Biolog* 1996; 68 (4): 523-537

Bamshad Richard Michael, Mastan Naidu, B.V. Ravi Prasad, P. Govinda Reddy, Arani Rasanayagam, Surinder S. Papiha, Toomas Kivisild, W. Scott Watkins, Mary E. Dixon, Chris E. Ricker, Baskara B. Rao, Genetic Evidence on the Origins of Indian Caste Populations. Genome Res. J. 2001 11: 994-1004;

Barbujani G, Magagni A, Minch E, & Cavalli-Sforza L. L An apportionment of human DNA diversity. Proceedings of the National Academy of Sciences. 1997; 94, 4516–4519.

Barnhart, Benjamin J. "DOE Human Genome Program". Human Genome Quarterly 1989;1: 1.

Belov, Katherine, Janine E. Deakin, Anthony T. Papenfuss, Michelle L. Baker, Sandra D. Melman, Hannah V. Siddle, Nicolas Gouin5, David L. Goode, Tobias J. Sargeant, Mark D. Robinson, Matthew J. Wakefield, Shaun Mahony, Joseph G. R. Cross, Panayiotis V. Benos, Paul B. Samollow, Terence P. Speed, Jennifer A. Marshall Graves, Robert D. Miller. "Reconstructing an Ancestral Mammalian Immune Supercomplex from a Marsupial Major Histocompatibility Complex". *PLoS Biol* 2006; 4(3) (e46).

Benacerraf,B. and McDevitt,H.O. Histocompatibility linked immune response genes. *Science:*1972: 175: 273-279.

Bhatnagar, D. P. *Immunization in Pregnant Women.*M.Sc. Diss. (unpublished) Punjabi University, Patiala (1972)

Bhattacharyya Nitai Pada, Basu Priyadarshi, Das Madhusudan, Banerjee Srimanta Pramanik, Rajat, Roy Bidyut, Roychoudhury Susanta ,and Majumder Partha P. Negligible Male Gene Flow Across EthnicBoundaries in India, Revealed by Analysis of Y-Chromosomal DNA Polymorphisms *Genome Res* 1999 9: 711-719

Black FL and Salzano FM, Evidence for heterosis in the HLA system. *Am. J. Hum. Genet.* 1981. 33: 894–899.

Blunt. E.A.H. *The Caste System of Northern India*, 1931, republished 1964, S. Chand, Delhi

Bodmer JG, Marsh SG, Albert ED, Bodmer WF, Bontrop RE, Dupont B, Erlich HA, Hansen JA, Mach B, Mayr WR, Parham P, Petersdorf EW, Sasazuki T, Schreuder GM, Strominger JL, Svejgaard A, Terasaki PI: Nomenclature for factors of the HLA system. Tissue Antigens 1998:53:407–446

Bowcock,A.M., Riuz-Linares,A., Tomfohrde,J., Minch,E., Kidd,J.R., Cavalli-Sfroza,L.L. High resolution of human evolutionary trees with polymorphic microsatellites. *Nature*: 1994: 368: 455-457.

Bowler JM, Johnston H, Olley JM, Prescott JR, Roberts RG, Shawcross W, Spooner NA New ages of human occupation and climatic change at Lake Mungo, Australia. Nature 2003; 421:837–840

Bowness P. HLA B27 in health and disease: a double-edged sword? Rheumatology 2002; 41: 857-868

Braun,L. Race, ethnicity, and health: Can genetics explain disparities?Perspectives in Biology and Medicine. 2002; 45, 159–174.

Browning,M. and McMichael,A. (Eds.) (1996): In HLA and MHC-genes, molecules and function. BIOS Scientific Publishers.

Calafell,F., Shuster,A., Speed,W.C., Kidd,J.R., Black,F.L., and Kidd,K.K. Genealogy Reconstruction From Short Tandem Repeat Genotypes In an Amazonian Population. *Am. J. Phys. Anthropol*: 1999: 108: 137-146.

Cann RL. Genetic clues to dispersal of Human populations: Retracing the past from the present. *Science,* 2001, 291:1742-1748.

Carmen Alaez, Miriam N Vazquez-García,Hilario Flores, Eduardo Infante,Andrea Munguía, Clara Gorodezky The genetic structure of Mexican Mestizos of different locations: Tracking back their origins through MHC genes, blood group systems and microsatellites. Human Inmunology 2001; 62: 979-991

Carrington M, Nelson GW, Martin MP, Kissner T, Vlahov D, Goedert JJ, Kaslow R, Buchbinder S, Hoots K, O'Brien SJ: HLA and HIV-1:Heterozygote advantage and B*35-Cw*04 disadvantage. Science1999, 283:1748–1752

Cavalli Sfroza Demic expansion and human evolution. *Science*:1993: 259: 639-646.

Cavalli Sfroza,L.L. and Piazza,A. Analysis of evolution: evolutionary rates, independence and treeness. *Theor. Popul. Biol*: 1994a: 8: 127-165.

Cavalli Sfroza,L.L. The DNA revolution in Population genetics. *Trends Genet*: 1998: 14: 60-65.

Cavalli Sfroza,L.L., Menozzi,P., Piazza,A. The history and geography of human genes. Princeton University Press, Princeton. 1994b

Cavalli-Sforza, L. L Demic Diffusion as the Basic Process of Human Expansions" Published by McDonald Institute for Archaeological Research, CambridgeExaming the Farming/Language Dispersal Hypothesis, 2003, chapter 7, pages 79-88.

Cavalli-Sforza, L. L. Demic Diffusion as the Basic Process of Human Expansions" Published by McDonald Institute for Archaeological Research, Cambridge Examining the Farming. Language Dispersal Hypothesis, eds. Bellwood, P. & Renfrew, C. (McDonald Institute Monographs, Cambridge, U.K.).2004

Cavalli-Sforza, L. L., Feldman, M. W. The Application of Molecular Genetic Approaches to the Study of Human Evolution" Nature Genetics Supplement, 2003, volume 33, pages 266-275

Ceman S, Sant AJ. The function of invariant chain in class II-restricted antigen presentation. *Semin Immunol.* 1995 Dec;7(6):373–387

Cerna M, Falco M, Friedman H, Raimondi E, Maccagno A, Fernandez-Vina M, Stastny_ Differences in HLA class II alleles of isolated South American Indian populations from Brazil and Argentina. Human Immunology 1993: 37: 213

Cerna M, Fernandez-Vina M, Ivaskova E,Stastny P Comparison of HLA class II alleles in Gypsy and Czech populations by DNA typing with oligonucleotides probes Tissue Antigens1992, vol. 39, pp. 111-116

Chakraborty,R., Kamboh,M.I., Nwankwo,M., Ferrell,R.E. Caucasian genes in American blacks: new data. *Am. J. Hum. Genet.* : 1992 50: 145-155.

Chandanayingyong D, Stephens HA, Klaythong R, Sirikong M, Udee S, Longta P, et al. HLA-A, -B, -DRB1, -DQA1, and -DQB1 polymorphism in Thais. HumImmunol.; 53: 1997; 174-182.

Chang CH, Flavell RA. Class II transactivator regulates the expression of multiple genes involved in antigen presentation. *J Exp Med.* 1995 Feb 1; 181(2): 765–767.

Charron. D. Anthropology : Contribution of HLA. Serves, France: Medical and Scientific international publisher. 344,1997

Chen, Y.S., Torroni, A., Excoffier, L., Santachiara-Benerecetti, A.S.,and Wallace, D.C.. Analysis of mtDNA variation in African populations reveals the most ancient of all human continent-specific haplogroups. *Am. J. Hum. Genet.*1995; 57: 133–149.

Cheroutre_H, Kronenberg_M, K Brorson, Hunt SW 3d, P Eghtesady, L Hood and DA Nickerson Analysis of MHC class I gene expression in adult bone marrow and fetal

liver of the BALB/c mouse The Journal of Immunology, Vol 146, Issue 10 3263-3272

Chhaya S.U. HLA-B27 polymorphism in Mumbai, Western India Tissue Antigens Volume 66 Issue 1 Page 48 - July 2005

Christensen, L. S., S. Schr, M. H. Schierup, B. F. Vestergaard and C. H. Mordhorst Sequence analysis of measles virus strains collected during the pre- and early-vaccination era in denmark reveals a considerable diversity of ancient strains. *Apmis* 2002; 110 (2): 113-122

Cizman BB, McKean ME, Kearns DJ, Heron SD, Wu J, Griffith BB,Argyros EG, Kamoun M, Zmijewski CM, Williams TM, Monos DS: DRB1*1120 allele; another example of the transition between the DR11 and DR13 families of alleles. Tissue Antigens 1996, 48:52–54

Comas, D., Calafell, F., Mateu, E., Perez-Lezaun, A., Bosch, E.,Martinez-Arias, R., Clarimon, J., Facchini, F., Fiori, G., Luiselli, D., et al.. Trading genes along the silk road: mtDNA sequences and the origin of Central Asian populations. *Am. J.Hum. Genet.* 1998: 63: 1824–1838.

Correa P.A, Whitworth W.C HLA- DR and DQB1 gene polymorphism in the North-Western Colombian populations. Tissue Antigens 2002,59:436-439

Crespí B, Milà J, Martínez-Pomar N, Etxagibel A, Muñoz-Saa, Priego, Luque A, Pons J, Picornell A, Ramon, Castro J.A, Matamoros N Tissue Antigens 2002, 60: 282-291 Spain Majorca

Cresswell P. Assembly, transport, and function of MHC class II molecules. *Annu Rev Immunol.* 1994;12:259–293.

Dalton ET.The descriptive ethnology of Bengal.Calcutta: Govt. Printing, 1872

Dani AH *Prehistory and Protohistory of Eastern India.* Calcutta: Firma KL Mukhopadhyay, 1960

Das BM, Walter H, Gilbert K, Lindenberg P, MalhotraKC, Mukherjee BN, Deka R, Chakraborty R 1987.

Das, K., K.C. Malhotra, B.N. Mukerjee et al.. Population structure and genetic differentiation among 16 tribal populations of Central India. Hum. Biology 1996

De Knijff L. An extensive analysis of Y-chromosomal microsatellite haplotypes in globally dispersed human populations. Am J Hum Genet, 2001: 68:990–1018.

Deka, R., Jin, L., Shriver, M.D., Yu, L.M., Saha, N., Barrantes, R.,Chakraborty, R., and Ferrell, R.E.. Dispersion of human Y chromosome haplotypes based on five microsatellites in global populations. *Genome Res.* 1996: 6: 1177–1184.

Destro-bisol,G., Maviglia,R., Caglia,A., Boschi,H., Spedini,G., Pascali,V., Clark,A., Tishkoff,S. Estimating European admixture in African-Americans by using microsatellites and a microsatellite haplotype (CD4/Alu). *Hum. Genet.*1999:104: 149-157.

Diamond JM Human genetics: Jewish lysosomes. Nature. 1994: 368:291–292

Diogo Meyer,Richard M. Single,Steven J. Mack,Henry A. Erlich and Glenys Thomson Signatures of Demographic History and Natural Selection in the Human Major Histocompatibility Complex Loci 2006 Genetics 173: 2121–2142

Djoulah S, Sanchez-Mazas A, Khalil I.,.Benhamamouch S. HLA DRB1, DQA1 and DQB1 DNA polymorphism in healthy Algerian and genetic relationships with other populations. Tissue Antigens 1994:43:102-109

Doherty DG,Vaughan RW,Donaldson PT,Mowat AP HLA DQA, DQB, and DRB genotyping by oligonucleotide analysis: distribution of alleles and haplotypes in British caucasoids. Hum Immunol. 1992 ; 34(1):53-63

Donadi E.A. et al. Proceedings of IHW Workshop, Seattle 2002.

Edidin, M. Function by Association? MHC antigens and membrane receptor complexes. *Immunol. Today*: 1988: 9: 218-219.

Edmonds, C., Lillie, A., Cavalli-Sforza, L. Mutations arising in the wave front of an expanding population" Proceedings of the National Academy of Science USA, 2004, volume 101, number 4, pages 975-979

Estoup,A., Solignac,M., Cornuet,J.M., Goudet,J., Scholl,A. Genetic differentiation of continental and island populations of *Bombus terrestris* in Europe. *Mol. Ecol.*1996: 5: 19-31.

Eswaran, V. *Curr. Anthropol.*2002: 43, 749–774. 32.

Evseeva I et al. HLA profile of three ethnic groups living in the North-Western region of Russia. Tissue Antigens 2002:59:38-43

Falcorner,D.S. & Mackay T., Introduction to Quantitative Genetics, Longman Scientific & Technical, 1996

Farjadian_S, Moqadam_F. A. &. Ghaderi A. HLA class II gene polymorphism in Parsees and Zoroastrians of Iran. International Journal of Immunogenetics. 2006: 33: 3, 185-191

Farrer LA. Intercontinental epidemiology of Alzheimer disease: global approach to gene hunting. JAMA 2001; 285:796-798.

Ferencik S, F.Gong, H.Grosse-Wilde HLA. 1998, Ed: P.I.Terasaki & DW Gjertson p:202-203

Ferencik S, H.Grosse-Wilde HLA 1997, p222-223 Editor: P.I.Teraski & D.W.Gjertson German Essen

Ferrara G.B, L. Delfino, A. Longo, A. Morabito, C.Pera in HLA 1998, Gjertson DW and Terasaki PI Eds.,published by ASHI Genetic variation of five blood group polymorphisms in ten population groups of Assam, India. *Int JAnthrop,* 2: 325-340.

Fienberg SE. The analysis of multidimensional contingency tables. Ecology 1970; 51: 419-433

Fiers W, Contreras R, De Wachter R, Haegeman G, Merregaert J IHGSH "Finishing the euchromatic sequence of the human genome." 2004. Nature 431: 931-945.

Fischer A., Wiebe V., Pääbo S., and Przeworski M. Evidence for a complex demographic history of chimpanzees. (2004) Molecular Biology and Evolution 21 (5): 799-808

Fischer C, Jock B, Vogel F.Interplay between humans and infective agents: a population genetic study. Hum Genet. 1998 Apr: 102(4):415-22.

Flajnik M F, C Canel, J Kramer, and M Kasahara Evolution of the major histocompatibility complex: molecular cloning of major histocompatibility complex class I from the amphibian Xenopus. Proc Natl Acad Sci U S A. 1991 January 15; 88(2): 537–541

Fort M, De Stefano GF, Cambon-Thomsen A, Giraldo-Alvarez P, Dugoujon JM, Ohayon E, Scano G, Abbal M HLA class II allele and haplotype frequencies in Ethiopian Amhara and Oromo populations. Tissue Antigens 1998 : 51,327,

Foster, M. W., & Sharp, R. R. Race, ethnicity, and genomics: Social classifications as proxies of biological heterogeneity. Genome

Francis L. Black and Philip W. Hedrick Random Mating and Selection in Families Against Homozygotes for HLA in South Amerindians 1997: 127 Issue 1-2 Page 51

Fu Y, Liu Z, Lin J, Jia Z, Chen W, Pan D, et al. HLA-DRB1,DQB1 and DPB1 polymorphism in the Naxi ethnic group of South-western China. Tissue Antigens 2003; 61: 179-183.

Gao X, Y.Sun, J.An, et al. DNA typing for HLA-DR, and DP alleles in a Chinese population using the polymerase chain reaction (PCR) and oligonucleotide probes. Tissue Antigens 1991:38:24-30

Gao X,HLA 1991-11th International Histocompatibility Conference,1997,pg 237

Glazko, G., V. Veeramachaneni, M. Nei, and W. Makalowski Eighty percent of proteins are different between humans and chimpanzees. *Gene,* 2005: 346 ,215-219.

Gomez-Casado E, P del Moral, J Martinez-Laso, A Garcia-Gomez, L Allende, C Silvera-Redondo, J Longas, M Gonzalez-Hevilla, M Kandil, J Moscoso, J Zamora, A Arnaiz-Villena HLA alleles in Moroccans. Tissue Antigens, 2000: 55: 239-249.

Gonem A. 1996. The encyclopedia of the peoples of the world.Jerusalem: Jerusalem Publishing House.

Gottelli,D., Sillero-Zubiri,C., Applebaum,G.D., Roy,M.S., Girman,D.J., Garcia-Moreno,J., Ostranders,E.A. and Wayne,R.K. Molecular genetics of the most endangered Canid: the Ethiopian wolf *Canis simensis*. *Mol. Eco,l*(1994):. 3: 301-312.

Grubic Z, Zunec R, A.Naipal Molecular analysis of HLA ClassII polymorphism in Croatians. Tissue Antigens 1995:46:293-29

Gruen,J.R. and Weissman,S.M. Evolving views of the major histocompatibility complex. *Blood*,1997:: 90: 4252-4265.

Guo,S.W. and Thompson,E.A. Performing the exact test of Hardy-Weinberg proportion for multiple alleles. *Biometrics*.1992: 45: 53-70.

Haldane,J.B.S An exact test for randomness of mating. *Journal of Genetics*: (1954): 52: 631-635.

Harding RM, Healy E, Ray AJ, Ellis NS, Flanagan N, Todd C, Dixon C, Sajantila A, Jackson IJ, Birch-Machin MA, Rees JL Evidence for variable selective pressures at MC1R. Am J Hum Genet,(2000) : 66:1351–1361

Harpending H, Rogers A Genetic perspectives on human origins and differentiation. Annu Rev Genomics Hum Genet ,2000 :1:361–385

Harpending,H. and Ward,R.H. Chemical systematics and human populations. In: Nitechi, M. H., (ed.) In Biochemical Aspects of Evolutionary Biology. University of Chicago Press, Chicago. 1982: 213-256.

Harpending,H.C., Batzer,M.A., Gurven,M., Jorde,L.B., Rogers,A.R., Sherry,S.T. (1998): Genetic traces of ancient demography. *Proc. Natl. Acad. Sci. U.S.A.*: 17: 1961-1967.

Hedrick PW and Loeschcke V, MHC and mate selection in humans? *Trends Ecol. Evol* : 1996 11: 24.

Hedrick PW, Evolutionary genetics of the major histocompatibility complex. *Am. Nat*,1994:.143:945–964.

Hedrick PW, Neutrality or selection at HLA? Am. *J. Hum. Genet.* (1983): 35: 1055–1057.

Herrnstein, R. J., & Murray, C. (1994). The Bell Curve: Intelligence and Class Structure in American Life. New York: Free Press.

Hickson Robert, Rebecca L. Cann Mhc Allelic Diversity and Modern Human Origins Journal of Molecular Evolution Volume 45, Number 6 / December, 1997

Hill A.V, Allsopp CE, Kwiatkowski D, Anstey NM, Twumasi P, Rowe PA, Bennett S, Brewster D, McMichael AJ, Greenwood BM. Common west African HLA antigens are associated with protection from severe malaria. Nature. 1991 Aug 15;352(6336):595-600

Hughes AL, Nei M. Pattern of nucleotide substitution at major histocompatibility complex class I loci reveals overdominant selection. Nature,1988. 335: 167-170. MEDLINE

Hughes, A. L, Hughes M. K, Howell C. Y, and Nei M Natural selection at the class II major histocompatibility complex loci of mammals. *Philos. Trans. R. Soc. Lond. B.* (1994) 346:359-367.

Hughes, A. L. and Nei M Evolution of the major histocompatibility complex: independent origin of nonclassical class I genes in different groups of mammals. Mol. Biol. Evol. 1989: 6:559-579.

Hughes, A. L. and Nei M Evolutionary relationships of the classes of major histocompatibility complex genes. Immunogenetics, 1993 37:337-346.

Hughes, A. L. and Nei M Maintenance of MHC polymorphism. Nature 1992 :355: 402-403.

Hughes,A.L. and Nei,M. Maintenance of MHC polymorphism. *Nature*: 1992: 355: 402-403.

Inman,K. and Rudin,N. (1997): In An Introduction to Forensic DNA analysis, CRC Press.

Itoh, T, Martin W, and Nei M Acceleration of genomic evolution caused by enhanced mutation rate in endocellular symbionts. *Proc. Natl. Acad. Sci. USA,* 2002: 99:12944-12948.

Ivanova M, Rozemuller E, Tyfekchiev N, Michailova A, Tilanus M, E.Naumova HLA polymorphism in Bulgarians defined by high-resolution typing methods in comparison with other populations.Tissue Antigens. 2002, 60,496

Jacob Pandian The Hindu Caste System and Muslim Ethnicity: The Labbai of a Tamil Village in South India ***Ethnohistory,*** Vol. 25, No. 2 (Spring, 1978), pp. 141-157

Jaini R, Kaur G, Mehra NK. Heterogeneity of HLA DRB1*04 and its associated haplotypes in the North Indian population. *Human Immonology.*2002: 63: 24-29.

Janeway CA, Travers P: Antigen recognition by T lymphocytes. Immunobiology, ed 3. New York and London, Garland Publishing, 1997, pp 41–46

Jensen,A.F. (1991): In India: Its culture and people. White plains, New York: Longmann.

Jin, L., Underhill, P. A., Doctor, V., Davis, R. W., Shen, P. D., Cavalli-Sforza,L. L. & Oefner, P. J. *Proc. Natl. Acad. Sci. USA*, 1999 93, 3796–3800.

Joaquim Fort, Toni Pujol & Luigi Luca Cavalli-Sforza Palaeolithic Populations and Waves of Advance" Cambridge Archaeological Journal, 2004, volume 15, pages 53-61

Jobling M.A., Hurles M.E., and Tyler-Smith C. (2004)Human Evolutionary Genetics: Origins, Peoples and Disease. Garland Publishing, New York. Chapter 5: How do we interpret genetic variation? Chapter 7: Humans as apes

Johnson-Donald, J., and Johnson,J.E. (1992): In Through Indian Eyes: the living Tradition, rev. eds. New York: Centre for International Training and Education

Jorde L.B., et al. "Immunogenetics." In *Medical Genetics.* 2nd ed. St. Louis: Moseby, 1999.

Jorde LB, Bamshad M. Questioning evidence for recombination in human mitochondrial DNA. Science. 2000 Jun 16;288

Jungerman M. Proceedings of 12th International Histocompatibility Workshop, Volume 2: 194 Poland Wielkopolska

Kapadia, K. 1995. Siva and her sisters: Gender, caste and class in rural South India. San Francisco: Oxford University Press .

Kapustin S, Lyshchov A, Alexandrova J, Imyanitov E., Blinov M. HLA class II molecular polymorphisms in healthy Slavic individuals from North-Western Russia Tissue Antigens 1999:54 :5: 517

Karve, I. 1961. *Hindu Society - An Interpretation.* Deshmukh Prakashan, Poona

Karve, Irawati. 1953. Kinship organization in India. Deccan College Monograph Series no. 11. Poona.

Kernan NA, Bartsch G, Ash RC, Beatty PG, Champlin R, Filipovich A, Gajewski J, Hansen JA, Henslee-Downey J, McCullough J: Analysis of 462 transplantations from unrelated donors facilitated by the National Marrow Donor Program. N Engl J Med 1993, 328:593–602

Kevles, D. 1995. In the name of eugenics: Genetics and the uses of human heredity. Harvard University Press, Cambridge, MA.

Kidd Kenneth K., Oota Hiroki, Pakstis Andrew J, Bonne-Tamir Batsheva , David Goldman Elena Grigorenko, Sylvester L. B. Kajuna, Nganyirwa J. Karoma, Selemani Kungulilo, Ru-Band Lu, Kunle Odunsi Friday Okonofua,Olga V. Zhukova,

Judith R. Kidd. The evolution and population genetics of the *ALDH2* locus: random genetic drift, selection, and low levels of recombination Annals of Human Genetics2004: 68 2 93 - March

Kidd KK, Pakstis AJ, Speed WC, Kidd JR.Understanding human DNA sequence variation. J Hered. 2004 Sep-Oct; 95(5): 406-20.

Kivisild, T, Rootsi, S, Metspalu, M, Mastana, S, Kaldma, K, Parik, J, Metspalu, E, Adojaan, M, Tolk, H.-V, Stepanov, V, Golge, M., Usanga, E, Papiha, S. S., Cinnioglu, C., King, R., Cavalli-Sforza, L., Underhill, P. A., Villems, R. The Genetic Heritage of Earliest Settlers Persists Both in Indian Tribal and Caste Populations" American Journal of Human Genetics, 2003, volume 72, pages 313-332

Kivisild,T, Bamshad,M.J, Kaldma,K, Metspalu,M, Metspalu,E, Reidla,M, Laos,S, Parik,J, Watkins,W.S, Dixon,M.E, Papiha,S.S, Mastana,S.S, Mir,M.R, Ferak,V, Villems,R. Deep common ancestry of Indian and western-Eurasian mtDNA lineages. *Curr. Biol.* 1999: 9: 1331-1334.

Klein J. (1990). ImmunologyBlackwell Scientific Publications, Boston, MA.

Klein RG (1999) The human career: human biological and cultural origins, 2nd ed. University of Chicago Press, Chicago

Klein, J. et al. Frozen haplotypes in MHC evolution. in Molecular Evolution of the Major Histocompatibility Complex (eds Klein, J. & Klein, D.) 261–286 (Springer-Verlag, 1991).

Kornguth, Steven; Bersu, Edward; Mack, Kenneth MHC Class I Gene expression Science, Volume 270, Issue 5237, pp. 720-721

Krings M, Stone A, Schmitz R, Krainitzki, H, Stoneking, M. Paabo, S, 1997. Neandertal DNA sequences and the origin of modern humans. Cell ,1997: 90, 19–30.

Krings M, Stone A, Schmitz RW, Krainitzki H, Stoneking M, Pääbo S Neandertal DNA sequences and the origin of modern humans. Cell ,1997 90:19–30

Kunz R, Kreutz R, Beige J, Distler A, and Sharma Association Between the Angiotensinogen 235T-Variant and Essential Hypertension in Whites A Systematic Review and Methodological Appraisal Hypertension, December 1, 1997; 30(6): 1331 – 1337

Lahr MM (1996) The evolution of modern human diversity: a study of cranial variation. Cambridge University Press, Cambridge, United Kingdom

Lahr MM, Foley RA Towards a theory of modern human origins: geography, demography, and diversity in recent human evolution. Am J Phys Anthropol Suppl 1998 27:137–176

Landsteiner,K. and Levine,P. (Further observations on individual differences of human blood. *Proc. Soc. Exp. Biol. Med* 1927,: 77: 402-405.

Lawlor David A, Cynthia D. Dickel†, William W. Hauswirth† & Peter Parham Ancient *HLA* genes from 7,500-year-old archaeological remains Nature, 1991: 349, 785 - 788

Lawlor, D. A, Ward F. E, Ennis P. D, Jackson A. P. & Parham, P HLA-A and B polymorphisms predate the divergence of humans and chimpanzees. Nature,1988: 335, 268–271

Lee KW, Oh DH, Lee C, Yang SY. Allelic and haplotypic diversity of HLA-A, -B, -C, -DRB1, and -DQB1 genes in the Korean population. Tissue Antigens 2005; 65: 437-447.

Leonard Lieberman, Rodney C. Kirk, and Alice Littlefield, "Perishing Paradigm: Race—1931-99," American Anthropologist 105, no. 1 (2003): 110-13.

Levene,H. On a matching problem in genetics. *Ann. Math. Stat.*1949: 20: 91-94.

Lewontin, R. C. (1997). The apportionment of human diversity. In E. N. Gates (Ed.), Critical race theory (pp. 7–24). New York: Garland.

Lewontin, R. C. The apportionment of human diversity. Evolutionary Biology, 1972 6, 381–398.

Lieberman, Hampton, Littlefield, and Hallead"Race in Biology and Anthropology: A Study of College Texts and Professors" in Journal of Research in Science Teaching, 1992: 29:301-321

Lindblom B. Proceedings of 11th International Histocompatibility Workshop Denmark

Little,A.M. and Parham,P. Polymorphism and evolution of HLA class I and II genes and molecules. *Reviews in Immunogenetics*: 1999 ,1: 105-123.

Liu Y, Liu Z., Y. F[I], Z. Jia, S. Chen & A. Xu Polymorphism of HLA class II genes in Miao and Yao nationalities of Southwest China. Tissue Antigens 2006: 67: 2: 157

Luo,M., Blanchard,J., Pan,Y. Brunham,K. and Brunham,R.C. High resolution sequence typing of HLA-DQA1 and DQB1 Exon2 DNA with taxonomy based sequence analysis (TSBA)allele assignment. *Tissue Antigens*: 1999: 54: 69-82.

Luzzatto L, Mehta A. Glucose-6-phosphate dehydrogenase deficiency. In: Scriver ChR, Beaudet AL, Sly WS, Valle D, eds. *The Metabolic and Molecular Basis of Inherited Disease.* New York: McGraw-Hill, 1995:pp 3677-3398,

Machulla H.K.G., D. Batnasan, F. Steinborn, F.A. Uyar, G. Saruhan-Direskeneli, F.S. Oguz, M.N. Carin, M.T. Dorak Genetic affinities among Mongol ethnic groups and their relationship to Turks Tissue Antigens 2003: 61:292

Majumdar, P. & Majumder, P. P. 1999 HAPLOPOP: A computer program package to estimate haplotype frequencies from genotype frequencies via the EM algorithm. *AHGU Tech Rep*. Indian Statistical Institute, Kolkata.

Majumder,P.P. (1980): ABO blood group gene frequencies in the Indian subcontinent Statistical study of patterns of variation. Ph.D. thesis. Indian Statistical Institute, Calcutta, India.

Malhotra,K.C. and Vasulu, T.S. (1993): Structure of Human Populations in India. Proc. Vol. Haldane centenary Conference on Human genetics. Majumdar, P. P. ed. Plenum Press, New York.

Manly,B.F.J. (1985): The statistics of natural selection. Chapman and Hall, London

Mark. Stoneking, Mack, S. J., T. L. Bugawan, P. V. Moonsamy, J. A. Erlich, E. A. Trachtenberg, Y. K. Paik, A. B. Begovich, N. Saha, H. P. Beck, , and H. A. Erlich.. Evolution of Pacific/Asian populations inferred from HLA class II allele frequency distributions. Tissue Antigens , 2000:55 383-400

Marsh SGE, Parham P, Barber LD: Evolution and anthropology of HLA. The HLA Facts Book. London, Academic Press, 2000, pp 73–83

Marsh Steven G. E, Bodmer Julia G, Albert Ekkehard D, Bodmer Walter F, Bontrop Ronald E. Dupont Bo, Erlich Henry A, Hansen John A, Mach Bernard, Wolfgang R. Mayr, Peter Parham, Effie W. Petersdorf, Takehiko Sasazuki, Geziena M. Th. Schreuder, Jack L. Strominger, Arne Svejgaard & Paul I. TerasakiNomenclature for factors of the HLA systemEuropean Journal of Immunogenetics , 2000 : 28 Issue 3 Page 377

Mason,P. and, Parham,P HLA class I region sequences. *Tissue Antigens*: . 1998: 51: 417-466.

Maynard V. Olson MOLECULAR EVOLUTION '99 When Less Is More: Gene Loss as an Engine of Evolutionary Change Am. J. Hum. Genet1999: 64:18–23,

McDougall I, Brown FH, Fleagle JG Stratigraphic placement and age of modern humans from Kibish, Ethiopia. Nature, 2005 : 433:733–736

Mehindate K, R al-Daccak, TJ Schall and W Mourad. Induction of chemokine gene expression by major histocompatibility complex class II ligands in human fibroblast-like synoviocytes. Differential regulation by interleukin-4 and dexamethasone. J. Biol. Chem., 1994 Vol. 269, Issue 51, 32063-32069, 12,

Mehra NK, Taneja V, Kailash S et al.. Distribution of HLA antigens in a sample of North Indian Hindu population. *Tissue Antigens*. 1986: 27: 64-74.

Migot-Nabias F, I.Fajardy, P.M.Danze HLA ClassII polymorphism in a Gabonese Banzabi population Tissue Antigens 1999:53:580-585.

Miller GH, Magee JW, Johnson BJ, Fogel ML, Spooner NA, McCulloch MT, Ayliffe LK Pleistocene extinction of Genyornis newtoni: human impact on Australian megafauna. Science, 1999 : 283:205–208

Mittal KK, Naik S, Sansonetti N et al. The HLA antigens in Indian Hindus. *Tissue Antigens.*1982: 20: 223-22

Mizuki N, Ohno S, Ando H, Sato T, Imanishi T, Gojobori T, et al. Major histocompatibility complex class II alleles in an Uygur population in the silk route of Northwest China. Tissue Antigens 1998; 51: 287-292.

Mizuki N, Ohno S, Ando H, Sato T, Imanishi T, Gojobori T,et al. Major histocompatibility complex class II alleles in Kazak and Han populations in the Silk Route of northwestern China. Tissue Antigens; 1997: 50: 527-534.

Mohyuddin A, Ayub Q, Khaliq S, Mansoor A, Mazhar K, Rehman S, et al . HLA polymorphism in six ethnic groups from Pakistan. Tissue Antigens 2002;59:492-501

Monsalve MV et al. Analysis of HLA Class I and Class II in Na-Dene and Amerindian Populations from British Columbia, Canada. Human Immunology 1998: 59: 48-55. Athabaskan

Monsalve,M.V., Booth,P.B., Cavalli-Sforza,L.L. Languages, Geography and HLA haplotypes in native American and Asian Populations. *Proc. R. Soc. Lond. B. Biol. Sci.*: 1999: 266: 2209-2216.

Mukherjee N, Nebel A, Oppenheim A, Majumder PP. High-resolution analysis of Y-chromosomal polymorphisms reveals signatures of population movements from Central Asia and West Asia into India. J Genet. 2001 Dec;80(3):125-35

Munkhbat B, T.Sato, M.Hagigara, K.Sato Molecular analysis of HLA polymorphism in Khoton Mongolians Tissue antigens 1997:50:124-134

Naipaul,V.S. (1992): in India a million mutinies now. New York: Viking Penguin, 1990: rep. 1992.

Nakajima, T, Wooding, S, Sakagami, T, Emi M, Tokunaga, K, Tamiya G, Ishigami T, Umemura S, Munkhbat B, Jin F, Guanjun, J, Hayasaka I, Ishida T, Saito, N, Pavelka, K, Lalouel J.M., Jorde, L. B, Inoue I Natural selection and population history in the human angiotensinogen gene (AGT): 736 complete AGT sequences in worldwide chromosomes. American Journal of Human Genetics ,2004: 74:.898-916

Nei, M. and A. K. Roychoudhury Evolutionary relationships of human populations on a global scale. Mol. Biol. Evol,1993: 10:927-943.

Nei, M. Selectionism and neutralism in molecular evolution. Mol. Biol. Evol. 2005 22:2318-2342.

NeumannH, Schmidt H, Cavalié A, JenneD, and Wekerle HMajor Histocompatibility Complex (MHC) Class I Gene Expression in Single Neurons of the Central Nervous System: Differential Regulation by Interferon (IFN) and Tumor Necrosis Factor (TNF)-α J. Exp. Med. Volume 185 January 1997 305-316

Nobuo Araki,D.Med.Sc. HLA1998 edited by Terasaki & Gjertson

Nordborg M On the probability of Neanderthal ancestry. Am J Hum Genet.1998 :63:1237–1240

Ober, C. and van der Ven, K. (1997) Immunogenetics of reproduction: an. overview. In Olding, L.B. (ed.), Current Topics in Microbiology and. Immunology

Oshima M, Deitiker P, Ashizawa T, Atassi M "Vaccination with a MHC class II peptide attenuates cellular and humoral responses against tAChR and suppresses clinical EAMG.". *Autoimmunity*,2002. 35 (3): 183-90

Paetkau,D. and Strobeck,C. Microsatellite analysis of genetic variation in Black bear populations. *Mol. Ecol.* (1994):.: 3: 489-495.

Palanichamy MG, Sun C, Agrawal S, Bandelt HJ, Kong QP, Khan F, Wang CY, Chaudhuri TK, Palla V, Zhang YP. Phylogeny of mitochondrial DNA macrohaplogroup N in India, based on complete sequencing: implications for the peopling of South Asia. Am J Hum Genet. 2004 Dec;75(6):966-78.

Papassavas E. C. Human Immunology 2000: 61: 615

Parham_P. and Ohta_T. "Population Biology of Antigen Presentation by MHC Class I Molecules.". *Science* 1996 : 272 -273

Partha P. Majumder Indian Caste Origins: Genomic Insights and Future Outlook Genome Research. 2001:Vol. 11, Issue 6, 931-932

Petrone A, T.l Bugowan, C.A Mesturino,L.Nistico The distribution of HLA Class II susceptible /protective haplotypes couls partially explain the low incidence of type 1 diabetes in continental Italy (Lazio region) Tissue antigens:2001:58:385-394

Petrovsky N, Brusic V. Virtual models of the HLA class I antigen processing pathway. Methods. 2004 Dec; 34(4): 429-35.

Pimtanothai N, C.K. Hurley, R. Leke, W. Klitz, A.H. Johnson HLA-DR and -DQ polymorphism in Cameroon Tissue Antigens 2001: 58: 1

Quintana-Murci, L, Semino O, Bandelt H. J, Passarino G, McElreavey K, & Santachiara-Benerecetti A. S. Genetic evidence of an early exit of Homo sapiens from Africa through eastern Africa. Nature Genetics, 1999. 23, 437–441.

Rapson EJ. People and languages. In: EJ Rapson (Ed.): *Cambridge History of India.* Vol.1. *Ancient India.* Delhi: S Chand and Co.1955: pp.33-57.

Rebecca L. Cann Genome Diversity: Applications in Human Population Genetics Am J Hum Genet. 2001; 69(3): 667–668

Reed E, Ho E, Lupu F et al. Polymorphism of HLA in the Romanian population. Tissue Antigens 1992:39:8-13

Reich,D.E. and Lander,E.S. On the allelic spectrum of human disease. *TrendsGenet.* 2001:17, 502–510.

RendineS, Borelli I, Barbanti,N Sacchi,S Roggero,ES Curtoni HLA polymorphisms in Italian Bone Marrow donors: a regional analysis Tissue Antigens 1998:52:135

Reveille JD, Arnett FC, Olsen ML, Moulds JM, Papasteriades CA, Moutsopoulos HM. HLA-class II alleles and C4 null genes in Greeks with systemic lupus erythematosus. Tissue Antigens 1995;46:417–21

Risley HH. 1915. The people of India. Calcutta: Thacker Spink.

Robert J. Sternberg, Elena L. Grigorenko, and Kenneth K. Kidd Intelligence, Race, and Genetics Yale University 2005

Rogers,J. and Comuzzie,A.G., When is ancient polymorphism a potential problem for molecular phylogenetics.*Am. J. Phys. Anthropo*,1995: *l*: 98: 216-218.

Rosenberg NA, Pritchard JK, Weber JL, Cann HM, Kidd KK, Zhivotovsky LA, Feldman MW.Genetic structure of human populations. Science. 2002 Dec 20;298(5602):2381-5.

Rosenberg, N. A, Pritchard, J. K, Weber, J. L, Cann, H. M, Kidd, K. K, Zhivotovsky, L. A, Feldman, M. W. Genetic Structure of Human Populations.Science, 2002, volume 298, pages 2381-2385

Roychoudhury, A. K, and Nei M (1988) Human Polymorphic Genes: World Distribution. Oxford University Press, Oxford and New York.

Ruemmele, Dionne, Levy & Seidman Dexamethasone inhibits IFNγ-induced MHC class II expression of intestinal epithelial cells independently of the TGF-β1 regulatory pathway. Alimentary Pharmacology & Therapeutics,1999; 5 Page 595

Rushton, J. P. (1995). Race, evolution, and man. Princeton, NJ: Princeton University Press.

Rzhetsky, A. and Nei M. Theoretical foundation of the minimum-evolution method of phylogenetic inference. Mol. Biol. Evol. 1993: 10:1073-1095.

Saito S, Ota S, Yamada E, Inoko H, Ota M. Allele frequencies and haplotypic associations defined by allelic DNA typing at HLA class I and class II loci in the Japanese population. Tissue Antigens 2000; 56:522-529.

Sanchez-Velasco P, E Gomez-Casado, J Martinez-Laso, J Moscoso, J Zamora, E Lowy, C Silvera, A Cemborain, F Leyva-Cobian, A Arnaiz-Villena HLA alleles in Basques and neighbouring populations. Tissue Antigens 2003:61: 384-392

Sanchez-Velasco P,.Escribano J de Diego, *J.E. Paz-Miguel, G. Ocejo-Vinyals, F. Leyva-Cobián*HLA-DR, DQ nucleotide sequence polymorphisms in the Pasiegos (Pas valleys, Northern Spain) and comparison of the allelic and haplotypic frequencies with those of other European populations Tissue Antigens: 1999:53 :1 - 65

Sankar P, Cho MK Toward a new vocabulary of human genetic variation. Science 2002: 298:1337–1338

Sartoris_S, MT Scupoli, L Scarpellino, F Paiola, M Jotterand-Bellomo, G Tridente and RS Accolla Inducible and constitutive MHC class II gene expression. Distinct tissue-specific genetic controls The Journal of Immunology, Vol 145, Issue 6 1960-1967

Saruhan-Direskeneli G, Uyar FA, Bakar S, Eraksoy M. Molecular analysis of HLA-DRB1, DQA1 and DQB1 polymorphism in Turkey.Tissue Antigens 2000: 55:2: 171

Saruhan-Direskeneli,G, Uyar,F.A, Bakar,S, Eraksoy M. Molecular Analysis of HLA-DRB1, DQA1 and DQB1 polymorphism in Turkey. *Tissue Antigens*: 2000: 55: 171-174.

Satta Y, Takahata N Out of Africa with regional interbreeding? Modern human origins. Bioessays. 2002: 24:871–875

SchaferP.H, PierceS.K, JardetzkyT.SThe structure of MHC class II: a role for dimer of dimers. *Seminars Immunol*: 1995: 7: 389-398.

Schanfield M, Gabriel MN, Andelinovic S, Reynolds RL, Ladd C, Lee HC, Primorac D. Allele frequencies for the 13 CODIS STR loci in a sample of southern Croatians. J Forensic Sci. 2002 May; 47(3): 669-70.

Schierup, M. H, Charlesworth D and Vekemans X The effect of hitch-hiking on genes linked to a balanced polymorphism in a subdivided population. *Genetical research*, 2000. 76: 63-73.

Schierup, M. H. The number of self-incompatibility alleles in a finite, subdivided population. *Genetics*, 1998. 149: 1153-1162.

Schreiber,A.B, Schlessinger,J, Edidin,M. Interaction between major histocompatibility complex antigens and epidermal growth factor receptors on human cells. *J. of Cell Biology,*1984:: 98: 725-731.

Schwartzberg,Joseph E.1968 "Caste Regions of North Indian Plains".In Milton Singer and Bernards. Cohn (Eds)'sStructure and Change in Indian Society.New York, Wenner-Gren Foundation forAnthropological Research Incorporated

Serre D, Langaney A, Chech M, Teschler-Nicola M, Paunovic M, Mennecier P, Hofreiter M, Possnert G G, Pääbo S No evidence of Neandertal mtDNA contribution to early modern humans. PLoS Biol. 2004: 2:313–317

Shankarkumar U, Pednakar SV, Gupte S, Ghosh K, Mohanty D HLA antigen distribution in Marathi speaking Hindu population from Mumbai, Maharastra, India. *J Hum Eco.* 1999:10(5-6): 367-372.

Shaw C.K, L.L.Chen, A.Lee, T.D.Lee Distribution of HLA gene and haplotype frequencies in Taiwan: acomparative study among Min-Man, Hakka,Aborigenes and Mainland Chinese. Tissue Antigens 1999:53:51-64

Shi L, Xu S. B, J. Ohashi, H. Sun, J. K. Yu, X. Q. Huang, Y. F. Tao, L. Yu, S. Hor, J. Y. Chu & K. Tokunaga HLA-A, HLA-B, and HLA-DRB1 alleles and haplotypes in Naxi and Han populations in Southwestern China (Yunnan province) Tissue Antigens2006: 67 1-38

Snell,G.D. Studies in histocompatibility. *Science,*1981: 10: 172-8.

Sonntag Kai-C Regulated Expression of an MHC Class II Gene from a Promoter-Inducible Retrovirus. Human Gene Therapy. Sep 2000, Vol. 11, No. 14 : 1961 -1969

Spinola H, Bettencourt B, Bruges-Armas J, Brehm A, Middleton D. HLA class I and II polymorphisms in Azores show different settlements in Oriental and Central islands Tissue Antigens 2005: 66: 217

Sternberg, R. J. (Ed.). (2004). International handbook of intelligence.New York: Cambridge University Press

Strachan, T. & Read A.P., Human Molecular Genetics, pp.597, John Wiley & Sons, 1996.

Stringer C (Modern human origins: progress and prospects. Philos Trans R Soc Lond B Biol Sci, 2002 :357:563–579

Su B, Xiao J, Underhill P, Deka R, Zhang W, Akey J, Huang W,Shen D, Lu D, Luo J, Chu J, Tan J, Shen P, Davis R, Cavalli-Sforza L, Chakraborty R, Xiong M, Du R, Oefner P, Chen Z,Jin L.. Y Chromosome evidence for a northward migration of modern humans in East Asia during the last ice age.Am J Hum Genet. 1999:65:1718–1724.

Svejgaard,A. and Ryder,L.P. Interaction of HLA molecules with non-immunological ligands as an explanation of HLA and disease association. *Lancet ii,*1976:: 547-549.

Swisher CC 3rd, Curtis GH, Jacob T, Getty AG, Suprijo A, Widiasmoro Age of the earliest known hominids in Java, Indonesia. Science, 1994: 263:1118–1121

Takahata N, Lee S, Satta Y Testing multiregionality of modern human origins. Mol Biol Evol. 2001:18:172–183

Tateno Y, Takezaki N, and Nei M Relative efficiencies of the maximum-likelihood, neighbor-joining, and maximum-parsimony methods when substitution rate varies with site. Mol. Biol. Evol, 1994. 11:261-277.

Taylor,A.C., Sherwin,W.B. and Wayne, R.K. Genetic variation of microsatellite loci in a bottlenecked species: the northern hairy-nosed wombat *Lasiorhinus krefftii. Mol. Ecol.* 1994: 3: 277-290.

Tethers Cox, Parham Peter, Mann Dean L, and Strominger Jack Structure of HLA Antigens: Amino-Acid and Carbohydrate Compositions and NH_2-Terminal Sequences of Four Antigen Preparations PNAS, 1976:73 : 3: 910-914

Thio CL, Thomas DL, Karacki P, Gao X, Marti D, Kaslow RA, Goedert JJ, Hilgartner M, Strathdee SA, Duggal P, O'Brien SJ, Astemborski J, Carrington M: Comprehensive analysis of class I and class II HLA antigens and chronic hepatitis B virus infection. J Virol. 2003 77: 12083-7,.

Thursz MR, Kwiatkowski D, Allsopp CEM, Greenwood BM, Thomas HC, Hill AVS. Association between an MHC class II allele and clearance of hepatitis B virus in the Gambia. *N Engl J Med* 1995; 332: 1065–9.

Thursz, M. R, Thomas H. C, Greenwood B. M, and Hill A. V. S. Heterozygote advantage for HLA class-II type in hepatitis B virus infection. Nature Genetics 1997. 17:11-12

Tishkoff SA, Kidd KK. Implications of biogeography of human populations for 'race' and medicine. Nat Genet. 2004 Nov;36(11 Suppl):S21-7. Review.

Tishkoff SA, Pakstis AJ, Stoneking M, Kidd JR, Destro-Bisol G, Sanjantila A, Lu R-b, Deinard AS, Sirugo G, Jenkins T, Kidd KK, Clark AG Short tandem-repeat polymorphism/Alu haplotype variation at the PLAT locus: implications for modern human origins. Am J Hum Genet, 2000 : 67:901–925

Tishkoff SA, Verrelli B Patterns of human genetic diversity: implications for human evolutionary history and disease. Annu Rev Genomics Hum Genet,2003 4:293–340

Tishkoff SA, Williams SM Genetic analysis of African populations: human evolution and complex disease. Nat Rev Genet,2002 3:611–621

Tishkoff, S. A., R. Varkonyi, N. Cahinhinan, S. Abbes, G. Argyrotween Haplotype diversity and linkage disequilibrium. at human G6PD: recent origin of alleles that confer malarial resistance. Science 293: 455–462.

Trachtenberg EA, Keyeux G, Bernal J, Noble JA, Erlich HA. Results of Expedicion Humana. II. Analysis of HLA class II alleles in three African American populations from Colombia using the PCR/SSOP: identification of a novel DQB1*02 (*0203) alleleTissue Antigens. 1996 Sep; 48(3): 192-8.

Trachtenberg, E.A, and H.A. Erlich. "DNA–based HLA typing for cord blood stem cell transplantation." *Journal of Hematotherapy* 5 (1996): 295–300.

Trinkaus E, Milota S, Rodrigo R, Mircea G, Moldovan O An early modern human from the Pestera cu Oase, Romania. Proc Natl Acad Sci USA ,2003:100:11231–11236

Trowsdale,J. In Both Man and Beast: a comparative organisation of MHC genes (Review). *Immunogenetics*,1995:: 41: 1-17.

Tsuneto LT, Probst CM, Hutz MH, Salzano FM, Rodriguez-Delfin LA, Zago MA, Hill K, Hurtado AM, Ribeiro-dos-Santos AKC, Petzl-Erler ML HLA class II diversity in seven Amerindian populations.Clues about the origins of the Ache tribe Tissue Antigens 2003:62, 512-526

Underhill P. A, Passarino G, Lin A. A, Shen P, Mirazo´n Lahr, M, Foley R. A, Oefner, P. J. & Cavalli-Sforza, L. L. (2001) *Ann. Hum. Genet.* 65, 43–62.

Underhill, Peter A. Inferring human history: Clues from Y-chromosome haplotypes" Cold Spring Harbor Symp. Quant. Biol.: The Genome of Homo sapiens, 2004, LXVIII, pages 487-494

Underhill,P.A, Shen,P, Wang F, Franco,C, YangW.H, Roxas,A, Sung,R, Lin,A.A, Hyman,R.W, Vollrath,D, Davis,R.W, Cavalli-Sfroza,L.L, Oefner,P.J. Population genetic implications from sequence variation in four Y chromosome genes. *Proc. Natl. Acad. Sci.USA*.:2000: 97: 7354-7359.

Undevia JV,Balakrishnan V, Kirk RL, Blake NM, Saha N, McDermid EM A population genetic study of the Vania Soni in Western India. Hum Hered. 1978; 28(2):104-21.

Vidyarthi, L.P. 1977 Rise of Anthropology in India: A Social Sciences Orientation (in two volumes) Delhi, Concept Publishing Company.

Vikrant Kumar, Debashis Basu,and B. Mohan Reddy Genetic heterogeneity in Northeastern India: Reflection of tribe–caste continuum in the genetic structure American Journal of Human Biology. 2004:16:334–345

Vinay Bahl Terminology, History and Debate: "Caste" Formation or "Class" Formation Journal of Historical Sociology Volume 17 Issue 2-3 Page 265 - June 2004

Waddel LA. 1975. The tribes of the Brahmaputra Valley.A contribution on their physical type and affinities.N. Delhi: Sanskaran Prakashak.

Wall JD Detecting ancient admixture in humans using sequence polymorphism data. Genetics. 2000: 154:1271–1279

Weinshilboum R. Inheritance and drug response. *New England Journal of Medicine.* 2003;348(6):529–537.

Westermarck E "A Short History of Marriage", Published by Macmillan and Co., London. 1926. (Book No: 7A/139)

White TD, Asfaw B, DeGusta D, Gilbert H, Richards GD, Suwa G, Howell FC Pleistocene Homo sapiens from Middle Awash, Ethiopia. Nature ,2003 : 423:742–747

Williams TM, Wu J, Foutz T, McAuley J, Troup GM: A new DRB1 allele (*DRB1*0811*) identified in Native Americans. Immunogenetics 1994, 40:314

Willman CL, Stewart CC, V Miller, Yi TL and Tomasi TB Regulation of MHC class II gene expression in macrophages by hematopoietic colony-stimulating factors (CSF). Induction by granulocyte/macrophage CSF and inhibition by CSF-1Journal of Experimental Medicine, Vol 170, 1559-1567

Wolpoff M, Hawks J, Frayer DW, Hunley K Modern human ancestry at the peripheries: a test of the replacement theory. Science, 2001 : 291:293–297

Wooding S, Ostler C, Prasad BV, Watkins WS, Sung S, Bamshad M, Jorde LB. Directional migration in the Hindu castes: inferences from mitochondrial, autosomal and Y-chromosomal data. Hum Genet. 2004 Aug;115(3):221-9.

Yu N, Chen FC, Ota S, Jorde LB, Pamilo P, Patthy L, Ramsay M, Jenkins T, Shyue SK, Li WH Larger genetic differences within Africans than between Africans and Eurasians. Genetics 2002 161:269–274

Zhong G, Romagnoli P, Germain RN. Related leucine-based cytoplasmic targeting signals in invariant chain and major histocompatibility complex class II molecules control endocytic presentation of distinct determinants in a single protein. *J Exp Med.* 1997 Feb 3;185(3):429–438

Zhou L, Lin B, Xie Y, Liu Z, Yan W, Xu A. Polymorphism of human leukocyte antigen-DRB1, -DQB1, and -DPB1 genes of Shandong Han population in China. Tissue Antigens 2005; 66: 37-43.

Zietkiewicz,E, Yotova,V, Jarnik,M, Korab-Laskowska,M, Kidd,K.K, Modiano,D, Scozzari,R, Tishkoff S, Batzer M, Labuda,D. Genetic Structure of The Ancestral Population of Modern Humans. *J. Mol.Evol*: 1998, 47: 146-155.

Appendices

Appendix A

Department of Medical Genetics

Sanjay Gandhi Post Graduate Institute of Medical Sciences

Distribution of HLA Antigens

S. No. Job. No. Study No. Date of collection Filled by

Name ---------- **Age** ---------- **Sex** ---------- **Occupation**

	Last	First	Middle	Age / Sex	Education/ Occupation
Index Case					
Spouse					

Permanent Address **Present residential Address** **Office Address**

Religion (H/M/S/C)	**Caste** (M/B/V/O)	**Sub caste – surname** (B/C/H)	**Gotra / Kuldevi etc.**	**Consanguinity in parents**	**Distt./ State of origin**	**Mother tongue**
Father						
Mother (Maiden)						
P. Grand Father						
P. Grand Mother (Maiden)						
M. Grand Father						
M. Grand Mother (Maiden)						

Medical History of index case

1. Are you suffering now or have suffered in the past from any major medical problem. Yes/ No.
2. If yes, specify symptoms with duration. From / to
   (a). ----------
   (b). ----------

3. Diagnosis, if known ---------- source of Diagnosis ----------

Collagen / Joint disease	Ankylosing/ Spondylitis/ Rheumatoid arthritis/ Betchet's disease Reiter's disease/ Acute anterior Uveitis/ Optic Neuritis/ Sicca Syndrome/ Systemic Lupus erythromatosus
Endocrine disease	Grave's disease Thyroditis/ Addisons disease/ Congenital adrenal Heperplasia
Skin disease	Psoriasis vulgaris/ Dermatitis herpeteformis
Other	Multiple Sclerosis, C2 deficiency, Pernicious anaemia, Myasthenia gravis

4. Do you suffer from
   Diabetes
   Hypertension
   CAD
   Stroke
   Cancer

5. Have you suffered from anyone of the following

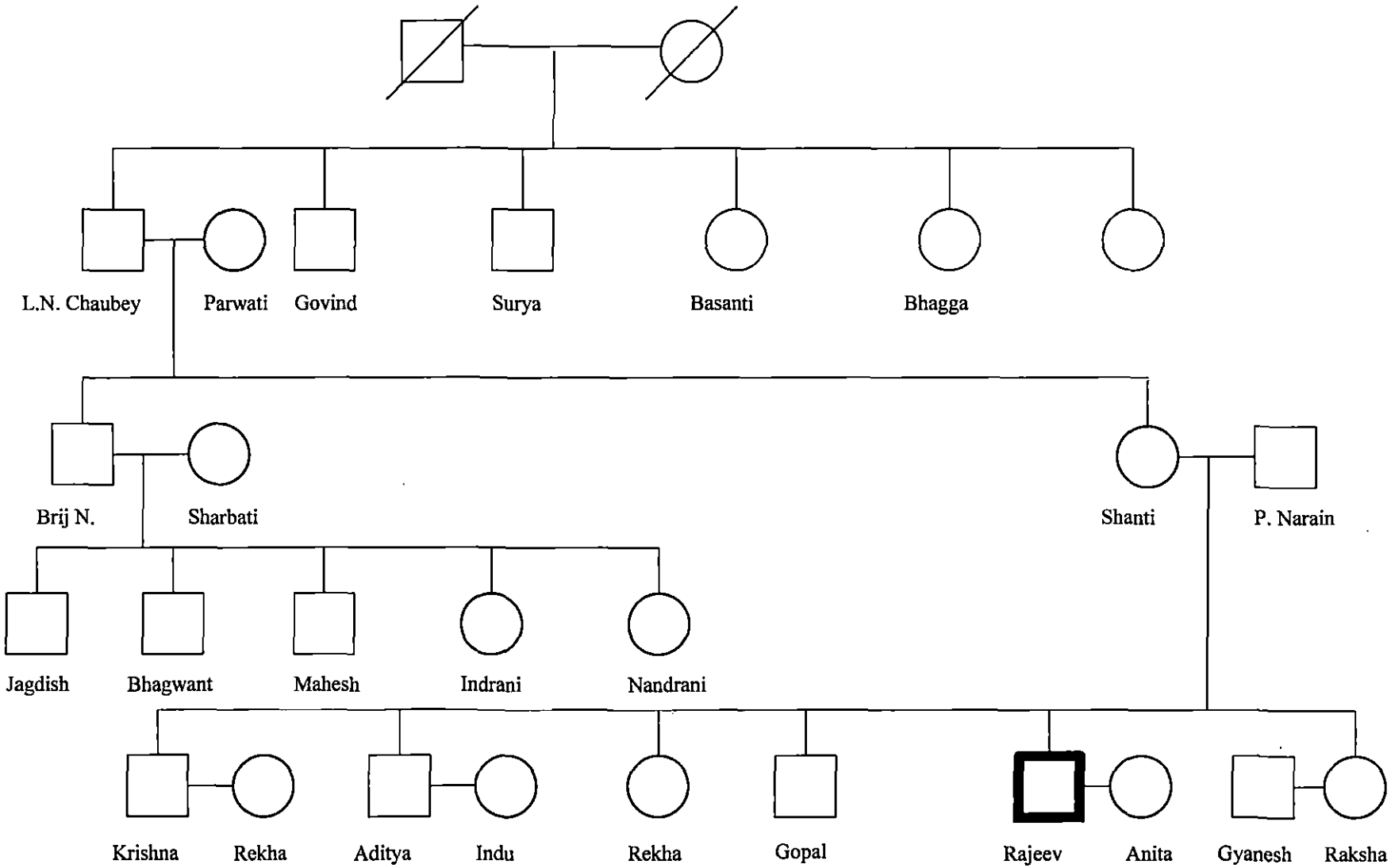
a	Loss of hairs	Y/ N	b	Bulging of eyes	Y/ N
c	Recurrent redness of eyes chronic oral ulcers	Y/ N	d	Loss of vision	Y/ N
e	Puffiness over face	Y/ N	f	General red swelling over body	Y/ N
g	Total lack of salivation and lacrimation	Y/ N	h	Painful swelling of thyroid	Y/ N
i	Skin lesions	Y/ N	j	Ambiguous genitalia	Y/ N
k	Pain in the back & hips, difficulty in walking stooped posture	Y/ N	l	Pain & swelling of Joints lasting more than 6 wk.	Y/ N
m	Chronic intractable diarrhoea	Y/ N	n	Burning in urination pus discharge	Y/ N
o	Extreme weakness	Y/ N	p	Marked wt loss without dieting	Y/ N
q	Prolonged low grade fever of unknown cause	Y/ N	r	Palpitation / Tremors	Y/ N
s	Episodes of paralysis	Y/ N	t	Blood in Sputum / urine	Y/ N

6. Blood Group

7. HLA type

Appendix – B

Example Pedigree:



Appendix-C

REAGENTS

0. 5 M EDTA:

186.1gms of EDTA.$2H_2O$ was added to 800ml of double distilled water. It was mixed vigorously on a magnetic stirrer. pH was adjusted to 8.0 with NaOH (~20g of NaOH pellets) Final volume was adjusted to 1000ml by adding remaining DDW, aliquoted and sterilized by autoclaving.

10% SDS (Sodium dodecyl sulphate or Sodium Lauryl sulphate):

Dissolved 100gms of SDS in 900ml of H_2O, heated to 68°C to assist dissolution. Adjusted the pH to 7.2 by adding a few drops of concentrated HCl. Adjusted the volume to 1000 ml with double distilled H_2O.

5M NaCl:

Dissolved 292.2 gms of NaCl in 800ml of H_2O. Adjusted the volume to 1 liter with H_2O, sterilized by autoclaving.

20xSSC:

Dissolved 175.3 gms of NaCl and 88.2 g of Sodium citrate in 800ml of double distilled H_2O, pH was adjusted to 7.5 with NaOH, final volume was adjusted to 1000 ml. by adding DDW.

20xSSPE:

Dissolved 175.3 gms of NaCl and 27.6 gms of Sodium dihydrogen thyosulphate-hydrated ($NaH_2PO_4.H_2O$) and 7.4 gms of EDTA in 800ml of double distilled H_2O. Adjusted the pH to 7.4 with a few drops of 10N NaOH. Adjusted the volume to 1 liter with H_2O, sterilized by autoclaving.

Ethidium Bromide (10mg/ml):

Added 100mg of ethidium bromide to 10ml of DDW water, stirred on a magnetic stirrer for several hours to ensure that dye was fully dissolved. Wrapped the container in aluminum foil or transferred the solution to a dark bottle and stored at room temperature.

Denhardt's Reagent (50X):

5gms Ficoll (type 400), 5gms of polyvinylpyrrolidone, 5gms of bovine serum albumin (fraction V) were dissolved in 500ml distilled water. Finally the solution was filtered and stored at -20°C.

Denatured, Fragmented Salmon Sperm DNA:

Salmon sperm DNA (Sigma type III sodium salt) was dissolved in DDW at a concentration of 10mg / ml .If necessary, the solution was stirred on a magnetic stirrer for 2-4 hours at room temperature to help the DNA to be dissolved fully. The concentration of NaCl was adjusted to 0.1 M and the solution was extracted once with phenol and once with Phenol: chloroform. The aqueous phase was recovered, and the DNA was sheared by passing it 12 times rapidly through a 17-gauge hypodermic needle. Adding 2 volumes of ice-cold ethanol precipitated the DNA. It was then recovered by centrifugation and re-dissolved at a concentration of 10mg/ml in water. The O. D. at 260nm of the solution was determined and the exact concentration of the DNA was calculated. The solution was then boiled for 10 minutes and stored at -20°C in small aliquots. Just before use, the solution was heated for 5 min in a boiling water bath and then chilled quickly in ice water.

Phenol-Chloroform:

Equal amounts of phenol and chloroform was mixed. The mixture was equilibrated by extracting several times with 0.1M Tris-Cl (pH = 7.6). This equilibriated mixture was stored under an equal volume of 0.01M Tris-Cl (pH = 7.6) at 4^0C in dark glass bottles.

Phenol-Chloroform: Isoamyl Alcohol (25:24:1):

A mixture consisting of equal parts of equilibrated phenol and chloroform:isoamyl-alcohol (24:1) was prepared and the mixture was stored under 100mM Tris.Cl (pH 8.0) in a dark brown tight bottle at 4 °C for periods upto one month.

Tris Borate Buffer:

5x concentrated stock:

54gm Tris base

27.5gm boric acid

20ml 0.5M EDTA (pH 8.0)

Dissolved the above constituents into 1000 ml of double distilled water.

Loading Dye-Bromophenol Blue (6X) stock solution:

0.25% Bromophenol blue ≡ 25mg

40 % Sucrose ≡ 4gms

Above constituents dissolved in 10.0ml of double distilled water.

10X PCR buffer 1.0mM

Supplied by manufacturers (Ms Bangalore Genei, India).

1 M Tris:

Added 121.4 g Tris (hydroxymethyl Aminomethane) in 800ml DDW. Adjusted the pH to 7.4 or 8.0 with 1N HCl and made up the volume to 1000 ml.

Red Cell Lysis Buffer (RCLB):

Dissolved the following constituents in 800ml distilled water

0.32 M Sucrose	≡	54.80g
1%Triton-X-100	≡	5ml
1mM $MgCl_2$	≡	0.51gms
12mM Tris	≡	0.73 gms

Final volume adjusted to 500 ml.

Proteinase K buffer:

EDTA(0.12mM)	≡	22.4gm
Sodium Chloride(0.375)	≡	11gm

Constituted the volume to 500ml in the double distilled water.

Denaturing Solution:

4N NaOH	≡	1ml
0.5MEDTA	≡	1ml

Added water to make up the volume to 10 ml

De-probing Solution:

0.2M NaOH	≡	5.0 ml of 4M NaOH
0.1% SDS	≡	1.0 ml of 10% SDS

Mixed the two solutions well. Made up the final volume to 100 ml with DDW.

10X Plynucleotide Kinase Buffer:

Supplied by manufacturers (Ms Bangalore Genei, India).

Hybridization Buffer:

6X SSPE	≡	30 ml of 20X SSPE
5X Denhardt's solution	≡	10 ml of 50X Denhardt's
0.5% SDS	≡	5 ml of 10% SDS
100 μg/ml Salmon sperm DNA.	≡	1 ml of 10mg/ml SS DNA

Volume made upto 100ml

Cold wash buffer :

2X SSPE.	≡	10ml of 20X SSPE
0.1% SDS.	≡	1ml of 10% SDS

Volume made upto 100ml

Critical wash buffer :

5X SSPE.	≡	30ml of 20X SSPE
1% SDS.	≡	10ml of 10% SDS

Volume made upto 100ml

Appendix D

Dilution of Primers and Probes:

The primers and probes (1OD each) were purchased from DNA agency USA. These were supplied in lyophilised state. The primers were diluted at a stock concentration of 200 pm / µl according to the following procedure.

1. Calculation of molecular weight (M) of the oligonucleotide:

The total numbers of A, T, G and C in each primer sequence were counted. The molecular weight of A=312.2, T= 303.2, G=328.2 and C= 288.2. Accordingly, the molecular weight (M) was calculated.

$M=n_A \times 312.2 + n_T \times 303.2 + n_G \times 328.2 + n_C \times 288.2$

Where,

n_A = no. of Adenine

n_T = no. of Thiamine

n_G = no. of Guanine

n_C = no. of cytosine

e.g. for a primer (2DRB amp A) with sequence

CCC CAC AGC ACG TTT CTT G

$n_A = 3$; $n_T = 5$; $n_C = 8$; $n_G = 3$

$M = 312.2 \times 3 + 303.2 \times 5 + 288.2 \times 8 + 328.2 \times 3$

$= 936.6 + 1516 + 2305.6 + 984.6$

$=$ **5742.8**

2. Calculation of water to be added:

1O.D. = 30µg of single stranded DNA (oligonucleotide) .

$= 30 \times 10^{-6}$ gm

$= 30 \times 10^{-6}$ / M moles of oligo

To make a solution of conc. 1 mole/ µl,

1 OD should be dissolved in $\frac{30 \times 10^{-6}}{M}$ µl

$\Rightarrow$ To make 10^{12} p moles / µl, water to be added = $\frac{30 \times 10^{-6}}{M}$ µl

$\Rightarrow$ To make 1 p mole / µl, water to be added = $\frac{30 \times 10^{-6} \times 10^{12}}{M}$ µl

$\Rightarrow$ To make 200 p mole / µl, water to be added = $\frac{30 \times 10^{6}}{M \times 200}$ µl

$= \frac{150 \times 1000}{M}$ µl

e.g. for the oligonucleotide with a mol. weight of 5742.8 (i.e. primer 2 DRB amp A)

Water to be added = 150 x 1000 / M

= 150 x 1000 / 5742.8 µl

= 26.12 µl

This reconstituted stock was aliquoted and kept at -80 ° C. This was further diluted in HPLC grade water to give a final concentration of 10pm /µl (the user solution).

Appendix-E

Primers for HLA class II gene amplification

Locus	Name	Codons	Sequence	Orientation
Generic				
DRB1	2DRB AMP-A	2-8	CCCCACGCACGTTTCTTG	5'
	2DRBAMP–B	87-94	CCGCTGCACTGTGAAGCTCT	3'
DQA1	2DQAAMP-A	11-18	ATGGTGTAAACTTGTACCAGT	5'
	2DQAAMP–B	80-87	TTGGTAGCAGCGGTAGAGTTG	3'
DQB1	2DQB AMP-A	13-20	CATGTGCTACTTCACCAACGG	5'
	2DQBAMP–B	78-87	CTGGTAGTTGTGTCTGCACAC	3'
Specific				
DR1 – DRB1	DRBAMP-1	8-14	TTCTTGTGGCAGCTTAAGTT	5'
	DRBAMP-B		Same as in generic Amplification	
DR2 – DRB1	2DRBAMP-2	7-13	TTCCTGTGGCAGCCTAAGAAG	5'
	DRBAMP-B		Same as in generic amplification	
DR4 – DRB1	DRBAMP-4	6-13	GTTTCTTGGAGCAGGTTAAC	5'
	DRBAMP-B		Same as in generic amplification	
DR52 – DRB1	2DRBAMP-3	2-10	CACGTTTCTTGGAGTACTCTAC	5'
	DRBAMP-B		Same as in generic amplification	

Appendix-F

Sanjay Gandhi Post Graduate Institute of Medical Sciences, Lucknow

Department of Medical Genetics

Immunogenetics / HLA Lab

PCR – SSOP Hybridisation based HLA typing

PCR – amplification record

Purpose : Population Study **Blot No. :** MSDQA1 **PCR vol. :** 25µl **No of blots :** 3

Date: 23.07.2003 **Primers used: Pr.** (1) 2 DQα amp A

(2) 2 DQα amp B

Sl No.	DNA No.	DNA / PCR (µl)	Amp	Vol / blot	Sl No.	DNA No.	DNA / PCR (µl)	Amp	Vol / blot
1	MS1	3.00	Y	5ul	49	MS49	1.50	Y	5ul
2	MS2	2.50	Y	5ul	50	MS50	1.50	Y	5ul
3	MS3	1.50	Y	5ul	51	MS51	1.50	Y	5ul
4	MS4	1.50	Y	5ul	52	MS52	2.50	Y	5ul
5	MS5	1.50	Y	5ul	53	MS53	2.00	Y	5ul
6	MS6	1.50	Y	5ul	54	MS54	2.50	Y	5ul
7	MS7	2.50	Y	5ul	55	MS55	2.50	Y	5ul
8	MS8	2.00	Y	5ul	56	MS56	2.50	Y	5ul
9	MS9	2.50	Y	5ul	57	MS57	2.00	Y	5ul
10	MS10	2.50	Y	5ul	58	MS58	1.50	Y	5ul
11	MS11	2.50	Y	5ul	59	MS59	3.00	Y	5ul
12	MS12	2.00	Y	5ul	60	MS60	2.50	Y	5ul
13	MS13	1.50	Y	5ul	61	MS61	1.50	Y	5ul
14	MS14	3.00	Y	5ul	62	MS62	1.50	Y	5ul
15	MS15	2.50	Y	5ul	63	MS63	1.50	Y	5ul
16	MS16	1.50	Y	5ul	64	MS64	1.50	Y	5ul
17	MS17	1.50	Y	5ul	65	MS65	2.50	Y	5ul
18	MS18	1.50	Y	5ul	66	MS66	2.00	Y	5ul
19	MS19	1.50	Y	5ul	67	MS67	2.50	Y	5ul
20	MS20	2.50	Y	5ul	68	MS68	2.50	Y	5ul
21	MS21	2.00	Y	5ul	69	MS69	2.50	Y	5ul
22	MS22	2.50	Y	5ul	70	MS70	2.00	Y	5ul
23	MS23	2.50	Y	5ul	71	MS71	1.50	Y	5ul
24	MS24	2.50	Y	5ul	72	MS72	1.50	Y	5ul
25	MS25	2.00	Y	5ul	73	MS73	1.50	Y	5ul
26	MS26	1.50	Y	5ul	74	MS74	1.50	Y	5ul
27	MS27	1.50	Y	5ul	75	MS75	1.50	Y	5ul
28	MS28	1.50	Y	5ul	76	MS76	2.50	Y	5ul
29	MS29	1.50	Y	5ul	77	MS77	2.00	Y	5ul
30	MS30	1.50	Y	5ul	78	MS78	2.50	Y	5ul
31	MS31	2.50	Y	5ul	79	MS79	2.50	Y	5ul
32	MS32	2.00	Y	5ul	80	MS80	2.50	Y	5ul
33	MS33	2.50	Y	5ul	81				
34	MS34	2.50	Y	5ul	82				
35	MS35	2.50	Y	5ul	83				
36	MS36	2.00	Y	5ul	84				
37	MS37	1.50	Y	5ul	85	C1	1.50	Y	5ul
38	MS38	3.00	Y	5ul	86	C2	2.50	Y	5ul
39	MS39	2.50	Y	5ul	87	C3	2.00	Y	5ul
40	MS40	1.50	Y	5ul	88	C4	2.50	Y	5ul
41	MS41	1.50	Y	5ul	89	C5	2.50	Y	5ul
42	MS42	1.50	Y	5ul	90	C6	2.50	Y	5ul
43	MS43	1.50	Y	5ul	91	C7	2.00	Y	5ul
44	MS44	2.50	Y	5ul	92	C8	1.50	Y	5ul
45	MS45	2.00	Y	5ul	93	C9	1.50	Y	5ul
46	MS46	2.50	Y	5ul	94	C10	1.50	Y	5ul
47	MS47	2.50	Y	5ul	95	C11	1.50	Y	5ul
48	MS48	2.50	Y	5ul	96	C12	1.50	Y	5ul

PCR details

dNTP used : Pharmacia-1-19 Breakwiper – real Taq used : Bangalore Genei Buf used : Bangalore Genei

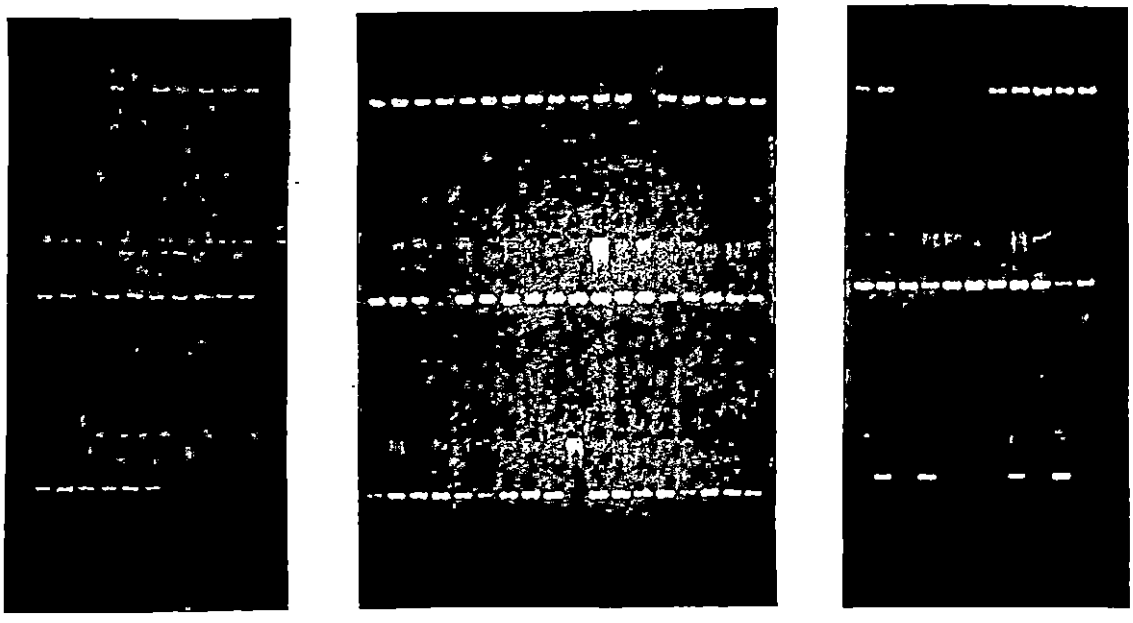
Primer used : Pr (1) 2DRB Amp A Sequence : Pr (1)
(2) 2DRB Amp B (2)

dNTP primer mix :

10 mM dNTP Mix (Breakwiper) → 20.0 µl
10× Buf (BG) → 100.0 µl
Water → 680.0 µl
Cr Red → 2.0 µl
Water → 198.0 µl
1600.0 µl

Gel Photo :

Gel Photo:



Blotting details

Blot Number **MSDQα1** Purpose Number of copies
Primers used: Pr (1) **2 DQα ampA**
Pr (2) **2 DQα ampB**
Protocol: **SSOP** Number of use: **IV** Satisfactory Results:
Tray Number: **2** Box Number: **4** Kept at:

Control DNA details:

Sl. No	DNA No.	Slot No.	Positive for probes	Sl. No	DNA No.	Slot No.	Positive for probes
1	C1	0101, 0104		1	C11	0301, 0301	
2	C2	0102, 0201		2	C12	0201, 0301	
3	C3	0101, 0101		3			
4	C4	0104, 0104		4			
5	C5	0201, 0201		5			
6	C6	0102, 0301		6			
7	C7	0501, 0401		7			
8	C8	0102, 0501		8			
9	C9	0104, 0601		9			
10	C10	0101, 0601		10			

Template:

MS1	MS2	MS3	MS4	MS5	MS6	MS7	MS8
MS9	MS10	MS11	MS12	MS13	MS14	MS15	MS16
MS17	MS18	MS19	MS20	MS21	MS22	MS23	MS24
MS25	MS26	MS27	MS28	MS29	MS30	MS31	MS32
MS33	MS34	MS35	MS36	MS37	MS38	MS39	MS40
MS41	MS42	MS43	MS44	MS45	MS46	MS47	MS48
MS49	MS50	MS51	MS52	MS53	MS54	MS55	MS56
MS57	MS58	MS59	MS60	MS61	MS62	MS63	MS64
MS65	MS66	MS67	MS68	MS69	MS70	MS71	MS72
MS73	MS74	MS75	MS76	MS77	MS78	MS79	MS80
				C1	C2	C3	C4
C5	C6	C7	C8	C9	C10	C11	C12

Appendix G

List of probes used for DRB1 generic typing

	Name	Sequence	Locus and allelic specificities
1	**2801**	CGGTTGCTGGAAAGATGC	DRB1*0101+0102+0103+0104
2	**1002**	AGCCTAAGAGGGAGTGTC	DRB1*1501+1502+1503+1601+1602+1603+1603 +1604+1606
3	**1008N**	GAGGAGGTTAAGTTTGAG	DRB1*1001
4	**7004**	GGCGCGGTGGACAACTAC	DRB1*03011+0302+0303+0304+1107
5	**5703**	GCCTGATGAGGAGTACTG	DRB1*0415+11011+11012+1102+1103+11041 +11042+1105+1106+1107+11081+11082+1109 +1110 +1111 +1112+1113+1411
6	**1003**	GTACTCTACGTCTGAGTG	DRB1*03011+03012+0302+0303+0304 +11011 +11012+1102+1103+11041+11042+1106 +1107 +11081 +11082+1112+1113+1301+1302+1303 +1304+1305+1306+1307+1308+1309+1310 +1311+ 312+1313+1401+1402+1403+1405+1406 +1407 +1408+ 1409+1412+1413+1414 +1416 +1417+0304
7	**1010**	GAGCTGCGTAAGTCTGAG	DRB3*0101 (DW24)
8	**1011**	GAGCTGCTTAAGTCTGAG	DRB3*0201+0202+0301 (DW25, 26)
9	**1004**	GAGCAGGTTAAACATGAG	DRB*0401+0402+0403+0404+0405+0406+0407 +0408+0409+0410+0411+0412+04 13 +0414 +0415 +0416 +0417 +0418+0419+1410
10	**1006**	TGGCAGGGTAAGTATAAG	DRB1* 0701
11	**1007**	GAAGCAGGATAAGTTTGA	DR B1*09011+0912
12	**2810**	GCGAGTGTGGAACCTGAT	DR B4*0101+0102+0103 (DRW 53 4, 7, 9)
13	**DRB12**	CAGGAGGAGCTCCTGCGC	DRB1*1201+1202+1203
14	**DRB8**	CGGGCCCTGGTGGACAC	DRB1*0412+0418 +0801+08021+08 022 +08031 +08032+08041+08042+0806+0807+0808 +0809 +0810+0811 +1313+1403+1412+1415 +1604
15	**DRB6**	CAGGAGGAGAACGTGCGC	DRB1*03011+03012+0302+0303+1109+1301 +1302 +1305+1306+1309+1310 +1402 +1403 +1406 +1409 +1412+1413+1417
16	**7031**	CTGGAAGACAAGCGGGCCG	DRB1*1303+1310
17	**DRB 14/1**	GGCCTGCTGCGGAGCACT	DRB1*1401+1404+1407+1410+1416+0808
18	**DRB13**	TGGAAGACGAGCGGGCCG	DRB1* 0103 +0402+0414 +1102+1103 +1301 +1302 +1304 +1308 +1111+1416
19	**7012**	ACCGCGGCCGGCCTCTGC	DRB1* 0101+0102+0104 +0404+405 +0408 +0410 +0417 +0419 +1402+1406+1409

List of probes used for DQA1 generic typing

	Name	Sequence	Locus and allelic specificities
1	2501	TGGCCAGTACACCCATGA	DQA1*0101+0102+0104+0401+0501+05012+05013 +0502
2	2502	TGGCCAGTTCACCATGA	DQA1*0103+0201+0601
3	2503	TGGGCAGTACAGCCATGA	DQA1*03011+03012+0302
4	3401	GAGATGAGGAGTTCTACG	DQA1*0101+0104
5	3402	GAGATGAGCAGTTCTACG	DQA1*0102+0103+05011+05012+05013+0502
6	3403	GAGACGAGCAGTTCTACG	DQA1*0401+0601
7	4101W	ACCTGGAGAGGAAGGAGA	DQA1*0101+0102+0104+0201+03011+03012+0302
8	4102	ACCTGGAGAAGAAGGAGA	DQA1*0103
9	4103W	ACCTGGGGAGGAAGGAGA	DQA1*0401+05011+05012+05013+0502+0601
10	5501	TCAGCAAATTTGGAGGTT	DQA1*0101+0102+0103+0104
11	5502W	TCCACAGACTTAGATTTG	DQA1*0201
12	5503	TCCGCAGATTTAGAAGAT	DQA1*0301+03012+0302
13	5504	TCAGACAATTTAGATTTG	DQA1* 0401+0601+05011+05012+05013+0502
14	5901W	TTTGACCCGCAATTTCGA	DQA1*0201+03011+03012+0302+0401+05011+050 12+05013+0601
15	5902W	TTTGACCGGCAATTTGCA	DQA1* 0502
16	6901	ATGGCTGTGGCAAAACAC	DQA1* 0101+0102+0103+0104
17	6902	ATCGCTGTGCTAAAACAT	DQA1* 0201+03011+0302
18	6903	ATCGCTGTCCTAAAACAT	DQA1* 03012+05011+05012+05013+0502
19	6904	ATCGCTGTGACAAAACAC	DQA1* 0401+0601
20	7502	CTTGAACATCCTGATTAA	DQA1* 0201+0401+0601
21	7504W	CTTGAACAGTCTGATTAA	DQA1* 05011+05012+05013+0502

List of probes used for DQB1 generic typing

	Name	Sequence	Locus and allelic specificities
1	2301	GACCGAGCTCTGGCGGGG	DQB1*0401
2	2302	AACGGGACCGAGCGCGTG	DQB1*0402+ 03031+ 0305
3	2601	CGGGGTGTGACCAGACAC	DQB1*0501+ 0502+ 05031+ 05032
4	2602	CGTTATGTGACCAGATAC	DQB1*0301+ 0304+ 06011+ 06012
5	2603	CGTCTGGTGACCAGATAC	DQB1*0302+ 03031+ 03032 +0602
6	2604	CGTCTTGTAACCAGACAC	DQB1*0603+ 0604+ 0607+ 0608
7	2605W	CGTCTTGTGAGCAGAAGC	DQB1*0201+0202
8	2606	CGTCTTGTAACCAGATAC	DQB1*06051+06052 + 0606/9
9	2607W	GGGTGTGACCAGATACAT	DQB1*0305+ 0401+ 0402+ 0504
10	3701	AGGAGTACGTGCGCTTCG	DQB1*0501+ 0502+ 05031+ 05032
11	3702	AGGAGGACGTGCGCTTCG	DQB1*06011+ 06012
12	3703W	TAACCGAGAAGAGTACGT	DQB1*0504
13	3704W	GGAGTACGCGCGCTTCGA	DQB1*0602+0603+0604+06051+06052 + 0606+0607+0608+0609+0305+0401+0402
14	3705W	GGAGTACGCACGCTTCGA	DQB1*0301+ 0302+ 03031+ 03032+ 0304
15	4501	GACGTGGAGGTGTACCGG	DQB1*0301+ 0304
16	4901	GGTGTACCGGGCAGTGAC	DQB1* 0501
17	4902W	GGTGTATCGGGCGGTGAC	DQB1*05031+ 05032+ 06011+ 06012 +0302+ 03031+ 03032+0305+ 0401 +0402
18	5701	GCGGCCTGTTGCCGAGTA	DQB1*0501+ 0604+ 06051+ 0606+ 0608
19	5702	GCGGCCTAGCGCCGAGTA	DQB1*0502+ 0504
20	5703	GCGGCCTGATGCCGAGTA	DQB1*05031+ 06011+ 06012
21	5704	GCGGCCTGATGCCTAGTA	DQB1*05032+ 0602+ 0603+ 0607
22	5705	GGCTGCCTGCCGCCGAGT	DQB1*0201+0202
23	5706	GGCCGCCTGACGCCGAGT	DQB1*0301+ 03031+ 03032
24	5707W	TGGGGCCGCCTGCCGCCG	DQB1*0302+ 0304+ 0305
25	5708	GCGGCTTGACGCCGAGTA	DQB1*0401+ 0402
26	7001W	TGGAGGGGGCCCGGGCGT	DQB1*0501+ 0502+ 05031+ 05032
27	7002	GACCCGAGCGGAGTTGGA	DQB1*06011+06012
28	7003	GAGGGGACCCGGGCGGAG	DQB1*0602+ 0603+ 0608
29	7004W	CTGGAGAGGACCCGGGCG	DQB1*0301+0302+ 03031+03032+ 0304 +0305+ 0604+ 06051 +06052 +0606 +0607 +0609
30	7005	GAAACGGGCGGCGGTGGA	DQB1*0201+0202
31	7006W	CTGGAGGAGGACCGGGCG	DQB1*0401+ 0402+ 0504
32	7007W	AGGACCCGGGCGGCGGTG	DQB1*0606
33	7008W	ACCCGGGCGGAGTTGGAC	DQB1*0301+0302+ 03031+ 03032+ 0304 +0305+ 0602+0603 +0604+06051+06052+ 0607+0608+0609
34	7009W	GGTGGACAGGGTGTGCAC	DQB1*0501+ 0504+ 0606

Probes used for DR2 subtyping

Probe	Sequence	Specificities
2813	GTTCCTGGACAGATACTT	DRB1*1501+1502+1601+1602+1603
2817W	TTCCTGGACAGACACTTC	DRB1*1503
3702W	AGATACTTCTATAACCAA	DRB1*1501+1502+1503 +1601+1602+1603
5706	GCCTGACGCTGAGTACTG	DRB1*1501+1503 +1601+1602+1603
7002	GACTTCCTGGAAGACAGG	DRB1*1601+1603 1604+1608
7003	GACCTCCTGGAAGACAGG	DRB1*1602
7011	GACATCCTGGAGCAGGCG	DRB1*1501+1502+1503+1506
7013W	GAAGACAGGGCCGCCGCG	DRB1*1603
7014W	GAAGACAGGCGCGCCGCG	DRB1*1601+1602
8601	AACTACGGGGTTGGTGAG	DRB1*1502+1601+1602+1603+1604+ 1605+1606+1607+1608
8603	AACTACGGGGTTGTGGAG	DRB1*1501+1503+1504+1505+1506
7010W	CATCCTGGAAGACAGGCG	DRB1*1605+1607

Appendix H

Probing Results

Blot No.: Ky (1) **Probe No.:** 2801 **Qty:** 1pm

Specificity: DRB1*0101+0102+0103+0104

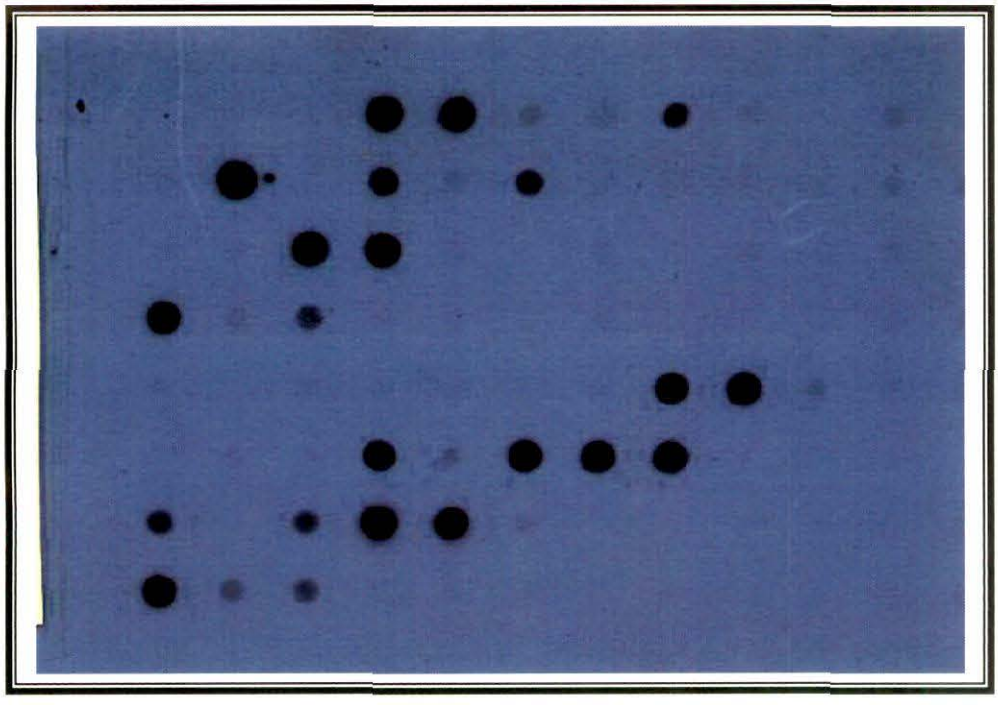
Hyb. Date: 16.02.2003 **Dev. Date:** 17.2.2003

Critical wash temp.:58°C

Scoring: **Positive Control:** H1

			+8	+8			+4				
	+8		+6		+4						
		+8	+8								
+8		+4									
							+8	+8			
			+8		+8	+8	+8				
+4		+4	+8	+8							
+8	+2	+2									

Auto-radiogram:



Probing Results

Blot No.: Ky DRB3 II (1) **Probe No.:** 2808 **Qly:** 1pm

Specificity: DRB3*0201+0202

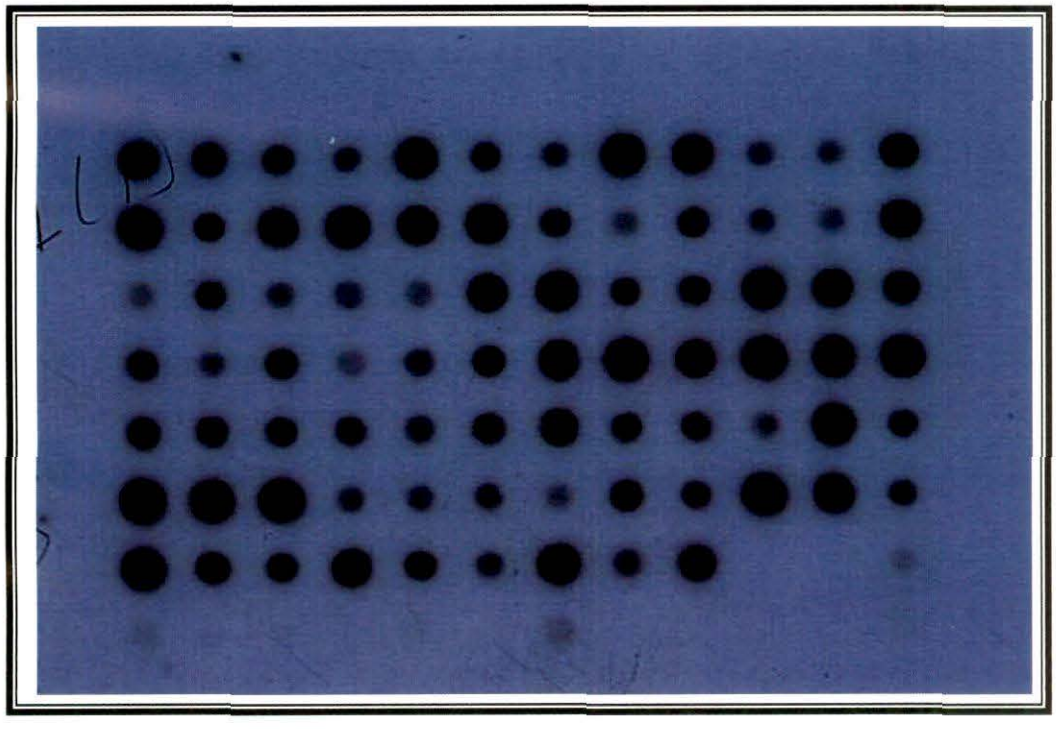
Hyb. Date: 11.10.2003 **Dev. Date:** 17.2.2003

Critical wash temp.: 53°C / 53.6

Scoring: **Positive Control:** H1

+8	+4	+4		+8			+8	+8			+8
+8		+8	+8	+8	+8						+8
	+4				+8	+8			+8	+8	+6
+4		+4			+6	+8	+8	+8	+8	+8	+8
+4		+4	+4		+6	+8				+8	
+8	+8	+8					+6		+8	+8	
+8	+4	+4	+8	+4		+8		+8			

Auto-radiogram:



Probing Results

Blot No.: Mr DRB1 (2) **Probe No.:** 7004 **Qty:** 1pm

Specificity: DRB1*03011+0302+0303+1107+0304

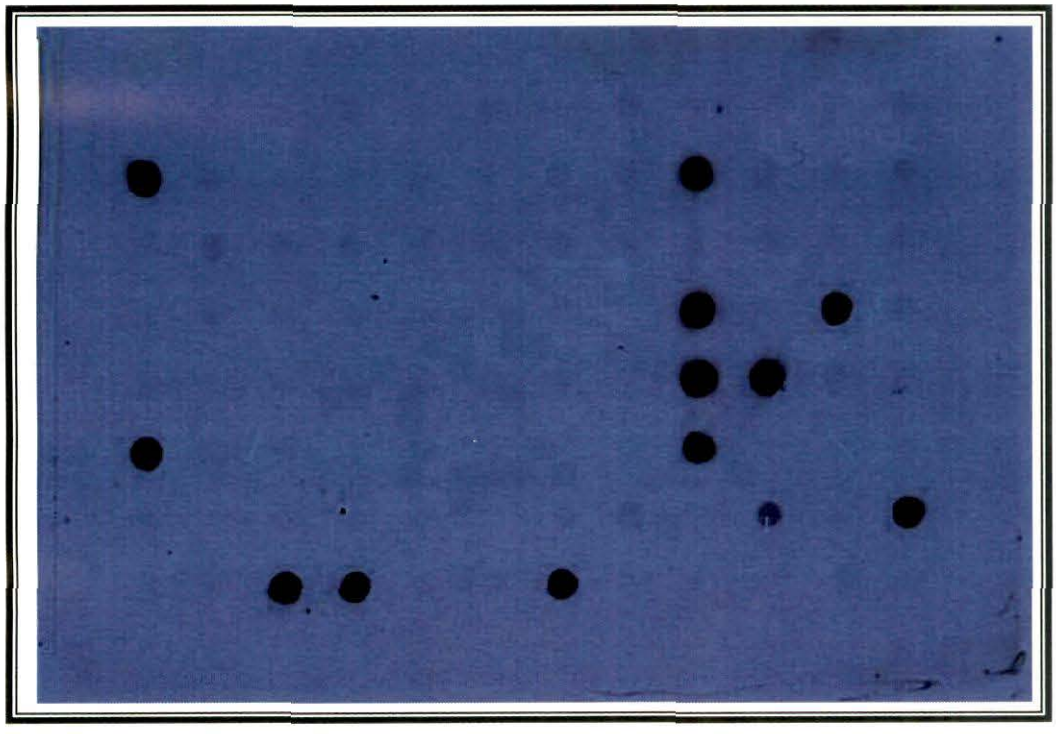
Hyb. Date: 17.02.2004 **Dev. Date:** 18.2.2004

Critical wash temp: 62°C

Scoring: **Positive Control:** H3,H4,H7

+8								+8			
								+8		+8	
								+8	+8		
+8								+8			
											+8
		+8	+8			+8					

Auto-radiogram:



Probing Results

Blot No.: Vs DQA I (2) **Probe No.:** 3401 **Qty:**

Specificity: DQA1*0101+0104

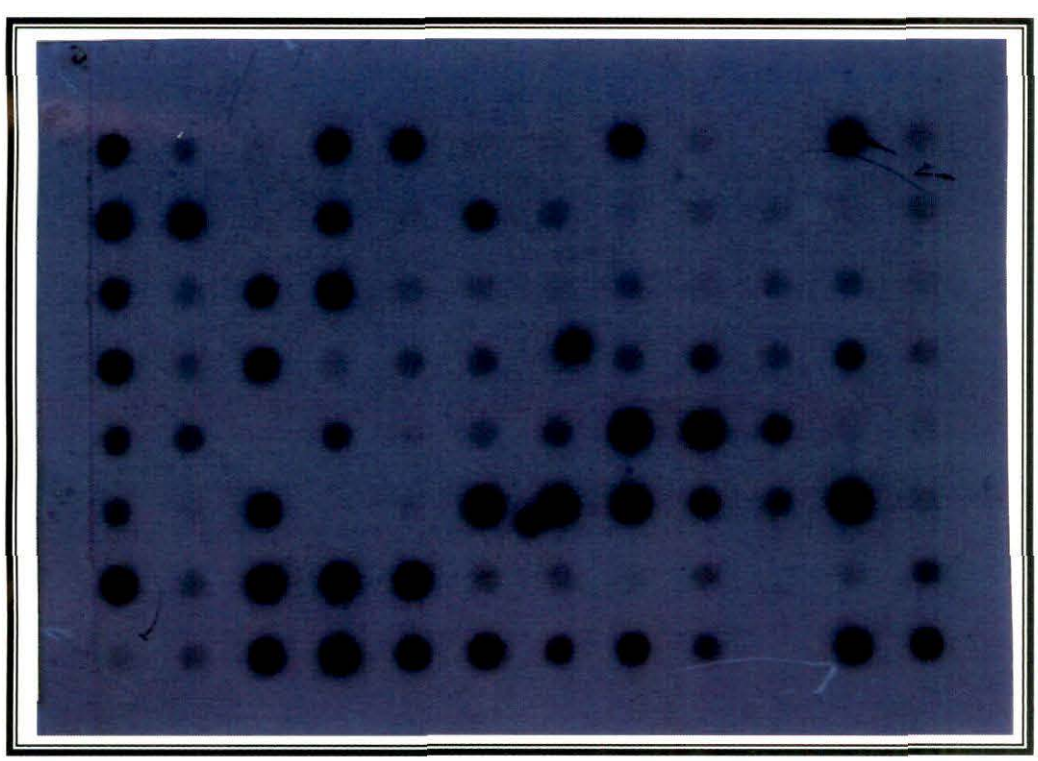
Hyb. Date: 15.06.2004 **Dev. Date:** 17.06.2004

Critical wash temp: 53.7°C

Scoring: **Positive Control:**

+6			+6	+6			+6			+6	
+8	+6		+6		+6						
+6		+6	+8								
+6		+6									
							+8	+8			
		+6			+8	+8	+8			+8	
+8		+8	+8	+8							
		+6	+8	+6	+6	+4	+6	+4		+8	+6

Auto-radiogram:



Probing Results

Blot No.: Vs DQB I (1) **Probe No.:** 2601 **Qty:**

Specificity: DQB1*0501+0502+05031+05032

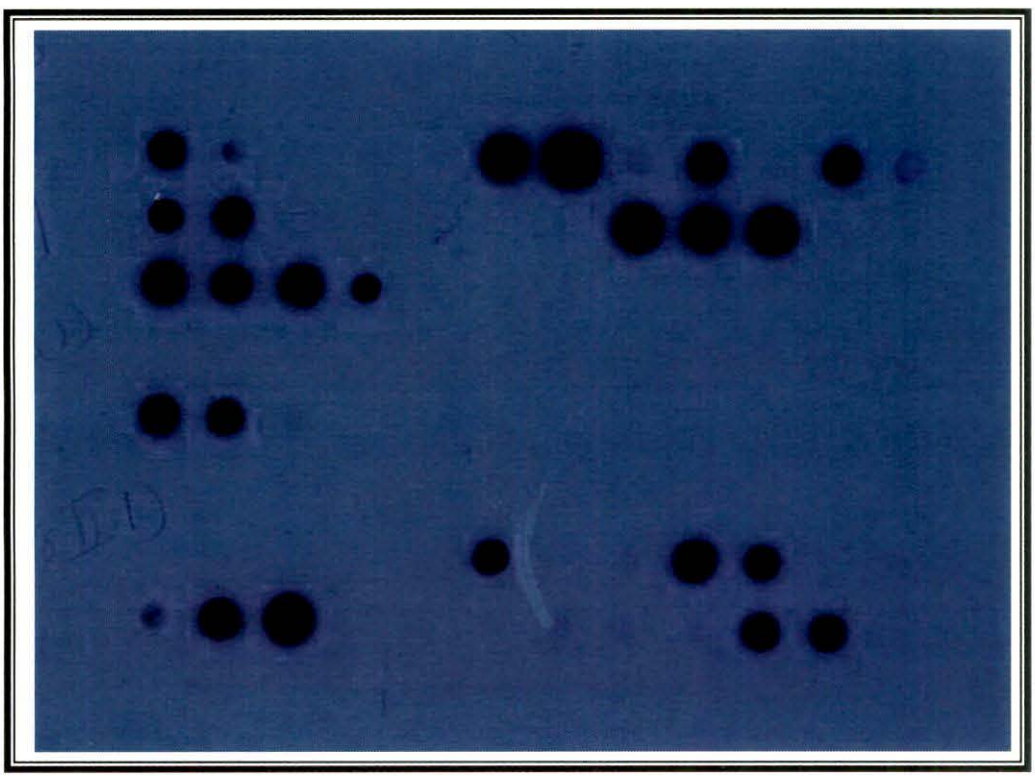
Hyb. Date: 21.06.2004 **Dev. Date:** 22.06.2004

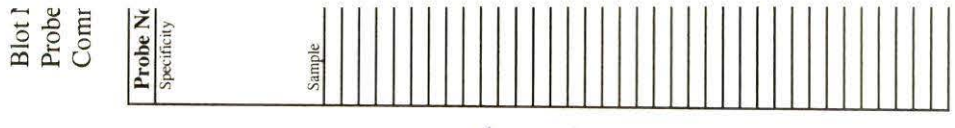
Critical wash temp: 59.7°C

Scoring: **Positive Control:**

+8					+8	+8		+8		+8	
+6	+8						+8	+8	+8		
+8	+8	+8	+4								
+8	+8										
						+6			+8	+6	
	+8									+8	+8

Auto-radiogram:





Appendix - I

Interpretation Table

Tissue Typing (DRB1)

Blot No.: Pg No.: Samples: Purpose: Primers:
Probes: Label : Controls: By :
Comments:

Probe No	2801	1002	1008N	7004	5703	1003	1010	1011	1004	1006	1007	2810	DRB12	DRB8	DRB6	7031	DRB 14/1	DRB13	7012	Typing
Specificity	DRB1* 0101/02/ 03/04	DRB1* 1501/02/03 +1601 /02/03/0 3/ 04/06	DRB1* 1001	DRB1* 03011/ 02/03/04 +1107	DRB1* 0415 +11011/012/02/ 03/041/042/05 /06/07/081/082/ 09/10/11/12/13 +1411	DRB1* 03011/012/02/03/ 04 +11011/012/02/03/ 041/042/06/07/081/08 2 /12/13/ +1301/02/03/04/05/0 6/07/08/09/10/11/12/ 13 +1401/02/03/05/06/0 7/08/09/12/13/14/16/ 17/ +0304	DRB3* 0101 (DW24)	DRB3* 0201/02 +0301 (DW25, 26)	DRB* 0401/02/ 03/04/05 /06/07/ 08/09/10 /11/12/ 13/14/15 /16/17/ 18/19 +1410	DRB1* 0701	DR B1* 09011/12	DR B4* 0101/02/03 (DRW 53 4, 7, 9)	DRB1* 1201/02/03	DRB1* 0412/18 +0801/021/ 022/031/032/ 041/042/ 06/07/08/09/ 10/11 +1313 +1403/12/15 +1604	DRB1* 03011/012/ 02/03 +1109 +1301/02/05/ 06/09/10 +1402/03/ 06/09/12/13/ 17	DRB1* 1303/1310	DRB1* 1401/4/7/10/16 +0808	DRB1* 0103 +0402/14 +1102/3 +1301/2/4/8 +1111 +1416	DRB1* 0101/02/04 +0404/05/ 08/10/17/ 19 +1402/06/ 09/	
Sample																				

Signature

Tissue Typing (DR2 ST)

Blot No.: Pg No.: Samples: Purpose: Primers:
Probes: Label : Controls: By :
Comments:

Probe No	2813	2817W	3702	5706	7002	7003	7011	7013W	7014W	8601	8603	7010W	Typing
Specificity Sample	DRB1* 1501/02 +1601/02/03	DRB1* 1503	DRB1* 1501/02/03 +1601/02/03	DRB1* 1501/03 +1601/02/03	DRB1* 1601/03/04/08	DRB1* 1602	DRB1* 1501/02/03/06	DRB1* 1603	DRB1* 1601/02	DRB1* 1502 +1601/02/ 03/04/05/ 06/07/08	DRB1* 1501/03/04/05/06	DRB1* 1605/07	

Signature

Tissue Typing (DQA1)

Blot No.: Pg No.: Samples : Purpose : Primers:
Probes : Label : Controls: By :
Comments:

Probe No	2501	2502	2503	3401	3402	3403	4101W	4102	4103W	5501	5502W	5503	5504	5901W	5902W	6901	6902	6903	6904	7502	7504W	Typing
Specificity	DQA1*0101 +0102+0104 +0401+0501 +05012 +05013 +0502	DQA1*0103 +0201 +0601	DQA1*03011 +03012 +0302	DQA1*0101 +0104	DQA1*0102 +0103 +05011 +05012 +05013 +0502	DQA1*0401 +0601	DQA1*0101 +0102+0104 +0201+03011 +03012+0302	DQA1*0103	DQA1*0401 +05011 +05012 +05013 +0502 +0601	DQA1*0101 +0102 +0103 +0104	DQA1*0201	DQA1*0301 +03012+0302	DQA1*0401 +0601 +05011 +05012 +05013 +0502	DQA1*0201 +03011 +03012 +0302+0401 +05011 +05012 +05013+0601	DQA1*0502	DQA1*0101 +0102 +0103 +0104	DQA1*0201 +03011 +0302	DQA1*03012 +05011 +05012 +05013 +0502	DQA1*0401 +0601	DQA1*0201 +0401 +0604	DQA1*05011 +05012 +05013 +0502	
Sample																						

Signature

Tissue Typing (DQB1)

Blot No.: Pg No.: Samples: Purpose: Primers:
Probes: Label : Controls: By :
Comments:

Probe No	2301	2302	2601	2602	2603	2604	2605W	2606	2607	3701	3702	3703W	3704W	3705W	4501	4901	4902W	5701	5702	5703	5704	5705	5706	5707	5708	7001W	7002	7003	7004W	7005	7006	7007	7008	7009	7010	Typing
Specificity	DQB1* 0401	DQB1* 0402+ 03031+ *0305*	DQB1* 0501+ 0502+ *05031+* 05032	DQB1* 0301+ 0304+ *06011+* 06012	DQB1* 0302+ 03031+ *03032* +0602	DQB1 * 0603+ *0604+* 0607+ 0608	DQB1* 0201/2	DQB1* 06051/ 52 + *060/59*	DQB1* 0305+ 0401+ *0402+* 0504	DQB1* 0501+ 0502+ *05031+* 05032	DQB1* 06011+ 06012	DQB1* 0504	DQB1* 0602+0603+ 0604+0605 1+0602+ 0606+0607+ 0608/9+ 0305+0401+ 0402	DQB1* 0301+ 0302+ 03031+ 03032+ 0304	DQB1* 0301+ 0304	DQB1* 0501	DQB1* 05031+ 05032+ 06011+ 06012+0302 + 03031+ 03032+0305 + 0401+ 0402	DQB1* 0501+ 0604+ 06051+ 0606+ 0608	DQB1* 0502+ 0504	DQB1* 05031+ 06011+ 06012	DQB1* 05032+ 0602+ 0603+ 0607	DQB1* 0201/2	DQB1* 0301+ 03031+ 03032	DQB1* 0302+ 0304+ 0305	DQB1* 0401+ 0402	DQB1* 0501+ 0502+ 05031+ 05032	DQB1* 06011/ 12	DQB1* 0602+ 0603+ 0608	DQB1* 0301/2+ 03031/32+ 0304/5+ 0604+ 06051/52+ 0606/7/9	DQB1* 0201+ 0202	DQB1* 0401+ 0402+ 0504	DQB1* 0606	DQB1* 0301/2+ 03031+ 03032+ 0304/5+ 0602/3 /4+0605 1/52+ 0607/8/9	DQB1* 0501+ 0504+ 0606	DQB1* 0401/2	
Sample																																				

Signature

Publication

Uddalak Bharadwaj, Faisal Khan, **Sanjeev Srivastava**, Himanshu Goel and Suraksha Agrawal. Phylogenetic Applications of HLA Class II Loci. Int J Hum Genet (in press)