Ancestry of Indian Population Determined By Y Chromosome Markers

ANCESTRY OF THE INDIAN POPULATION DETERMINED BY Y CHROMOSOME MARKERS

By

JHILIK DE, B.Sc.

A Thesis

Submitted to the School of Graduate Studies

In Partial Fulfillment of the Requirements

For the Degree

Master of Science

McMaster University

MASTER OF SCIENCE
(Biology)

McMASTER UNIVERSITY
Hamilton, Ontario

TITLE: Ancestry of the Indian Population Determined By Y Chromosome Markers.

AUTHOR: Jhilik De, B.Sc.
University of Calcutta, India.

SUPERVISOR: Dr. Rama S. Singh.

NUMBER OF PAGES: x, 153

ABSTRACT

Seven microsatellites, DYS 388, DYS 389I, DYS 389II, DYS 390, DYS 391, DYS 392 and DYS 393 and a biallelic locus DYS 271 were analyzed in the Indian population with the aim of understanding the relationships of the Indo-European and the Dravidian population of India with those of the European /Central Asian populations. The 116 Indian samples, used in the study, enjoy a wide geographical distribution and represents well the different religions and caste groups of India. Supporting previously published results, most variation is found between populations within continents than between continents. No significant differences were found between the different religious and caste groups of India with those of the World populations. All the alleles of the different markers are well represented in the different ethnic groups of India, and thus, does not support the popular belief that the highest diversity exists within the middle caste.

Significant differences in the distribution of alleles are found in different regions of India especially between the North and the South. A phylogenetic relationship appears to exist between the Sindhis from South Pakistan with those of the South Indian population. This lends support to the belief that the present population from Sind might represent a remnant population of the original inhabitant of the region that was mainly pushed further to the Southern part of Subcontinent by later invasions. A phylogenetic relationship also exists between the Indian population and the Burushaski population

from Northwest Pakistan and with the Kazakhs, Kirgiz and the Uighurs from Central Asia. This is in agreement with previously published results, and supports the occurrence of two main population movements from Central Asia into India at different time intervals that gave birth to the two main language families-the Dravidian and the Indo-European languages in India.

Acknowledgements

The task of this M.Sc. program could not have been possible without the guided advice, direction, suggestions and inspirations from my supervisors and advisors, Dr. Rama S. Singh and Dr. Richard Morton. With the deepest gratitude and respect, I would like to thank them.

There aren't enough words to thank Aaron Thomson. I consider myself very lucky to have a good friend like Aaron. He has been my friend, philosopher and guide for the last few years and has provided me all the reasons to smile when things did not look right. He had encouraged and supported me in all possible way in trying times.

I thank Sujatha Tampi who assisted me to learn about polyacrylamide gels and PCRs, and analyzed one of the biallelic loci YAP. Thanks are also due to Charlotte Lombardo for analyzing the DYS 19 locus, and Rob Kulathinal and Santosh Jagadeshan for creating a congenial atmosphere in the Lab. I appreciate the help of Martha and Dorothy from Dr. Campos Lab. They helped me to learn bacterial works.

I like to thank everyone in the Biology Department and especially Pat, for their advice in coping with a graduate student's life.

Last but not the least; I like to thank my husband, Deb. Without him, I would have starved during most of my research and course works. I appreciate his mental strength in putting up with me and fulfilling my every weir demands.

Dedicated to my parents

Whose Inspiration, Encouragement and Support have made me what I am today

Table of Contents

Table of Contents (contd.)

List of Figures

List of Tables

INTRODUCTION

Questions about the origin, migration and admixture pattern of the human race have been a matter for debate by scientists around the world in the last century. To seek answers to these questions, they have turned to archaeology, anthropology, linguistics, as well as other fields. Recently, anthropologists have used molecular techniques such as RFLPs and RAPDs and in particular microsatellites to study human movement patterns.

Microsatellites are short tandem repeat blocks of 1 to 6 base pairs, and are widely distributed in the eukaryotic genome. They are preferred for high resolution genetic mapping. In fact, microsatellites are fast replacing RFLPs and RAPDs in population and demographic studies. The genetic basis of microsatellite variability is due to the difference in the number of repeat units. They can be easily scored by primers that amplify well defined repeat region between unique binding sites. Studies on microsatellites indicate that they are very unstable and have some of the highest mutation rates (compared to other markers used for phylogenetic studies). Human pedigree studies (Weber and Wong 1993) show that the average mutation rate for dinucleotide and tetranucleotide repeat units is 0.001. Tetranucleotide repeat units are more mutable than dinucleotide (Goldstein and Pollock 1997). This high mutation rate is attributed to polymerase slippage during DNA replication (Levinson and Gutman 1987, Moxon and Wills 1999). Various mathematical and statistical models have been developed that take into account the rate and mode of microsatellite evolution. High variability have not only made microsatellites a favorite among the forensic and clinical experts characterizing

human genetic individuality but also among molecular anthropologists analyzing human population and evolution.

Mode of Microsatellite Mutation

Microsatellites usually mutate by single step or by a single repeat unit, though in few cases mutation does occur by larger units (Amos and Rubinsztein 1996, Henderson and Petes 1992). Mutations resulting in larger repeat changes (Ashley and Warren, 1995) are most often observed in trinucleotide repeats where the number of repeats can exceed more than 1000 as in some neurological disorders (eg, Huntington, Fragile X syndrome). But with the exception of trinucleotides, repeat lengths of more than 60 are uncommon. Thus the mutation process may not be as unconstrained as would be expected with such a high mutation rate and with the large number of loci that have been characterized. This may be due to the mutation process itself or natural selection that sets a bound above which alleles are rarely observed, so that the allele size distribution eventually approaches equilibrium.

Microsatellites and the Y Chromosome

The study of microsatellites on the Y chromosome is gaining increasing importance for the study of paternal lineages in human evolution. Except for a small portion on the tips of the p and the q arm of the Y chromosome, the first and the second pseudoautosomal region (Freije *et.al.* 1992), the Y chromosome does not undergo recombination and thus persists from generation to generation as a haploid lineage.

Relative to the autosomes, the Non Recombining Region of the Y chromosome (NRY) has a smaller effective population size so that drift and founder effects are more enhanced with the Y chromosomal genes than on the autosomal genes. The spectrum of polymorphism found on the Y chromosome ranges from large scale insertions, deletions, duplications and rearrangements as well as micro and minisatellites. The level of polymorphism in the Y chromosome is high even in closely related populations. The estimated time of the most common ancestor calculated for Y chromosome (100,000-200,000 years Before Present (BP)) agrees with that of mtDNA (Gibbons 1997, Watson *et. al.* 1997). These features make the Y chromosome a sensitive index of population study and allow the application of powerful phylogenetic tools.

India as a Case Study

The subcontinent of India was chosen for the study of the paternal aspects of human evolution for a variety of reasons. Not only is the Indian subcontinent isolated from rest of the world by its unique physical features, but it also boasts of a culture and social system that is unique. Archeological and anthropological evidence shows that people from different ethnicity, cultural and linguistic background have made India their home from a very ancient time and have contributed to the present day gene pool of the subcontinent (Papiha 1996). Thus, it was expected that such a country as India would further contribute to our knowledge and understanding of human migration pattern. In order to appreciate India's place in prehistoric human migration patterns one must first have a better understanding of India's location and history.

Geographic Location of India

The Indian subcontinent is made up of India, Pakistan, Bangladesh, Nepal and 'Ceylon' or modern day Sri Lanka (Figure 1). To the Northwest of the subcontinent, lies the West Asian world that stretches from Persia to Egypt, and to the Northeast lies the East Asian World of Central Asia and China. Both have contributed to different extent to the history of the subcontinent. Topogeograpically the subcontinent is isolated from the rest of the world by the towering Himalayas that merge into the Hindukush, Sulaiman and Kirthar ranges in the West, the great Himalayas in the North, and the Khasi, Jaintia and Chin Hills in the East. The hilly region in the East is covered with dense tropical forests that presently isolate India from East Asia. The Western mountain ranges are transversed by a few harsh and inaccessible passes like the Khyber and the Bolan. These passes have served as passageways into India from very ancient times. The southern region of the subcontinent is surrounded by the Arabian Sea in the West, the Bay of Bengal in the East and the Indian Ocean in the South. The fertile Indo-Gangetic plains spreads about 1900 miles beginning at the Indus Delta which is drained by the River Indus from which India gets its name. To the South and West of the Himalayas lies the Gangetic plain that is drained by the River Ganga and its innumerable tributaries, and is the perennial gift of its snow and ice. This fertile plain has cradled the Dravidian civilization. South of this fertile plain lies the rugged Deccan plateau that stretches into the Oceans (Figure 2).

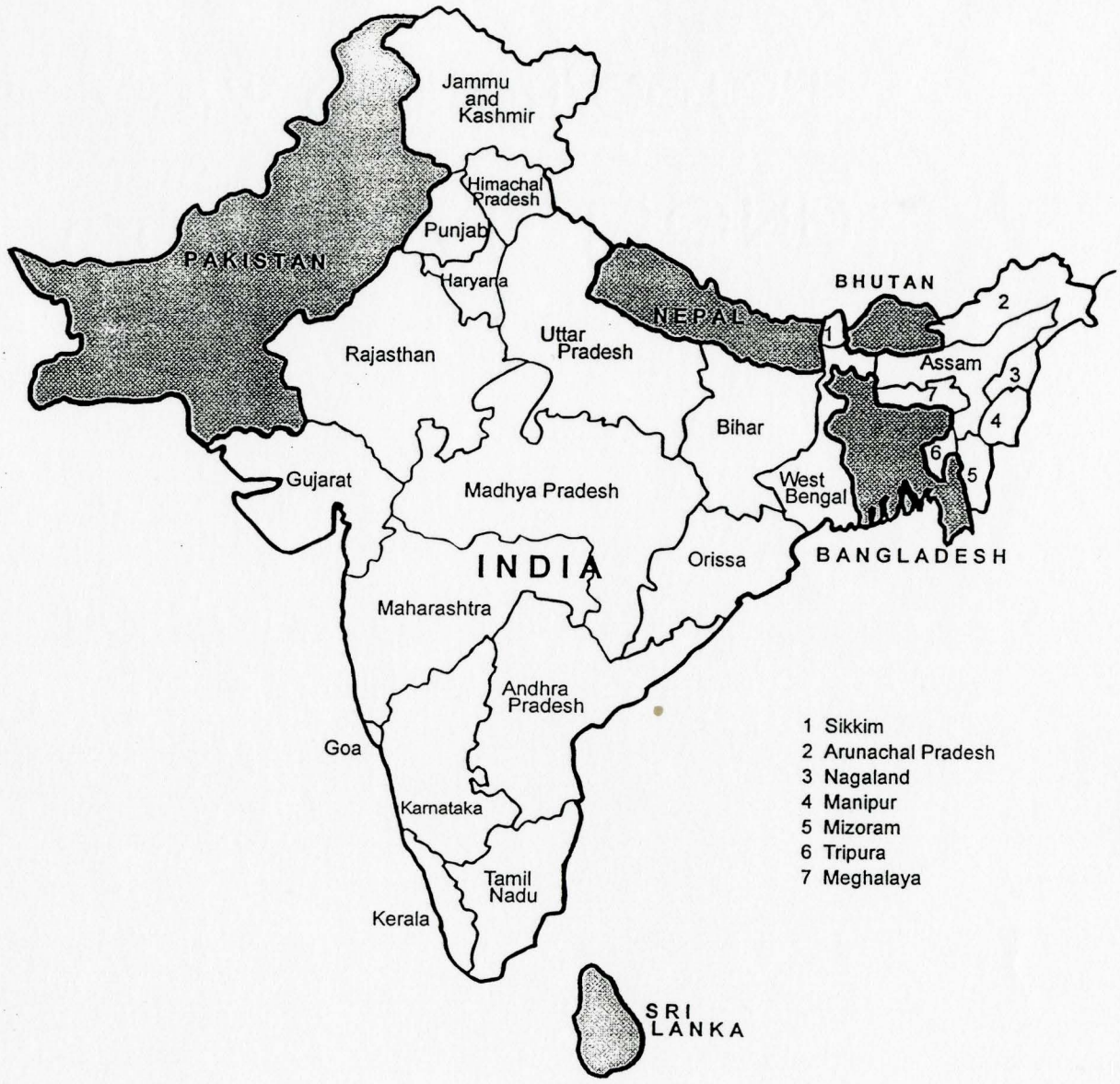


Figure 1: The Indian Subcontinent.

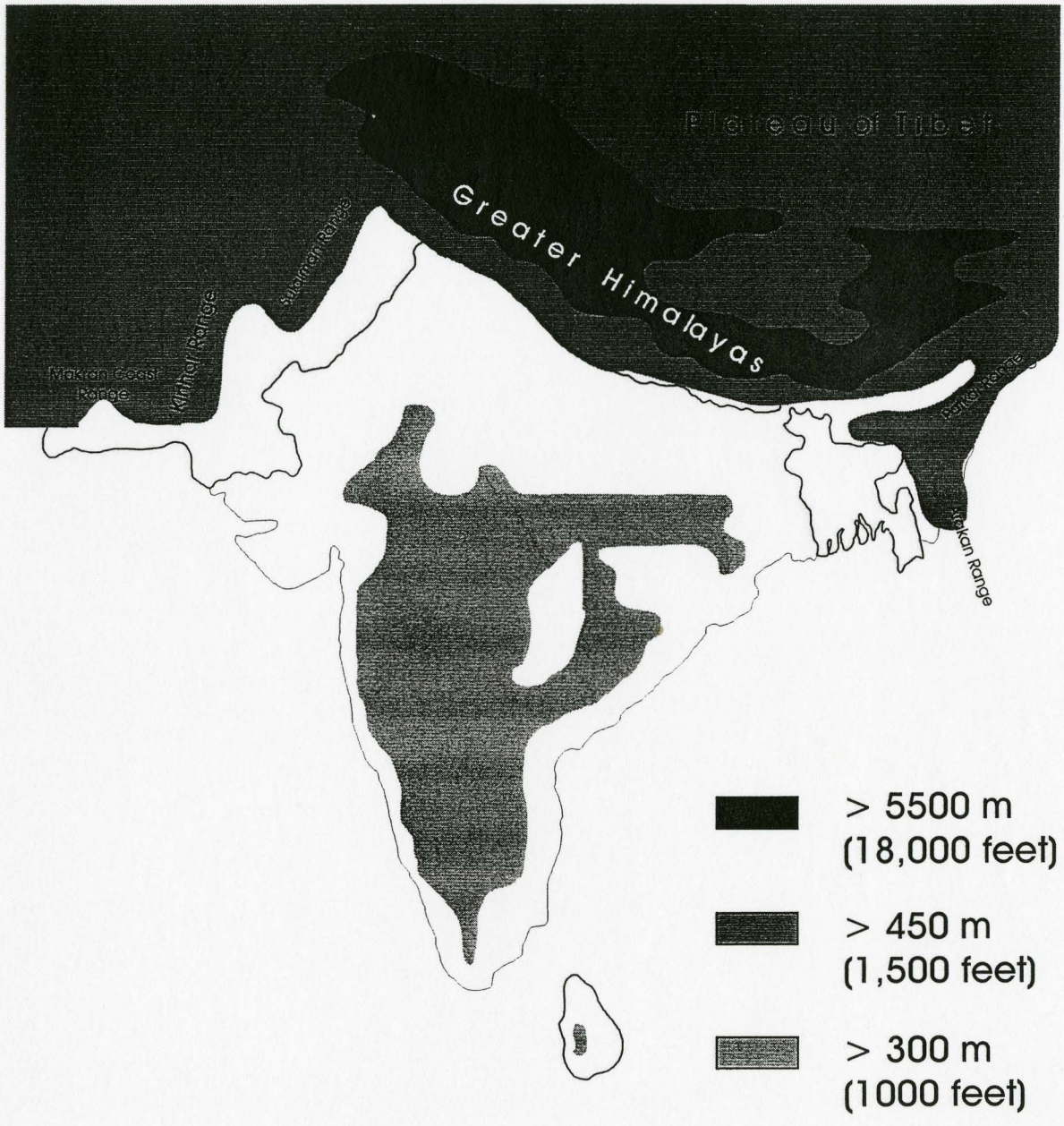


Figure 2: Physical Features of the Indian Subcontinent.

Human Habitation in India

Prehistoric evidence of hominids presences, as recognized from artifacts, is found in all regions of India of diverse environmental conditions. The lower Paleolithic sites dates from 500,000 to 50,000 years BP (Ghosh 1989). The first fossilized remains of a human skull to be discovered in India were found in the middle Pleistocene deposits of the Hathnora region of Narmada valley. It is 0.2- 0.7 million years old. This skull or the 'Narmada Man' as it is commonly called, has been assigned as an 'Evolved *Homo erectus*' by Lumley and Sonakia (1985) or as an ancient form of *H. sapiens* by Kennedy *et.al.* (1991) and Sankhyan (1997). Even though other, older skeletons have been found in India of Pleistocene antiquity, all but for a few have been destroyed (Kennedy 1980) and the few that remain have never been studied to determine their place in history (Kennedy 1980). Thus, it is yet unknown when *H. sapiens* first habituated this subcontinent. The first clear evidence of the presences of 'anatomically modern *H. sapiens*' that inhabited the subcontinent comes from fossilized skeletal remains that are about 12,000 years old. These remains have been found in the upper Pleistocene layer, in a cave at Batadomba Lena of the Sabaragamuva Province of Sri Lanka (Kennedy *et. al* 1987, Kennedy and Deraniyagala 1989).

Even though a large number of ancient relics as well as fossilized skeletal remains of ancient man have been found in the subcontinent, little is known about his way of living. This uncertain and dim activities of man is for the first time replaced by the first historical evidence that gives one a glimpse of his life style in the elaborate remains of the two cities of Mahenjodaro and Harappa. The richness of archeological remains in

these two cities are found, in the elaborate street-planning ability of its people, its drainage system, its small palatial buildings made of well-burnt bricks, public bathing facilities, granaries for storage of crops and innumerable seals inscribed with pictographic signs of animals. These remains provide a glimpse of how these early inhabitants lived. Although the first writing of Indian civilization is yet undeciphered, the archeological remains point to a highly sophisticated and technologically advanced group of people. These people were among the first to spin cotton into yarn and weave it into cloth. This civilization that flourished along the River Indus and ranged from the Simla Hills in the East to the Suktagen–dor (Baluchistan coast) in the West and from Dabarkot in the North to the Tapi and Narmada rivers in the South (Gokhale 1982), has come to be known in history as the Indus Valley Civilization (ca. 2500-1600 B.C.) and was contemporary to other world civilizations like those of the Nile and the Tigris-Euphrates valleys. It is not known how this civilization came to an end although the popular believe is that a series of floods caused by tectonic earth movements brought about its end (Wolpert 1989).

Advent of Aryans and the Emergence of Hinduism

It seems that the first group of the original Indo-European speaking, seminormadic tribes known to history as Aryans first entered India about 1500 B.C. These people originally lived between the Caspian and the Black seas, and were probably driven from their original homeland by some calamity. Because archeological evidence is scare, most of our knowledge of India's Aryan age between 1500 to 1000 B.C. comes from religious books such as the 'Vedas' or the 'Book of Knowledge', the oldest of

which is the Rig Veda. Although the Rig Veda, does not give any account of the journey of the Aryans from their original homeland to India, it does mention how the fair skinned Aryans (the term Aryan is synonymous with 'noble' or 'highborn', but now basically denotes a linguistic family) were victorious over the dark-skinned people or the 'dasas'. The superiority of the Aryans over their conquered tribes was in their advanced weapons, possible because of their knowledge of Iron Age Technology, and horses. The Rig Veda was composed during the era when the Aryans occupied the 'Land of the Seven Rivers' or 'Sapta Sindhava' consisting the region between the Indus Valley and Punjab (Wolpert 1989). The essence of Hinduism or the 'Aryan Way of Life' was already being laid out during this time. The Aryan conquest was marked by a gradual assimilation and sociocultural integration between the conquering tribe and their more civilized "pre-Aryan slaves". This is evident in the fact that although the Aryans worshipped nature gods like 'Indra' the 'Lord of Water', 'Agni' the 'Lord of Fire', etc., worship of gods of non-Aryan origin like 'Siva' (whose worship was prevalent in the Indus Valley Civilization prior to the advent of Aryans) was also predominant (Moreland and Chatterjee 1960). This ability of Hinduism to integrate different aspects and rituals of other religions and cults into itself might have been a factor in absorbing the Dravidians (denotes a linguistic group but loosely refers to the tribe, the 'Dasyus' or dark skinned people that the Aryan conquered and pushed further South) within its folds. Thus Hinduism must be regarded as both Aryan and Dravidian but little is known about Dravidian practice and rituals at the time the Rig Vedas were written (Moreland and Chaterjee 1960). Hinduism might have been accepted in South India in later times (first

century B.C.), when there were many powerful Hindu kings that ruled South India. They might have helped in the spread and acceptance of Hinduism in South India. The invasion of the Aryans was thus, a very important event in Indian history as they brought with them not only their Caucasian genes but also a new language – Sanskrit, and profoundly influenced the history of the country (Moreland and Chaterjee 1960, Wolpert 1989).

Foreign Invasions and its Impact

The next foreign invasion of India came during the rule of an Aryan king, Chandragupta Maurya. The Greeks under Alexander (326 B.C.) marched from the Northwest side of India conquering the land as they passed. Although his army did not move further east of Punjab, the impact of the Greek conquest was felt for many generations. The Greeks continued to control India west of Punjab for a long time. A number of foreign powers from Central Asia also invaded India. Of these the Scythians and the Pallavas were the first to invade around 50 B.C. They ruled Gujarat and the surrounding regions. Next were the Kushans, invading India in the middle of the first century A.D. Their empire comprised of Bacteria in Asia, and included Punjab, Sind, Kashmir, parts of Malwa and Gujarat and areas west of Gangetic plains and may also have included Bihar, Orissa and Bengal. After the collapse of the Kushans not much is known about the political history of India. The next phase of foreign invaders that conquered India and controlled its political history, (came during the rule of the great Gupta emperor Samudragupta) were the Huns (200 A.D.), a tribe from Central Asia (Wolpert 1989).

The fate of these foreign invaders who stayed and established their rule in India is unknown. These once-foreign invaders may have been absorbed into Hindu caste though nothing more can be said about which caste they represent after more than two millennia. A plausible example might be the Gujars, a tribe from western India. In Sanskrit literature, a tribe name Gujara is associated with the Huns or may have entered India at the time when Huns invaded India. In the eight century A.D., there were various Gurjara kingdoms in the north and the west of India. Gurjara is synonymous with Gujar and the country that were formerly held by some of these tribes, still bear the name Gujarat. A present day Gujara caste is also present in Punjab and Bombay. Gujars are Indians and some of them are Hindus and some have been converted to Islam. It may be tempting to speculate that this is a caste originating from the fifth century as an invading tribe but in the absence of any definite scientific conclusion on the foreign origin of this caste nothing can be definitely said (Moreland and Chatterjee 1960).

Muslim Invasion and the Impact of Islam on India

The next great impact on the Indian history and culture came with the Muslim invasion of India. Turks from Afghanistan under Mahmud of Ghazni at around 997 A.D. conquered India. It was not until 1175 A.D. that the Afghans, under Mohammad of Ghur and his general, establish their rule in India. Thus the Delhi Sultanate that ruled the whole Subcontinent for the next few centuries was established. In the year 1526 A.D. India was once again attacked. This time it was the Mughals, also from Afghanistan. For the next two centuries, the whole of India came under the Mughal rule. This Islamic invasion had

a profound influence on India's history, its people and its culture. It introduced a new religion and culture to India. There was mass conversion of Hindus into Islam, forcefully or otherwise. Inter-religion marriages between the two religions were also encouraged in order to convert the common people into Islam. This had the effect of diluting the contribution of the original Muslim invader to the gene pool of the people of the subcontinent (Papiha 1996).

Christianity in India

It is uncertain when Christianity was first introduced in the land. Although claims are made that Christianity was introduced in India around first century A.D. by Apostle Thomas, little can be said with certainty. The second line of evidence suggesting the origin of Christianity and Christian Churches comes from the works of a Greek traveler, Indicopleustes who traveled India between 525 to 530 A.D. He reported the presence of Christians in South India and Sri Lanka (Neill 1970). Early Christians were probably Persian merchants who came to India to trade spices. This group did not actively convert Indians into Christianity but remained as a secluded community. Although from time to time different missionaries came to India, Christianity did not gain a stronghold in India until 1498 A.D. when, a small group of Portuguese merchants led by Vasco da Gama came to India. With the arrival of the Portuguese in South India and their control over some major ports in South India, Christianity in India entered a new phase. Efforts were made to translate the Gospel into native languages and people were converted to Christianity in exchange for protection and jobs. With the decay of Moghul power in

India at the end of eighteen century, a new power arose in the political scene of India. The English, who first came to India as merchants, slowly began to gain control over the political situation of India. With the victory of Robert Clive, the then viceroy of India from England in 1757, over the Muslim ruler in eastern India, the East Indian Company representing the Queen of England established their rule in India. With a Christian ruler at the political head of India, India saw a steady increase in the number of converts adopting the new faith. At present Christians of India forms the second largest minority community after Muslims with a population of twenty million representing 2.5% of the total population (Census of India 1991 excluding state of Jammu and Kashmir where census was not conducted due to political tension).

Birth of Other Religious Sects in India

Many new religions were established by different religious teachers as a revolt against traditional polytheism and elaborate rituals of Hinduism. Jainism was founded by Mahavira. Buddhism was founded by Siddhartha Gautama who later came to be known as Buddha. Both teachers came from the Kshatriya caste. Both preached a life of nonviolence or 'ahimsa'. But with time, Buddhism (560-490 B.C.) and Jainism (599-527 B.C.) were absorbed into Hinduism (websites-The major world religions, Introduction to Jainism).

With the impact of Islam in India in the 11th century, a new religious order was born that was the fusion of Hindu and Islamic laws. Nanak, the founder of Sikhism was born in a Hindu family of Punjab. Since converts to Sikhism were mostly from the

Hindus not much genetic variation is found between the Hindu and the Sikh population (R Singh Per. comm.). At present, the Sikhs represent about 2.34 % of the total population (Census Of India, 1991) with the majority living in Punjab.

Tribal Populations of India

The tribal populations (also called scheduled tribes) of India are scattered throughout India and represent about 8.8% (Census of India, 1991) of the total Indian population. At present there are about 427 tribes in India (Cavalli–Sforza *et. al.* 1994). It is thought that some of the tribes descended from the original inhabitants of the subcontinent while others are more recent immigrants. Even the same tribe residing in different regions has different gene frequencies due to varying level of admixture with the local population (Cavalli–Sforza *et. al.* 1994). Some of the tribes also belong to different linguistic groups and do not fall within the prescribed caste system as described in Hinduism. The tribes were generally nomadic and hunters but at present many are involved with agriculture (Cavalli–Sforza 1994). Genetic studies shows tribes are more closely related to one another, followed by non-tribal Indian populations (Cavalli-Sforza 1994). Study using blood markers have shown that the tribes do not have an African or Australian Aboriginals origin (Roychoudhury 1984) although appearance of some tribe bears close resemblance to these populations. Another study points to the fact that some tribes share a common origin (Labie *et. al.* 1989) even when large geographic distance separates them.

Indian Caste System

The concept of caste might have originated when the fair-skinned Aryans conquered the dark skinned 'dasas'. With further conquest and penetration of the Aryans into the India territories, the caste system may have become more and more complex. The original Vedic concept of the caste system was based on the principle of livelihood or the four 'varnas' (meaning color or class). It was a unique system of social organization where people of a particular caste performed a particular job in order to ensure the proper running of the system. The four main Varnas of ancient India were the brahmans (the priests who looked after the spiritual lives of the people and performed rituals for their good), the kshatriyas (the warriors whose primary responsibility was to ensure the proper administration of the people and to protect them from invaders), the vaishyas (the merchants), and the shudras or the 'dasas' (the slaves). The latter might have performed tasks that were considered unclean such as the work of tanners, associated with animal carcasses, and sweepers especially among the ashes of cremation grounds. Later, a fifth class emerged of those of the untouchable or 'panchamas' and their habits or occupations were so unclean that even the sudra did not touch them. Marriages were endogamous within each caste but 'lineages' within a caste group were exogamous and based on "*gotras*"' (Bamshad *et.al.* 1996). '*Gotras*' are basically family names. These were taken by the families after a saint or teacher or 'guru' they followed. The concept probably originated when brahmins and the kshatriyas went to stay with a teacher as students at his residence for their education (the four stages of life) (R. Singh per. comm). Intercaste marriages under certain circumstances were allowed where women could marry men of

higher caste but marriage of women from a higher caste to lower caste men was discouraged. The son of a marriage between a high caste man with a lower caste woman was allowed to retain the caste of his father but the son from a higher caste mother with the lower caste father may belong to a different caste or became untouchable (Tambia 1973).

Population Structure as Revealed by Blood, Protein, and Serum Markers

Studies on blood, protein and serum markers have shown that some important features are present in the different caste groups. In some regions genetic distance between the different castes are great with considerable genetic distance existing between the highest and the lowest caste. If castes of different geographical regions are compared then within a given geographic area different castes are closer to each other as compared to caste belonging to different geographic regions.

Studies on blood, serum and protein markers have also shown that gene frequency clines are present from North to South India (Papiha 1996). Genetic similarities also are present between India and other world populations. In general Northwest and Northeast Indian population shows similarities with the European and other Southeast Asian populations respectively. 'Genetic variation is larger within caste and tribes than within populations (consisting of these caste and tribes)' (Bamshad *et. al.*1996). A few of the clines and genetic variation seen in India with blood, protein serum and DNA markers and their correlation with other world populations are listed below.

MN Ss blood groups show that M allele is found in higher frequency in schedule tribes. The Mongoloid populations also have a higher frequency than any other populations. The Ns haplotypes are higher in Northwest and Central States with some of the highest frequencies in Maharastra and Orissa (Papiha 1996).

Of the Rh system, the gene frequencies are highest in Tibeto-Chinese and Austro-Asiatic speakers and low in Indo-European speakers. The CDe haplotype of the Rh antigen system shows an increase from 50% in the north to 70% in the Mongoloid speakers in east India (Papiha 1996).

The *cde* haplotype shows a high frequency from North, Northwest and along the Western coast of India but is low in central and eastern India.

Of the FY blood system, FY*A allele frequency is less than 50% in Indian Austro–Asiatic and Indo-Aryan speakers but is more than 60% in other linguistic groups (Papiha 1996).

A study was conducted on 5 populations from Himachal Pradesh, four from the Kanet region and one from the Koli region of the extent of genetic variation in the immunoglobilin allotypes (GM, KM, and AM). The haplotype GM*A G showed a higher frequency in the Kanet (40-60%) population than the other populations under study, and are comparable to other Asian populations. The IG haplotypic data suggests different origin for the Kanet and the northeastern populations of Nepal from populations of central India represented by samples from Delhi (Papiha 1996).

Studies on complement component of five loci C2, C3, C4A, C4b and BF of the Hindu population of Andhra Pradesh and Bengali speaking Muslim population of

Bangladesh have been done. The C3*F and BF*F alleles shows a range of frequency variation in different ethnic groups of India. This variation range of the C3*F allele is intermediate between European whites and Southeast Asians but the variation range of BF*F allele frequencies places Indians between European whites and African blacks. The frequency of C2*B among Indians are slightly higher than in other European populations (Ad'hiah and Papiha 1996).

Studies on human serum group-specific component (GC) (also known as Vitamin D binding protein) in 30 Indian populations showed that variant 1C11 and 1C21 previously reported in the French population and 1C36 reported for Nepalis were for the first time reported in an Indian sample. Higher frequency of GC*1F allele was observed in the eastern region of India differentiating ethnic groups from eastern India with other ethnic groups from rest of India. The frequency of GC2 has a positive correlation with latitude, with a clinal increase from South to North India (Mastana *et. al* 1996).

DNA Markers

mtDNA and Y chromosome loci are highly polymorphic and are thus more informative in revealing population structure. mtDNA also has a high mutation rate and lack of recombination (like the Y chromosome), and is maternally inherited. A few general conclusions can be drawn using these markers, 1. Genetic diversity is higher in African population compared to other world populations (Hammer 1997), 2. phylogenetic tree analysis using such data places the first split between African and non-African populations (Seislstad *et. al.* 1999), 3. In the case of India, clinal affinities are present

between North specially Northwest and South India but no such affinity is observed between Northeast and South India attesting to the fact that migration from Northeast frontier countries if it happened was minimal.

Studies were conducted for Y chromosomal and mtDNA variation in 250 individuals from 12 populations from Andhra Pradesh and South India to estimate the genetic distance among the populations. Individuals were classified as the upper (Brahmin, Kshatriya, and Vysya), middle (kapu, Yadav, Jahari and Wadabahija), and lower (Relli, Madiga, and Mala) castes. Of the 182 unique mtDNA haplotypes, four were shared between the upper, middle and lower caste, seven between lower and middle caste, and three between middle and upper caste. No haplotypes were shared exclusively between the upper and lower caste. The genetic distance between the upper and lower caste was reported to be 1.5 times greater than between upper and middle or middle and lower caste. Thus the genetic distance between highest and the lowest castes were greater than castes that are closer together. Thus, based on mtDNA distances, caste of different status are stratified according to social rank, and mtDNA haplotype points to a higher level of female gene flow between more closely ranked castes. From this it can be suggested that the genetic stratification of the Hindu caste system is driven by social mobility of women. In contrast to the above data, the genetic distance based on Y chromosomal loci is 0.0054 between upper and lower caste and is similar to the genetic distance between the upper and middle (0.0062) but is lower for middle and lower caste (0.0005). Thus no concrete conclusions can be obtained for the Y chromosomal variation (Bamshad *et. al.* 1998). In another study, haplotypes constructed with one biallelic and

five microsatellite markers for 125 Indian Y chromosome sample representing 10 ethnic groups gave 81 distinct haplotypes. Y chromosomal data in this case showed that there is very little gene flow across different castes and tribes (Bhattacharyya *et. al.* 1999).

Studies done on samples from Indian subcontinent as well as other world populations show Caucasoid as well as African influences on the Indians. A study on a group 36 Hindus from Andhra Pradesh, South India representing 4 castes, Brahmin (representing upper caste), Yadav (upper middle caste), Kapu (lower middle caste) and Relli (untouchables), and individuals from Africa, Asia and Europe using the HVS2 (hypervariable segment 2) region of mitochondria, showed that the genetic diversity is second highest for India (0.98) next to Africa (0.99). Of the Indian castes the highest diversity was found in Yadav (0.98) and the lowest in the Kapu (0.90). The diversity is intermediate in Brahmins (0.92) and the Relli (0.93). Nucleotide diversity is significantly higher in Africans (0.027) as well as in Indians (0.014) than that of Europeans. The Gst value of the three continents and the Indian Caste groups is 0.15 meaning 15% of the total variance of this region of the mitochondria occurs between these populations. The study concludes that, 1) there is significant geographical structuring among continental population with respect to mtDNA, 2) mtDNA variation is high within Indian castes groups, and 3) phylogenetic analysis using neighbor-joining network shows that Indians groups closely with Africans than with other populations (Bamshad *et. al* 1996). Studies on the presence of intergenic COII/tRNALys 9-bp deletion in human mtDNA have shown that 1) individuals from different castes are closer to the Caucasoid populations, and 2) the tribal population of India with the deletion is most closely related to tribal

population without the deletion than to any other caste group (Gibbons 1998, Watkins *et. al.* 1999).

Studies done on mtDNA RFLPs on samples from the India and Europe showed a marked Caucasoid influence with some amount of Asian admixture on the people of India. There is sharing of mtDNA haplotype between the Indian and Caucasoid populations with the northern Indian population showing a more recent admixture of the Caucasoid mtDNA type (Semino *et. al.* 1991, Barnabas *et. al* 1996). Another study on the Indian population using the D-Loop region of mtDNA shows lineages that predate the divergence of Euroasian population. The three groups examined in this study are the Brahmin, the untouchables or the 'Harijan' and a tribal population. Pairwise nucleotide distributions in this study also suggests to a faster growth rate for the Brahmin population, with the 'Harijan' maintaining a constant growth rate (Mountain *et. al* 1995).

Another study on 200 samples from India belonging to various ethnic and linguistic groups was conducted. It shows that some haplotypes are unique to India, some are shared between India, Africa and Europe, while some are shared between India and Africa. But no such sharing was evident between India and other Southeast Asian countries. On the basis of these results these researchers proposes two hypothesis that, 1) India might have served as a route for migration from 'Out Of Africa' to Southeast Asia, with no interaction of Indian and Southeast Asian populations afterwards resulting in haplotype divergence between the two populations (Southern Route as suggested by Cavalli-Sforza 1988). A study on Y chromosome biallelic markers (Su *et.al.* 1999) indicates that these new migrants might have first settled in Eastern Asia followed by a

more easterly and northernly migration into Asia. 2) Haplotype sharing between Africa, Europe and India might have been due to migration from West Asia, with no influence from East Asia and so are not related to East Asians. (Barnabas *et. al.* 2000).

A study on the mtDNA RFLPS markers (*Hpa*I, *Bam*HI, *Hae*II, *Msp*I, *Ava*II, *Hinc*IIin and sites *Alu*$I_{7,025}$, *Dde*$I_{10,394}$, and *Alu*$I_{10,397}$) on samples from India, Europe and Africa showed European and African influences on India. Common haplotypes are present in Caucasians and Northwest Indian population. A frequency cline decrease, for the presence of the sites *Dde*$I_{10,394}$, and *Alu*$I_{10,397}$ (+ +), was seen from South to North India. The presence of these sites defines the predominant East Asian specific haplogroup M. This haplogroup is absent in the Caucasians. Of all the African populations, this haplogroup is present in high frequencies in the Ethiopian population (Passarino *et. al.* 1998). The Caucasian specific haplotype *Dde*$I_{10,394}$, and *Alu*$I_{10,397}$ (--,+ -) shows a frequency cline decrease from North to South India. Thus, these results show 1. Caucasoid influence is most predominant in North India, 2. pre-caucasoid features are also present in the Indian population (Passarino *et. al.* 1996) that might possibly represent Ethiopian influence (Passarino *et.al.* 1998). The latter conclusion supports the hypothesis that the migration of 'anatomically modern humans' from Africa to Southeast Asia might have taken place through Ethiopia, the middle east and through India (Passarino *et. al.* 1998, Santachiara–Benerecetti 1999, Cavalli-Sforza *et. al.* 1993) and not through North Africa, as has been suggested by archeological findings (Lahr and Foley 1994).

On the other hand, European influence on the Indian Subcontinent might be older than was previously thought and may not be the result of a single massive Aryan invasion

(Kivisild *et.al.* 1999). Kivisild *et.al.* 1999 sequenced mitochondrial hypervariable region I in Dravidians and Indo-European (Indo-Aryan Language speakers) samples. According to this group, the East Asian specific haplogroups are most frequent in India followed by Caucasian specific haplogroups. The frequency of the Caucasian specific haplogroup does not show any North to South, or language based gradient among the Indian populations. Of the Caucasian specific haplogroup, haplogroup U is most frequent in India. The subcluster U2 (of the U haplogroup) is found in both India and Europe. The time of divergence between Indian and West Euroasian U2 lineages is estimated to be around 53,000+/- 4,000 yr BP. The coalescence age of U2, U5 (the most frequent and ancient subcluster of U) and M (East Asian specific haplogroup) is estimated to be identical. Inspite of the fact that these haplogroups have identical coalescence age, U haplogroup is not found in the Orientals and M haplogroup is not found in the Caucasians. Taken together these result may point to the fact that two separate late Pleistocene human migration might have brought in Modern Man into India. The first migration may have carried the M haplogroup into the subcontinent. From here Modern Man might have spread further East and Southeast. The second migration may have introduced the haplogroup U into India.

Unpublished results by Thomson and Singh (2000) in our own laboratory using mtDNA RFLP haplotypes scattered throughout the coding regions in 187 Indian samples was recently completed. Of these 44.9% have haplotypes most common in Europe. The Northwest Indian population is significantly different from the rest of Indian populations and has more European characteristic haplotypes than the Northcentral, Northeast and

South India. 50.8% of the samples show haplotypes found in India that are similar to the Siberian populations and 2.7% of the populations have haplotypes unique to India. African haplotypes present in India population are rare. Central Asian populations were also shown to share haplotypes with Europe, India and Siberia. Phylogenetic analysis shows that Indians are more similar to Siberians than to other East and Southeast Asian groups. This finding also supports the linguistic similarities that exist between the Dravidians and the Uralic-Altaic language family of the world (Zvelebil 1990) and especially the linguistic similarity that exist with the Elam population from Southwest Iran (Renfrew 1989).

Objective of the Present Study

Phylogenetic studies done on Indian and other different world population, as well as studies in our own laboratory using mtDNA point to the fact that similarities exist between India and other European and Central Asian populations. This study was conducted to investigate the following points,

1. To see whether the Y Chromosome microsatellites shows any significant similarities exist between Indian and other Caucasian and/or Central Asian populations.
2. To verify whether the NW Indian population differs significantly from the rest of the Indian population. Study in our own laboratory shows that Northwest population of India is significantly different from the rest of Indian population (Thomson and Singh 2000, unpublished results).

3. To determine the presence of any heterogeneity in the distribution of different microsatellite allele frequency in the four regions.
4. To determine how diverse are the Indian populations, the genetic diversity of Indian populations will be compared to other world populations. Previous studies (Jorde 2000) on autosomal, Y chromosomal microsatellites, mitochondrial hypervariable regions 1 and 2, and for *Alu* polymorphism has shown that genetic diversity is highest for the African populations.
5. To examine whether there exist any significant differences between different religions, caste and linguistic groups.

MATERIALS AND METHODS

Microsatellites and Biallelic Markers Used in this Study

The Y chromosome microsatellite loci selected for this study are two trinucleotides DYS 388 and DYS 392, and four tetranucleotides DYS 389, DYS 390, DYS 391, and DYS 395 and a biallelic marker DYS 271 (Table 1). The DYS 389 locus consists of two regions with two stretches of different number of TCTG/TA repeat units (Rolf *et. al.* 1998). The data from this study will be used in conjuncture with the data of the already typed DYS 19 and DYS 287 (presence or absence of Alu insertion – YAP^+/YAP^-) locus, as well as results obtained in the study of other World populations (table 5).

Samples

Blood samples were obtained from immigrants from the Indian subcontinent or from first generation Canadians of Indian descent living in Southern Ontario. They were asked to fill out information on consent forms such as name, gender, place of birth, parents' place of birth, native language, and religion (optional) (Table 2,3, and 4, and in Figure 3 and 4). Caste was assigned to these individuals based on their surnames by knowledgeable individuals. Qualified technicians drew out blood in 5 or 7.5ml Vaccutainers (containing 0.05 ml of 15% EDTA solution) and were stored in 1ml aliquots at -80^0 C.

Table 1: Characteristic of Y Chromosomal Microsatellite and Biallelic Markers

Locus	Repeat Motif	Number of Repeats	Size Range of alleles (bp)
DYS 388	(ATA)n	11-15	126-138
DYS 389I	(CTG/AT)n	7-13	239-263
DYS 389II	(CTG/AT)n	23-31	353-385
DYS 390	(CTG/AT)n	18-27	191-227
DYS 391	(CTAT)n	8-13	275-295
DYS 392	(ATT)n	7-16	236-263
DYS 393	(GATA)n	9-15	108-132

Data from de Kniff *et. al.* 1997, Seielstad *et.al.* 1994.

Table 2: Distribution of Samples within the Indian Subcontinent

Region and State/ Country	No. of samples
NORTH	
North-West	
PAKISTAN	7
JAMMU & KASHMIR	2
HIMACHAL PRADESH	1
HARIYANA	1
PUNJAB	16
GUJRAT	3
	30
North-Central	
NEPAL	1
UTTAR PRADESH	37
BIHAR	1
MAHARASTRA	3
	42
North-East	
WEST BENGAL	13
BANGLADESH	5
MEGHALAYA	1
	19
SOUTH	
ANDRA PRADESH	2
KARANATAKA	2
KERALA	7
TAMIL NADU	13
	24
UNKNOWN LOCATION	1
Total Samples	**116**

Table 3: Distribution of samples in the Indian Subcontinent according to religion and caste

Religion/Caste	No. of samples
HINDUS:	73
BRAHMINS	24
KSHATRIYA	16
VAISHYA	12
UNKNOWN	21
	73
SIKHS	19
MUSLIM	12
CHRISTIAN	10
JAIN	1
UNKNOWN	1
Total Samples	**116**

Table 4: Linguistic ancestry of the samples

Language	No. of samples
Dravidian	24
Indo-European	90
Unknown	2
Total Samples	**116**

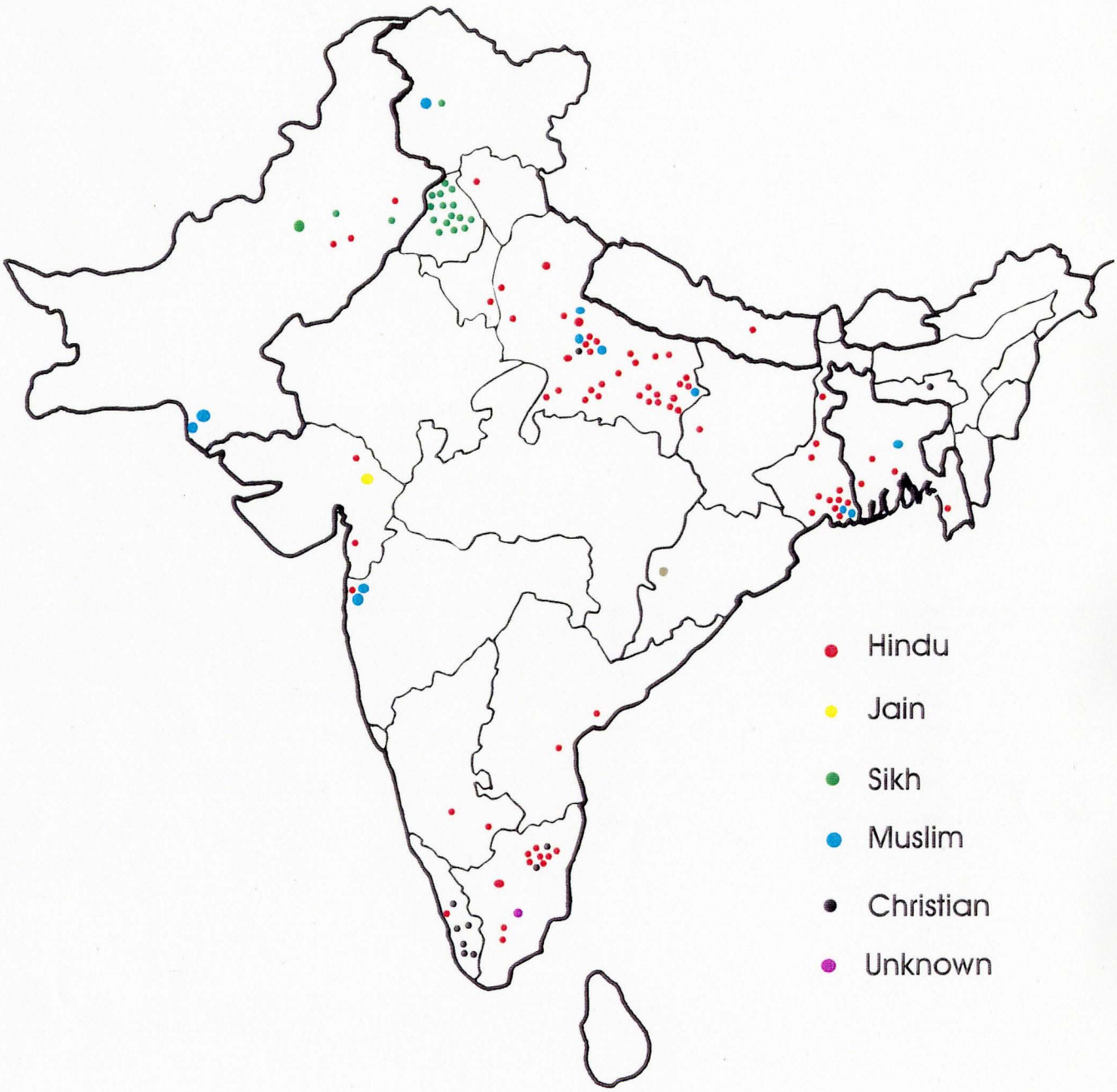


Figure 3: Distribution of Samples by their Religions.

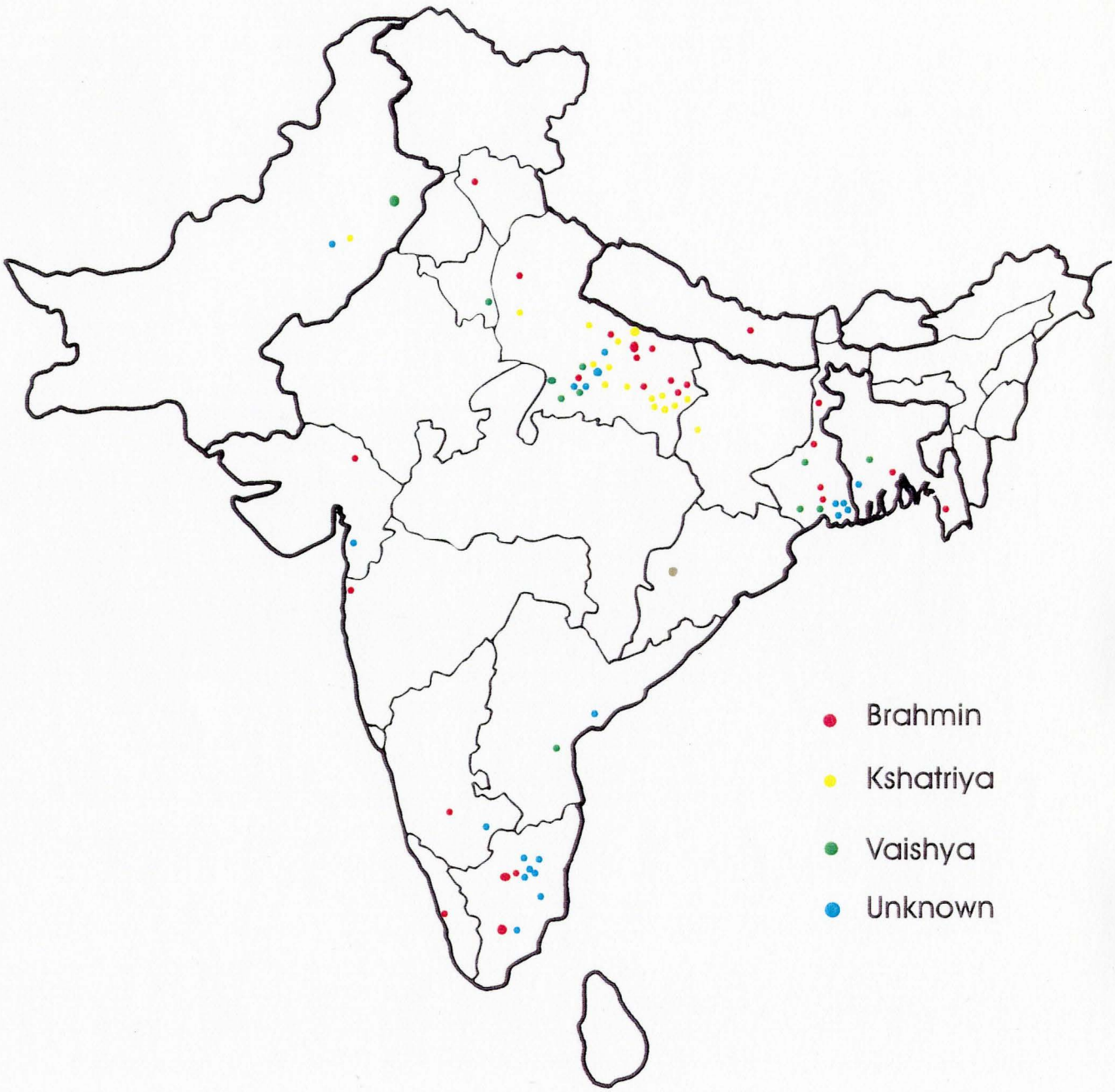


Figure 4: Distribution of Hindu Samples by their Caste.

Table 5: The Different World Populations used for this Study

POPULATIONS	SAMPLE SIZE	PAPERS
AFRICA		
Arabs from NW Africa	45	Bosch *et.al.* 1999
CAR Pygmy	20	Prichard *et.al.* 1999
Majangir	10	same as above
Moroccan Berbs	12	Bosch *et.al.* 1999
Mzabites	68	Bosch *et.al.* 1999
Nyangatom	11	Prichard *et.al.* 1999
Omega San	15	Same as above
Saharawis	29	Same as above
Sekele San	14	Same as above
Sotho	17	Same as above
Surma	11	Same as above
Tahelhits	44	Same as above
Tswana	12	Same as above
Zulu	15	Same as above
EUROPE		
Basques	51	Perez- Lezaun *et.al.* 1997
Catalans	28	Perez- Lezaun *et.al.* 1997
Finns	280	Kittle *et.al.* 1998
ASIA		
Central Asia		
Highland Kirgiz	43	Perez- Lezaun *et.al.* 1999
Kazakhs	49	Perez- Lezaun *et.al.* 1999
Lowland Kirgiz	41	Perez- Lezaun *et.al.* 1999
Uighurs	39	Perez- Lezaun *et.al.* 1999

POPULATIONS	SAMPLE SIZE	PAPERS
West Asia		
Burushaski	24	Prichard *et.al.* 1999
Sindhi	15	Prichard *et.al.* 1999
East Asia		
Cambodia	16	Prichard *et.al.* 1999
Japan	10	Prichard *et.al* 1999
Taiwan	15	Prichard *et.al* 1999
OCEANIA		
New Guinea	12	Prichard *et.al.* 1999
NATIVE AMERICANS		
Ingano	9	Prichard *et.al.* 1999
Ticuna	24	Prichard *et.al.* 1999
Wayuu	12	Prichard *et.al.* 1999

Samples Preparation

DNA Extraction

Genomic DNA was extracted using Purgene Genomic DNA isolation kit (Gentra). The extracted DNA was amplified using the polymerase chain reaction using suitable primers (Table 6). T_m was calculated in all cases by the method of Breslauer *et. al.* 1986. PCR amplification (Appendix B) was carried out in thin walled GeneAmp^{TM} (Perkin Elmer) reaction tubes. 4 µl of amplified products were mixed with 4 µl of 5X loading solution and were run in a 2% agarose gel containing 1 µg of ethidium bromide and visualized under ultraviolet light to assess the amount of PCR product. The buffer used for agarose gel electrophoresis was 1X TBE. The DNA fragments were sized as well as the DNA concentration was estimated by comparison with a 100 bp ladder (GIBCO-BRL).

Polyacrylamide Gel Electrophoresis

The size of each microsatellite allele was determined in a 6% polyacrylamide gel containing 8M urea (Kayer *et. al.* 1997). Depending on the amount of the amplified products as assessed by the agarose gels, the PCR samples were diluted upto 100 times with 50 mM EDTA (pH 8.0). 4µl of these diluted samples were mixed with 3µl of the loading solution. The samples were denatured by heating at 95^0C for 5 minutes before loading. From this, 3 µl were than loaded onto the wells of the polyacrylamide gel and

Table 6: PCR Primers

PRIMER	SYNTHESIS NEMBER AND SEQUENCE (5′ – 3′)[1]	MOLECU-LAR WEIGHT	MP (^{O}C)[2]
DYS 388	AB19111 GTGAGTTAGCCGTTTAGCGA	6182	**57.9**
	AB19112 CAGATCGCAACCACTGCG	5450.6	**62.5**
DYS 389	AB19994 CCAACTCTCATCTGTATTATCTAT	7212.8	**51.0**
	AB19995 TCTTATCTCCACCCACCAGA	5957	**57.8**
DYS 390	AB16783 TATATTTTACACATTTTTGGGCC	7075	**50.2**
	AB16784 TGACAGTAAAATGAACACATTTGC	7145	**49.7**
DYS 391	AB16785 CTATTCATTCAATCATACACCCA	6967	**48.8**
	AB16786 GATTCTTTGTGGTGGGTCTG	6281	**49.1**
DYS 392	AB12693 TCATTAATCTAGCTTTTAAAAACAA	7607	**53.8**
	AB12694 AGACCCAGTTGATGCAATGT	6137	**57.7**
DYS 393	AB12695 GTGGTCTTCTACTTGTGTCAATAC	7320.8	**54.4**
	AB12696 CTCAAGTCCCAAAAATGAGG	6116	**56.9**
DYS 271	AB19992 AGGCACTGGTCAGAATGAAG	6209	**57.6**
	AB19993 AATGGAAAATACAGCTCC	6077	**57.7**
T3 PROMOTER	AB15454 ATTAACCCTCACTAAAGGGA	6174	**51.9**

[1]The sequence for all primer except T3 promoter is from de Kniff *et. al.* 1997. The sequence for the T3 promoter primer is from Stratagene. [2]As predicted by Tm Determination.

separation was achieved with 1X TBE buffer. In addition to the samples, the product of the sequencing reaction of the pBluescript XL1 vector (Stratagene) was also loaded and run. This sequencing ladder was used to determine the allelic size of the microsatellite. The gels were run on the Life Technologies Inc. Model S2 sequencing apparatus at 3000 volts, 45 watts and 115 mAmps. The gels were run for approximately 3-6 hrs depending on the size of the markers or until the xylene cyanol had crossed the middle of the gel. Finally the gels were stained using silver nitrate. A permanent record of the gels were kept using the Automatic Processor Compatible (APC) films. The films were exposed for 2-3 minutes depending on the background of the gels and were then developed (Figure 5a and b).

DYS 271

The biallelic marker DYS 271 is an A to G polymorphism on Y chromosome. This transition occurs at position 168. 'G' allele digest produces bands at 144 and 65 bps, whereas 'A' allele digest produces bands at 102, 65 and 42 bps. The restriction site sequence of the locus is given below (restriction recognition site is in lower case and the polymorphic site is indicated by lower case and is italicized, primers are bolded),

5'**AGGCACTGGTCAGAATGAA**GTGAATGGCACACAGGACAAGTCCAGACCCAGGA
AGGTCCAGTAAcatgGGAGAAGAACGGAAGGAGTTCTAAAATTCAGGGCTCCCTTGGGC
TCCCCTGTTTAAAAATGTAGGTTTTATTATTATATTTCATTGTTAACAAAAGTCcatgA
GATCTGTGGAGGATAAAG**GGGGAGCTGTATTTTCCATT** 3'

The restriction site of the amplified product was cut using *Nla*III restriction enzyme. To the reaction mixture 1 unit of enzyme was added and was digested at 37^0C for more than 2 hours. 20µl of the reaction solution was mixed with 5 µl loading buffer and a 100 bp ladder was run in a 4% Agarose gel for 3-4 hours. The gels were ethilium bromide stained and visualized under UV light (Seiestad *et. al.* 1994).

Data Analysis

Data from this study, as well as other published data, which included the same sites, were used for haplotype and allele frequency distribution analysis. India was separated into four regions (Northcentral India, Northwest India, Northeast India and South India for the purpose of some analyses)(Figure 6).

Haplotype Analysis

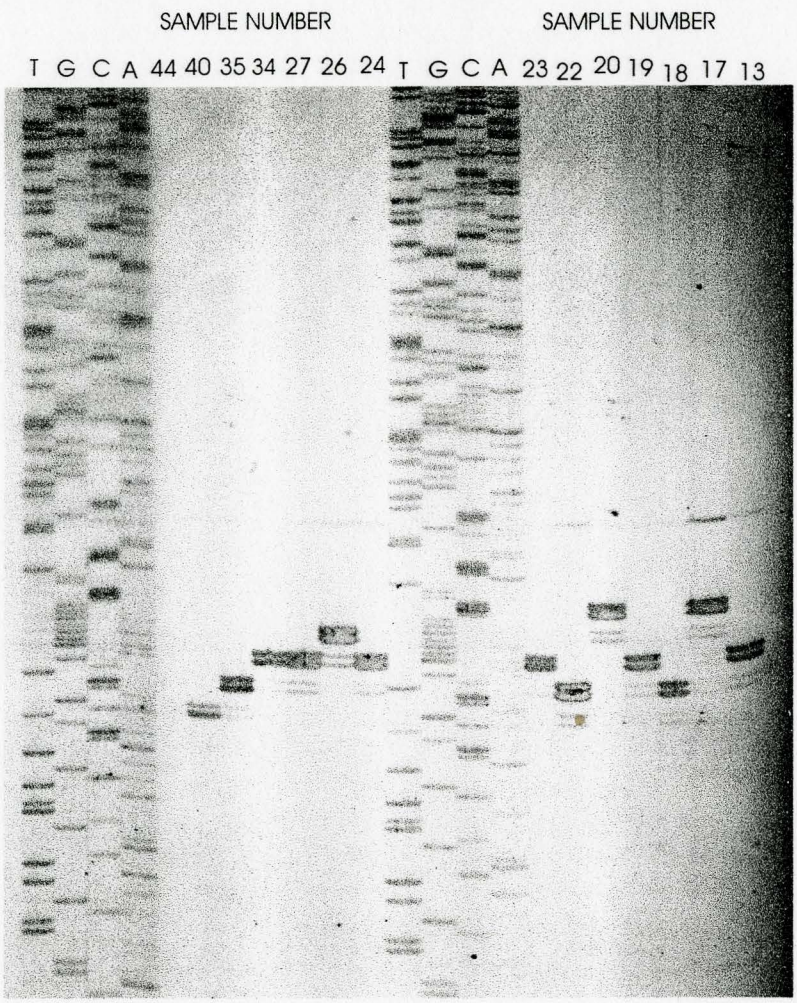
The data was sorted into haplotypes. "Pie plots" were constructed to demonstrate whether there was any sharing of haplotypes in different regions, religions and the caste system of the Indian subcontinent.

Allele Frequency Analysis

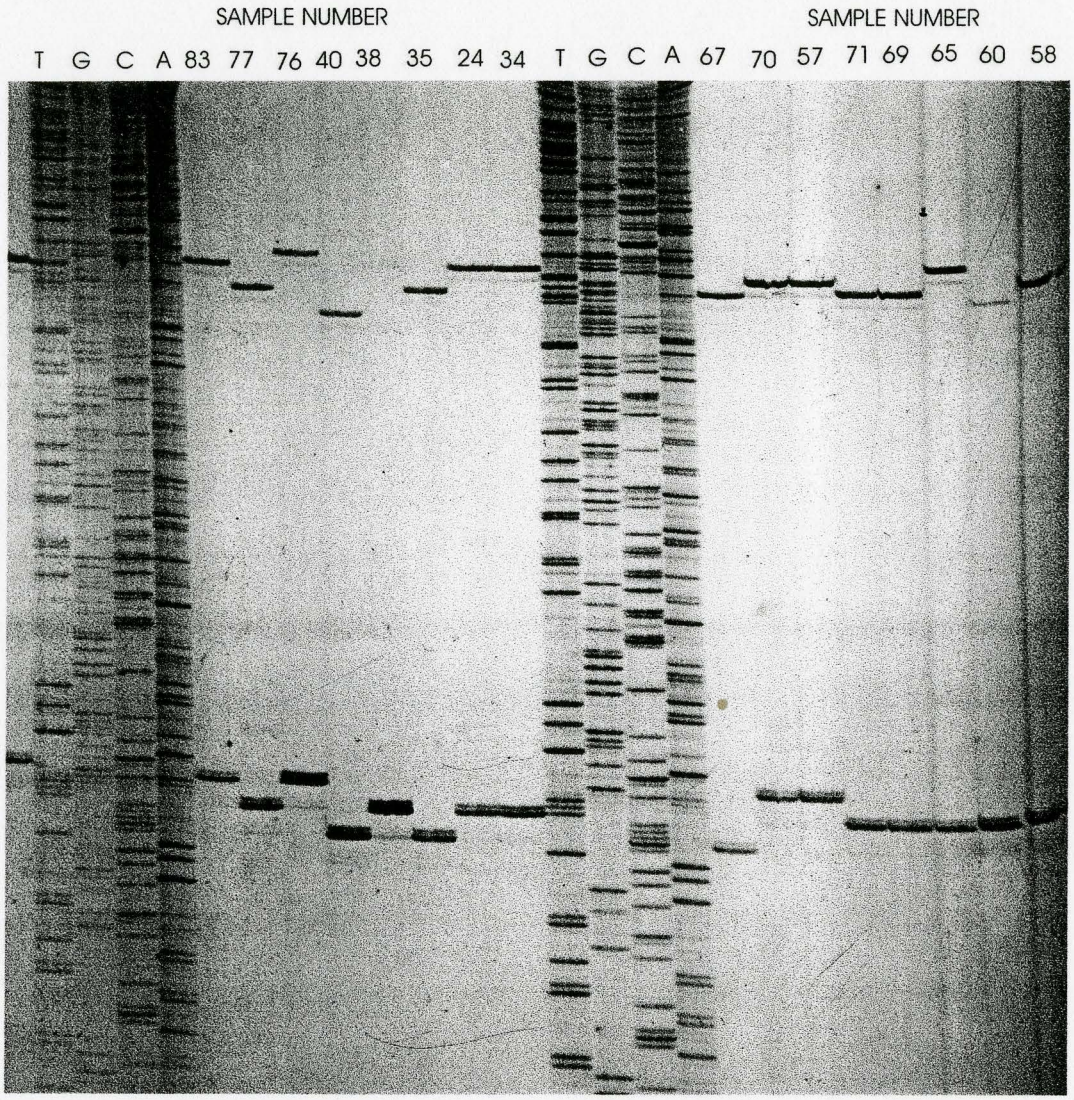
Allele frequencies for each marker were calculated for each of the Indian populations. Pie plots were constructed to show the distribution of alleles in Indians as well as the World populations.

Figure 5: Photograghs of Two Polyacrylamide Gels:

a) For Marker DYS 393



b) For Marker DYS 389



Gene diversity ($\hat{H}$) was estimated based on allele frequency data for India and the world populations as,

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

where x_i is the frequency of the 'i'th allele in the system, n is the number of gene copies, and k is the number of alleles (Jorde *et. al.* 2000, ARLEQUIN, 1997). Variance of gene diversity, $V(\hat{H})$, was estimated (ARLEQUIN, 1997) as,

$$V(\hat{H}) = \frac{2}{n(n-1)}\left\{2(n-2)\left[\sum_{i=1}^{k} x_i^3 - \left(\sum_{i=1}^{k} x_i^2\right)^2\right] + \sum_{i=1}^{k} x_i^2 - \left(\sum_{i=1}^{k} x_i^2\right)^2\right\} \qquad (2)$$

The genetic structure of the population was determined using nested Analysis of Molecular Variance (AMOVA). AMOVA was performed using allele frequencies at each locus. A nested AMOVA assumes a linear model (Excoffier *et. al.* 1992, Sokal and Rohlf 1995) where individuals are arranged into populations and populations are nested within continents (groups), which are chosen on nongenetic criteria. The linear model can be written as (Excoffier *et. al.* 1992, Sokal and Rohlf 1995),

$$\mathbf{p}_{jig} = \mathbf{p} + \mathbf{a}_g + \mathbf{b}_{ig} + \mathbf{c}_{jig} \qquad (3)$$

where, $\mathbf{p}_{jig}$ is the frequency of the 'j'th allele in the 'i'th population in the 'g'th continent. '**p**' is the parametric mean of $\mathbf{p}_{jig}$, averaged over the whole study. In Equation (3), 'a_g' is

the random contribution for the 'g'th continent, 'b_{ig}' is the random contribution for the 'i'th population of the 'g'th continent and 'c_{jig}' is the random contribution of the 'j'th allele in the 'i'th population of the 'g'th continent. The contributions (a_g, b_{ig}, c_{jig}) are assumed to be additive, random, uncorrelated and normally distributed with a mean zero. A nested (hierarchical) AMOVA partitions the total variance (σ^2) into variance components due to intra-individual differences within populations (σ_c^2), inter-individual differences among populations within continents (σ_b^2), and inter-population differences among continents (σ_a^2). Thus, the total molecular variance (σ_T^2) can be written as (Excoffier *et. al.* 1992)

$$\sigma_T^2 = \sigma_a^2 + \sigma_b^2 + \sigma_c^2 \qquad (4)$$

For AMOVA, the total sum of squared deviation, SSD(T) was partitioned into (1) sum of squared deviation among continents, SSD(AG), (2) sum of squared deviation among populations, within continents, SSD(AP/WG), and (3) sum of squared deviation within populations SSD(WP). The corresponding MSD (mean squared deviations) are obtained by dividing each SSD by the appropriate degrees of freedom. The variance components (σ^2 's) of each hierarchical level were extracted by equating the MSDs (mean squared deviations) to their expectations.

The variance components (σ^2 's) were used to compute fixation indices, defined by Wright (1951, 1965) as,

$$F_{CT} = \frac{\sigma_a^2}{\sigma_T^2}, F_{SC} = \frac{\sigma_b^2}{\sigma_b^2 + \sigma_c^2}, \text{and } F_{ST} = \frac{\sigma_a^2 + \sigma_b^2}{\sigma_T^2} \qquad (5)$$

"F_{ST} is defined as correlation of random alleles within populations, relative to that of random pairs of alleles drawn from the whole species, F_{CT} as the correlation of random alleles within a group of populations, relative to that of random pairs of alleles drawn from the whole species, and F_{SC} as the correlation of gene diversity of random alleles within populations, relative to that of random pairs of alleles drawn from the region", (Arlequin).

Under the null hypothesis, samples are considered drawn from a global population, with variation due to random sampling in the construction of populations. The null distribution was obtained by allocating each individual to a randomly chosen population, keeping the sample size constant. This results in random permutation of the rows and the corresponding columns of the squared-distance matrix (Excoffier *et. al.*, 1992). The variance components are estimated from a large number of permuted matrices. This is used to test the significance of F_{ST} and σ_c^2.

Contingency Analysis

Contingency tables were used to test the heterogeneity of allele frequency in different loci among populations. In a R X C contingency table, each row represents a population and each column an allele. The null hypothesis tested was Ho: no heterogeneity of allele frequency among populations i.e., row and column variables are independent of each other. Under the null hypothesis, Raymond and Rousset (1995) reported that the probability (π) of the observed table could be expressed as

$$\pi = \frac{\prod_{i=1}^{p}(N_{i\cdot}!)\prod_{j=1}^{k}(N_{\cdot j}!)}{(N!)\prod_{i=1}^{p}\prod_{j=1}^{k}(N_{ij}!)} \qquad (6)$$

where $N_{i\cdot}$ the sample size of population i, N_{ij} total number of alleles j in population i, N sum of all cells, p number of populations and k number of alleles. The null hypothesis can be tested by an exact probability test via the Classical Fisher test for R X C contingency tables. In the Classical Fisher test (Fisher 1935), the exact value of type-one error probability (P) is the sum of the probabilities of all tables having same or smaller probabilities than the observed probability (π) and with same row and column sums. As Gail and Mantel (1977) reported that in practice, this exact test is not suitable (if not impossible) for most of the typical data sets as the total number of cases to be considered increases rapidly.

Raymond and Rousset (1995) reported a Markov chain method that can compute an unbiased approximation of the Fisher's exact probability, P. In the Markov chain method, all possible contingency tables with the same marginal totals are explored and the unbiased estimate of P is then obtained from the proportion of tables with a lower or equal probability than the observed probability (π). Exact tests such as Fisher's exact probability test has several advantages: (a) it is accurate and unbiased, even for small samples or low frequency alleles, (b) it provides test results for each locus, allowing detection of aberrant loci, and (c) it is independent of the ploidy level. For realistic data sets, the Markov chain method can provide the unbiased estimate of Fisher's exact P-

values within reasonable computational time and is faster than the permutation procedures (Raymond and Rousset 1995).

Statistical analysis of the data was performed using ARLEQUIN (Excoffier *et.al.* 1992; Schneider *et. al.* 1997) and Genepop (Minch *et. al.* 1992).

Phylogenetic Analysis

Genetic distance measures are indicators of relatedness among populations and are thus useful for constructing the historic and the phylogenetic relationships among such groups. For phylogenetic studies, the allele frequencies of seven loci were used. The Locus DYS 389II was not used, as it has not been analyzed in some world populations. Genetic distance analysis has shown extreme divergence for the Finnish populations, therefore for phylogenetic studies the Finnish population was thus excluded (Kittle *et. al.* 1998).

Three different models of distance calculation were used for the purpose of the present study. Cavalli-Sforza chord distances (Cavalli-Sforza and Edwards 1967) were generated using the GENEDIST program. A tree was then generated by the FITCH program. The Global option was turned on, so that subtrees would be removed from the generated trees and put back on in all possible ways and thus more trees are examined.

The second distance was generated using Stephens's *et. al.* (1994) proportion of allele sharing. Proportion of shared alleles for individual pairwise comparisons was estimated by;

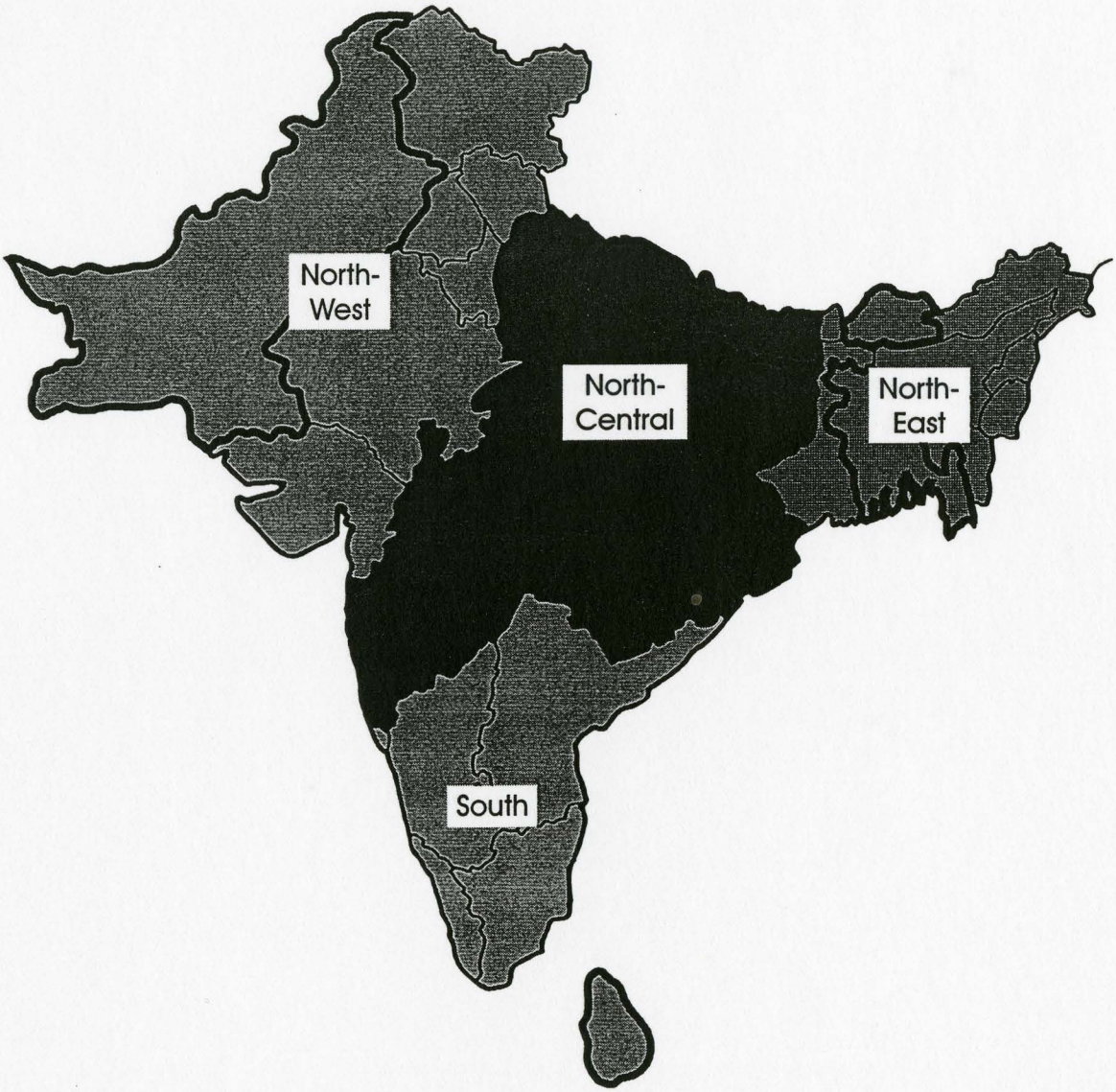


Figure 6: The Four Main Subdivision of the Indian Subcontinent used for the Study

$$P_{SA_I} = \frac{\sum_U S}{2U} \qquad (7)$$

where, s the number of shared alleles summed over all loci u. Distance between individuals (D_{SA_I}) is estimated (Bowcock *et. al.* 1994) by;

$$D_{SA_I} = 1 - P_{SA_I} \qquad (8)$$

The distance between populations (D_{SA_B}) (Chakraborty and Jin 1993) can be calculated as;

$$D_{SA_B} = 1 - \frac{2\bar{P}_{SA_B}}{\bar{P}_{SA_1} + \bar{P}_{SA_2}} \qquad (9)$$

D_{AS} is used for reconstructing phylogenies when taxa are closely related. The accuracy of this measure for short distances stems from the use of information available in the degree of overlap between the allele frequency distribution of the two populations. The measure becomes less accurate at greater distances. (Goldstein and Pollock 1997).

The third distance is based on a model for microsatellite mutation. This is a more appropriate measure of genetic distance as it takes into account the fact that mutation in microsatellites increases or decreases the repeat number by one or a few units (Stepwise mutation model proposed by Ohta and Kimura 1973). This distance (D_g) proposed by Goldstein *et. al.* (1995) is based on the following formula;

$$D_g = (\text{delta}\,\mu)^2 = (\mu_A - \mu_B)^2 \qquad (10)$$

where μ_A and μ_B are the means of allele sizes in populations A and B respectively. In Equation (10), D_g is independent of sample size (n).

In reality, one must estimate $(\text{delta}\,\mu)^2$ (distance between the populations A and B), based on the samples from the populations. $(\text{delta}\,\mu)^2$ can also be expressed as (Goldstein *et. al.* 1995),

$$(\text{delta}\,\mu)^2 = \text{ASD} - \sigma_A^2 - \sigma_B^2 \qquad (11)$$

where σ_A^2 and σ_B^2 are the variances of allelic size of populations A and B respectively and ASD is the average squared distance between the populations. From Equations (10) and (11), one can write

$$D_g = (\text{delta}\,\mu)^2 = (\mu_A - \mu_B)^2 = \text{ASD} - \sigma_A^2 - \sigma_B^2 \qquad (12)$$

Goldstein *et. al.* (1995a) reported that the ASD calculated from the samples of the populations is an unbiased estimator of ASD for the populations. However, sample variances are not unbiased estimators of the population variances. Thus, an unbiased estimator of $(\text{delta}\,\mu)^2$ can be obtained as (Goldstein *et. al.*, 1995),

$$(\text{delta}\,\mu)^2 = \text{ASD} - \hat{\sigma}_A^2 - \hat{\sigma}_B^2 \qquad (13)$$

where $\hat{\sigma}_A^2$ and $\hat{\sigma}_B^2$ denote the unbiased estimates of the population variances σ_A^2 and σ_B^2 respectively, based on the variance in the samples for the populations A and B. In this analysis, Equation (10) is used to compute (delta μ)^2 for the populations A and B. Goldstein and Pollock (1997) also reported that unless both the sample size and the level of differentiation is small, (delta μ)^2 (calculated from Eqn. (10)) does not differ significantly from (delta μ)^2, calculated from the sample means (Eqn. 7).

Both D_{AS} and (delta μ)^2 were computed using the Microstat program (*Minch et. al.* 1992). The allele frequency data was also bootstrapped 1000 times. The bootstrapped distance matrices were then used in the Fitch program to give 1000 bootstrapped trees. The global option was turned on. The CONSENSUS program was used to get a consensus of the bootstrapped trees. The distance matrix and the consensus tree were then used together in the FITCH program to give a tree that had the distances incorporated in it. The tree construction was done by using the PHYLIP Software Package (Felsenstein 1993).

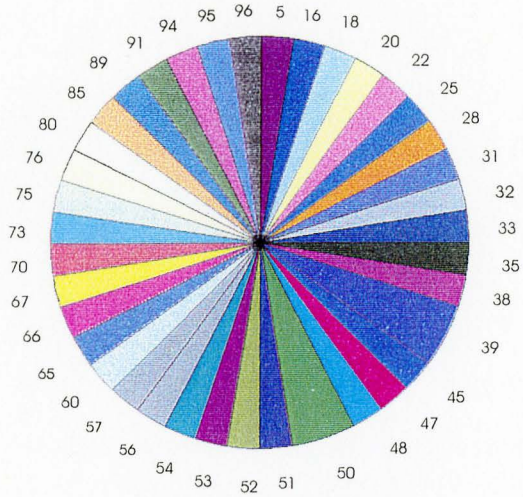
RESULTS

Of the total one hundred and fourteen samples, ninety-seven unique haplotypes were found using all the loci. Results of the distribution of the different haplotypes in the different regions, different religions and caste of India are shown Figure 7,8 and 9 respectively. As there were so many haplotypes that are unique in the different region, religion and caste, one could not rule out the possibility that when two haplotypes are shared between different region, religion and caste, it might simply be due to chance and not due to some phylogenetic reasons. As a result it was thought that a more meaningful result would be obtained if analysis were done on the allele frequencies distributions for the different loci.

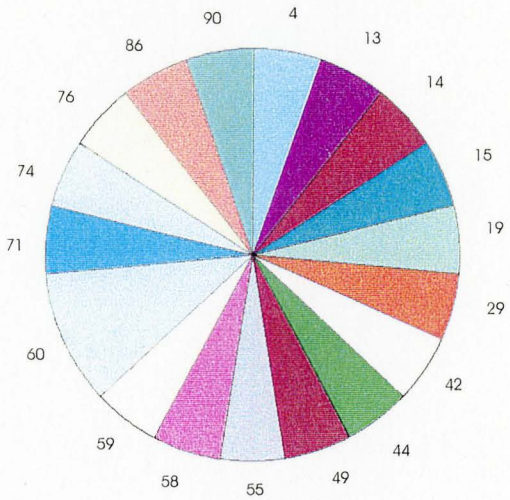
The distribution of allele frequencies has been shown in Table 7. The distribution of the different alleles frequencies of 8 loci in the Indian as well as in the different World populations are shown in figures 10 through 17, for Loci DYS 19, and 388 through 393. It is seen that certain alleles of a particular marker may be present in low frequency or be totally absent in South as compared to North India. This data was then used for further analysis. The loci DYS 389II was not used for subsequent analysis, as some of the populations were not analyzed for this locus. All the alleles of these loci have a worldwide distribution, though the frequencies of the alleles vary among the different populations.

Genetic Diversity and its variance were estimated for the Indian and the different World populations and are shown in Table 8 and 9 respectively. The highest and the

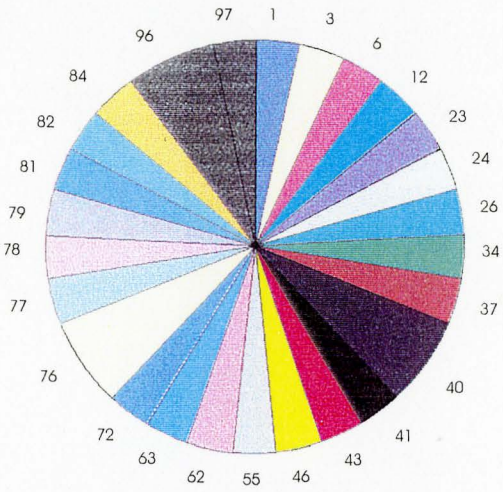
Figure 7: Distribution of Different Haplotypes within the Four Regions of the Indian Subcontinent.



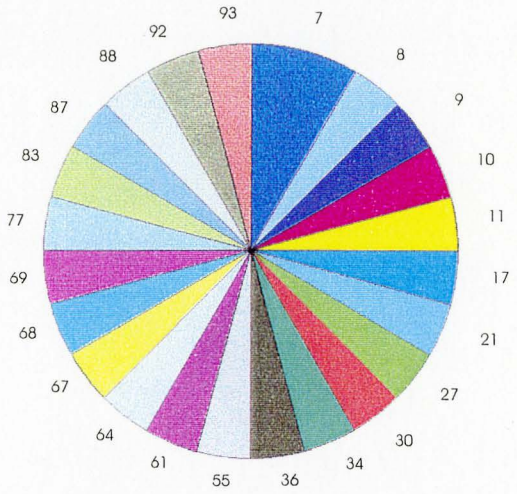
NORTHCENTRAL INDIA (n=40)



NORTHEAST INDIA (n=19)

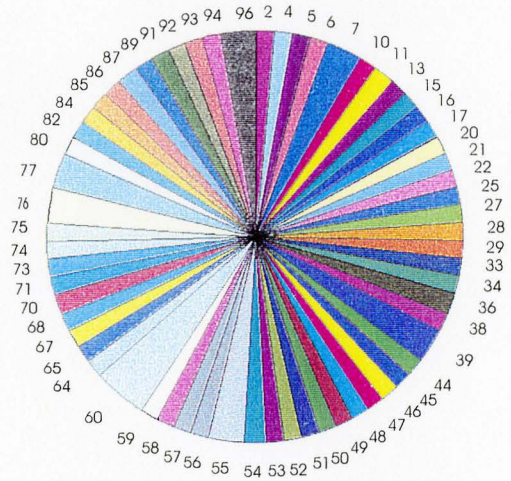


NORTHWEST INDIA (n=29)

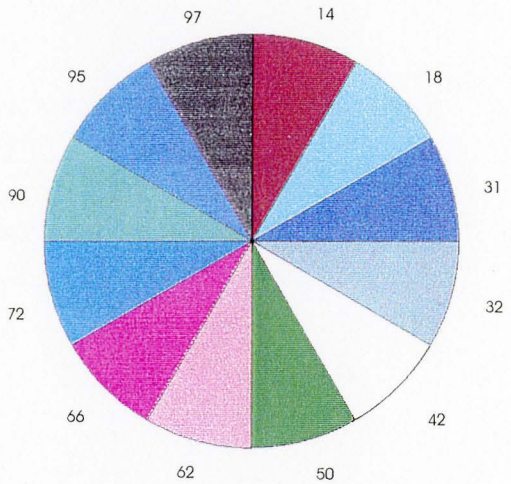


SOUTH INDIA (n=24)

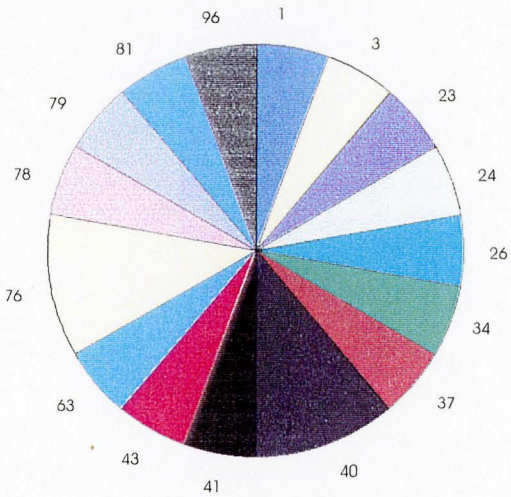
Figure 8: Distribution of Different Haplotypes within the Major Religions of the Indian Subcontinent.



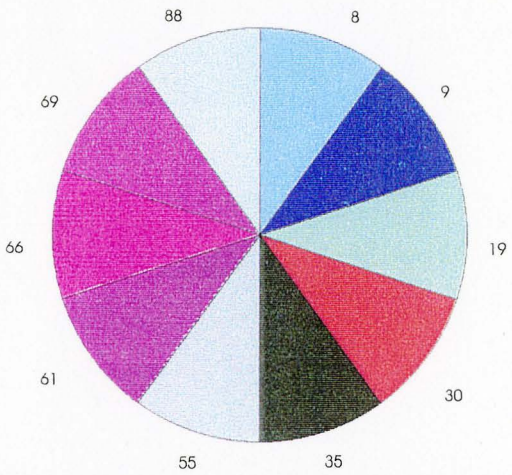
HINDUS
(n=71)



MUSLIMS
(n=12)

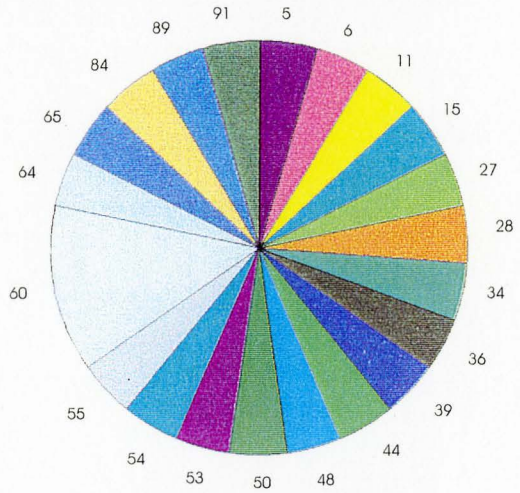


SIKHS
(n=18)

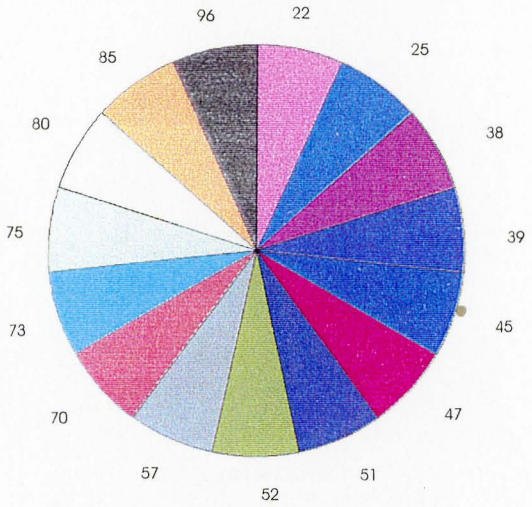


CHRISTIANS
(n=10)

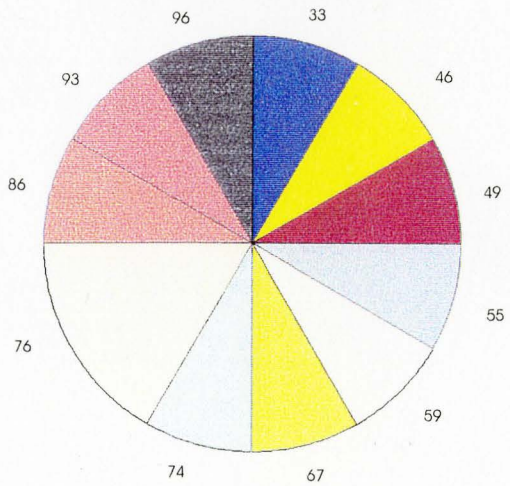
Figure 9: Distribution of Different Haplotypes within the Three Major Castes



BRAHMIN
(n=23)



KSATRIYA
(n=15)



VAISYA
(n=12)

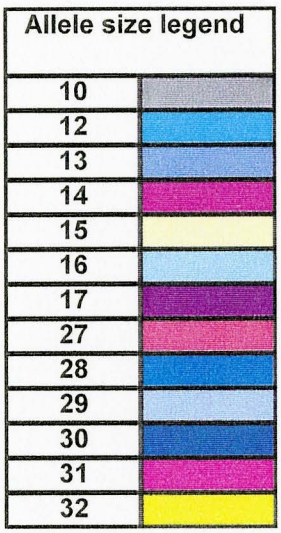
Table 7: Allele Frequencies in Different Ethnic Groups of India.

DYS	Allele	Total	Hindu	Christian	Muslim	Sikh	Brahmin	Ksatriya	Vaishya
19	13	0.044	0.042	0	0	0.111	0.043	0	0
	14	0.254	0.250	0.4	0.333	0.167	0.217	0.133	0.083
	15	0.395	0.403	0.5	0.333	0.389	0.609	0.467	0.417
	16	0.263	0.278	0.1	0.167	0.278	0.130	0.333	0.417
	17	0.044	0.028	0	0.167	0.056	0	0.067	0.083
388	12	0.741	0.764	0.7	0.583	0.833	0.696	0.867	0.75
	13	0.103	0.097	0	0.167	0.056	0.174	0.067	0
	14	0.026	0.014	0	0	0.111	0	0.067	0
	15	0.086	0.083	0.3	0.083	0	0.087	0	0.167
	16	0.026	0.014	0	0.167	0	0.043	0	0
	17	0.017	0.028	0	0	0	0	0	0.083
389I	9	0.181	0.194	0.4	0.167	0.056	0.261	0.067	0.167
	10	0.414	0.403	0	0.417	0.722	0.261	0.667	0.417
	11	0.379	0.375	0.5	0.417	0.222	0.478	0.267	0.417
	12	0.026	0.028	0.1	0	0	0	0	0
389II	24	0.017	0.009	0	0.009	0	0	0	0
	25	0.147	0.096	0.035	0	0.018	0.130	0.133	0.167
	26	0.267	0.175	0.009	0.026	0.053	0.348	0.4	0.083
	27	0.293	0.167	0.035	0.035	0.052	0.261	0.2	0.417
	28	0.172	0.114	0.009	0.026	0.026	0.217	0.067	0.25
	29	0.060	0.044	0	0.009	0.009	0	0.2	0.083
	30	0.043	0.04	0	0	0.009	0.043	0	0
390	21	0.009	0.014	0.3	0	0	0	0	0
	22	0.267	0.278	0.3	0.250	0.278	0.304	0.2	0.25
	23	0.284	0.236	0.2	0.417	0.222	0.217	0.333	0.083
	24	0.172	0.181	0.2	0.167	0.167	0.087	0.2	0.417
	25	0.216	0.222	0	0.167	0.278	0.348	0.2	0.167
	26	0.052	0.069	0	0	0.056	0.043	0.067	0.083
391	9	0.026	0.014	0	0.167	0	0	0	0
	10	0.681	0.681	0.8	0.667	0.611	0.696	0.733	0.667
	11	0.267	0.292	0.2	0.167	0.333	0.304	0.267	0.333
	12	0.017	0.014	0	0	0.056	0	0	0

Table 7 contd.

DYS	Allele	Total	Hindu	Christian	Muslim	Sikh	Brahmin	Ksatriya	Vaishya
392	10	0.095	0.083	0	0.25	0.056	0.087	0.133	0
	11	0.690	0.736	0.5	0.667	0.611	0.783	0.8	0.917
	12	0.069	0.069	0.1	0.083	0.056	0.087	0	0.083
	13	0.026	0.014	0.1	0	0.056	0	0	0
	14	0.095	0.069	0.3	0	0.167	0.043	0.067	0
	15	0.026	0.028	0	0	0.056	0	0 0	
393	11	0.060	0.083	0	0	0.056	0.130	0	0
	12	0.328	0.306	0.6	0.333	0.278	0.217	0.467	0.417
	13	0.466	0.486	0.4	0.417	0.556	0.435	0.467	0.583
	14	0.129	0.097	0	0.25	0.111	0.174	0.067	0
	15	0.017	0.028	0	0	0	0.043	0	0

Figure 10:Pie-Plots Showing the Worldwide Distribution of the Locus DYS 19



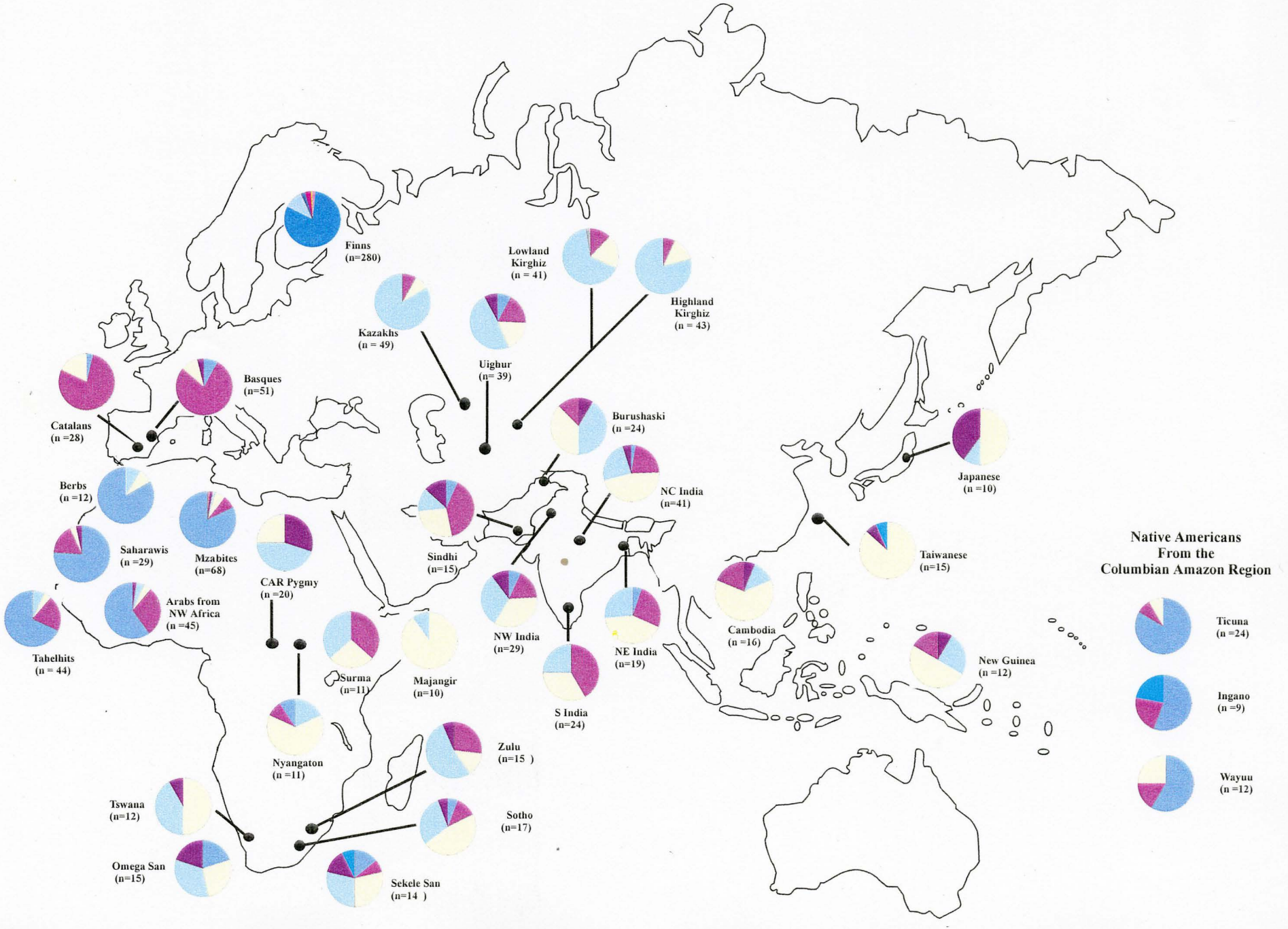
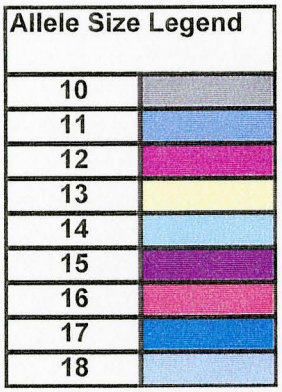

Finns
(n=280)
Lowland
Kirghiz
(n = 41)
Highland
Kirghiz
(n = 43)
Kazakhs
(n = 49)
Uighur
(n= 39)
Basques
(n=51)
Catalans
(n =28)
Burushaski
(n =24)
Japanese
(n =10)
Berbs
(n =12)
NC India
(n=41)
Saharawis
(n =29)
Mzabites
(n=68)
Sindhi
(n=15)
Taiwanese
(n=15)
CAR Pygmy
(n =20)
Cambodia
(n =16)
Arabs from
NW Africa
(n =45)
NW India
(n=29)
NE India
(n=19)
New Guinea
(n =12)
Tahelhits
(n = 44)
Surma
(n=11)
Majangir
(n=10)
S India
(n=24)
Nyangaton
(n =11)
Zulu
(n=15)
Tswana
(n=12)
Sotho
(n=17)
Omega San
(n=15)
Sekele San
(n=14)
Native Americans
From the
Columbian Amazon Region
Ticuna
(n =24)
Ingano
(n =9)
Wayuu
(n =12)

Figure 11:Pie-Plots Showing the Worldwide Distribution of the Locus DYS 388



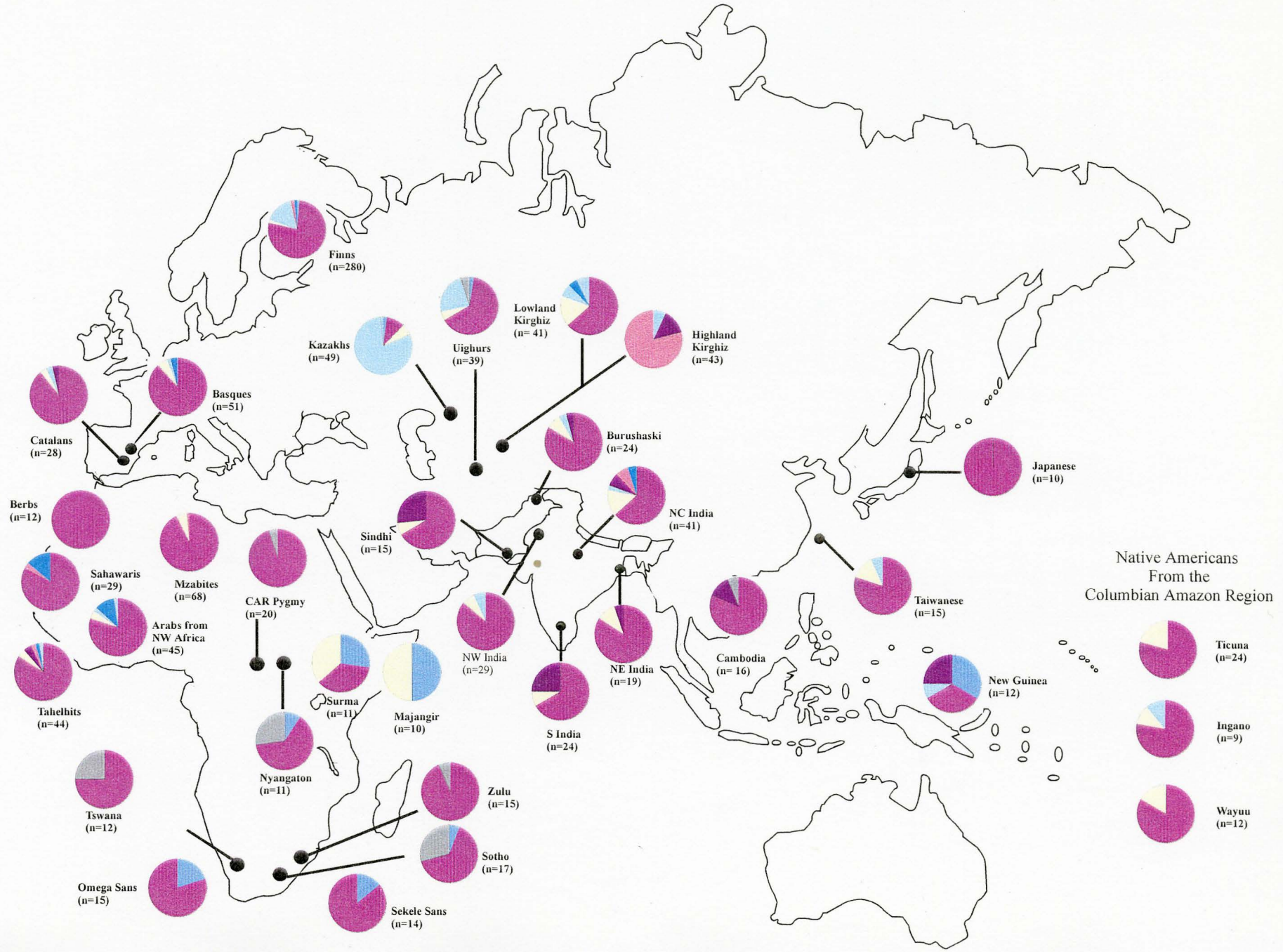
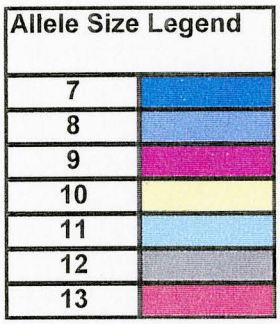

Finns
(n=280)
Kazakhs
(n=49)
Uighurs
(n=39)
Lowland
Kirghiz
(n= 41)
Highland
Kirghiz
(n=43)
Basques
(n=51)
Catalans
(n=28)
Burushaski
(n=24)
Japanese
(n=10)
Berbs
(n=12)
Sindhi
(n=15)
NC India
(n=41)
Sahawaris
(n=29)
Mzabites
(n=68)
CAR Pygmy
(n=20)
Taiwanese
(n=15)
Arabs from
NW Africa
(n=45)
NW India
(n=29)
Cambodia
(n= 16)
NE India
(n=19)
Tahelhits
(n=44)
Surma
(n=11)
Majangir
(n=10)
New Guinea
(n=12)
S India
(n=24)
Nyangaton
(n=11)
Tswana
(n=12)
Zulu
(n=15)
Sotho
(n=17)
Omega Sans
(n=15)
Sekele Sans
(n=14)
Native Americans
From the
Columbian Amazon Region
Ticuna
(n=24)
Ingano
(n=9)
Wayuu
(n=12)

Figure 12:Pie-Plots Showing the Worldwide Distribution of the Locus DYS 389I



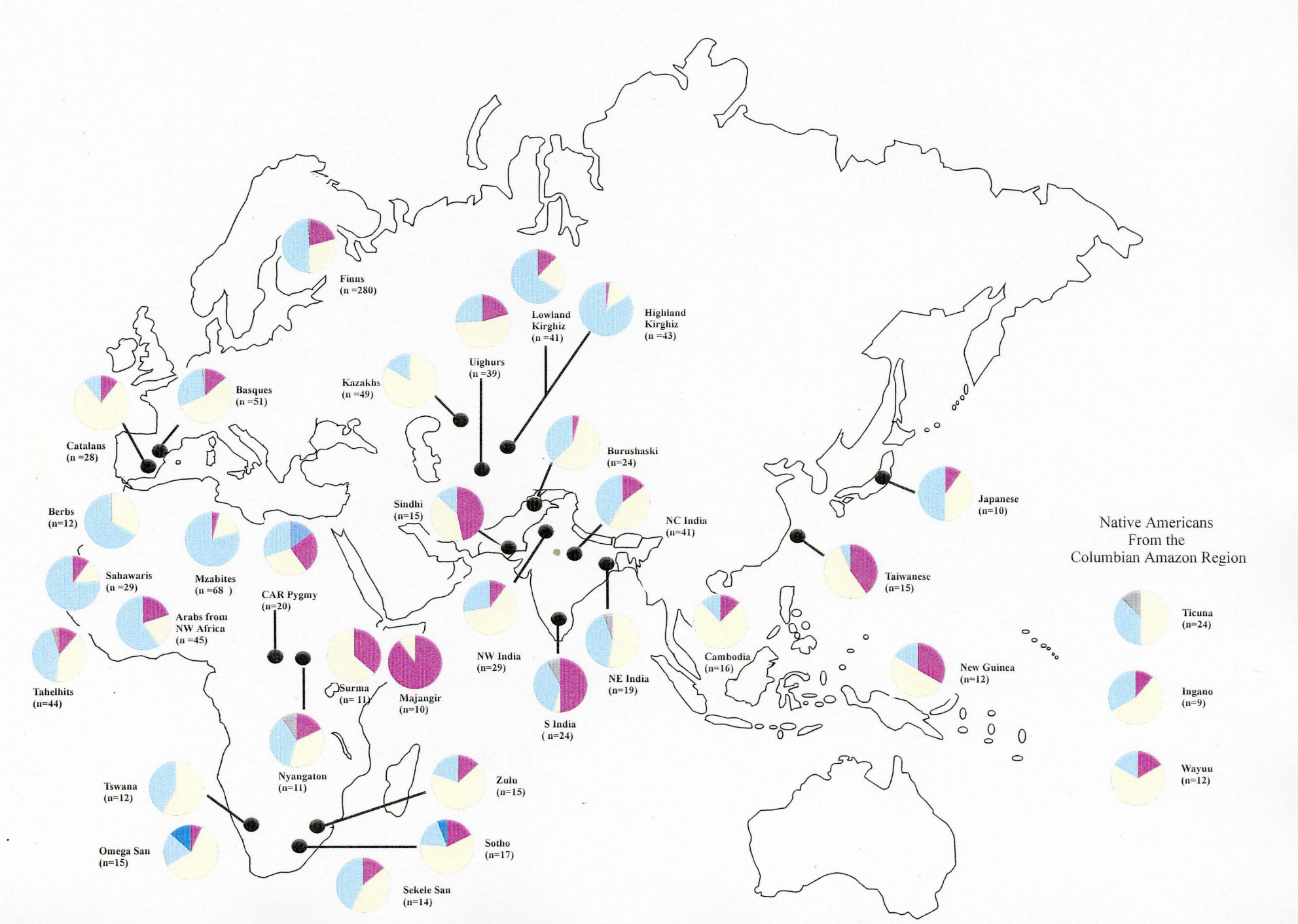
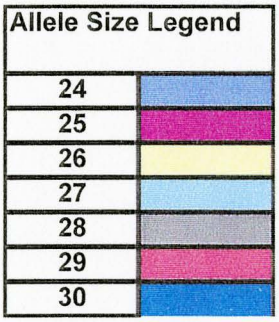

Finns
(n =280)
Lowland
Kirghiz
(n =41)
Highland
Kirghiz
(n =43)
Uighurs
(n =39)
Kazakhs
(n =49)
Basques
(n =51)
Catalans
(n =28)
Burushaski
(n=24)
Sindhi
(n=15)
NC India
(n=41)
Japanese
(n=10)
Berbs
(n=12)
Sahawaris
(n =29)
Mzabites
(n =68)
CAR Pygmy
(n=20)
Taiwanese
(n=15)
Arabs from
NW Africa
(n =45)
Tahelhits
(n=44)
Surma
(n= 11)
Majangir
(n=10)
NW India
(n=29)
S India
(n=24)
NE India
(n=19)
Cambodia
(n=16)
New Guinea
(n=12)
Nyangaton
(n=11)
Tswana
(n=12)
Zulu
(n=15)
Omega San
(n=15)
Sotho
(n=17)
Sekele San
(n=14)
Native Americans
From the
Columbian Amazon Region
Ticuna
(n=24)
Ingano
(n=9)
Wayuu
(n=12)

Figure 13:Pie-Plots Showing the Worldwide Distribution of the Locus DYS 389II



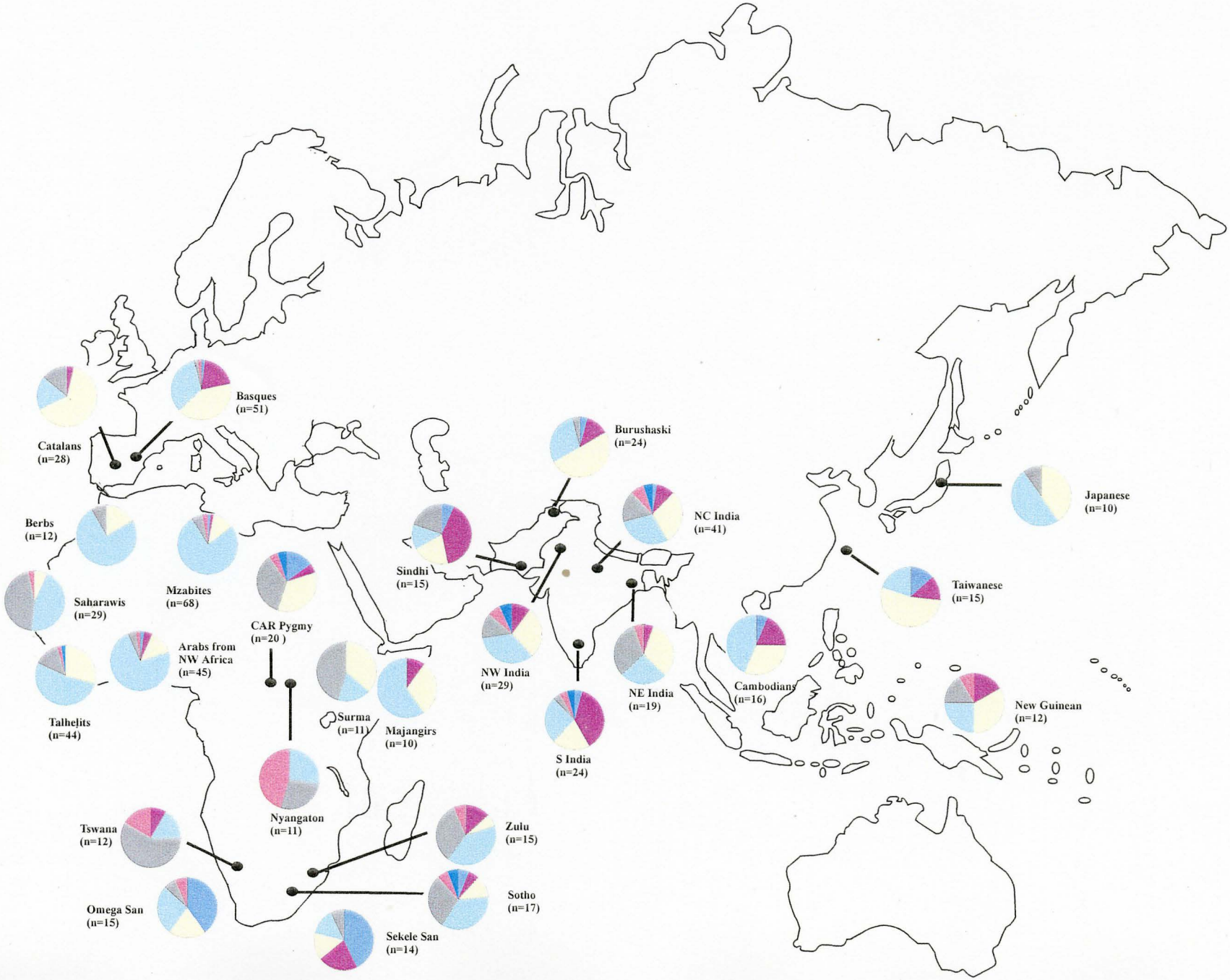
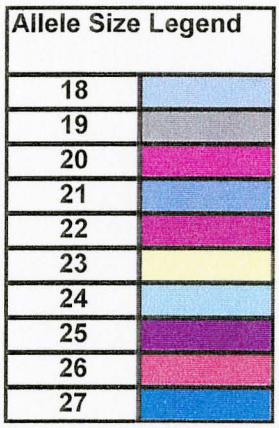

Basques
(n=51)
Catalans
(n=28)
Berbs
(n=12)
Saharawis
(n=29)
Mzabites
(n=68)
Arabs from
NW Africa
(n=45)
Talhejits
(n=44)
CAR Pygmy
(n=20)
Nyangaton
(n=11)
Surma
(n=11)
Majangirs
(n=10)
Sindhi
(n=15)
Burushaski
(n=24)
NC India
(n=41)
NW India
(n=29)
NE India
(n=19)
S India
(n=24)
Japanese
(n=10)
Taiwanese
(n=15)
Cambodians
(n=16)
New Guinean
(n=12)
Tswana
(n=12)
Omega San
(n=15)
Sekele San
(n=14)
Zulu
(n=15)
Sotho
(n=17)

Figure 14:Pie-Plots Showing the Worldwide Distribution of the Locus DYS 390



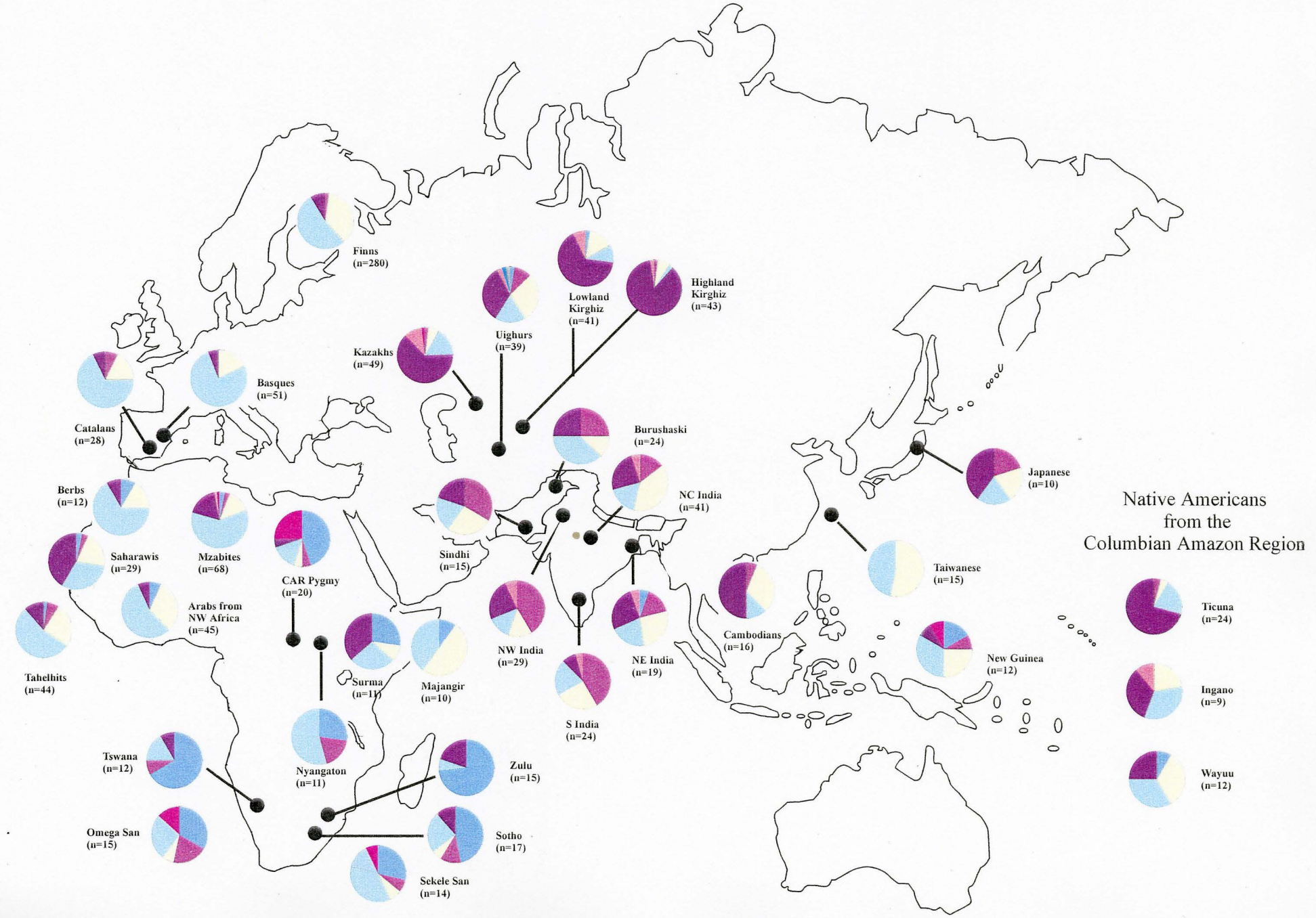
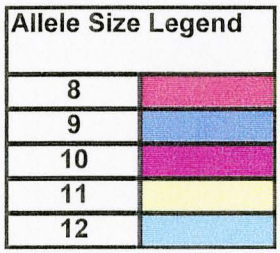

Finns
(n=280)
Lowland
Kirghiz
(n=41)
Highland
Kirghiz
(n=43)
Uighurs
(n=39)
Kazakhs
(n=49)
Basques
(n=51)
Catalans
(n=28)
Burushaski
(n=24)
Japanese
(n=10)
Berbs
(n=12)
NC India
(n=41)
Saharawis
(n=29)
Mzabites
(n=68)
Sindhi
(n=15)
CAR Pygmy
(n=20)
Taiwanese
(n=15)
Arabs from
NW Africa
(n=45)
Cambodians
(n=16)
Tahelhits
(n=44)
NW India
(n=29)
NE India
(n=19)
New Guinea
(n=12)
Surma
(n=11)
Majangir
(n=10)
S India
(n=24)
Tswana
(n=12)
Nyangaton
(n=11)
Zulu
(n=15)
Omega San
(n=15)
Sotho
(n=17)
Sekele San
(n=14)
Native Americans
from the
Columbian Amazon Region
Ticuna
(n=24)
Ingano
(n=9)
Wayuu
(n=12)

Figure 15:Pie-Plots Showing the Worldwide Distribution of the Locus DYS 391



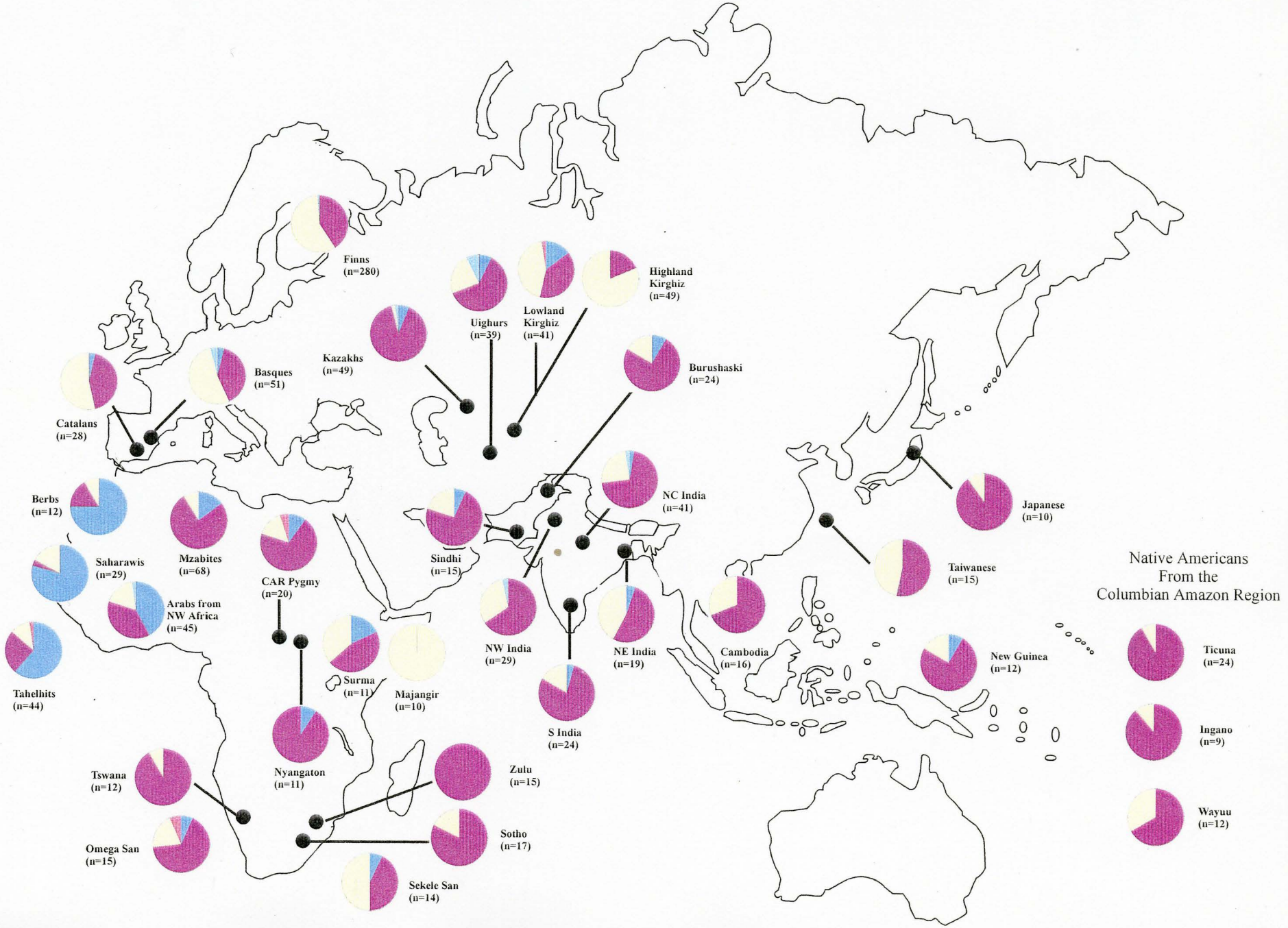
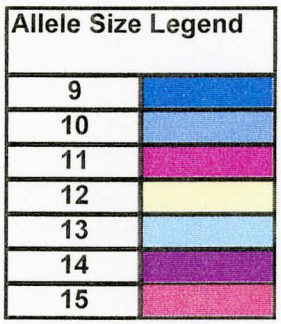

Finns
(n=280)
Highland
Kirghiz
(n=49)
Uighurs
(n=39)
Lowland
Kirghiz
(n=41)
Kazakhs
(n=49)
Basques
(n=51)
Burushaski
(n=24)
Catalans
(n=28)
NC India
(n=41)
Berbs
(n=12)
Japanese
(n=10)
Sindhi
(n=15)
Saharawis
(n=29)
Mzabites
(n=68)
CAR Pygmy
(n=20)
Taiwanese
(n=15)
Arabs from
NW Africa
(n=45)
Cambodia
(n=16)
NW India
(n=29)
NE India
(n=19)
New Guinea
(n=12)
Tahelhits
(n=44)
Surma
(n=11)
Majangir
(n=10)
S India
(n=24)
Nyangaton
(n=11)
Zulu
(n=15)
Tswana
(n=12)
Sotho
(n=17)
Omega San
(n=15)
Sekele San
(n=14)
Native Americans
From the
Columbian Amazon Region
Ticuna
(n=24)
Ingano
(n=9)
Wayuu
(n=12)

Figure 16:Pie-Plots Showing the Worldwide Distribution of the Locus DYS 392



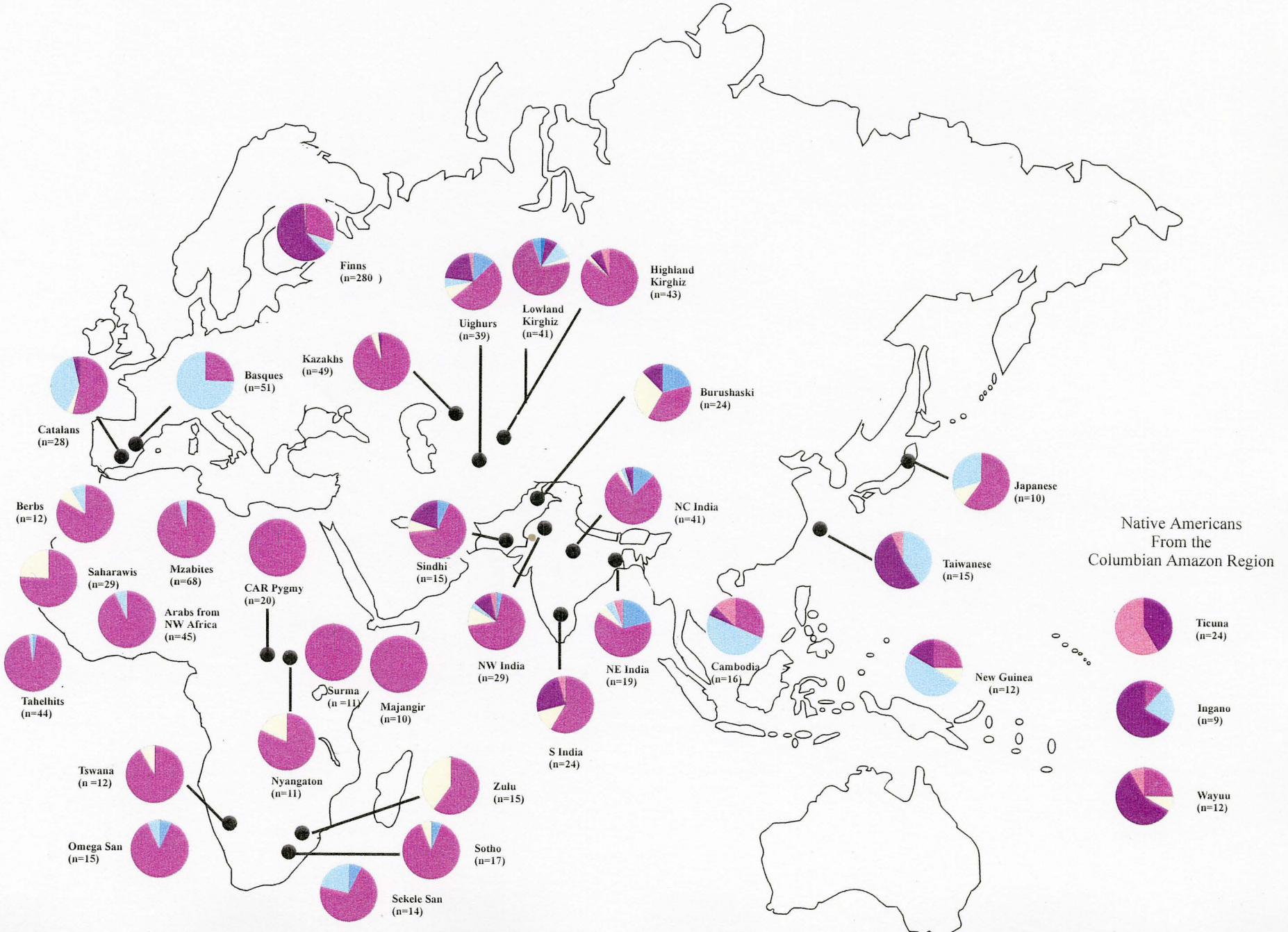
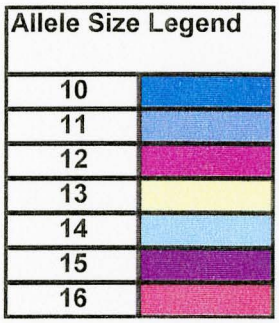
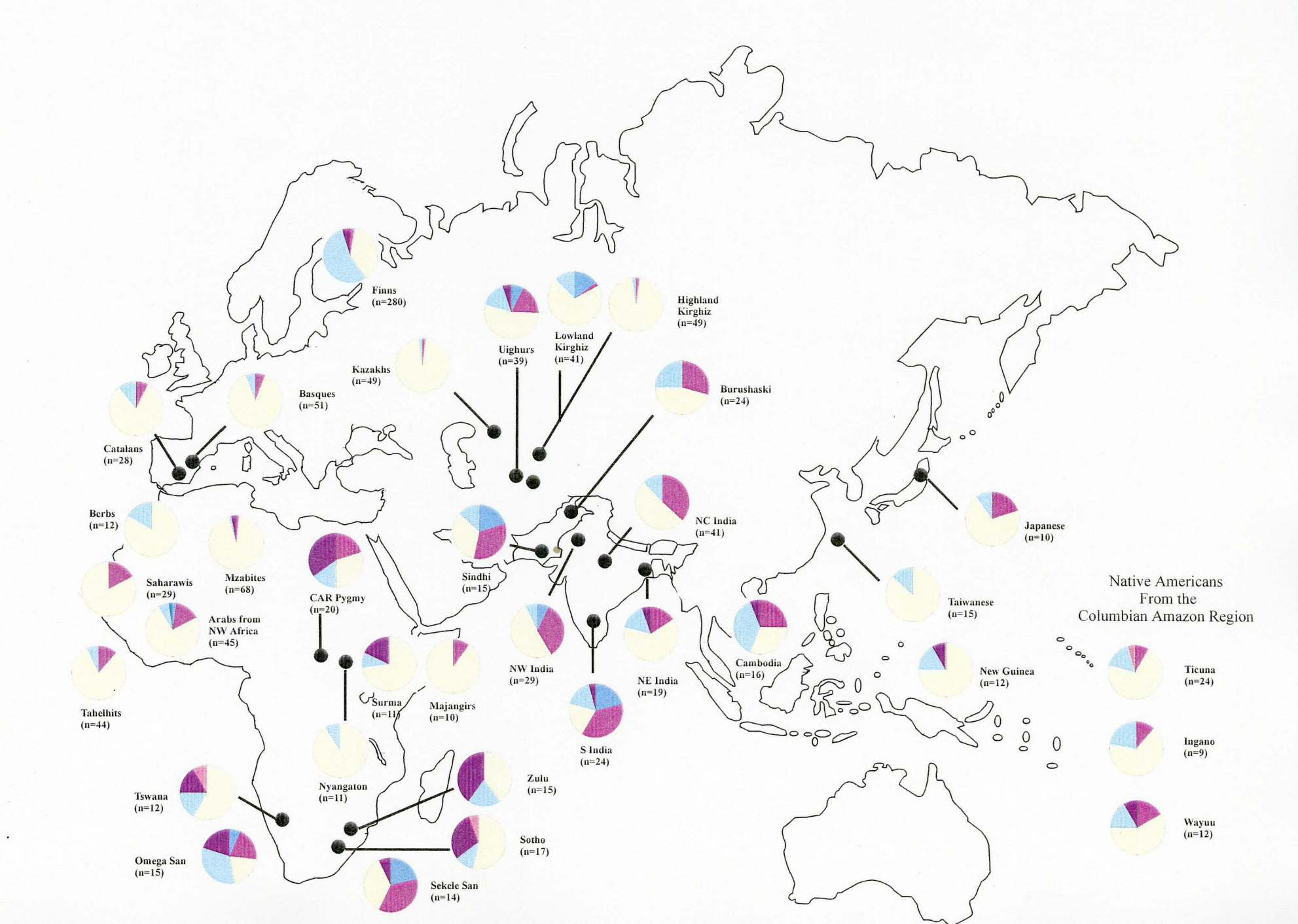

Finns
(n=280)
Highland
Kirghiz
(n=43)
Uighurs
(n=39)
Lowland
Kirghiz
(n=41)
Kazakhs
(n=49)
Basques
(n=51)
Catalans
(n=28)
Burushaski
(n=24)
Japanese
(n=10)
Berbs
(n=12)
NC India
(n=41)
Saharawis
(n=29)
Mzabites
(n=68)
Sindhi
(n=15)
CAR Pygmy
(n=20)
Taiwanese
(n=15)
Arabs from
NW Africa
(n=45)
Tahelhits
(n=44)
NW India
(n=29)
NE India
(n=19)
Cambodia
(n=16)
New Guinea
(n=12)
Surma
(n =11)
Majangir
(n=10)
S India
(n=24)
Tswana
(n =12)
Nyangaton
(n=11)
Zulu
(n=15)
Omega San
(n=15)
Sotho
(n=17)
Sekele San
(n=14)
Native Americans
From the
Columbian Amazon Region
Ticuna
(n=24)
Ingano
(n=9)
Wayuu
(n=12)

Figure 17:Pie-Plots Showing the Worldwide Distribution of the Locus DYS 393




Finns
(n=280)
Highland
Kirghiz
(n=49)
Uighurs
(n=39)
Lowland
Kirghiz
(n=41)
Kazakhs
(n=49)
Basques
(n=51)
Burushaski
(n=24)
Catalans
(n=28)
Berbs
(n=12)
NC India
(n=41)
Japanese
(n=10)
Sindhi
(n=15)
Saharawis
(n=29)
Mzabites
(n=68)
CAR Pygmy
(n=20)
Taiwanese
(n=15)
Arabs from
NW Africa
(n=45)
NW India
(n=29)
NE India
(n=19)
Cambodia
(n=16)
New Guinea
(n=12)
Tahelhits
(n=44)
Surma
(n=11)
Majangirs
(n=10)
S India
(n=24)
Nyangaton
(n=11)
Zulu
(n=15)
Tswana
(n=12)
Sotho
(n=17)
Omega San
(n=15)
Sekele San
(n=14)
Native Americans
From the
Columbian Amazon Region
Ticuna
(n=24)
Ingano
(n=9)
Wayuu
(n=12)

lowest values of gene diversity and its variance have been bolded and italicized respectively. There is no world population that has the highest or the lowest diversity value for all the loci. However most of the highest diversity values belong to the African population. The genetic diversity of the Indian populations do not differ greatly from that of the other world populations and generally show a tendency towards a higher diversity value. In fact, the NE Indian population has the highest diversity value for locus DYS 390. In Table 9, the variance of gene diversity shows the spread in the distribution of gene diversity in a population for a given locus. A variance of zero means that there is no spread in the distribution of gene diversity, i.e. only one type of allele is present in that locus for that particular population. Higher variance implies wider spread in the distribution of gene diversity i.e. more types of alleles are present in that locus for that particular population.

AMOVA analysis of the 33 populations (New Guinea was not included in this calculation because it was not certain in which continent it belongs) shows that the within population variation is higher than among the continents or the among population within continent variation (Table 10), which is in accordance with published results (Jorde *et.al.*2000). The Fst values computed for the four continents (Asia, Africa, Europe and America) at different loci are also shown. The Fst values are significantly different from 0 ($P< 0.05$) for all the loci tested except for the loci DYS 388 and 389I. At locus DYS 19 has the highest Fst (0.48366) value, whereas for Locus DYS 390, it is the lowest (Fst = 0.2046).

Heterogeneity of allele frequency in different loci among populations was tested

TABLE 8: AN ESTIMATION OF GENE DIVERSITY IN DIFFERENT WORLD POPULATIONS.

POPULATIONS	n	DYS19	DYS388	DYS389I	DYS390	DYS391	DYS392	DYS393	Average
NC INDIA	41	0.6915	0.5512	0.6293	0.7537	0.4500	0.3817	0.6037	0.5801
NE INDIA	19	0.7193	0.2924	0.5731	**0.8304**	0.5731	0.5789	0.5789	0.5923
NW INDIA	29	0.766	0.2562	0.5468	0.7537	0.4901	0.5172	0.6256	0.5750
S. INDIA	24	0.6812	0.5072	0.6268	0.7428	0.3587	0.6051	0.7754	0.6138
BASQUES	51	0.3749	0.2204	0.6047	*0.3882*	0.5741	0.3875	0.2188	0.4193
CATALANS	28	0.3624	0.2063	0.373	0.5159	0.5476	0.5767	0.3201	0.4131
SAHARAWIS	29	0.4064	0.3054	0.4089	0.7118	0.3522	0.3793	0.2956	0.4083
MZABITES	68	0.3169	0.14	0.3464	0.5944	0.3916	0.0856	0.1146	0.2842
ARABS	45	0.5646	0.3525	0.5727	0.599	0.6616	0.1273	0.4091	0.4695
BERBERS	12	0.3182	*0*	0.4848	0.5606	0.4394	0.3182	0.303	0.3463
TAHELHITS	44	0.4979	0.2939	0.6469	0.6364	0.5603	0.0899	0.3203	0.4350
INGANO	9	0.6667	0.4167	0.6389	0.8056	0.2222	0.5556	0.5556	0.5516
TICUNA	24	0.3043	0.3442	0.6196	0.471	0.1594	0.5072	0.4819	0.4125
WAYUU	12	0.6212	0.303	0.5455	0.7727	0.4848	0.6364	0.6515	0.5735
KAZAKHS	49	0.2925	0.3588	0.2789	0.5697	0.1556	0.119	0.0808	0.2650
H. KIRGIZ	43	0.3588	0.3588	0.2857	0.2591	0.3101	0.258	0.0919	0.2746
L. KIRGIZ	41	0.5256	0.5695	0.5159	0.5427	0.6488	0.4573	0.4744	0.5334
UIGHURS	39	0.7045	0.5439	0.6181	0.7841	0.4966	0.6869	0.6626	0.6423
FINNS	280	0.3367	0.3792	0.6349	0.581	0.497	0.5262	0.5521	0.5010
BURUSHASKI	24	0.692	0.308	0.5399	*0.75*	0.4203	**0.7464**	0.6703	0.5895
CAR PYGMY	20	0.6789	0.1	**0.7737**	0.7421	0.5	0.5	0.7632	0.5797
CAMBODIA	16	0.5917	0.3417	0.4333	0.675	0.4583	0.675	0.7417	0.5595
JAPANESE	10	0.6444	*0*	0.6444	0.8	0.2	0.6	0.5111	0.4857
MAJANGIR	10	*0.2*	0.5556	*0.2*	0.6444	*0*	*0*	*0.2*	0.2571
NYANGATOM	11	0.6	0.5636	0.7636	0.6545	0.1818	0.3273	0.1818	0.4675
OMEGA SAN	15	0.7905	0.3429	0.619	0.8095	0.5429	0.2571	0.819	0.5972
S.SAN	14	**0.8462**	0.2637	0.6593	0.8132	0.6044	0.4725	0.7473	0.6295
SINDHI	15	0.781	0.5143	0.6476	0.7905	0.4476	0.5429	**0.7714**	0.6421
SOTHO	17	0.7132	0.5221	0.625	0.7353	0.3088	0.2279	0.6544	0.5409
SURMA	11	0.7273	**0.7273**	0.5091	0.7818	**0.6909**	0.6909	0.4727	0.6571
TAIWANESE	15	0.2571	0.3619	0.5905	0.5333	0.5333	0.5905	0.2476	0.4448
TSWANA	12	0.6212	0.4091	0.5303	0.5606	0.1667	0.1667	0.6515	0.4437
ZULU	15	0.6667	0.1333	0.5333	0.4476	*0*	*0*	0.6857	0.3523

TABLE 9:VARIANCE OF GENE DIVERSITY.

POPULATIONS	DYS19	DYS388	DYS389I	DYS390	DYS391	DYS392	DYS393
NC INDIA	0.0454	0.0861	*0.0352*	0.0372	0.0736	0.0907	0.0428
NE INDIA	0.0567	0.1274	0.0614	0.0433	0.0614	0.1138	0.1122
NW INDIA	0.0419	0.102	0.0756	0.0516	0.0714	0.1059	0.0607
S INDIA	0.0391	0.0929	0.0624	0.0554	0.1096	0.0849	0.047
BASQUES	0.0815	0.0756	0.0482	0.0742	0.0394	0.0608	0.0745
CATALANS	0.0995	0.1005	0.1062	0.0989	0.0448	0.0552	0.1062
SAHARAWIS	0.1012	0.1006	0.1021	0.0459	0.0983	0.0842	0.093
MZABITES	0.0713	0.0556	0.065	0.0576	0.0662	0.0454	0.0525
ARABS(NW AFRICA)	0.0608	0.0868	0.0599	0.0538	0.0333	0.0646	0.0824
MOROCCAN BERBERS	**0.1637**	*0*	0.1059	**0.154**	0.1581	0.1637	0.1475
TAHELHITS	0.0761	0.089	0.04	0.0593	0.0642	0.0586	0.0843
INGANO	0.1318	**0.1907**	0.1258	0.0889	**0.1662**	**0.1653**	**0.1653**
TICUNA	0.1149	0.0987	0.0566	0.1051	0.0945	0.0445	0.1106
WAYUU	0.1176	0.1475	0.1436	0.0686	0.1059	0.1277	0.1327
KAZAKHS	0.08	0.0836	0.0716	0.0728	0.068	0.062	0.0531
HIG KIRGHIZ	0.0843	0.0843	0.081	0.0863	0.0752	0.0854	0.0598
LOW KIRGHIZ	0.076	0.0796	0.072	0.0827	0.0383	0.0923	0.0836
UIGHURS	0.0571	0.0763	0.0516	*0.0362*	0.0711	0.0631	0.066
FINNS	*0.0349*	0.033	0.0154	0.0174	0.0122	0.023	0.0171
BURUSHASKI	0.0554	0.118	0.0619	0.0427	0.11	0.0439	0.0473
CAR PYGMY	0.0517	0.088	0.0394	0.0744	0.1222	0.1222	*0.046*
CAMBODIA	0.1217	0.1403	0.1382	0.0853	0.0954	0.0853	0.0569
JAPANESE	0.1012	*0*	0.1012	0.0888	0.1541	0.1305	0.1643
MAJANGIR	0.1541	0.0745	**0.1541**	0.1012	*0*	*0*	0.1541
NYANGATOM	0.1539	0.134	0.0833	0.1115	0.1436	0.1533	0.1436
OMEGA SAN	0.0514	0.1278	0.1196	0.0589	0.1327	0.1416	0.0562
SEKELE SAN	0.0614	0.136	0.0724	0.0737	0.0759	0.1358	0.0659
SINDHI	0.074	0.1164	0.0716	0.0514	0.1345	0.1327	0.0585
SOTHO	0.0827	0.1009	0.1084	0.0853	0.1222	0.1295	0.0891
SURMA	0.0679	0.0679	0.1008	0.0749	0.0861	0.0861	0.1617
TAIWANESE	0.1416	0.1448	0.0771	0.0515	0.0515	0.0771	0.1307
TSWANA	0.0867	0.1333	0.0764	**0.154**	0.1343	0.1343	0.1327
ZULU	0.0991	0.1123	0.1259	0.1345	*0*	0.069	0.0615

TABLE 10:RESULTS Of AMOVA ANALYSIS, ITS ASSOCIATED VARIANCE AND Fst VALUES

Percentage of variation

	DYS19	DYS388	DYS389I	DYS390	DYS391	DYS392	DYS393	Average
Among Continent	23.52	-1.57	-3.67	9.24	7.37	7.07	3.78	6.518
Among pop./within Continent	25.22	27.41	17.27	11.22	18.69	21.02	18.42	19.9
Within populations	51.63	74.16	86.4	79.54	73.94	71.9	77.8	73.6

variance components

	DYS19	DYS388	DYS389I	DYS390	DYS391	DYS392	DYS393
Among Continent(Va)	0.10613	-0.00373	-0.01167	0.03551	0.02236	0.0223	0.01115
Among pop /within Conti.(Vb)	0.11561	0.06495	0.05495	0.04311	0.05673	0.06626	0.05432
Within populations (Vc)	0.23671	0.17573	0.27495	0.30564	0.22437	0.22663	0.2294

Fst Values

	DYS19	DYS388	DYS389I	DYS390	DYS391	DYS392	DYS393	Average
Four Continents	0.48366	0.25837	0.136	0.2046	0.26061	0.28097	0.22204	0.26375

All values are significant except for the DYS 388 and 389I loci.

TABLE 11: Pairwise Contingency Table Analysis of Allele Frequency Distribution in Different Regions within Indian Subcontinent at various loci.
(Significant test: P (Probability) < 0.05)

Locus	Population Pair	Unbiased Estimate of Fisher's Exact Probability By Markov Chain Methods		
		Probability (P)	Standard Error	Significant*?
DYS19	NC India & NE India	0.91896	0.00031	No
	NC India & NW India	0.59522	0.00080	No
	NC India & S India	0.38100	0.00091	No
	NE India & NW India	0.69834	0.00063	No
	NE India & S India	0.59709	0.00071	No
	NW India & S India	0.15218	0.00068	No
DYS388	NC India & NE India	0.83920	0.00051	No
	NC India & NW India	0.23001	0.00070	No
	NC India & S India	0.27620	0.00083	No
	NE India & NW India	0.50131	0.00057	No
	NE India & S India	0.23064	0.00049	No
	NW India & S India	0.00797	0.00011	Yes
DYS389I	NC India & NE India	0.14833	0.00062	No
	NC India& NW India	0.30316	0.00066	No
	NC India & S India	0.00024	0.00002	Yes
	NE India & NW India	0.21878	0.00068	No
	NE India & S India	0.00003	0.00001	Yes
	NW India & S India	0.00001	0.00001	Yes
DYS389II	NC India & NE India	0.90519	0.00042	No
	NC India & NW India	0.99841	0.00002	No
	NC India & S India	0.17717	0.00080	No
	NE India& NW India	0.69145	0.00069	No
	NE India & S India	0.03027	0.00033	Yes
	NW India & S India	0.23676	0.00089	No
DYS390	NC India & NE India	0.79235	0.00060	No
	NC India & NW India	0.07447	0.00047	No
	NC India & S India	0.11119	0.00056	No
	NE India & NW India	0.35449	0.00092	No
	NE India & S India	0.29154	0.00090	No
	NW India & S India	0.52390	0.00079	No
DYS391	NC India & NE India	0.36644	0.00086	No
	NC India & NW India	0.93793	0.00022	No
	NC India& S India	0.86137	0.00035	No
	NE India& NW India	0.53727	0.00084	No
	NE India & S India	0.15138	0.00052	Yes
	NW India & S India	0.30114	0.00089	No

(contd.) Locus	Population Pair	Unbiased Estimate Of Fisher's Exact Probability By Markov Chain Method.		
		Probability (P)	Standard Error	Significant*?
DYS392	NC India & NE India	0.41889	0.00102	No
	NC India & NW India	0.31020	0.00084	No
	NC India & S India	0.00789	0.00016	Yes
	NE India & NW India	0.30294	0.00078	No
	NE India & S India	0.01187	0.00017	Yes
	NW India & S India	0.71372	0.00069	No
DYS393	NC India & NE India	0.17245	0.00066	No
	NC India & NW India	0.50140	0.00075	No
	NC India & S India	0.00307	0.00009	Yes
	NE India & NW India	0.24216	0.00077	No
	NE India & S India	0.01889	0.00020	Yes
	NW India & S India	0.08949	0.00051	No

TABLE 12: Pairwise Contingency Table Analysis of Allele Frequency Distribution in Different Regions within Indian Subcontinent across all loci. (Significant test: P (Probability) < 0.05)

Population Pair	Unbiased Estimate Of Fisher's Exact Probability By Markov Chain Method.	
	Probability (P)	Significant ?
NC India & NE India	0.6827	No
NC India & NW India	0.4039	No
NC India & S India	0.0000	Yes
NE India & NW India	0.6111	No
NE India & S India	0.0000	Yes
NW India & S India	0.0000	Yes

TABLE 13: Contingency Table Analysis of Allele Frequency Distribution in Four Regions (NC, NE, NW and S India) within Indian Subcontinent at various loci.
(Significant test: P (Probability) < 0.05)

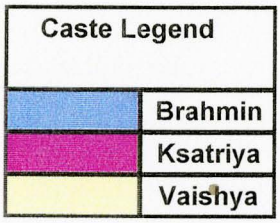
Locus	Unbiased Estimate Of Fisher's Exact Probability By Markov Chain Method		
	Probability (P)	Standard Error	Significant* ?
DYS19	0.60928	0.00105	No
DYS388	0.12845	0.00082	No
DYS389I	0.00001	0.00000	Yes
DYS389II	0.41887	0.00127	No
DYS390	0.20142	0.00102	No
DYS391	0.55161	0.00109	No
DYS392	0.04181	0.00043	Yes
DYS393	0.01250	0.00022	Yes

TABLE 14: Contingency Table Analysis of Allele Frequency Distribution in 31 World Populations (Burushaski,CAR Pygmy, Cambodia, Japan, Majangir, New Guinea, Nyangatom, Omega San, Sekele San, Sindhi, Sotho, Surma, Taiwan, Tswana, Zulu,Finns, Kazakhs, Highland Kirgiz, Lowland Kirgiz, Uighurs, Saharawis, Mzabites, Arabs from NW Africa, Berb, Tahelhits, Ingano, Ticuna, Wayuu,Basques, Catalans and Indians) at various loci.
(Significant test: P (Probability) < 0.05)

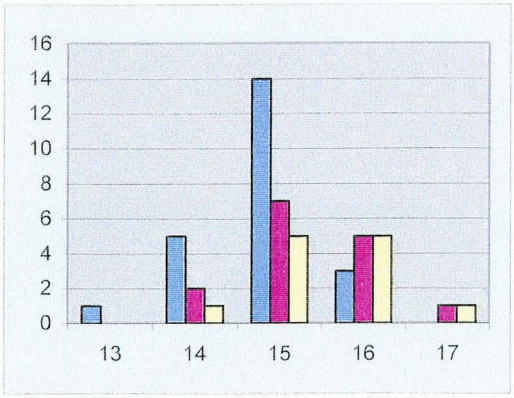
Locus	Unbiased Estimate of Fisher's Exact Estimate By Markov Chain Method		
	Probability (P)	Standard Error	Significant* ?
DYS19	0.0000	0.0000	Yes
DYS388	0.0000	0.0000	Yes
DYS389I	0.0000	0.0000	Yes
DYS390	0.0000	0.0000	Yes
DYS391	0.0000	0.0000	Yes
DYS392	0.0000	0.0000	Yes
DYS393	0.0000	0.0000	Yes

Figure 18: Distribution of Allele Frequencies of the Different Y Chromosomal Microsatellite Markers In The Different Castes.

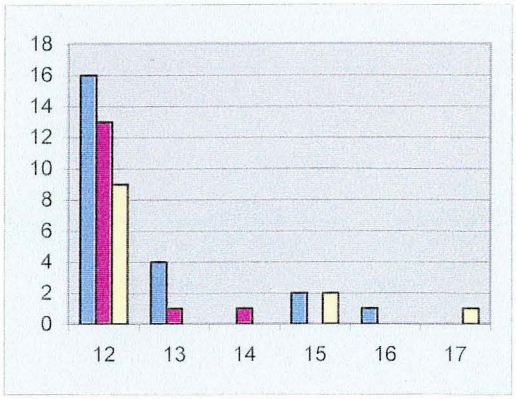
Y axis= Number of Individusals
X axis=Allele Size.



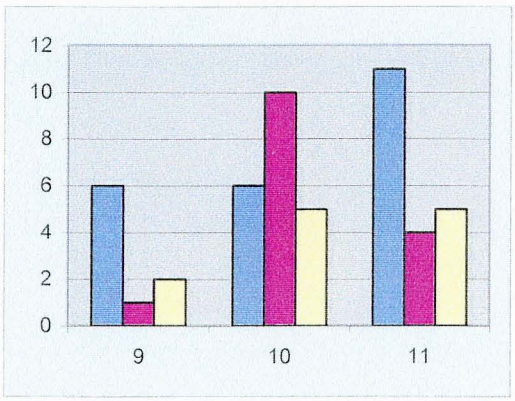
DYS19



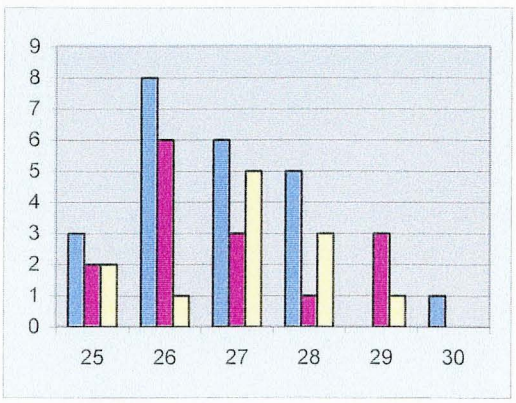
DYS388



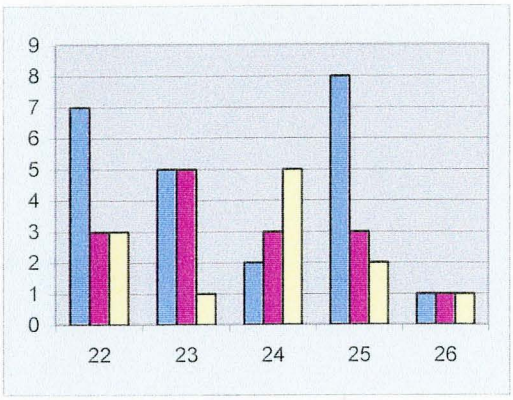
DYS 389I



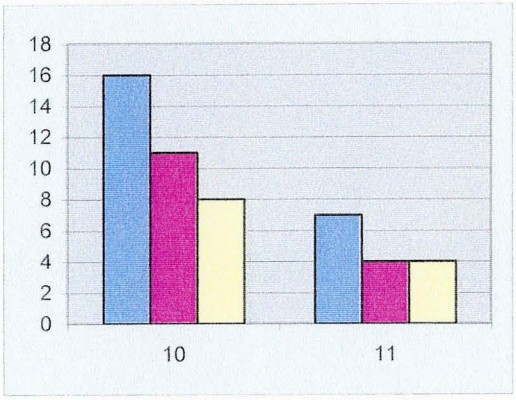
DYS389II



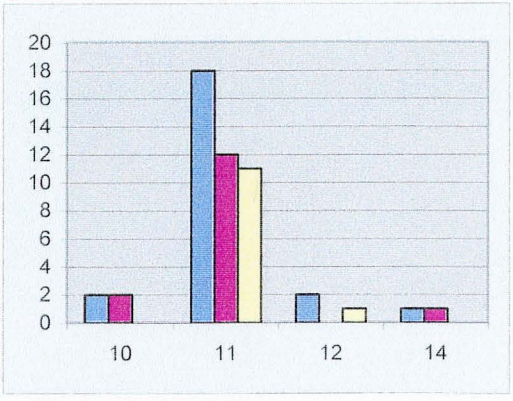
DYS390



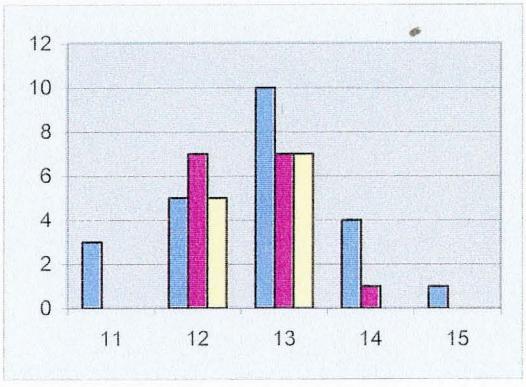
DYS391



DYS 392



DYS393



by pairwise contingency table analysis and the results are presented in Tables 11 through 14. The null hypothesis was H_O: no heterogeneity of allele frequency in populations. To test the null hypothesis, the unbiased approximation of the exact value of type-one probability (P, obtained by Fisher's exact test) was computed by a Markov chain method (see the section, Contingency Table analysis before). Based on 95% confidence intervals ($P < 0.05$), Tables 11 though 13 report the presence or absence of heterogeneity of allele frequency among populations within the Indian subcontinent. Table 14 shows significant heterogeneity ($P < 0.05$) of allele frequency in each of the seven loci among 31 world populations, which is obvious. Dememorization steps are necessary in the Markov chain method in order to obtain unbiased estimates of Fisher's exact P-values. Guo and Thompson (1992) reported that if the chain starts from an arbitrary state (e.g. observed state) instead of a random state, the estimate would be biased. Dememorization forces the chain to start from an initial arbitrary state (such as observed state) and run for a long time in order to forget its initial state. Thus, via the process of dememorization prior to the computation of estimates of Fisher's exact P-values, the chain is forced to start from a random state. Initially, a small number of dememorization steps (5000) were used for the Contingency Tables 11 through 14. Though 5000 dememorization steps were sufficient for the contingency Tables 11 through 13, it produced heterogeneity (significant, $P<0.05$) of allele frequency in each of the seven loci among 31 world populations in Table 14. In spite of increasing the number of dememorization steps from 5000 to 30000 (upper limit in GENEPOP), the results in Table 14 remained unchanged.

Phylogenetic trees based on Cavalli-Sforza chord distance method and the proportion of shared alleles distance models shows similar results (Figure 19 and 20). Goldstein's distance model based on Single Step Mutation gave a somewhat different result (Figure 21). In the Cavalli-Sforza method the four Central Asian Populations of Kazakhs, Highland and Lowland Kirgiz and Uighurs group together. This is not the case with the other models where Uighurs groups differently and goes with the Indian and other West Asian populations like the Burushaski. All the African populations are not as closely related as might have been expected in the methods based on these models. Most notably, the Northwest African populations of Arabs, Mzabites, Tahelhits, Saharawis and Moroccan Berbs always separate out in all the three models. Also the Native American populations of Ingano, Ticuna and Wayuu always group together in all three models and in some models also group with some Non- African populations. In the Cavalli-Sforza and D_{AS} models the Indian populations appears to be closer to the other Central Asian populations but in the Goldstein the Northcentral Indian population groups with the three Central Asian populations of Kazakhs, Highland and Lowland Kirgiz. The South Indian population group with the Sindhi population from West Asia in all the three cases. The Spanish population of Catalans and Basques group together in the Cavalli-Sforza and Stephens models but shows a different clustering in the Goldstein's model. Among the East Asian populations, Japan is closer to Indian Populations than any other East Asian populations namely, New Guinea, Taiwan or Cambodian.

The range of the bootstrap values is different in the three distances. Although the bootstrap values are not significant, some high bootstrapped values are evident for the

NW African and the Spanish populations in the Cavalli-Sforza and Stephens distances. Bootstrapped values are high with the Cavalli-Sforza and D_{AS} distance for the South Indian and Sindhi population from West Asia but is 48.3% with the Goldstein's model. Bootstrapped values are high for the Central Asian population of Kazakhs, Highland and Lowland Kirgiz when Goldstein's distance is used. Overall the bootstrapped values are lower when based on Goldstein's distance model than the values that are obtained with the other distance method and thus the result obtained with the Cavalli-Sforza and Stephens method differ from that of the Goldstein results.

Figure 18 is discussed in detail in the next section.

Figure 19: **Fitch–Margolish Tree based on Cavalli-Sforza Chord Distance Method on the Different World Population with that of the Indian Population** (Bootstraps out of 1000)

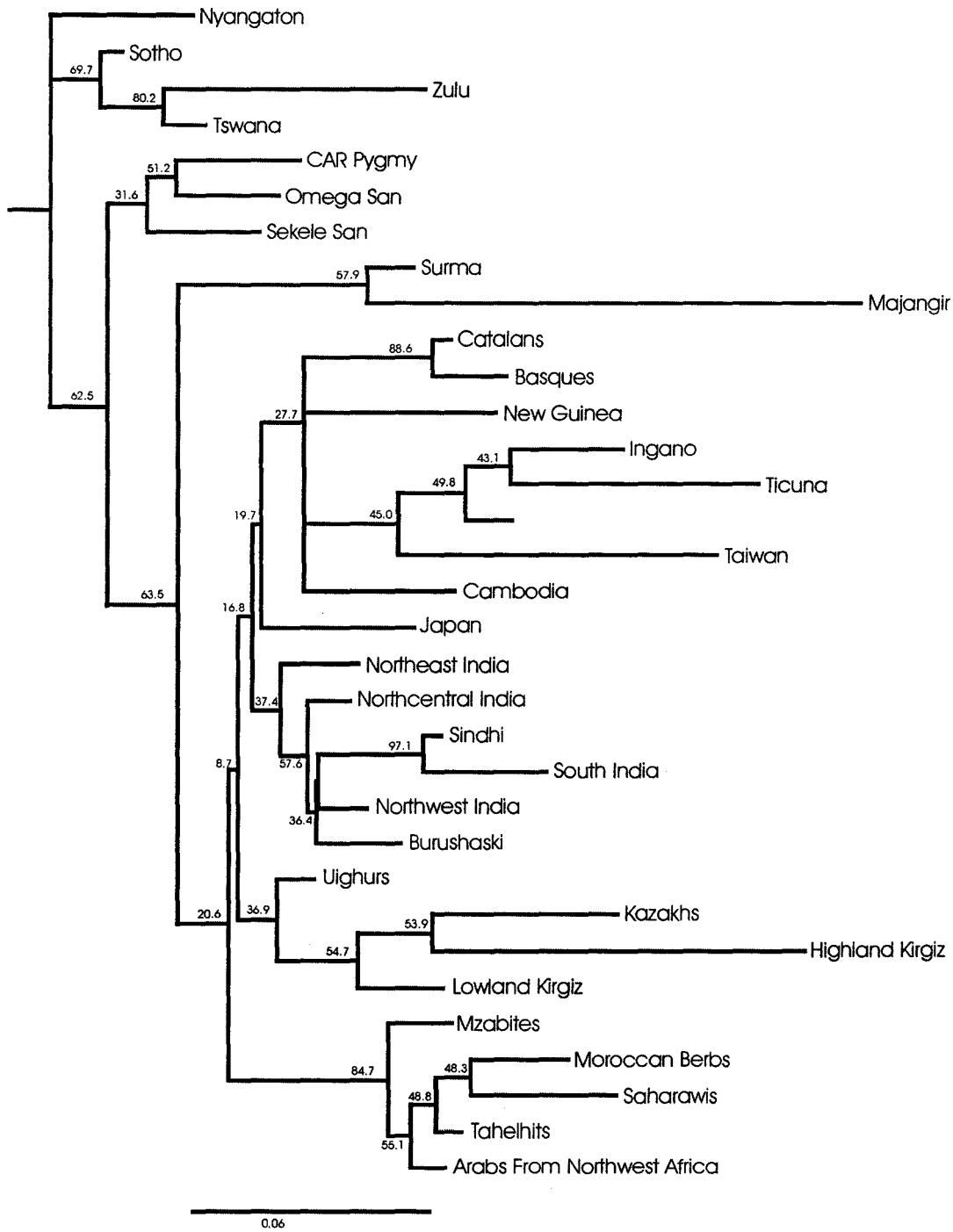


Table 20: Fitch–Margolish Tree Based On Proportion Of Shared Alleles on the Different World Populations with that of the Indian Populations.
(Bootstraps out of 1000)

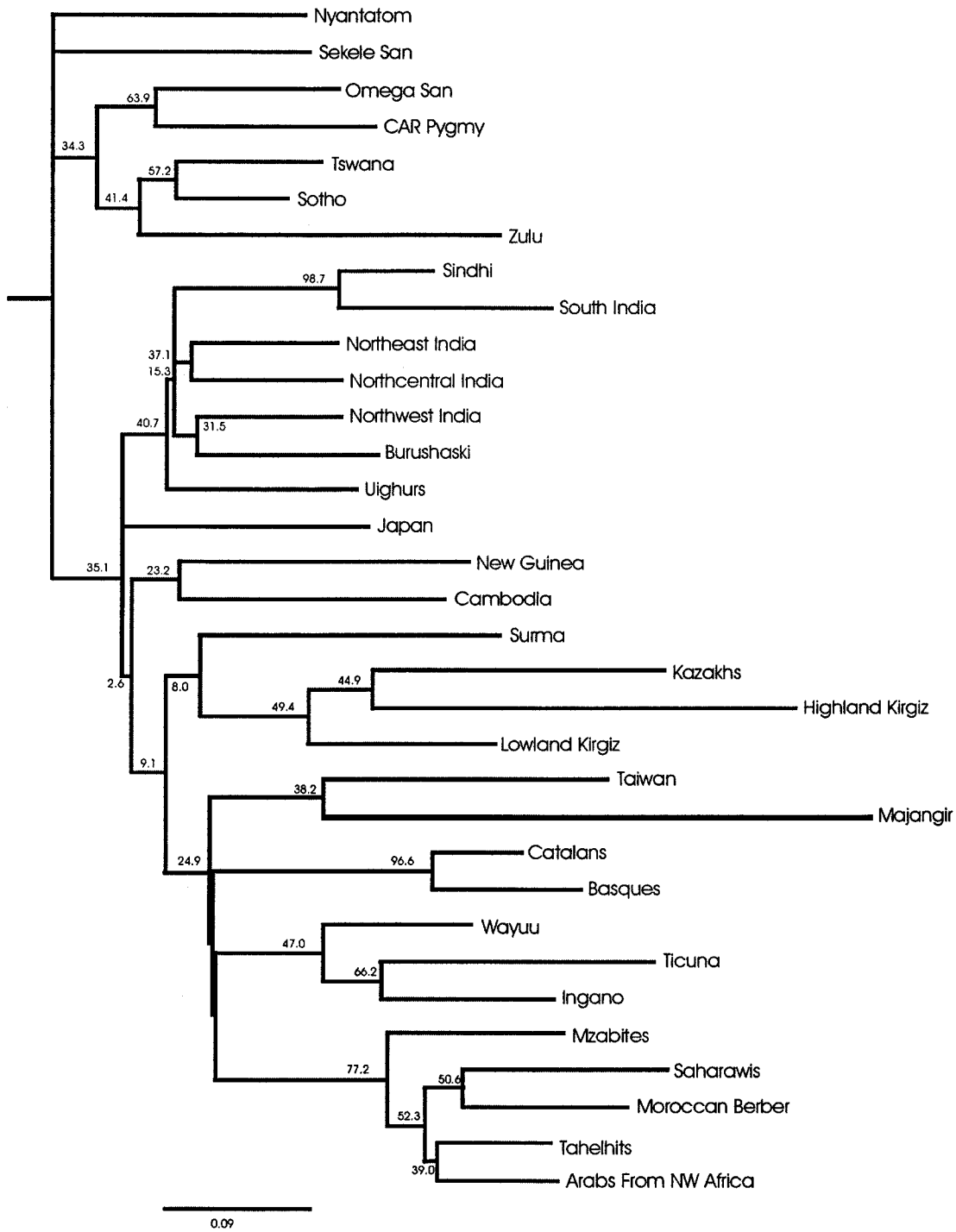
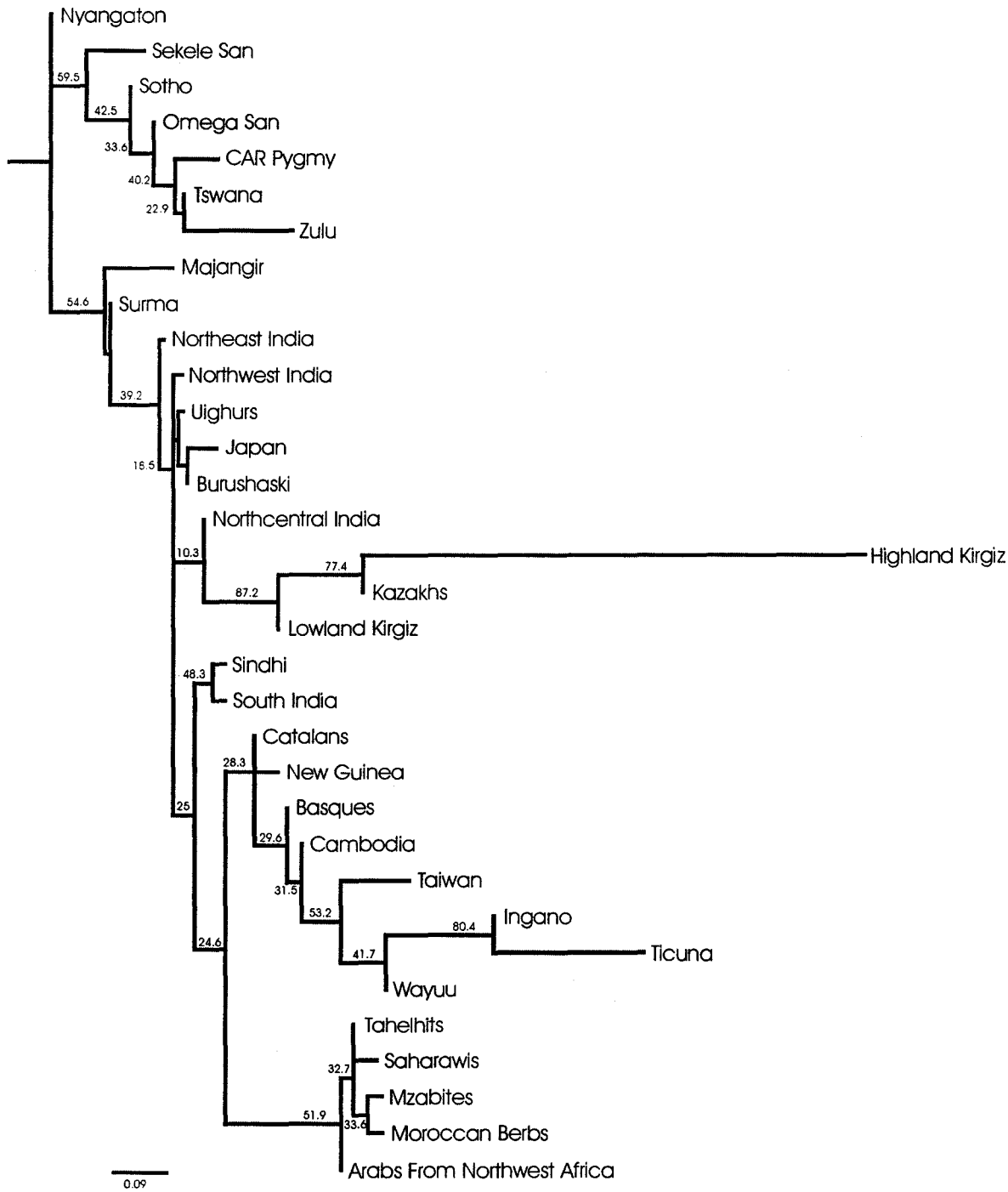


Figure 21: Fitch- Margolish Tree Based On Goldstein (delta μ)^2 Distance Method on the Different World Populations with that of the Indian Population (Bootstraps out of 1000).



DISCUSSION

Haplotype Sharing between India Populations

Extensive Y chromosome diversity was found in a study by Majumder *et. al.* (1999) who constructed 81 distinct haplotypes using 6 markers from 125 Indian samples (similar results was found in this study). A total of 97 haplotypes were constructed in this study from 113 samples using 8 markers. Sharing of haplotype within the same region is evident in this study. This study shows that the sharing of haplotypes (Figure 7) within different regions of India is minimal. Only 7.2% haplotypes are shared between the different regions. The frequency of haplotype sharing is 0.01 between Northcentral - South India, Northcentral - Northwest India and Northcentral - Northeast India and 0.02 between Northwest - South India. The frequency of haplotype sharing between three regions like the Northwest, Northeast and South India and between Northcentral, Northwest and Northeast India is also 0.01. Sharing of haplotypes between different regions is not uniform with some regions enjoying a comparatively (of the small haplotypes shared) higher frequency of a particular haplotype than other region.

Minimal sharing of haplotypes (Figure 8) is also evident among the different religions found in India. Of the total 96 distinct haplotypes formed among the different religions only 6 (6.25%) haplotypes are shared. Hindus shares haplotypes with the Sikhs, Muslims and Christians. Highest sharing of haplotypes is found between the Hindus and

the Sikhs as opposed to other religions (0.03 as compared to 0.01 haplotype sharing between Hindus and Muslims and Hindus and Christians). This is not surprising as it was the customs in Punjab that made it compulsory to convert the first born son from Hinduism to Sikhism for a long time. Most of the haplotype sharing exists between Hindus and the other major religions. This might be because the original religion of India was Hinduism and historically speaking the conversion to other religions occurred mostly among the Hindus than otherwise. Thus minority religions contain subsets of haplotype from the major religious group. The only other instance of haplotype sharing is between Muslims and Christians (0.01). Both these religions have been introduced in India recently and thus haplotype sharing is low. Since the sample size in these cases is small, there might be some amount of bias in haplotype sharing.

Of the total 50 samples that were classified into different castes 44 distinct haplotypes (Figure 9) were obtained of which only 3 (6%) were shared among the different castes (the upper caste Brahmin shared one haplotype with the middle caste Ksatriya and another with the lower caste Vaishya; and 1 haplotype is shared between the Ksatriya and Vaishya). Thus, there is minimal sharing of haplotypes among the ethnic groups studied, as has been recently suggested (Majumder *et. al.* 1999). However, since the sample size is small a certain amount of bias arising from small sample size cannot be ignored. The sample also does not contain any tribal populations and this could also affect the results and thus must be taken into account. Sharing of the same haplotype within the same ethnic group also occurs.

Sharing of haplotypes between the Indian and world populations is also minimal. The number of haplotyes formed between India and other world population is too large that it is impossible to show any form of relationship between the populations.

Allele frequency Distribution within India and Other Population

There exist differences in the distribution of alleles of the different markers between different regions of India especially between the North and South India as seen in contingency table 11, 12 and 13. This is true even when each marker is compared between the different regions (Table 11), or when all the markers taken together is compared between the different regions (Table 12), or when individual markers are analyzed over the four regions in the Indian population (Table 13) where DYS 391, 392 and 389I are significant. Even when Sequential Bonferroni Correction was done on the P values obtained from the Fisher's test there were not significantly different from the original values. A similar study (Lahermo *et. al.* 1996) showed that the allele frequency distribution in two Finno-Ugric speaking populations, the Finns and the Lapps were significantly different for all mini and microsatellite markers used in that study. These two populations were culturally and linguistically separated, as is the case with the Aryans (North) and the Dravidian (South) speaking groups in India. Furthermore it has been shown that North South clines exist for mitochondrial haplotype frequencies (Passarino *et. al.* 1996) as well as for blood markers, a Northwest to Southeast cline exists. Skin color differences that exist between North and South India have also been

used as identifiable genetic differences between the two populations (Papiha 1996). However, this may have resulted from adaptation.

Thus, the existence of significant differences in allele frequency distribution between the Northern and the Southern population points to the fact that the Indian population is not a homogenous entity and that the Northern population differs from its Southern counterpart. These results may provide support for the popular belief that the Dravidians were the original inhabitant of India and the Aryans later displaced them (discussed later in details). The most recent view about these two Indian populations, is that, both the Dravidians and the Aryans migrated from Central Asia but the timing of these two migration were separated by some 10,000 yrs BP and were thus Central Asian populations. Thus, the significant differences that exist in allele frequency distribution might reflect this difference in the time of migration.

Allele size distribution of the different ethnic groups in Indian population is shown in Figure 18 and the allele frequency distribution is shown in Table 7. Not all alleles are present in all ethnic groups. This might also be due to small sample size in some of the groups. Brahmin seems to harbor most of the alleles present at different loci. The most frequent allele of a locus is present but not in equal magnitude in all the ethnic groups. Previous studies indicate that the highest level of haplotypic diversity exist in the middle caste (Ksatriya) attesting to the fact that historically the social boundary of the middle caste was not that rigid (Majumder *et. al.* 1999). This was not supported apparently in our study, as most of the alleles seem to be present in the different groups.

This might be due to a small sample size of the ethnic populations (the populations used for comparison have 23 Brahmin, 15 Ksatriya and 12 Vaishyas).

India and Other World Populations

DYS 271

The Indian samples were found to be monomorphic having an A allele at position 168 for the biallelic marker DYS 271. The A to G transition is found mostly in African samples (Seielstad *et. al.* 1994). This biallelic marker has been shown to exhibit 'maximum linkage disequilibrium' with Alu insertional polymorphism (YAP - Y Alu polymorphic element- presence or absence of the insertion in the Y chromosome). Y chromosome with a G at position 168 in the marker DYS271 without YAP insertion has not been observed yet (Seielstad *et. al.* 1994). Studies in our laboratory for the presence or absence of YAP insertion found it to be absent in all Indian samples but one (unpublished result, Tampi and Singh). This is in agreement with recent studies done (Majumder *et. al.* 1999). The frequency of YAP+ insertion is highest in Sub-Saharan Africans followed by Northern Africans, Europeans, Oceanians and Asian populations. Among the Asians, high frequency of this insertion is found in the Japanese populations (Spurdle *et. al.* 1994, Passarino *et. al.* 1998, Seielstad *et. al.* 1994, Hammer 1994, Hammer and Horai 1995). Studies on Austroasiatic tribal populations of India have shown that human specific Alu elements are present in the nuclear genome at frequencies similar to that of African populations (Majumder 1998, Stoneking *et. al.* 1997). The Austroasiatic tribal population of India resides in the Central and Northeastern region of

India. The individual that had the YAP insertion came from the Northeastern region of India. Other than the individuals location no other information is available on his ancestry and so no conclusions can be drawn. The present day Indian population consists of about 8% tribal population. The tribal population of India has been considered by many to be the first inhabitant of the subcontinent (Cavalli-Sforza *et. al.* 1994). As this study lacks tribal samples, the possible link between Indians and Africans could not be explored and future sampling of these populations might yield interesting results. The absence of the YAP element in the general population and the presence of A at loci DYS 271 in the Indian population points to the fact that Indians are more closely related to the Europeans than to the African populations (Majumder 1998).

Genetic Diversity

As compared to other species like the Chimpanzees, genetic diversity is low in human populations. This has also been proved by protein polymorphism and craniometric analysis (Relethford and Harpending 1994, Lewontin 1972). This points to the recent origin of human population and a small size of ancestral population (Li and Sadler 1991, Kaessmann *et. al.* 1999, Crouau-Roy *et. al.* 1996).

A study based on protein polymorphism has shown that diversity among human population is only 15% (Lewontin 1972, Nei and Livshits 1990). Genetic diversity studies on mtDNA (Merriwether *et. al.* 1991), autosomal polymorphism (Jorde *et. al.* 1995, Bowcock *et. al.* 1994, Deka *et. al.* 1995, Stoneking *et. al.* 1997) and Y Chromosomal polymorphism (Hammer *et. al.* 1997) points to the fact that most genetic

diversity exists within human populations and not among populations, as is also seen here with AMOVA. Diversity values shown in table 8 are consistent with published values, with African populations showing some of the higher diversity values (Jorde *et. al.* 2000). Indian populations also show high diversity values. This is in accordance with published results on serum proteins and enzyme polymorphisms, which have shown that ethnic population of India has higher genetic diversity than populations from other geographic regions (Majumder 1998). Higher diversity might not necessarily imply that the population is ancient but might also point to past demographic events such as larger effective size of the population and lack of bottlenecks (Relethford 1995, Relethford and Jorde 1999). As expected, within population variation is also higher than between continents. Slightly negative values are obtained for between continent variations. This is because the estimation of variation between continents is obtained by subtraction (Table 10) (Jorde *et. al.* 2000).

Phylogenetic Relationship

This study also shows that India is more similar to Central Asian populations followed by Europe and East Asia (figure 19 through 21). Interestingly in all three figures the South Indian population is closely related to the Sindhi population belonging to the Sind province of Pakistan. The Sind province has been known to be an important archeological site for the Indus Valley Civilization. The phylogenetic relationship might signify a close tie between the South Indians and the Sindhi group of people from Pakistan. It is believe that the older inhabitants of India were displaced by newer

immigrants from Central Asia. The older inhabitants, having been displaced moved and settled in the Southern part of the continent. The Sindhi people might represent this remnant population. However, it must be considered that there might be a misrepresentation in this phylogenetic relationship because of small sample size of these two populations (the Sindhi sample consist of 15 individuals and the South Indian samples consist of 24 individuals). Phylogenetic relationship also exists between populations from Northern India with Uighurs, Lowland Kirgiz, and the Kazakhs populations from Central Asia. These relationships strengthened the hypothesis that the original inhabitants from Central Asia moved and possibly came to settle in India in two successive waves. The older Dravidian population of India might represent an older migration from Central and West Asia (Kivisild *et. al.* 1999). The linguistic differences that exist in India between the North and the South, the North predominantly speaking an Indo-European language and the South speaking the Dravidian dialet is interesting. However studies has shown that the Dravidian language shows similarities with the Uralaltaic language groups that include the Turkish, Elamite, Mongolian and Tungusic language groups as well as Brahui language of Baluchistan and might also include the Japanese and Korean languages (Zvelebil 1990, Renfrew 1989). These languages are mainly spoken by the people living in Central and West Asia as well as by people from Southwest Iran. So taken together all these evidences, existence of phylogenetic relationship between the Indians and the Central Asian populations might not be surprising. These studies thus, support the hypothesis that there might be two migrations of late Pleistocene humans to India (Kivisild *et. al.* 1999). The first wave might have

brought in the ancestors of the modern Dravidian people and the second wave the Indo-Aryans. It is also interesting to note that the later group also probably left their homeland around the Black and Caspian Sea lying between East Europe and Central Asia (Quintana-Murci *et. al.* 2001). Thus, these results are also consistent with the historical evidence that two separate migrations contributed to the present day gene pool of the modern Indian population (Barnabas *et. al.* 1996). The fact that the ancestors of Dravidians population of India whose original homeland was also Central Asia might be the population that was responsible for Indus Valley Civilization is supported by Quintana-Murci *et. al.* (2001). mtDNA haplotype analysis in our own laboratory has also supported the claim of this thesis that India is more closely related to Central Asian populations of Uzbekistan and Kazakhstan (unpublished results, Thomson and Singh).

The Northwest Indian population also shows phylogenetic relationship with the Burushaski. These groups of people inhabit three rugged mountainous region of Northwest Pakistan - Hunza, Nagar, and Yasin valley. The exact origin of these people is still under speculation (website-Sadie Anderson). Popular legends have it that soldiers of Alexander settled in this region. Another legend says that Burusho people were driven from Northwest India by Indo-Aryan warriors. The language spoken by the Burusho is not related to any known language of the world. A phylogenetic relationship between this group and individuals from Northwest India might signify a genetic relationship, as Northwest India has been a known passageway for migrants from Europe to India. The sample size for these two groups are as large as most of the other populations in the study (Burushaski is represented by 24 individuals and Northwest India sample in represented

by 29 individuals). It is also seen that the general Indian population shows relationship with European populations of Basques and the Catalans followed by Southeast Asian populations. In this study, only two populations from Europe - the Basques and the Catalans from Spain, have been used to show the phylogenetic relationship of Europe with India. The Basques people speak a language that is not related to any known Indo-European Language and is considered to be a pre-Indo European language (website-FAQs about Basques and the Basques). The Catalans belongs to the Hispanic group of languages. These two European populations are not an ideal representative of all the present European populations. And as there is no Y chromosomal microsatellite data of the markers used in the study, available for other populations in Europe, it was not possible to explore further the relationship of Indian populations with Europe. However previously published linguistic, classical markers, mitochondrial as well as Y chromosome biallelic studies have shown that such a relationship does exists (Cavalli-Sforza *et. al.* 1994, Roychoudhury and Nei 1985, Barnadas *et. al.* 1996, Kivislid *et. al.* 1999).

Hammer (1998) has also shown using Y chromosome biallelic markers that the Mongolian populations followed by West Siberian populations displays the most similarities with South Indian populations. Another study on mtDNA and nuclear DNA has placed Indians between Caucasoid and Mongolian populations (Majumder 1998). Studies in our own lab also points to a possible relationship between India and Siberia (unpublished results, Thomson and Singh). In the absence of any published Y

chromosome microsatellite data on Mongolian and Siberian populations, this study could not shed any light whether such a relationship does exist.

In this analysis the Southeast Asian population was represented by Japan, Cambodia, New Guinea and Taiwan. Based on phylogenetic studied of these Southeast Asian countries India displays most similarities with the Japanese population. The exact origin of the Japanese population is controversial and is under study (Piazza 1998, Oota *et. al.* 1999). Linguistic similarities have also been shown between the Dravidian people in the South and the Japanese (Zvelebil 1990). This might imply a possible relation of the Indian population with that of the Japanese population. Similarity existing in these two populations is thus interesting and further studies need to be done. But the Japanese sample size is not high (only 10 individuals) and so no definite conclusion can be reached.

The migration route of the modern human population to Southeast Asia has been proposed to be through India along the Southern Coast and not through a 'Northern Route' through Central Asia (Su *et. al.* 1999, Chu *et. al.* 1998). Since the Indian population exhibit similarities with Central Asian population and less so with Southeast Asian populations, then for the 'Southern Route' hypothesis to hold true, the original genepool of the Indian population during the 'Out of Africa' era might have been totally replaced by more recent migration from Europe and Central Asia. But the lack of tribal population in this study and reported presence of genomic YAP+ insertion in Austroasiatic tribal population in India in frequency similar to the Africans, and the presence of Y *Alu* insertion in Japanese and Africans but not in Indian population (tribal

population not included in the study) needs to be further examined. Studies have also reported on the similarities between the Australian Aborigines and the tribal populations of India. Thus further study on the tribal population could shed light on the proposed India-Australia migration theory of Redd and Stoneking (1999) and also to the proposal by anthropologists that the tribals are the oldest inhabitants of the continent (Cavalli-Sforza *et. al.* 1994). The 'Northern Route' hypothesis is well supported in this study due to the similarities that exist between the Indian and the Central Asian populations.

Religious and Caste differentiation With Indian and Comparison with other World Populations

As noted earlier very little haplotype sharing is evident between the different religious and caste groups in India. As well whatever sharing there is, is mainly between the Hindus and the other religions. Since the effective population size of the Y chromosome is only ¼ of the effective autosomal population this may be due to drift effects. Little haplotype sharing may also show that the prevailing cultural norm prevents any male gene flow across boundaries. Phylogenetic analysis of the caste group and different world populations (Figures not shown) shows that the three caste groups are closer to the Central Asian populations of Uighurs, Kazakhs and Lowland Kirgiz and also to the Buruskaski population from Northwest Pakistan. This is not unexpected as the caste system was created by the Aryans, and since the Indo-Aryan population of India is believed to have originated from Central Asia such similarities lends further support to the hypothesis. Since the European population used for the study may not be a true

representation of Europe, we were unable to confirm Watkins *et. al.* 1999 claims that castes in India are closer to European populations. Another interesting point that emerges is that the Brahmin seems to be closer to the Sindhi population from Pakistan. It has already been shown that the South India population enjoys similarities with the Sindhi population. Thus, such a similarity existing between the Brahmin and the Sindhi population is at first seems to be counterintuitive, but if the Dravidian population originated from Central Asia (as is supported by linguistic and molecular markers), such a possible correlation might not be unexpected. Our samples did not contain the lowest caste groups, the Sudras and the Untouchable and so a full understanding of the intrinsic relationship of the caste with each other as well as with the other world populations is not possible at this time. The claim of Majumder (1998) that the Brahmins forming the upper caste group are closer to the Muslim population from India proving that conversion from Hinduism to other religions occurred even from higher caste (and was not restricted to lower caste) is not supported in our study.

From the phylogenetic analysis of the Y chromosome microsatellite markers of the different religious groups in India with that of the World populations does not show any significant results. The Sikhs in most cases group with the Central Asian populations and the Christians with the Sindhi population of Pakistan. The pairing of the Sikh population from the Northwest part of the Subcontinent with the Central Asian Population and the pairing of Christians with the Sindhis probably reflects geographic location of the samples and that religious conversion rates differing by regions and not because of any phylogenetic differences intrinsic to the groups. This is further supported

by the fact that 70% of Christians in India live in the four Southern provinces and about 85% of Sikhs live in Punjab (Gupta, 1961).

CONCLUSION

The Indian population appears to be more closely related to populations from West Asia and then to populations from Central Asia followed by Europe. Of the Southeast Asian countries, India shows similarities with the Japanese population. Contingency analysis showed that the two main sections of the Indian populations, representing the Indo-Aryans in the North and the Dravidians from the South, does differ from each other as far as allele frequency distribution of Y Chromosomal microsatellite markers are concerned. But the phylogenetic similarities existing between the Sindhi and the South Indian populations as well as the linguistic similarities existing between the Dravidians and Uralaltaic populations of Central Asia points to the fact that the Dravidian population of India might represent an older migration from Central Asia. Phylogenetic relationship also exists between Northern Indian population with that of Central Asian population of Uighurs, Kirgiz and Kazakhs. Thus it is possible that two different migrations from West and Central Asia into the Subcontinent might be responsible for the present Indian population. Thus, the differences in the distribution of alleles between the Northern and Southern Indian populations might be due to the difference in the time of migration of these populations from Central Asia. As a result the Dravidic population of South India do not differ greatly from the Indo-Aryans population from North India whose original homeland is also believed to be Central Asia.

Population from Northwest part of India is not significantly different from Northcentral or Northeast Indian populations implying that whatever migrations that

might have taken place from the Northwest part of the country must have dispersed over time to the rest of the country. Religious groups appear to be the subsets from the general populations with respect to microsatellite markers. Some of the caste groups shows similarities with Central Asian populations possibly attesting to the fact that the caste was originally an Indo-Aryan concept and became popular with the advent of a more recent migration from Central Asia.

Appendix A: Solutions And Buffers

Binding Solution:

Concentration	0.5%
Glacial Acetic Acid	1 mls
95% Ethanol	199 mls

To 1mls of this solution add, 3µl of Methacryloxypropyltrimethoxysilane (Sigma). Store methacryloxypropyltrimethoxysilane at 4^0C.

10% Ammonium persulfate:

Ammonium persulfate	0.5 gms
Double-Distilled Water	5 mls

Loading Solution:

Concentration	2X
NaOH	10mM
Formamide	95%
Bromophenol blue	0.05%
Xylene cyanol FF	0.05%

Fixative Solution:

Concentration	10%
Glacial Acetic Acid	200 mls
ddWater	1800 mls

Staining Solution:

Silver Nitrate	2 gms
ddWater	2000 mls
37% Formaldehyde	3 mls

Developing Solution:

Sodium Carbonate	60 gms

ddWater	2000 mls

Just before use add,

37% Formaldehyde	3 mls
Sodium Thiosulphate (10 mg/ml)	400µls

TBE :

Concentration	10X
Amount	1 litre
Autoclave	Yes
Tris base	108 gms
Boric Acid	55 gms
0.5M EDTA (pH 8.0)	20 mls

Add the above ingredients in 800 mls of water. Then bring the volume to 1 litre with ddwater. Autoclave.

EDTA:

Concentration	0.5 M
pH	8.0
Autoclave	Yes
Disodium Ethylenediaminetetraacetate.$2H_2O$	186.1 gms
Water	800 mls

Concentration	50mM
pH	8.0
Autoclave	Yes
Disodium Ethylenediaminetetraacetate.$2H_2O$	18.61 gms
Water	800 mls

Adjust pH with NaOH.
Bring the volume to 1 litre.

Arylamide Solution:

Concentration	6%
10X TBE	7.5 mls
Urea	36.03 gms.
Acrylamide: bisacrylamide(19:1)(Biorad#161-0144)	11.25 mls

Bring the volume to 75 ml. The freshly prepared acrylamide solution is filtered by using a 0.2 micron filter.

Just before pouring the gel, 500µl of freshly prepared ammonium persulfate is added followed by 50 µl of TEMED.

Luria-Bertani Medium:

Autoclave	Yes
ddWater	950 mls
Bacto-tryptone	10 gms
Bacto-yeast extract	5 gms
NaCl	10 gms

Adjust pH 7.0 with,

5N NaOH	200µls

Adjust volume to 1 litre

LB plates:

Autoclave	Yes
Bacto-agar	7 gms / litre

.

To this add 1 litre of LB medium and mix. Cool the mixture to room temperature and add the appropriate antibiotics.

$CaCl_2$

Autoclave	Yes
pH	7.0
Storage	4^0C
60mM $CaCl_2$	12 ml 0.5M stock Solution
15% Glycerol	18.75 ml 80% stock Solution
10mM PIPES Buffer	1 ml of 1M stock Solution

Bring to volume 100 mls with ddwater.

Appendix B: Protocols and Cycling Conditions

DNA extraction Procedure:

Genomic DNA was extracted from 116 Indian males using the Puregene Genomic DNA isolation kit (Gentra).

1. To 300 µl of blood, 900 µl of the RBC lysis solution was added to lyse Red Blood Cells in a 1.5 ml microfuge tube. The tube was inverted to mix and incubated for 10 minutes during which time it was inverted once.
2. The above solution was then centrifuged at 13,000-16,000 g for at least 20 seconds. The supernatant was then remove leaving behind the white pellet in about 10–20 µl the residual solution.
3. The tube containing the pellet in the solution was then vortex to resuspend the cells. This facilitates cell lysis.
4. To lyse the White Blood cells, 300 µl of Cell Lysis Solution was added to the above solution and was pipetted up and down to lyse the cells. If clumps of cells are still visible after this procedure the solution is incubated at 37^0C until the solution became homogeneous.
5. To this cell lysate add 1.5µl RNase A solution. The solution was inverted repeatedly to mixed and incubated for 15-60 minutes at 37^0C.

6. The samples were cooled to room temperature and 100µl of Protein Precipitation solution was added to it. The samples were vortexed at high speed for 20 seconds. The samples were then incubated on ice for 5 minutes. The protein was then precipitated by centrifuging at 13,000-16,000 x g for 3 minutes. The precipitated protein formed a tight brown pellet.
7. The supernatant was then transferred into a fresh microfuge containing 300 µl of 100 % isopropanol. The samples were inverted to mix about 50 times. It was then centrifuged at 13,000-16,000 x g for 1 minute to precipitate the DNA. The DNA was visible as a white pellet.
8. The supernatant was poured and the tube poured briefly on clean Kleenex paper.
9. The precipitated DNA pellet was then rinsed with 300 µl of 70 % ethanol. The sample was then centrifuged at 13,000-16,000 x g for a minute and the supernatant drained. This procedure was repeated twice.
10. The DNA was then dried and and dissolved in 30 µl of autoclaved water.
11. The DNA samples were quantified using a flurometer. The aliquots of the concentrated DNA samples were diluted to 20 ng/µl using double distilled water. These diluted samples were used for Polymerase Chain Reaction.

Protocols For the Preparation Of pBluescript Plasmid:

A. Preparation of Competent Cells:

1. *E.coli* strain XL 1 blue was steaked on LB plates.

2. The plates were incubated at 37^0C overnight (not more than 16 hours).
3. A single colony was picked from the overnight culture and was grown in 5 mls of 1X LB at 37^0C with constant shaking. Care was taken to prevent anaerobic growth of the bacteria.
4. Next day 50 µl of the overnight culture was transferred to 50 ml of 1X LB and grown to a minimum $OD_{600} = 0.375$. It was grown to 0.4.
5. The cells were chilled for 5 min. Pellet cells in culture bottles for 7 min at 3000rpm (at 4^0C) without the brakes on to prevent resuspending.
6. Resuspend in 10 ml ice cold $CaCl_2$ solution and leave on ice for 30 min
7. Pellet cells for 5 min at 2500rpm and 4^0C without brakes.
8. Resuspend in 2 ml of ice cold $CaCl_2$ solution
9. Dispense into 400 µl aliquots and store immediately in -80^0C.

B. Transformation of Competent Cells:

1. The XL1 blue competent cells are thawed on ice for about 10 min.
2. Into prechilled 6 mls falcon tubes 100 µl of cells are transferred.
3. 10 ng of pBluescript plasmid DNA is then added and is incubated for 30 min.
4. The cells are then put into water bath (42^0C) for 90 sec without shaking to heat shock.
5. The cells are then kept on ice for 2 min for recovery.
6. The cells were grown in 900µl of 1X LB for 45 min at 37^0C on a rotary shaker to allow for the expression of antibiotic resistance gene.

7. The cells were then steaked on LB plates that have the antibiotic ampicillin (100ng/ml).
8. The cells were grown from a single colony in 250 mls of 1X LB (containing 250µl of ampicillin) overnight.
9. DNA was extracted using the Qiagen Plasmid purification kit.

C. Extraction of the Plasmid (adapted from Qiagen midi Plasmid Purification Kit):

1. The bacterial cells were harvested by centrifugation at 6000 x g for 15 minutes at 4^0C.
2. The pellet was resuspended in 4 ml of Buffer P1.
3. To the above solution, 4 ml of Buffer P2 was then added. The solution was then mixed by inverting several times and is incubated at room temperature for 5 minutes.
4. To this add 4 ml of chilled Buffer P3. The solution was mixed immediately and gently by inverting 4-6 times.
5. The lysate was immediately poured into the Qiafilter cartridge and was incubated at room temperature for 10 minutes. This helps the precipitation formed from protein, genomic DNA and detergents to float and form the top layer of the solution. If the precipitation has not formed after 10 minutes, a sterile pipette tip is carefully run around the cartridge wall to dislodge it.
6. The Qiagen-tip 100 is equilibrated by adding 4 ml of Buffer QBT and allowing the buffer to empty by gravitational force.
7. The plunger is into the Qiafilter Midi Cartridge and the cell lysate is then filtered into the previously equilibrated Qiagen-tip.

8. The clear lysate is allowed to enter the resin by gravitational force. The Qiagen–tips are then washed with 10 ml of Buffer QC. This is repeated twice.
9. The DNA is eluted in a clean 10 ml tube by 5ml 0f Buffer QF.
10. The DNA is precipitated by adding 3.5 ml of isopropanol. The solution is mixed and centrifuged at 15,000 x g for 30 minutes at 4^0C.
11. The DNA pellet was washed with 70% ethanol and centrifuged at 15,000 x g for 10 minutes. The supernatant was then poured off.
12. The pellet was air dried and then dissolved 5 ml of water. The DNA was then aliquoted into 1 ml tubes and stored in -80^0C.

Preparation of Arcylamide gel:

1. The plates are meticulously cleaned with soap and water. These are then dried and again wiped clean twice with ethanol to remove any residues left by the detergents.
2. Gloves should be worn during preparation of the plates and at all times during the procedure. Fresh gloves should be worn every time during preparation of the plates to avoid contamination between the glass plates.

a. Preparation of the short plate:

   1. Freshly prepared binding solution is poured onto the plate. The solution is wiped on the plate with KimWipe®, first in one direction and then perpendicular to the first direction to ensure that the plate is completely covered with the solution.

2. The plate is allowed to dry for 5minutes. It is then wiped with clean KimWipe® twice to remove excess amount of the binding solution.

b. Preparation of the long plates:

1. The long plate is wiped with 5ml of SigmaCote® solution using the same procedure as in the case of the short plate.
2. The plate was air dried for 5-10 minutes and wiped twice with clean KimWipe® tissues. Excess SigmaCote® causes inhibition of staining.

It is important to change gloves and avoid touching both plates wearing same gloves. This ensures that gel from tearing. The above procedure should be preformed in the fumehood.

c. Assembling and pouring of Gels between the Glass plates:

1. The glass plates are assembled by placing two clean 0.4 mm side spacers between them. The plates were clamped on all sides to hold the plated together. The assembled plates are than inclined at a 10^0 angle. Care should be taken to avoid allowing the two plates to touch one another.
2. To the freshly prepared 75 ml of acrylamide solution, 500 µl of freshly prepared 10% Ammonium persulphate was first added followed by 50 µl TEMED solution.
3. The freshly prepared acrylamide solution containing the TEMED and ammoniunm persulphate solution is then poured between the glass plates with the help of 60 ml syringe (no needle) moving gently from one side. After

pouring, the gels are inclined at a 10^0 angle. Acrylamide is extremely toxic and double gloves should be worn at all times when handling acrylamide.

4. Two combs (about 6 mm) are then inserted at the top of the gels.
5. The remaining acrylamide solution is left in the syringe as control.
6. The gel is allowed to polymerize overnight. To prevent drying of the gels, paper towels soaked in TBE are wrapped around the ends of the combs. Plastic wrap is then placed around the wells to prevent the paper towels from drying.

Polyacryamide Gel Electrophoresis:

1. The polymerized acrylamide gel was removed from the clamps and the exposed sided of the glass plates were cleaned with paper towel saturated with deionized water.
2. Excess polyacrylamide was shaved off away from the comb. The combs were then removed.
3. To the bottom clamber of the electrophoresis apparatus, 1X TBE was added.
4. The glass plates were lowered into the buffer with the longer plate facing out and the well side on the top. The plates were secured on to the sequencing gel apparatus.
5. The rest of the 1X TBE buffer was added on the top buffer chamber of the electrophoresis apparatus.

6. A syringe filled with buffer was flushed in the well area to remove air bubbles and small pieces of acrylamide trapped on top of the gel. This is important step otherwise sample solution will not settle well onto the well.
7. The gels were pre-run for 11/2 hours.
8. After the pre-run, syringe filled with buffer was again flushed in the well area to remove urea. The sharkstooth comb was then carefully inserted into the gel to about 1-2mm (approximately).
9. The samples were then loaded onto the wells. Care was taken to see that the loading process did not take more than 20 minutes to prevent cooling of the gels.
10. The gel was run for 3-6 hours depending on the size of the marker.

The plates can be soaked for 1 hour in 10% sodium hydroxide solution to clean the plates of the gels and the solutions.

Silver Staining the Sequencing Gels:

1. After electrophoresis, the plates are separated using a metal spatula. The gel remains attached to the short plate.
2. The gel is then transferred into plastic trays containing fixative solution. It is kept overnight without agitation and covered to prevent evaporation. The fixative solution is saved to be used to terminate the developing reaction.

3. The gel is then rinsed for 5minutes with 2000 ml of double distilled water, three times using agitation.
4. The gel is then transferred to the freshly made staining solution and agitated for 30 minutes.
5. The gel is then rinsed with 2000 ml of double distilled water briefly for 5 sec to reduce background. It is drained and placed immediately in 1000 ml of prechilled (10^0C) developing solution. Just before using the developing solution, 3 ml of 37% formaldehyde and 400µls of sodium thiosulphate was added to the solution.
6. The gel is agitated until the first template bands start to appear. The gel is then transferred to the remaining 1000 ml of developing solution.
7. The gel is then developed for about 2-3 minutes until all bands are visible. Care was taken to prevent prolonged developing to reduce background. The bands appear initially light but will intensify on drying.
8. The developing reaction was terminated by adding 1l of the fixative solution.
9. The gels were then washed twice with 2000 ml of double distilled water for 2 minute each. (Merril *et al* 1981, Bassam *et al* 1991).

PCR Cycling Conditions:

DYS 390

95°C for 5 mins, 80°C for 10 mins, 95°C for 1 mins	1 cycle
95°C for 1 mins, 53°C for 1 mins, 72°C for 1 mins	2 cycles
95°C for 1 mins, 52°C for 1 mins, 72°C for 1 mins	2 cycles
95°C for 1 mins, 51°C for 1 mins, 72°C for 1 mins	2 cycles
95°C for 1 mins, 50°C for 1 mins, 72°C for 1 mins	2 cycles
95°C for 1 mins, 49°C for 1 mins, 72°C for 1 mins	2 cycles
95°C for 1 mins, 48°C for 1 mins, 72°C for 1 mins	2 cycles
95°C for 1 mins, 47°C for 1 mins, 72°C for 1 mins	2 cycles
95°C for 1 mins, 46°C for 1 mins, 72°C for 1 mins	26 cycles

DYS 391

95°C for 5 mins, 80°C for 10 mins, 95°C for 1 mins	1 cycle
95°C for 1 mins, 59°C for 1 mins, 72°C for 1 mins	2 cycles
95°C for 1 mins, 58°C for 1 mins, 72°C for 1 mins	2 cycles
95°C for 1 mins, 57°C for 1 mins, 72°C for 1 mins	2 cycles
95°C for 1 mins, 56°C for 1 mins, 72°C for 1 mins	2 cycles
95°C for 1 mins, 55°C for 1 mins, 72°C for 1 mins	2 cycles
95°C for 1 mins, 54°C for 1 mins, 72°C for 1 mins	2 cycles
95°C for 1 mins, 53°C for 1 mins, 72°C for 1 mins	2 cycles
95°C for 1 mins, 52°C for 1 mins, 72°C for 1 mins	26 cycles

DYS 388

95°C for 5 mins, 80°C for 10 mins, 95°C for 1 mins	1 cycle
95°C for 1 mins, 55°C for 1 mins, 72°C for 1 mins	30 cycles

DYS 392& 395

95°C for 5 mins, 80°C for 10 mins, 95°C for 2 mins	1 cycle
95°C for 1 mins, 57°C for 1 mins, 72°C for 1 mins	2 cycles
95°C for 1 mins, 50°C for 1 mins, 72°C for 1 mins	2 cycles
95°C for 1 mins, 49°C for 1 mins, 72°C for 1 mins	2 cycles
95°C for 1 mins, 48°C for 1 mins, 72°C for 1 mins	2 cycles
95°C for 1 mins, 47°C for 1 mins, 72°C for 1 mins	2 cycles
95°C for 1 mins, 52°C for 1 mins, 72°C for 1 mins	26 cycles

DYS 389

95^{0}C for 5 mins, 80^{0}C for 10 mins, 95^{0}C for 2 mins	1 cycle
95^{0}C for 1 mins, 60^{0}C for 1 mins, 72^{0}C for 1 mins	2 cycles
95^{0}C for 1 mins, 59^{0}C for 1 mins, 72^{0}C for 1 mins	2 cycles
95^{0}C for 1 mins, 58^{0}C for 1 mins, 72^{0}C for 1 mins	2 cycles
95^{0}C for 1 mins, 57^{0}C for 1 mins, 72^{0}C for 1 mins	2 cycles
95^{0}C for 1 mins, 56^{0}C for 1 mins, 72^{0}C for 1 mins	2 cycles
95^{0}C for 1 mins, 55^{0}C for 1 mins, 72^{0}C for 1 mins	2 cycles
95^{0}C for 1 mins, 54^{0}C for 1 mins, 72^{0}C for 1 mins	2 cycles
95^{0}C for 1 mins, 53^{0}C for 1 mins, 72^{0}C for 1 mins	20 cycles

DYS 271

95^{0}C for 5 mins, 80^{0}C for 10 mins, 95^{0}C for 2 mins	1 cycle
95^{0}C for 1 mins, 60^{0}C for 1 mins, 72^{0}C for 1 mins	2 cycles
95^{0}C for 1 mins, 59^{0}C for 1 mins, 72^{0}C for 1 mins	2 cycles
95^{0}C for 1 mins, 58^{0}C for 1 mins, 72^{0}C for 1 mins	2 cycles
95^{0}C for 1 mins, 57^{0}C for 1 mins, 72^{0}C for 1 mins	2 cycles
95^{0}C for 1 mins, 56^{0}C for 1 mins, 72^{0}C for 1 mins	2 cycles
95^{0}C for 1 mins, 55^{0}C for 1 mins, 72^{0}C for 1 mins	2 cycles
95^{0}C for 1 mins, 54^{0}C for 1 mins, 72^{0}C for 1 mins	20 cycles

pbluescript T3

95^{0}C for 2 mins. Hot start	
95^{0}C for 30 secs, 42^{0}C for 30 secs, 70^{0}C for 1mins	60 cycles.
Soak at 4^{0}C.	

APPENDIX C: INDIAN Y CHROMOSOME MICROSATELLITE DATA

MICROSATELLITE MARKERS ON THE Y CHROMOSOMES

CODE	19	388	389I	389II	390	391	392	395	257	271	L	R	C
1	16	12	10	27	25	11	11	13	0	A	NE	H	V
13	15	12	10	27	25	11	11	13	0	A	NC	H	B
14	15	12	10	26	25	11	11	13	0	A	NE	H	B
16	15	12	11	28	25	10	11	13	0	A	NE	H	B
18	15	12	11	27	23	10	11	12	0	A	NC	H	
19	15	12	11	26	25	11	11	13	0	A	NC	H	B
20	14	12	10	27	23	10	10	15	0	A	NE	H	B
21	16	12	12	27	25	10	11	14	0	A	SI	H	
22	15	12	10	27	22	10	11	12	0	A	NW	H	V
23	15	12	11	28	25	10	11	13	0	A	NE	H	B
24	15	12	10	29	25	10	11	13	0	A	NC	H	K
25	14	12	9	25	23	10	14	12	0	A	SI	C	
26	16	13	10	26	23	0	11	14	0	A	NC	H	B
27	17	13	10	28	22	10	11	13	0	A	NW	M	
34	15	12	10	28	24	10	11	13	0	A	NE	H	V
35	14	12	9	26	22	10	11	12	0	A	NW	J	
36	14	12	9	26	22	10	11	11	ND	A	SI	H	B
37	14	15	9	24	23	10	11	12	ND	A	SI	H	B
38	15	12	10	25	23	11	11	12	0	A	NC	H	K
39	15	12	9	26	22	11	11	14	ND	A	SI	H	B
40	14	12	9	25	22	10	12	11	ND	A	NW	H	B
41	16	15	9	25	24	10	11	12	ND	A	SI	H	V
43	15	15	9	25	24	10	11	12	ND	A	SI	C	
44	15	12	10	27	22	10	14	12	0	A	NC	H	K
46	13	16	11	27	24	10	11	14	0	A	NC	H	B
47	16	12	11	28	24	11	11	13	0	A	NW	H	B
48	14	12	9	25	22	10	14	13	ND	A	SI	C	
51	16	13	10	26	23	10	11	14	0	A	NC	H	B
52	16	12	10	27	26	11	11	13	ND	A	SI	H	
55	15	13	11	26	23	10	12	14	ND	A	SI	H	B
57	15	12	11	27	22	10	11	12	0	A	NW	H	V
58	16	12	10	26	22	10	11	12	0	A	NW	M	
60	14	12	10	25	23	11	10	14	0	A	NE	H	
61	15	13	11	30	25	10	11	13	0	A	NC	H	B
62	15	12	11	27	22	10	10	12	0	ND	NC	H	B

65	15	12	10	28	25	11	11	13	0	A	NC	H	B
67	14	15	9	26	23	10	11	12	0	A	NC	H	B
68	15	12	11	28	25	10	11	13	0	A	NC	H	B
69	15	12	10	26	22	10	11	12	0	A	NC	H	B
70	16	13	11	27	26	11	11	13	0	A	NC	H	B
71	16	12	10	26	23	10	11	12	0	A	NC	H	K
72	0	13	11	27	23	10	11	12	ND	A	NC	H	K
73	15	12	10	26	26	10	11	13	0	A	NC	H	K
74	16	12	10	29	24	11	11	13	0	A	NC	H	K
75	15	12	11	28	23	10	11	13	0	A	NC	H	K
76	16	12	11	29	25	11	11	13	0	A	NC	H	K
77	15	12	10	26	22	10	11	12	0	A	NC	H	K
78	17	12	11	27	22	10	11	12	0	A	NW	H	K
79	14	13	10	26	23	10	10	13	0	A	NC	H	K
81	15	12	11	25	24	10	10	14	0	A	NC	H	K
82	16	12	9	26	23	10	11	12	0	A	NC	H	K
83	14	12	11	28	23	10	11	13	0	A	NC	H	
86	16	12	10	27	25	11	11	13	0	A	NC	H	V
87	14	17	11	28	23	10	12	13	0	A	NC	H	V
88	14	16	9	27	23	10	11	12	0	A	NC	M	
89	16	12	10	26	25	11	11	13	0	A	NC	H	K
90	15	12	9	25	23	10	14	13	0	A	NC	C	
91	14	12	10	30	24	12	13	13	0	A	NC	H	
92	14	16	11	28	23	10	11	13	0	A	NC	M	
93	14	14	10	27	24	10	11	12	0	A	NC	H	K
94	17	12	11	27	22	10	11	12	0	A	NC	H	V
95	14	12	11	27	23	10	10	14	0	A	NC	M	
96	17	12	11	27	22	9	11	12	0	A	NC	M	
98	16	12	11	29	26	11	11	13	0	A	NE	H	V
99	16	12	10	26	24	10	11	13	0	A	NE	H	V
100	15	15	9	25	24	10	11	12	0	A	NC	H	V
101	15	12	11	27	22	10	11	12	0	A	NE	H	B
102	13	12	10	28	25	10	11	13	1	A	UN	H	
103	16	13	11	27	23	10	10	14	0	A	NE	M	
104	15	12	10	26	24	11	11	13	0	A	NE	M	
105	14	12	10	26	23	9	10	14	0	A	NE	M	
109	13	13	12	28	22	10	15	13	0	A	NE	H	
117	16	12	10	26	21	11	11	12	0	A	NE	H	
118	15	12	11	28	23	10	12	12	0	A	NE	H	
119	14	15	10	26	22	10	11	13	0	A	NE	H	
123	16	12	11	27	23	10	11	14	0	A	SI		
125	16	13	11	30	22	10	12	12	0	A	SI	H	
126	15	15	11	26	24	10	12	12	0	A	SI	C	
127	14	12	9	25	22	10	14	11	0	A	SI	H	
130	14	12	9	25	24	10	14	11	0	A	SI	H	

133	14	12	9	25	22	10	14	11	0	A	SI	H	
134	16	12	12	27	22	10	11	12	0	A	SI	C	
135	14	12	11	26	22	9	15	13	0	A	SI	H	
136	14	12	11	29	23	10	11	15	0	A	SI	H	
139	15	12	9	25	22	10	14	11	0	A	NW	S	
140	15	12	9	25	22	10	14	11	0	A	SI	H	B
144	15	12	11	29	25	10	12	13	0	A	NW	M	
145	15	15	9	24	24	10	11	12	0	A	NC	M	
146	15	12	11	28	25	11	11	13	0	A	SI	C	
147	15	12	10	26	24	12	11	13	0	A	NW	S	
148	14	15	11	27	23	10	11	12	ND	A	SI	C	
149	15	12	11	28	24	11	11	13	0	A	NE	H	V
150	15	12	10	28	25	11	11	13	0	A	NC	M	
151	15	12	11	27	22	10	11	12	ND	A	SI	C	
152	16	12	10	28	26	11	11	13	0	A	NW	S	
153	16	12	10	30	25	11	11	13	0	A	NW	H	
154	15	12	11	30	25	11	11	13	0	A	NW	S	
157	16	12	10	27	25	11	11	13	0	A	NW	S	
158	16	12	10	29	25	11	11	13	0	A	NW	S	
159	14	14	10	27	23	10	11	14	0	A	NW	S	
161	16	12	10	27	26	11	11	13	0	A	NW	H	
166	16	12	10	28	25	11	11	13	0	A	NW	S	
167	0	12	11	28	23	10	10	14	0	A	NW	S	
168	15	12	10	26	22	10	14	12	0	A	NW	S	
169	14	12	11	27	25	11	13	13	0	A	NE	C	
170	16	12	10	27	25	11	11	13	0	A	NW	S	
171	15	12	10	26	22	10	14	12	0	A	NW	S	
172	15	12	10	26	24	10	11	13	0	A	NW	S	
175	13	12	11	27	22	10	15	13	0	A	NW	S	
177	15	12	10	25	23	10	10	14	0	A	NW	S	
179	14	14	11	27	23	10	11	12	0	A	NW	S	
180	14	13	10	26	24	10	13	12	0	A	NW	S	
181	13	12	10	26	23	10	12	13	0	A	NW	S	
183	17	12	11	27	22	10	11	12	0	A	NW	S	
187	16	17	11	26	23	10	10	13	0	A	NC	H	
189	15	15	9	25	24	11	11	13	0	A	SI	H	

APPENBIX D: HAPLOTYPES MADE FROM INDIAN SAMPLES

HT	freq	DYS MARKERS								L	R	C
		19	388	389I	389II	390	391	392	395			
1	1	13	12	10	26	23	10	12	13	NW	S	
2	1	13	12	10	28	25	10	11	13	UN	H	
3	1	13	12	11	27	22	10	15	13	NW	S	
4	1	13	13	12	28	22	10	15	13	NE	H	
5	1	13	16	11	27	24	10	11	14	NC	H	B
6	1	14	12	9	25	22	10	12	11	NW	H	B
7	2	14	12	9	25	22	10	14	11	2SI	2H	
8	1	14	12	9	25	22	10	14	13	SI	C	
9	1	14	12	9	25	23	10	14	12	SI	C	
10	1	14	12	9	25	24	10	14	11	SI	H	
11	1	14	12	9	26	22	10	11	11	SI	H	B
12	1	14	12	9	26	22	10	11	12	NW	J	
13	1	14	12	10	25	23	11	10	14	NE	H	
14	1	14	12	10	26	23	9	10	14	NE	M	
15	1	14	12	10	27	23	10	10	15	NE	H	B
16	1	14	12	10	30	24	12	13	13	NC	H	
17	1	14	12	11	26	22	9	15	13	SI	H	
18	1	14	12	11	27	23	10	10	14	NC	M	
19	1	14	12	11	27	25	11	13	13	NE	C	
20	1	14	12	11	28	23	10	11	13	NC	H	
21	1	14	12	11	29	23	10	11	15	SI	H	
22	1	14	13	10	26	23	10	10	13	NC	H	K
23	1	14	13	10	26	24	10	13	12	NW	S	
24	1	14	14	10	27	23	10	11	14	NW	S	
25	1	14	14	10	27	24	10	11	12	NC	H	K
26	1	14	14	11	27	23	10	11	12	NW	S	
27	1	14	15	9	24	23	10	11	12	SI	H	B
28	1	14	15	9	26	23	10	11	12	NC	H	B
29	1	14	15	10	26	22	10	11	13	NE	H	
30	1	14	15	11	27	23	10	11	12	SI	C	
31	1	14	16	9	27	23	10	11	12	NC	M	
32	1	14	16	11	28	23	10	11	13	NC	M	
33	1	14	17	11	28	23	10	12	13	NC	H	V
34	2	15	12	9	25	22	10	14	11	NW SI	S	
35	1	15	12	9	25	23	10	14	13	NC	C	
36	1	15	12	9	26	22	11	11	14	SI	H	B

37	1	15	12	10	25	23	10	10	14	NW	S	
38	1	15	12	10	25	23	11	11	12	NC	H	K
39	2	15	12	10	26	22	10	11	12	NC	H	B
										NC	H	K
40	2	15	12	10	26	22	10	14	12	2NW	2S	
41	1	15	12	10	26	24	10	11	13	NW	S	
42	1	15	12	10	26	24	11	11	13	NE	M	
43	1	15	12	10	26	24	12	11	13	NW	S	
44	1	15	12	10	26	25	11	11	13	NE	H	B
45	1	15	12	10	26	26	10	11	13	NC	H	K
46	1	15	12	10	27	22	10	11	12	NW	H	V
47	1	15	12	10	27	22	10	14	12	NC	H	K
48	1	15	12	10	27	25	11	11	13	NC	H	B
49	1	15	12	10	28	24	10	11	13	NE	H	V
50	2	15	12	10	28	25	11	11	13	2NC	H	B
											M	
51	1	15	12	10	29	25	10	11	13	NC	H	K
52	1	15	12	11	25	24	10	10	14	NC	H	K
53	1	15	12	11	26	25	11	11	13	NC	H	B
54	1	15	12	11	27	22	10	10	12	NC	H	B
55	3	15	12	11	27	22	10	11	12	NW	H	V
										NE	H	B
										SI	C	
56	1	15	12	11	27	23	10	11	12	NC	H	
57	1	15	12	11	28	23	10	11	13	NC	H	K
58	1	15	12	11	28	23	10	12	12	NE	H	
59	1	15	12	11	28	24	11	11	13	NE	H	V
60	3	15	12	11	28	25	10	11	13	2NE	2H	2B
										NC	H	B
61	1	15	12	11	28	25	11	11	13	SI	C	
62	1	15	12	11	29	25	10	12	13	NW	M	
63	1	15	12	11	30	25	11	11	13	NW	S	
64	1	15	13	11	26	23	10	12	14	SI	H	B
65	1	15	13	11	30	25	10	11	13	NC	H	B
66	1	15	15	9	24	24	10	11	12	NC	M	
67	1	15	15	9	25	24	10	11	12	SI	C	
										NC	H	V
68	1	15	15	9	25	24	11	11	13	SI	H	
69	1	15	15	11	26	24	10	12	12	SI	C	
70	1	16	12	9	26	23	10	11	12	NC	H	K
71	1	16	12	10	26	21	11	11	12	NE	H	
72	1	16	12	10	26	22	10	11	12	NW	M	
73	1	16	12	10	26	23	10	11	12	NC	H	K
74	1	16	12	10	26	24	10	11	13	NE	H	V
75	1	16	12	10	26	25	11	11	13	NC	H	K

76	4	16	12	10	27	25	11	11	13	NE	H	V
										NC	H	V
										2NW	2S	
77	2	16	12	10	27	26	11	11	13	SI	H	
										NW	H	
78	1	16	12	10	28	25	11	11	13	NW	S	
79	1	16	12	10	28	26	11	11	13	NW	S	
80	1	16	12	10	29	24	11	11	13	NC	H	K
81	1	16	12	10	29	25	11	11	13	NW	S	
82	1	16	12	10	30	25	11	11	13	NW	H	
83	1	16	12	11	27	23	10	11	14	SI		
84	1	16	12	11	28	24	11	11	13	NW	H	B
85	1	16	12	11	29	25	11	11	13	NC	H	K
86	1	16	12	11	29	26	11	11	13	NE	H	V
87	1	16	12	12	27	25	10	11	14	SI	H	
88	1	16	12	12	27	22	10	11	12	SI	C	
89	1	16	13	10	26	23	10	11	14	NC	H	B
90	1	16	13	11	27	23	10	10	14	NE	M	
91	1	16	13	11	27	26	11	11	13	NC	H	B
92	1	16	13	11	30	22	10	12	12	SI	H	
93	1	16	15	9	25	24	10	11	12	SI	H	V
94	1	16	17	11	26	23	10	10	13	NC	H	
95	1	17	12	11	27	22	9	11	12	NC	M	
96	3	17	12	11	27	22	10	11	12	NW	H	K
										NW	S	
										NC	H	V
97	1	17	13	10	28	22	10	11	13	NW	M	

ABBRECIATION USED IN APPENDIX C AND D:

L=LOCATION
R=RELIGION
C=CASTE
NW=NORTHWEST INDIA
NC=NORTHCENTRAL INDIA
NE=NORTHEAST INDIA
SI=SOUTH INDIA
H=HINDU
S=SIKH
M=MUSLIM
C=CHRISTIAN
J=JAIN
UN=UNKNOWN
B=BRAHMIN
K=KSATRIYA

V=VAISHYA.
HT=HAPLOTYPES
FREQ.=FREQUENCY

APPENDIX E: World Data Used In This Study.

POPULATIONS	DYS MARKERS							
	19	**388**	**389I**	**389II**	**390**	**391**	**392**	**393**
Burushaski	16	12	10	26	24	10	11	13
Burushaski		16	12	10	26	25	11	12	13
Burushaski		15	12	11	27	24	10	10	14
Burushaski		15	12	11	27	22	11	11	12
Burushaski		15	12	11	26	24	10	10	14
Burushaski		16	12	10	26	22	10	14	12
Burushaski	16	12	10	26	25	10	12	13	
Burushaski	16	12	10	26	22	10	14	12	
Burushaski	16	12	11	27	22	10	14	12	
Burushaski	15	13	10	28	22	10	11	12	
Burushaski	14	12	10	26	23	9	12	13	
Burushaski	15	12	11	26	24	10	10	14	
Burushaski	15	12	11	27	24	10	10	14	
Burushaski	16	12	11	27	25	10	12	13	
Burushaski	17	12	10	23	25	10	11	13	
Burushaski	17	12	10	25	25	10	11	13	
Burushaski	15	12	10	26	23	10	12	13	
Burushaski	16	12	10	27	24	11	11	13	
Burushaski	16	13	11	27	22	9	11	14	
Burushaski	15	12	11	25	24	10	10	14	
Burushaski	14	12	10	26	23	10	12	13	
Burushaski	16	12	10	26	25	11	12	13	
Burushaski	15	15	9	25	24	10	11	12	
Burushaski	14	14	10	26	24	10	11	12	
CAR Pygmy	17	12	11	28	21	10	11	15	
CAR Pygmy	17	12	9	25	23	11	11	14	
CAR Pygmy	17	12	11	28	21	10	11	15	
CAR Pygmy	16	12	10	26	21	10	11	13	
CAR Pygmy	15	12	10	28	21	11	11	13	
CAR Pygmy	17	12	11	28	21	10	11	15	
CAR Pygmy	15	12	8	24	24	9	11	12	
CAR Pygmy	17	12	11	28	21	10	11	15	
CAR Pygmy	16	12	10	26	21	10	11	13	
CAR Pygmy	16	12	8	24	22	10	11	12	
CAR Pygmy	16	12	9	26	20	10	11	14	
CAR Pygmy	16	10	11	29	24	10	11	13	

CAR Pygmy	15	12	10	28	21	11	11	13
CAR Pygmy	17	12	10	28	20	10	11	15
CAR Pygmy	16	12	11	30	21	10	11	13
CAR Pygmy	16	12	9	26	20	10	11	15
CAR Pygmy	16	12	9	26	20	10	11	15
CAR Pygmy	16	12	9	26	20	10	11	14
CAR Pygmy	15	12	8	24	24	9	11	12
CAR Pygmy	15	12	10	26	25	8	11	12
Cambodian	15	12	10	26	25	11	13	14
Cambodian	15	12	10	26	25	10	13	14
Cambodian	15	12	10	27	23	10	13	12
Cambodian	14	10	9	24	25	10	13	12
Cambodian	14	15	10	26	23	10	11	13
Cambodian	14	15	10	26	23	10	11	13
Cambodian	15	12	10	27	25	10	11	13
Cambodian	15	12	10	27	25	11	13	14
Cambodian	15	12	10	25	25	10	13	15
Cambodian	16	12	10	27	25	11	13	13
Cambodian	15	12	9	25	23	10	14	12
Cambodian	17	12	11	26	23	10	11	14
Cambodian	15	12	10	27	24	11	15	14
Cambodian	15	12	10	27	24	11	15	14
Cambodian	16	12	10	25	25	10	13	13
Cambodian	15	12	11	27	22	10	11	12
Japanese	15	12	9	27	23	10	12	12
Japanese	15	12	11	28	23	10	11	13
Japanese	17	12	10	26	24	10	13	12
Japanese	17	12	10	27	25	10	11	13
Japanese	15	12	11	26	22	10	13	13
Japanese	15	12	10	27	25	10	11	13
Japanese	17	12	11	26	24	10	11	13
Japanese	15	12	11	26	22	10	13	13
Japanese	17	12	11	27	25	11	11	13
Japanese	16	12	10	27	25	10	11	14
Majangir	15	11	9	27	23	11	11	13
Majangir	15	11	9	27	23	11	11	13
Majangir	15	11	9	27	23	11	11	13
Majangir	15	13	9	26	24	11	11	13
Majangir	15	11	9	27	23	11	11	13
Majangir	15	11	9	27	23	11	11	13
Majangir	15	13	9	26	24	11	11	13
Majangir	15	13	10	27	24	11	11	13
Majangir	15	13	9	26	24	11	11	13
Majangir	16	13	9	25	21	11	11	12
New Guinea	15	11	10	26	23	10	13	13

New Guinea	15	12	10	28	24	11	14	13
New Guinea	15	12	9	27	23	10	14	13
New Guinea	17	12	9	25	23	10	13	13
New Guinea	15	11	10	26	24	9	13	13
New Guinea	14	11	10	26	22	10	13	13
New Guinea	15	15	10	27	21	10	11	14
New Guinea	16	14	10	27	25	11	12	13
New Guinea	14	11	9	26	24	10	13	13
New Guinea	16	12	9	25	24	10	13	13
New Guinea	16	15	11	28	21	10	11	15
New Guinea	15	15	11	29	20	10	11	14
Nyangatom	15	12	10	27	21	10	11	13
Nyangatom	15	12	9	27	21	10	11	13
Nyangatom	14	11	9	27	22	9	11	13
Nyangatom	13	12	11	28	24	10	11	14
Nyangatom	16	12	10	29	24	10	12	13
Nyangatom	15	12	11	28	22	10	11	13
Nyangatom	16	12	10	29	24	10	12	13
Nyangatom	15	10	11	29	24	10	11	13
Nyangatom	15	12	10	28	21	10	11	13
Nyangatom	15	10	12	29	24	10	11	13
Nyangatom	15	10	11	29	24	10	11	13
Omega San	15	12	11	29	21	10	11	13
Omega San	16	12	11	26	20	10	10	13
Omega San	13	12	7	24	24	10	11	14
Omega San	13	12	9	23	21	10	13	11
Omega San	17	12	10	26	21	8	11	13
Omega San	16	12	10	27	21	10	11	15
Omega San	16	11	10	27	22	11	11	14
Omega San	17	12	10	26	20	10	11	15
Omega San	15	12	10	22	24	10	11	12
Omega San	13	12	7	24	23	10	11	14
Omega San	16	11	10	27	22	11	11	14
Omega San	15	12	10	22	24	9	11	12
Omega San	16	11	10	27	22	11	11	14
Omega San	15	12	10	22	24	10	11	12
Omega San	17	12	11	28	21	10	11	15
Sekele San	15	12	10	22	23	11	11	12
Sekele San	17	11	9	25	18	11	10	13
Sekele San	15	12	10	22	24	11	11	12
Sekele San	16	12	11	23	24	10	11	12
Sekele San	16	12	11	27	20	9	11	13
Sekele San	15	12	10	28	21	11	11	13
Sekele San	16	12	11	23	24	10	11	12
Sekele San	12	12	11	26	21	10	13	11

Sekele San	13	12	10	25	21	10	13	11
Sekele San	14	12	11	23	24	10	11	15
Sekele San	16	12	11	23	24	10	11	12
Sekele San	15	12	10	27	21	11	11	13
Sekele San	13	12	10	26	22	11	13	11
Sekele San	17	11	9	25	18	11	11	13
Sindhi	14	12	9	25	22	10	14	11
Sindhi	14	12	10	26	23	10	10	14
Sindhi	14	13	9	24	23	10	11	12
Sindhi	15	12	10	28	25	10	11	13
Sindhi	14	12	11	27	24	10	11	14
Sindhi	16	15	9	25	23	9	12	12
Sindhi	15	15	9	26	24	11	11	13
Sindhi	17	12	11	25	22	10	11	12
Sindhi	14	12	9	26	22	10	14	11
Sindhi	17	12	10	27	24	10	11	13
Sindhi	14	12	9	25	22	10	14	11
Sindhi	16	12	10	28	25	11	11	13
Sindhi	13	15	10	25	23	10	11	12
Sindhi	15	15	9	25	22	10	11	12
Sindhi	15	12	10	28	25	11	11	13
Sotho	15	12	10	27	22	10	11	14
Sotho	15	10	11	29	24	11	11	13
Sotho	15	10	10	28	25	10	11	13
Sotho	14	12	11	28	21	10	11	15
Sotho	14	12	9	25	25	10	11	13
Sotho	16	12	9	26	21	10	11	13
Sotho	15	10	10	28	23	10	11	13
Sotho	15	12	10	27	21	10	11	15
Sotho	16	12	10	27	21	10	12	15
Sotho	15	12	10	27	21	10	11	15
Sotho	15	10	11	30	24	10	11	13
Sotho	16	12	10	28	21	10	11	13
Sotho	15	10	10	28	24	10	11	13
Sotho	13	12	7	24	24	11	11	14
Sotho	16	12	10	27	21	10	11	15
Sotho	16	11	9	26	22	11	10	13
Sotho	17	12	10	27	21	10	11	16
Surma	14	12	10	28	25	10	11	13
Surma	15	11	10	27	21	9	11	15
Surma	16	13	9	26	24	11	11	13
Surma	16	13	9	26	23	11	11	13
Surma	14	12	10	28	25	10	11	13
Surma	16	13	9	26	24	11	11	13
Surma	15	11	10	28	21	9	11	14

Surma	16	13	9	26	24	11	11	13
Surma	15	11	10	27	21	10	11	15
Surma	14	12	10	28	25	10	11	13
Surma	14	12	10	28	25	10	11	13
Taiwan	15	12	10	26	23	11	14	13
Taiwan	15	12	10	27	23	11	14	13
Taiwan	15	12	10	26	23	11	14	13
Taiwan	15	12	10	26	23	11	14	13
Taiwan	15	12	10	26	23	11	14	13
Taiwan	15	12	10	26	23	11	14	13
Taiwan	15	12	10	26	23	11	14	13
Taiwan	15	12	9	27	24	10	13	14
Taiwan	15	14	9	24	24	10	13	13
Taiwan	15	12	9	24	23	10	15	13
Taiwan	15	12	9	27	24	10	13	14
Taiwan	15	12	9	26	24	10	14	13
Taiwan	12	12	11	26	24	10	13	13
Taiwan	17	13	9	25	24	10	13	13
Taiwan	15	13	10	25	24	10	13	13
Tswana	16	12	10	27	21	10	11	15
Tswana	15	12	10	28	21	10	11	13
Tswana	15	12	11	28	21	10	11	13
Tswana	15	10	11	29	24	10	12	13
Tswana	16	12	10	25	21	10	11	16
Tswana	17	10	10	28	24	10	11	13
Tswana	15	12	10	28	21	11	11	13
Tswana	16	12	11	28	21	10	11	15
Tswana	15	10	11	29	25	10	11	13
Tswana	16	12	10	27	22	10	11	14
Tswana	15	12	10	28	21	10	11	13
Tswana	16	12	11	28	21	10	11	14
Zulu	16	12	10	28	21	10	11	14
Zulu	17	12	10	26	21	10	12	15
Zulu	16	12	10	27	21	10	12	14
Zulu	16	12	11	28	21	10	11	14
Zulu	14	12	9	25	25	10	11	13
Zulu	16	12	10	27	21	10	12	15
Zulu	15	12	10	28	21	10	11	13
Zulu	14	12	9	25	25	10	11	13
Zulu	15	12	11	29	21	10	11	13
Zulu	16	12	10	28	21	10	12	15
Zulu	16	12	10	27	21	10	12	15
Zulu	16	12	10	27	21	10	11	15
Zulu	16	12	10	27	21	10	12	15
Zulu	14	12	11	27	25	10	11	13

Zulu	14	10	10	28	24	10	11	13
Saharawis	14	17	10	27	23	11	11	12
Saharawis	13	12	11	27	24	9	11	13
Saharawis	13	12	11	27	23	9	11	13
Saharawis	14	17	10	27	23	11	11	12
Saharawis	13	12	11	28	25	9	11	13
Saharawis	13	12	11	28	25	9	11	13
Saharawis	17	12	9	27	22	9	12	13
Saharawis	13	12	11	28	25	9	12	13
Saharawis	13	12	11	28	25	9	12	13
Saharawis	13	12	11	28	25	9	11	13
Saharawis	13	12	11	28	25	9	12	13
Saharawis	13	12	11	27	24	9	11	13
Saharawis	13	12	11	28	25	9	11	13
Saharawis	14	17	9	26	23	11	11	12
Saharawis	13	12	10	26	24	9	11	13
Saharawis	13	12	11	28	25	9	12	13
Saharawis	13	12	11	27	24	9	11	13
Saharawis	13	12	11	27	24	9	11	13
Saharawis	13	12	11	27	24	9	11	13
Saharawis	13	12	11	27	24	9	11	13
Saharawis	15	12	9	27	21	10	11	13
Saharawis	14	17	10	27	23	11	11	12
Saharawis	13	12	11	28	24	9	11	13
Saharawis	13	12	11	28	25	9	12	13
Saharawis	13	12	11	28	25	9	12	13
Saharawis	13	12	11	27	24	9	11	13
Saharawis	14	16	11	29	23	11	11	12
Saharawis	13	12	11	28	25	9	11	13
Saharawis	13	12	11	28	25	9	11	13
Mzabites	13	12	11	27	24	10	11	13
Mzabites	13	12	10	26	24	9	11	13
Mzabites	13	12	11	27	24	10	11	13
Mzabites	13	12	11	27	24	10	11	13
Mzabites	13	12	11	27	23	10	11	13
Mzabites	13	12	9	25	24	10	11	13
Mzabites	13	12	11	28	24	10	11	13
Mzabites	13	12	10	26	24	10	11	13
Mzabites	15	12	10	23	23	10	13	13
Mzabites	13	12	11	27	24	10	11	13
Mzabites	13	12	11	27	24	10	11	13
Mzabites	13	12	11	27	24	10	11	13
Mzabites	13	12	11	27	24	10	11	13
Mzabites	13	12	11	27	24	10	11	13
Mzabites	13	13	11	27	23	9	11	13

Mzabites	15	12	10	27	21	11	11	13
Mzabites	15	12	11	27	26	11	13	13
Mzabites	13	12	11	27	25	10	11	13
Mzabites	13	13	11	27	23	9	11	13
Mzabites	13	13	11	27	23	9	11	13
Mzabites	14	16	10	27	22	11	11	12
Mzabites	13	12	11	27	24	10	11	13
Mzabites	13	12	11	27	24	9	11	13
Mzabites	13	12	11	27	25	10	11	13
Mzabites	13	12	11	27	24	10	11	13
Mzabites	13	12	11	27	24	10	11	13
Mzabites	13	12	11	27	24	10	11	13
Mzabites	13	12	11	27	25	10	11	13
Mzabites	13	12	11	27	25	10	11	13
Mzabites	14	12	9	26	25	10	11	13
Mzabites	13	12	11	27	24	10	11	13
Mzabites	13	12	11	27	24	10	11	13
Mzabites	13	12	11	27	24	9	11	13
Mzabites	13	12	10	26	24	10	11	13
Mzabites	13	12	11	27	23	10	11	13
Mzabites	14	12	11	29	20	10	11	14
Mzabites	13	12	11	27	23	10	11	13
Mzabites	13	12	11	27	25	10	11	13
Mzabites	13	12	11	27	24	10	11	13
Mzabites	18	12	11	27	25	10	11	13
Mzabites	13	12	11	27	24	9	11	13
Mzabites	13	12	11	27	24	9	11	13
Mzabites	13	12	11	27	24	10	11	13
Mzabites	13	12	10	26	24	10	11	13
Mzabites	13	12	11	27	24	10	11	13
Mzabites	13	12	10	26	24	9	11	13
Mzabites	14	12	10	29	25	10	11	15
Mzabites	13	12	10	26	24	9	11	13
Mzabites	13	12	11	28	24	10	11	13
Mzabites	13	12	11	27	23	10	11	13
Mzabites	13	12	11	27	23	10	11	13
Mzabites	13	12	11	27	24	10	11	13
Mzabites	13	12	11	27	24	10	11	13
Mzabites	13	12	11	27	25	10	11	15
Mzabites	13	12	11	27	25	10	11	13
Mzabites	13	13	11	27	25	10	11	13
Mzabites	13	12	11	27	25	10	11	13
Mzabites	13	12	11	27	24	10	11	13
Mzabites	13	12	11	27	24	10	11	13
Mzabites	13	12	11	28	24	10	11	13

Mzabites	13	12	11	27	24	10	11	13
Mzabites	13	12	11	28	24	10	11	13
Mzabites	13	12	11	27	24	10	11	13
Mzabites	17	12	11	27	24	11	13	13
Mzabites	13	12	11	27	24	10	11	13
Mzabites	15	12	9	26	22	11	11	13
Mzabites	14	12	11	27	24	10	11	13
Arabs(NW Africa)	14	17	10	27	23	11	11	12
Arabs(NW Africa)	13	12	11	27	24	10	11	13
Arabs(NW Africa)	14	12	10	27	24	11	13	13
Arabs(NW Africa)	13	12	11	27	24	9	11	13
Arabs(NW Africa)	13	12	11	27	24	9	11	13
Arabs(NW Africa)	13	13	11	27	23	9	11	13
Arabs(NW Africa)	13	12	10	27	24	9	11	13
Arabs(NW Africa)	13	12	9	26	24	10	11	13
Arabs(NW Africa)	13	12	11	27	25	9	11	13
Arabs(NW Africa)	13	12	11	27	24	9	11	13
Arabs(NW Africa)	13	12	11	27	24	9	11	13
Arabs(NW Africa)	13	12	11	27	24	9	11	13
Arabs(NW Africa)	14	12	9	24	23	10	11	14
Arabs(NW Africa)	13	12	11	27	24	9	11	13
Arabs(NW Africa)	13	12	11	27	24	9	11	13
Arabs(NW Africa)	14	17	10	27	23	11	11	12
Arabs(NW Africa)	14	18	10	27	23	12	11	12
Arabs(NW Africa)	13	12	11	27	24	9	11	13
Arabs(NW Africa)	13	12	11	27	24	9	11	13
Arabs(NW Africa)	13	12	11	27	23	11	11	13
Arabs(NW Africa)	14	14	11	29	25	10	11	12
Arabs(NW Africa)	14	12	11	27	24	10	11	13
Arabs(NW Africa)	13	12	11	28	24	9	11	13
Arabs(NW Africa)	13	12	11	27	24	9	11	13
Arabs(NW Africa)	15	12	9	27	21	10	11	14
Arabs(NW Africa)	13	12	11	27	24	9	11	13
Arabs(NW Africa)	14	12	11	27	24	9	11	13
Arabs(NW Africa)	13	12	11	27	24	10	11	13
Arabs(NW Africa)	15	12	9	27	21	10	11	14
Arabs(NW Africa)	13	12	9	26	23	10	11	10
Arabs(NW Africa)	14	12	9	26	23	10	11	13
Arabs(NW Africa)	17	13	10	25	23	10	11	13
Arabs(NW Africa)	13	12	11	27	24	9	11	13
Arabs(NW Africa)	13	12	11	27	24	10	11	13
Arabs(NW Africa)	16	12	11	28	21	10	11	13
Arabs(NW Africa)	14	17	10	27	23	11	11	12
Arabs(NW Africa)	13	12	11	27	24	11	11	13
Arabs(NW Africa)	13	12	11	27	24	9	11	13

Arabs(NW Africa)	13	12	9	26	23	10	11	13
Arabs(NW Africa)	16	12	11	27	24	11	13	13
Arabs(NW Africa)	14	17	10	26	23	11	11	12
Arabs(NW Africa)	14	17	10	27	23	10	11	12
Arabs(NW Africa)	14	12	9	25	24	10	13	13
Arabs(NW Africa)	13	12	11	27	25	9	11	13
Arabs(NW Africa)	13	12	9	26	23	10	11	13
Tahelhits	15	12	10	27	25	10	11	13
Tahelhits	13	12	10	27	24	9	11	13
Tahelhits	13	12	11	27	24	9	11	13
Tahelhits	13	12	9	26	23	10	11	14
Tahelhits	16	12	10	26	25	11	13	13
Tahelhits	13	13	10	26	24	9	11	13
Tahelhits	13	12	11	27	24	9	11	13
Tahelhits	13	12	11	27	25	9	11	13
Tahelhits	14	12	11	27	24	9	11	13
Tahelhits	13	12	10	26	23	9	11	13
Tahelhits	13	12	11	27	23	8	11	13
Tahelhits	13	12	10	30	22	9	11	13
Tahelhits	15	12	10	28	24	10	11	13
Tahelhits	13	12	10	26	24	9	11	13
Tahelhits	13	12	11	27	24	9	11	13
Tahelhits	14	12	11	27	24	9	11	13
Tahelhits	14	12	11	27	24	9	11	13
Tahelhits	13	13	10	28	25	10	11	13
Tahelhits	14	12	11	27	24	9	11	13
Tahelhits	13	12	12	28	23	9	11	13
Tahelhits	16	12	11	27	24	9	11	13
Tahelhits	16	12	9	26	21	10	11	14
Tahelhits	13	12	13	29	23	9	11	13
Tahelhits	13	12	11	27	24	9	11	14
Tahelhits	13	12	11	27	24	10	11	13
Tahelhits	13	12	9	26	24	10	10	13
Tahelhits	13	12	10	28	22	9	11	13
Tahelhits	13	12	11	27	24	9	11	13
Tahelhits	14	15	9	26	23	10	11	12
Tahelhits	13	12	11	27	24	11	11	13
Tahelhits	14	17	10	26	23	11	11	12
Tahelhits	13	12	9	26	25	10	11	13
Tahelhits	13	12	10	28	24	10	11	13
Tahelhits	13	12	11	27	24	9	11	13
Tahelhits	14	15	10	26	24	10	11	12
Tahelhits	13	12	11	27	24	9	11	13
Tahelhits	13	12	11	27	24	9	11	13
Tahelhits	13	12	10	26	24	9	11	13

Tahelhits	13	12	10	26	23	9	11	13
Tahelhits	13	12	10	28	22	9	11	13
Tahelhits	13	12	11	27	23	9	11	13
Tahelhits	14	16	10	27	23	11	11	12
Tahelhits	14	18	10	27	24	11	11	12
Tahelhits	13	12	11	27	23	9	11	13
Berbs	13	12	11	27	24	9	11	13
Berbs	15	12	10	26	25	11	13	14
Berbs	13	12	11	27	24	9	11	13
Berbs	13	12	11	27	23	9	11	13
Berbs	13	12	11	27	24	9	12	13
Berbs	13	12	11	27	24	9	11	13
Berbs	13	12	10	26	24	9	11	13
Berbs	13	12	11	27	24	9	11	13
Berbs	16	12	10	28	21	10	11	14
Berbs	13	12	10	27	24	10	11	13
Berbs	13	12	11	27	24	9	11	13
Berbs	13	12	11	27	23	9	11	13
Basque	13	12	11	27	24	9	11	13
Basque	14	12	9	25	24	11	13	13
Basque	14	12	11	29	24	10	13	13
Basque	14	17	9	25	24	10	11	12
Basque	14	17	9	25	24	11	13	13
Basque	13	12	10	26	24	10	13	13
Basque	14	12	10	26	24	11	11	13
Basque	14	12	10	26	24	11	13	13
Basque	14	12	11	27	23	11	13	14
Basque	15	13	10	26	25	10	11	13
Basque	14	12	10	26	25	10	13	12
Basque	14	12	10	25	24	11	13	13
Basque	14	12	10	26	24	10	13	13
Basque	15	12	10	26	24	11	13	13
Basque	14	12	11	28	24	10	13	13
Basque	14	12	10	26	24	10	13	13
Basque	14	12	10	26	23	10	13	14
Basque	15	12	10	27	23	11	13	13
Basque	14	12	11	27	24	11	13	13
Basque	13	13	9	25	24	12	11	13
Basque	14	12	10	25	23	11	13	14
Basque	14	12	10	26	24	11	13	13
Basque	14	12	10	26	24	11	13	13
Basque	14	12	11	27	23	11	13	13
Basque	17	13	12	27	23	10	11	13
Basque	14	12	10	27	24	10	13	13
Basque	14	12	10	26	24	11	13	13

Basque	14	12	9	25	24	11	13	13
Basque	14	12	10	26	24	11	13	13
Basque	14	12	11	27	24	11	13	13
Basque	14	12	10	26	24	11	13	13
Basque	13	12	10	26	25	10	13	12
Basque	14	12	10	26	24	11	11	13
Basque	14	12	8	24	24	10	13	13
Basque	14	12	10	26	23	10	13	13
Basque	14	12	10	26	24	10	13	13
Basque	17	13	10	25	24	9	11	13
Basque	14	12	10	27	24	11	13	13
Basque	14	12	9	25	24	11	13	13
Basque	14	12	10	26	24	12	13	13
Basque	14	12	10	25	24	10	13	13
Basque	14	12	11	27	24	10	13	13
Basque	14	12	11	27	24	11	13	13
Basque	14	12	11	27	24	11	11	13
Basque	14	12	10	26	24	11	11	13
Basque	14	12	11	27	24	11	13	13
Basque	14	12	11	27	24	10	13	13
Basque	15	12	11	27	23	10	11	13
Basque	14	12	11	27	23	10	11	13
Basque	14	12	10	26	24	11	11	13
Basque	15	12	11	27	24	11	13	13
Catalans	14	12	10	26	24	11	13	13
Catalans	14	12	10	26	24	10	13	13
Catalans	14	14	9	25	23	10	11	13
Catalans	14	11	10	28	25	10	11	12
Catalans	14	12	10	26	24	11	13	13
Catalans	14	12	10	26	24	11	13	13
Catalans	15	13	9	26	22	10	11	14
Catalans	14	12	10	26	24	11	11	13
Catalans	14	12	10	26	24	11	11	13
Catalans	14	12	10	26	24	11	13	13
Catalans	15	12	10	26	24	10	13	13
Catalans	15	15	10	28	23	9	11	13
Catalans	14	12	11	27	24	11	11	13
Catalans	15	12	9	27	22	10	11	13
Catalans	14	12	10	26	23	10	13	13
Catalans	14	12	10	26	23	11	14	14
Catalans	14	12	10	26	24	10	11	13
Catalans	14	12	10	27	24	10	11	13
Catalans	14	12	10	26	24	11	11	13
Catalans	14	12	10	26	24	11	12	12
Catalans	13	12	10	27	24	10	11	13

KAZAKHS	16	14	10	ND	25	10	11	13
KAZAKHS	16	14	10	ND	25	10	11	13
KAZAKHS	16	14	10	ND	25	10	11	13
KAZAKHS	14	13	10	ND	23	10	12	13
KAZAKHS	16	12	11	ND	24	9	11	13
KAZAKHS	16	14	11	ND	25	10	11	13
KAZAKHS	0	12	10	ND	25	12	11	13
KAZAKHS	16	14	10	ND	20	10	11	13
KAZAKHS	16	14	10	ND	24	10	11	13
KAZAKHS	16	14	10	ND	26	10	11	13
KAZAKHS	16	14	10	ND	22	10	11	13
KAZAKHS	16	14	10	ND	25	10	11	13
H.KIR	15	12	11	ND	25	11	11	13
H.KIR	16	12	11	ND	25	11	11	13
H.KIR	16	12	11	ND	25	11	11	13
H.KIR	16	12	11	ND	25	11	11	13
H.KIR	16	12	11	ND	25	11	11	13
H.KIR	16	12	11	ND	25	11	11	13
H.KIR	15	12	11	ND	25	11	11	13
H.KIR	16	12	11	ND	25	11	11	13
H.KIR	16	12	11	ND	25	11	11	13
H.KIR	16	12	11	ND	25	11	11	13
H.KIR	16	12	11	ND	25	10	11	13
H.KIR	16	14	10	ND	25	10	11	13
H.KIR	15	12	11	ND	25	11	11	13
H.KIR	16	12	11	ND	25	11	11	13
H.KIR	16	12	11	ND	25	11	11	13
H.KIR	16	12	11	ND	25	11	11	13
H.KIR	16	12	11	ND	25	11	11	13
H.KIR	16	12	11	ND	25	11	11	13
H.KIR	16	12	11	ND	25	11	11	13
H.KIR	16	12	11	ND	25	11	11	13
H.KIR	16	12	11	ND	25	11	11	13
H.KIR	16	12	11	ND	25	11	11	13
H.KIR	16	12	11	ND	25	11	11	13
H.KIR	16	12	11	ND	25	11	11	13
H.KIR	16	12	11	ND	25	11	11	13
H.KIR	14	12	10	ND	23	10	15	13
H.KIR	14	12	10	ND	23	10	15	13
H.KIR	16	12	11	ND	25	11	11	13
H.KIR	16	12	11	ND	25	11	11	13
H.KIR	16	12	11	ND	25	11	11	13
H.KIR	16	12	11	ND	25	11	11	13
H.KIR	15	14	10	ND	25	10	11	13
H.KIR	14	12	10	ND	23	10	14	13

H.KIR	16	12	11	ND	25	11	11	13
H.KIR	15	12	9	ND	24	10	14	14
H.KIR	16	12	11	ND	25	10	11	13
H.KIR	16	12	11	ND	25	11	11	13
H.KIR	16	12	11	ND	25	11	11	13
H.KIR	16	12	11	ND	25	11	12	13
H.KIR	16	12	11	ND	25	11	11	13
H.KIR	15	12	10	ND	22	11	14	12
H.KIR	16	12	11	ND	25	11	11	13
H.KIR	16	12	11	ND	26	11	11	13
L.KIR	15	12	10	ND	25	11	11	13
L.KIR	16	12	11	ND	25	11	11	13
L.KIR	16	13	9	ND	25	10	12	11
L.KIR	16	12	11	ND	25	10	11	13
L.KIR	16	12	11	ND	25	11	11	13
L.KIR	15	13	10	ND	24	9	11	13
L.KIR	15	13	10	ND	24	9	11	13
L.KIR	15	12	11	ND	25	10	11	13
L.KIR	16	12	11	ND	26	11	11	13
L.KIR	16	12	11	ND	25	11	11	13
L.KIR	10	12	11	ND	25	8	11	13
L.KIR	16	12	11	ND	25	10	11	13
L.KIR	15	12	11	ND	25	11	11	13
L.KIR	16	12	11	ND	25	10	11	13
L.KIR	16	14	10	ND	25	10	11	13
L.KIR	16	12	11	ND	25	11	11	13
L.KIR	16	14	10	ND	25	10	11	13
L.KIR	14	13	11	ND	23	9	10	14
L.KIR	16	12	11	ND	26	11	11	13
L.KIR	16	12	11	ND	25	9	11	13
L.KIR	16	14	10	ND	21	10	11	13
L.KIR	14	12	11	ND	23	9	14	14
L.KIR	16	12	11	ND	25	11	11	13
L.KIR	15	18	11	ND	25	11	11	13
L.KIR	16	18	10	ND	24	11	11	13
L.KIR	16	12	11	ND	25	10	11	13
L.KIR	16	18	9	ND	26	10	13	11
L.KIR	16	17	11	ND	25	11	11	13
L.KIR	14	12	11	ND	23	10	14	14
L.KIR	15	12	9	ND	23	10	10	11
L.KIR	16	12	11	ND	25	11	11	13
L.KIR	16	12	11	ND	25	10	9	11
L.KIR	16	12	11	ND	25	11	11	13
L.KIR	15	13	10	ND	24	9	11	13
L.KIR	16	12	11	ND	25	11	11	13

L.KIR	14	17	10	ND	23	11	11	12
L.KIR	14	12	11	ND	23	10	14	14
L.KIR	16	12	11	ND	25	11	11	14
L.KIR	16	13	9	ND	25	10	13	11
L.KIR	16	13	9	ND	25	10	13	11
L.KIR	16	12	11	ND	25	11	13	13
UIGURS	14	12	10	ND	25	11	13	12
UIGURS	16	14	9	ND	25	10	11	13
UIGURS	15	12	11	ND	23	10	14	11
UIGURS	16	14	9	ND	24	10	14	14
UIGURS	15	12	11	ND	22	10	14	11
UIGURS	14	12	11	ND	23	9	10	14
UIGURS	17	13	11	ND	22	10	10	12
UIGURS	16	12	10	ND	25	11	11	13
UIGURS	16	14	10	ND	25	10	11	13
UIGURS	0	0	0	ND	0	12	0	0
UIGURS	0	0	0	ND	25	12	0	13
UIGURS	16	14	10	ND	25	10	11	13
UIGURS	16	12	10	ND	23	10	14	12
UIGURS	15	12	9	ND	27	10	11	13
UIGURS	15	12	10	ND	26	11	11	13
UIGURS	16	14	10	ND	25	11	11	15
UIGURS	0	0	0	ND	0	12	0	13
UIGURS	13	12	10	ND	25	10	14	13
UIGURS	17	11	10	ND	24	10	11	13
UIGURS	16	14	10	ND	25	11	11	13
UIGURS	14	12	10	ND	22	10	14	14
UIGURS	13	14	10	ND	24	10	11	13
UIGURS	14	13	10	ND	23	11	14	11
UIGURS	16	12	10	ND	22	10	11	12
UIGURS	15	12	11	ND	19	10	11	12
UIGURS	16	14	10	ND	24	10	11	13
UIGURS	16	12	10	ND	23	10	11	13
UIGURS	16	12	9	ND	25	10	10	15
UIGURS	16	12	10	ND	25	11	15	14
UIGURS	15	12	9	ND	23	10	12	13
UIGURS	16	12	9	ND	23	10	12	12
UIGURS	17	12	10	ND	24	11	11	13
UIGURS	14	12	11	ND	23	10	12	14
UIGURS	16	12	11	ND	23	9	11	13
UIGURS	14	10	9	ND	24	10	14	12
UIGURS	16	12	11	ND	23	9	10	14
UIGURS	16	12	10	ND	23	11	11	13
UIGURS	15	12	11	ND	21	10	10	13
UIGURS	16	14	10	ND	25	10	11	13

UIGURS	16	12	10	ND	25	11	11	13
UIGURS	13	12	9	ND	25	10	13	13
UIGURS	14	10	11	ND	24	10	11	13
INGANO2	13	12	11	ND	24	10	14	13
INGANO3	14	14	9	ND	23	10	14	14
INGANO4	13	12	10	ND	25	10	11	14
INGANO5	12	12	10	ND	25	10	13	13
INGANO6	14	12	10	ND	23	10	13	13
INGANO7	12	12	11	ND	25	10	14	13
INGANO8	13	13	11	ND	26	11	14	12
INGANO9	13	12	10	ND	24	10	14	13
INGANO10	13	12	10	ND	24	10	14	13
WAYUU1	13	13	9	ND	24	11	12	13
WAYUU2	13	12	10	ND	25	10	14	13
WAYUU3	13	12	10	ND	25	11	11	13
WAYUU4	13	13	10	ND	24	10	11	13
WAYUU5	13	12	11	ND	24	10	14	13
WAYUU6	13	12	9	ND	24	10	14	12
WAYUU7	15	12	10	ND	21	11	15	15
WAYUU8	14	12	10	ND	23	10	14	13
WAYUU9	13	12	11	ND	23	11	14	14
WAYUU10	15	12	10	ND	23	10	14	12
WAYUU11	15	12	10	ND	25	10	11	14
WAYUU12	14	12	10	ND	23	10	14	13
TICUNA1	13	12	11	ND	24	10	14	12
TICUNA2	14	12	10	ND	25	10	14	16
TICUNA3	13	13	10	ND	25	10	14	13
TICUNA4	13	12	11	ND	24	10	15	13
TICUNA5	13	12	11	ND	24	10	15	13
TICUNA6	15	13	10	ND	25	10	14	14
TICUNA7	13	12	10	ND	25	10	15	13
TICUNA8	13	12	11	ND	24	10	15	13
TICUNA9	13	12	11	ND	24	10	15	14
TICUNA10	13	13	10	ND	25	10	14	13
TICUNA11	13	12	11	ND	25	10	15	13
TICUNA12	14	13	10	ND	25	10	14	14
TICUNA14	13	13	10	ND	25	10	14	13
TICUNA15	13	12	11	ND	25	10	15	13
TICUNA16	13	12	10	ND	25	10	15	12
TICUNA17	13	12	10	ND	25	10	15	13
TICUNA18	13	12	10	ND	25	10	15	13
TICUNA19	15	12	11	ND	25	10	15	13
TICUNA20	13	12	10	ND	22	10	14	13
TICUNA21	13	12	10	ND	25	11	15	13
TICUNA22	13	12	12	ND	25	11	15	13

TICUNA23	13	12	12	ND	25	10	14	13
TICUNA24	13	12	11	ND	23	10	14	13
Finn1	28	12	10	ND	25	11	14	14
Finn2	31	12	11	ND	24	10	14	14
Finn3	28	12	11	ND	24	11	15	14
Finn4	30	12	11	ND	24	11	14	14
Finn5	28	12	11	ND	24	12	15	14
Finn6	28	12	10	ND	24	10	14	14
Finn7	29	12	11	ND	24	10	14	14
Finn8	28	12	12	ND	23	10	14	14
Finn9	28	12	12	ND	24	10	14	15
Finn10	28	12	10	ND	23	11	14	13
Finn11	29	12	10	ND	23	11	14	14
Finn12	28	12	10	ND	24	11	14	13
Finn13	28	12	11	ND	23	11	14	15
Finn14	28	12	11	ND	24	11	14	13
Finn15	28	13	10	ND	23	11	14	13
Finn16	28	13	9	ND	23	10	11	13
Finn17	29	14	9	ND	22	10	11	13
Finn18	29	14	9	ND	22	10	12	13
Finn19	28	16	11	ND	24	10	14	14
Finn20	28	17	9	ND	23	10	11	13
Finn21	29	17	10	ND	23	10	11	13
Finn22	31	12	10	ND	24	11	11	13
Finn23	32	12	10	ND	25	10	11	13
Finn24	30	12	10	ND	25	10	11	13
Finn25	30	13	10	ND	24	10	11	13
Finn26	29	12	9	ND	23	10	11	13
Finn27	27	12	10	ND	24	10	11	13
Finn28	28	12	11	ND	23	11	11	14
Finn29	28	11	10	ND	24	11	13	13
Finn30	28	12	10	ND	24	11	13	14
Finn31	28	12	10	ND	25	11	13	13
Finn32	28	11	9	ND	23	10	11	13
Finn33	28	12	10	ND	24	10	11	13
Finn34	30	12	11	ND	24	11	11	14
Finn35	31	12	10	ND	25	11	14	14
Finn36	29	12	11	ND	24	11	14	14
Finn37	31	12	11	ND	24	11	14	14
Finn38	32	12	11	ND	24	11	14	14
Finn39	28	12	11	ND	24	12	14	14
Finn40	28	12	10	ND	23	10	14	13
Finn41	29	12	10	ND	23	10	14	14
Finn42	28	12	10	ND	23	11	14	12
Finn43	28	12	11	ND	23	11	14	13

Finn44	28	14	9	ND	22	10	11	13
Finn45	27	14	9	ND	23	10	11	13
Finn46	31	14	9	ND	23	10	11	13
Finn47	28	16	11	ND	24	10	14	15
Finn48	28	17	11	ND	24	10	14	14
Finn49	29	12	10	ND	25	10	11	14
Finn50	29	13	10	ND	24	11	11	13
Finn51	29	12	10	ND	23	10	11	13
Finn52	28	12	8	ND	24	10	13	13
Finn53	29	12	10	ND	23	10	13	12
Finn54	28	12	11	ND	24	10	14	13
Finn55	27	11	10	ND	22	10	11	13
Finn56	31	12	10	ND	25	11	11	13
Finn57	31	12	10	ND	25	11	14	14
Finn58	29	12	11	ND	24	11	14	14
Finn59	31	12	11	ND	24	11	14	14
Finn60	32	12	11	ND	24	11	14	14
Finn61	28	12	11	ND	24	12	14	14
Finn62	28	12	10	ND	23	10	14	13
Finn63	29	12	10	ND	23	10	14	14
Finn64	28	12	10	ND	23	11	14	12
Finn65	28	12	11	ND	23	11	14	13
Finn66	28	14	9	ND	22	10	11	13
Finn67	27	14	9	ND	23	10	11	13
Finn68	31	14	9	ND	23	10	11	13
Finn69	28	16	11	ND	24	10	14	15
Finn70	28	17	11	ND	24	10	14	14
Finn71	29	12	10	ND	25	10	11	14
Finn72	29	13	10	ND	24	11	11	13
Finn73	29	12	10	ND	23	10	11	13
Finn74	28	12	8	ND	24	10	13	13
Finn75	29	12	10	ND	23	10	13	12
Finn76	28	12	11	ND	24	10	14	13
Finn77	27	11	10	ND	22	10	11	13
Finn78	31	12	10	ND	25	11	11	13
Finn79	28	14	9	ND	23	10	12	13
Finn80	28	16	10	ND	23	10	11	12
Finn81	28	12	10	ND	24	10	13	13
Finn82	28	12	11	ND	24	11	13	13
Finn83	29	12	10	ND	25	10	11	13
Finn84	30	12	10	ND	25	11	11	13
Finn85	28	14	9	ND	23	10	12	13
Finn86	28	16	10	ND	23	10	11	12
Finn87	28	12	10	ND	24	10	13	13
Finn88	28	12	11	ND	24	11	13	13

Finn89	29	12	10	ND	25	10	11	13
Finn90	30	12	10	ND	25	11	11	13
Finn91	28	14	9	ND	23	10	12	13
Finn92	28	16	10	ND	23	10	11	12
Finn93	28	12	10	ND	24	10	13	13
Finn94	28	12	11	ND	24	11	13	13
Finn95	29	12	10	ND	25	10	11	13
Finn96	30	12	10	ND	25	11	11	13
Finn97	28	12	11	ND	24	11	14	15
Finn98	28	12	11	ND	25	11	14	14
Finn99	28	12	11	ND	23	10	14	14
Finn100	29	14	9	ND	23	10	11	13
Finn101	28	12	10	ND	24	11	13	13
Finn102	29	12	10	ND	25	11	11	13
Finn103	28	12	11	ND	24	11	14	15
Finn104	28	12	11	ND	25	11	14	14
Finn105	28	12	11	ND	23	10	14	14
Finn106	29	14	9	ND	23	10	11	13
Finn107	28	12	10	ND	24	11	13	13
Finn108	29	12	10	ND	25	11	11	13
Finn109	28	12	11	ND	24	11	14	15
Finn110	28	12	11	ND	25	11	14	14
Finn111	28	12	11	ND	23	10	14	14
Finn112	29	14	9	ND	23	10	11	13
Finn113	28	12	10	ND	24	11	13	13
Finn114	29	12	10	ND	25	11	11	13
Finn115	28	12	11	ND	24	11	14	15
Finn116	28	12	11	ND	25	11	14	14
Finn117	28	12	11	ND	23	10	14	14
Finn118	29	14	9	ND	23	10	11	13
Finn119	28	12	10	ND	24	11	13	13
Finn120	29	12	10	ND	25	11	11	13
Finn121	28	12	10	ND	24	11	14	15
Finn122	28	12	10	ND	23	10	14	14
Finn123	28	12	10	ND	24	11	14	15
Finn124	28	12	10	ND	23	10	14	14
Finn125	28	12	10	ND	24	11	14	15
Finn126	28	12	10	ND	23	10	14	14
Finn127	28	12	10	ND	24	11	14	15
Finn128	28	12	10	ND	23	10	14	14
Finn129	28	12	10	ND	24	11	14	15
Finn130	28	12	10	ND	23	10	14	14
Finn131	28	12	10	ND	24	11	14	14
Finn132	28	12	10	ND	24	11	14	14
Finn133	28	12	10	ND	24	11	14	14

Finn134	28	12	10	ND	24	11	14	14
Finn135	28	12	10	ND	24	11	14	14
Finn136	28	12	10	ND	24	11	14	14
Finn137	28	12	11	ND	23	11	14	14
Finn138	28	12	9	ND	23	10	11	13
Finn139	28	12	11	ND	23	11	14	14
Finn140	28	12	9	ND	23	10	11	13
Finn141	28	12	11	ND	23	11	14	14
Finn142	28	12	9	ND	23	10	11	13
Finn143	28	12	11	ND	23	11	14	14
Finn144	28	12	9	ND	23	10	11	13
Finn145	28	12	11	ND	23	11	14	14
Finn146	28	12	9	ND	23	10	11	13
Finn147	28	12	11	ND	23	11	14	14
Finn148	28	12	9	ND	23	10	11	13
Finn149	28	12	11	ND	23	11	14	14
Finn150	28	12	9	ND	23	10	11	13
Finn151	28	12	11	ND	24	10	14	14
Finn152	28	12	11	ND	24	10	14	14
Finn153	28	12	11	ND	24	10	14	14
Finn154	28	12	11	ND	24	10	14	14
Finn155	28	12	11	ND	24	10	14	14
Finn156	28	12	11	ND	24	10	14	14
Finn157	28	12	11	ND	24	10	14	14
Finn158	28	12	11	ND	24	10	14	14
Finn159	28	12	10	ND	23	11	14	14
Finn160	28	12	10	ND	23	11	14	14
Finn161	28	12	10	ND	23	11	14	14
Finn162	28	12	10	ND	23	11	14	14
Finn163	28	12	10	ND	23	11	14	14
Finn164	28	12	10	ND	23	11	14	14
Finn165	28	12	10	ND	23	11	14	14
Finn166	28	12	10	ND	23	11	14	14
Finn167	28	12	10	ND	23	11	14	14
Finn168	28	12	10	ND	23	11	14	14
Finn169	28	12	10	ND	23	11	14	14
Finn170	28	14	9	ND	23	10	11	13
Finn171	28	14	9	ND	23	10	11	13
Finn172	28	14	9	ND	23	10	11	13
Finn173	28	14	9	ND	23	10	11	13
Finn174	28	14	9	ND	23	10	11	13
Finn175	28	14	9	ND	23	10	11	13
Finn176	28	14	9	ND	23	10	11	13
Finn177	28	14	9	ND	23	10	11	13
Finn178	28	14	9	ND	23	10	11	13

Finn179	28	14	9	ND	23	10	11	13
Finn180	28	14	9	ND	23	10	11	13
Finn181	28	14	9	ND	23	10	11	13
Finn182	28	14	9	ND	23	10	11	13
Finn183	28	14	9	ND	23	10	11	13
Finn184	28	14	9	ND	23	10	11	13
Finn185	28	14	9	ND	23	10	11	13
Finn186	28	14	9	ND	23	10	11	13
Finn187	28	14	9	ND	23	10	11	13
Finn188	28	14	9	ND	23	10	11	13
Finn189	28	14	9	ND	23	10	11	13
Finn190	28	14	9	ND	23	10	11	13
Finn191	28	14	9	ND	23	10	11	13
Finn192	28	14	9	ND	23	10	11	13
Finn193	28	14	9	ND	23	10	11	13
Finn194	28	14	9	ND	23	10	11	13
Finn195	28	14	9	ND	23	10	11	13
Finn196	28	14	9	ND	23	10	11	13
Finn197	28	14	9	ND	23	10	11	13
Finn198	28	14	9	ND	23	10	11	13
Finn199	28	14	9	ND	23	10	11	13
Finn200	28	12	11	ND	24	11	14	14
Finn201	28	12	11	ND	24	11	14	14
Finn202	28	12	11	ND	24	11	14	14
Finn203	28	12	11	ND	24	11	14	14
Finn204	28	12	11	ND	24	11	14	14
Finn205	28	12	11	ND	24	11	14	14
Finn206	28	12	11	ND	24	11	14	14
Finn207	28	12	11	ND	24	11	14	14
Finn208	28	12	11	ND	24	11	14	14
Finn209	28	12	11	ND	24	11	14	14
Finn210	28	12	11	ND	24	11	14	14
Finn211	28	12	11	ND	24	11	14	14
Finn212	28	12	11	ND	24	11	14	14
Finn213	28	12	11	ND	24	11	14	14
Finn214	28	12	11	ND	24	11	14	14
Finn215	28	12	11	ND	24	11	14	14
Finn216	28	12	11	ND	24	11	14	14
Finn217	28	12	11	ND	24	11	14	14
Finn218	28	12	11	ND	24	11	14	14
Finn219	28	12	11	ND	24	11	14	14
Finn220	28	12	11	ND	24	11	14	14
Finn221	28	12	11	ND	24	11	14	14
Finn222	28	12	11	ND	24	11	14	14
Finn223	28	12	11	ND	24	11	14	14

Finn224	28	12	11	ND	24	11	14	14
Finn225	28	12	11	ND	24	11	14	14
Finn226	28	12	11	ND	24	11	14	14
Finn227	28	12	11	ND	24	11	14	14
Finn228	28	12	11	ND	24	11	14	14
Finn229	28	12	11	ND	24	11	14	14
Finn230	28	12	11	ND	24	11	14	14
Finn231	28	12	11	ND	24	11	14	14
Finn232	28	12	11	ND	24	11	14	14
Finn233	28	12	11	ND	24	11	14	14
Finn234	28	12	11	ND	24	11	14	14
Finn235	28	12	11	ND	24	11	14	14
Finn236	28	12	11	ND	24	11	14	14
Finn237	28	12	11	ND	24	11	14	14
Finn238	28	12	11	ND	24	11	14	14
Finn239	28	12	11	ND	24	11	14	14
Finn240	28	12	11	ND	24	11	14	14
Finn241	28	12	11	ND	24	11	14	14
Finn242	28	12	11	ND	24	11	14	14
Finn243	28	12	11	ND	24	11	14	14
Finn244	28	12	11	ND	24	11	14	14
Finn245	28	12	11	ND	24	11	14	14
Finn246	28	12	11	ND	24	11	14	14
Finn247	28	12	11	ND	24	11	14	14
Finn248	28	12	11	ND	24	11	14	14
Finn249	28	12	11	ND	24	11	14	14
Finn250	28	12	11	ND	24	11	14	14
Finn251	28	12	11	ND	24	11	14	14
Finn252	28	12	11	ND	24	11	14	14
Finn253	28	12	11	ND	24	11	14	14
Finn254	28	12	11	ND	24	11	14	14
Finn255	28	12	11	ND	24	11	14	14
Finn256	28	12	11	ND	24	11	14	14
Finn257	28	12	11	ND	24	11	14	14
Finn258	28	12	11	ND	24	11	14	14
Finn259	28	12	11	ND	24	11	14	14
Finn260	28	12	11	ND	24	11	14	14
Finn261	28	12	11	ND	24	11	14	14
Finn262	28	12	11	ND	24	11	14	14
Finn263	28	12	11	ND	24	11	14	14
Finn264	28	12	11	ND	24	11	14	14
Finn265	28	12	11	ND	24	11	14	14
Finn266	28	12	11	ND	24	11	14	14
Finn267	28	12	11	ND	24	11	14	14
Finn268	28	12	11	ND	24	11	14	14

Finn269	28	12	11	ND	24	11	14	14
Finn270	28	12	11	ND	24	11	14	14
Finn271	28	12	11	ND	24	11	14	14
Finn272	28	12	11	ND	24	11	14	14
Finn273	28	12	11	ND	24	11	14	14
Finn274	28	12	11	ND	24	11	14	14
Finn275	28	12	11	ND	24	11	14	14
Finn276	28	12	11	ND	24	11	14	14
Finn277	28	12	11	ND	24	11	14	14
Finn278	28	12	11	ND	24	11	14	14
Finn279	28	12	11	ND	24	11	14	14
Finn280	28	12	11	ND	24	11	14	14

Abbreviations Used in this appendix.

ND= Not Done
H.KIR= Highland Kirgiz
L. KIR= Lowland Kirgiz

Appendix F: Sequence of the pBluescript Vector.

ATTAACCCTC	**ACTAAAGGGA**	ACAAAAGCTG	GAGCTCCACC	GCGGTGGCGG
CCGCTCTAGA	ACTAGTGGAT	CCCCCGGGCT	GCAGGAATTC	GATATCAAGC
TTATCGATAC	CGTCGACCTC	GAGGGGGGGC	CCGGTACCCA	ATTCGCCCTA
TAGTGAGTCG	TATTACGCGC	GCTCACTGGC	CGTCGTTTTA	CAACGTCGTG
ACTGGGAAAA	CCCTGGCGTT	ACCCAACTTA	ATCGCCTTGC	AGCACATCCC
CCTTTCGCCA	GCTGGCGTAA	TAGCGAAGAG	GCCCGCACCG	ATCGCCCTTC
CCAACAGTTG	CGCAGCCTGA	ATGGCGAATG	GGACGCGCCC	TGTAGCGGCG
CATTAAGCGC	GGCGGGTGTG	GTGGTTACGC	GCAGCGTGAC	CGCTACACTT
GCCAGCGCCC	TAGCGCCCGC	TCCTTTCGCT	TTCTTCCCTT	CCTTTCTCGC
CACGTTCGCC	GGCTTTCCCC	GTCAAGCTCT	AAATCGGGGG	CTCCCTTTAG
GGTTCCGATT	TAGTGCTTTA	CGGCACCTCG	ACCCCAAAAA	ACTTGATTAG
GGTGATGGTT	CACGTAGTGG	GCCATCGCCC	TGATAGACGG	TTTTTCGCCC
TTTGACGTTG	GAGTCCACGT	TCTTTAATAG	TGGACTCTTG	TTCCAAACTG
GAACAACACT	CAACCCTATC	TCGGTCTATT	CTTTTGATTT	ATAAGGGATT
TTGCCGATTT	CGGCCTATTG	GTTAAAAAAT	GAGCTGATTT	AACAAAAATT
TAACGCGAAT	TTTAACAAAA	TATTAACGCT	TACAATTTAG	GTGGCACTTT
TCGGGGAAAT	GTGCGCGGAA	CCCCTATTTG	TTTATTTTTC	TAAATACATT
CAAATATGTA	TCCGCTCATG	AGACAATAAC	CCTGATAAAT	GCTTCAATAA
TATTGAAAAA	GGAAGAGTAT	GAGTATTCAA	CATTTCCGTG	TCGCCCTTAT
TCCCTTTTTT	GCGGCATTTT	GCCTTCCTGT	TTTTGCTCAC	CCAGAAACGC
TGGTGAAAGT	AAAAGATGCT	GAAGATCAGT	TGGGTGCACG	AGTGGGTTAC
ATCGAACTGG	ATCTCAACAG	CGGTAAGATC	CTTGAGAGTT	TTCGCCCCGA
AGAACGTTTT	CCAATGATGA	GCACTTTTAA	AGTTCTGCTA	TGTGGCGCGG
TATTATCCCG	TATTGACGCC	GGGCAAGAGC	AACTCGGTCG	CCGCATACAC
TATTCTCAGA	ATGACTTGGT	TGAGTACTCA	CCAGTCACAG	AAAAGCATCT
TACGGATGGC	ATGACAGTAA	GAGAATTATG	CAGTGCTGCC	ATAACCATGA
GTGATAACAC	TGCGGCCAAC	TTACTTCTGA	CAACGATCGG	AGGACCGAAG
GAGCTAACCG	CTTTTTTGCA	CAACATGGGG	GATCATGTAA	CTCGCCTTGA
TCGTTGGGAA	CCGGAGCTGA	ATGAAGCCAT	ACCAAACGAC	GAGCGTGACA
CCACGATGCC	TGTAGCAATG	GAACAACGTT	GCGCAAACTA	TTAACTGGCG
AACTACTTAC	TCTAGCTTCC	CGGCAACAAT	TAATAGACTG	GATGGAGGCG
GATAAAGTTG	CAGGACCACT	TCTGCGCTCG	GCCCTTCCGG	CTGGCTGGTT
TATTGCTGAT	AAATCTGGAG	CCGGTGAGCG	TGGGTCTCGC	GGTATCATTG
CAGCACTGGG	GCCAGATGGT	AAGCCCTCCC	GTATCGTAGT	TATCTACACG
ACGGGGAGTC	AGGCAACTAT	GGATGAACGA	AATAGACAGA	TCGCTGAGAT
AGGTGCCTCA	CTGATTAAGC	ATTGGTAACT	GTCAGACCAA	GTTTACTCAT
ATATACTTTA	GATTGATTTA	AAACTTCATT	TTTAATTTAA	AAGGATCTAG
GTGAAGATCC	TTTTTGATAA	TCTCATGACC	AAAATCCCTT	AACGTGAGTT
TTCGTTCCAC	TGAGCGTCAG	ACCCCGTAGA	AAAGATCAAA	GGATCTTCTT

GAGATCCTTT TTTTCTGCGC GTAATCTGCT GCTTGCAAAC AAAAAAACCA
CCGCTACCAG CGGTGGTTTG TTTGCCGGAT CAAGAGCTAC CAACTCTTTT
TCCGAAGGTA ACTGGCTTCA GCAGAGCGCA GATACCAAAT ACTGTCCTTC
TAGTGTAGCC GTAGTTAGGC CACCACTTCA AGAACTCTGT AGCACCGCCT
ACATACCTCG CTCTGCTAAT CCTGTTACCA GTGGCTGCTG CCAGTGGCGA
TAAGTCGTGT CTTACCGGGT TGGACTCAAG ACGATAGTTA CCGGATAAGG
CGCAGCGGTC GGGCTGAACG GGGGGTTCGT GCACACAGCC CAGCTTGGAG
CGAACGACCT ACACCGAACT GAGATACCTA CAGCGTGAGC TATGAGAAAG
CGCCACGCTT CCCGAAGGGA GAAAGGCGGA CAGGTATCCG GTAAGCGGCA
GGGTCGGAAC AGGAGAGCGC ACGAGGGAGC TTCCAGGGGG AAACGCCTGG
TATCTTTATA GTCCTGTCGG GTTTCGCCAC CTCTGACTTG AGCGTCGATT
TTTGTGATGC TCGTCAGGGG GGCGGAGCCT ATGGAAAAAC GCCAGCAACG
CGGCCTTTTT ACGGTTCCTG GCCTTTTGCT GGCCTTTTGC TCACATGTTC
TTTCCTGCGT TATCCCCTGA TTCTGTGGAT AACCGTATTA CCGCCTTTGA
GTGAGCTGAT ACCGCTCGCC GCAGCCGAAC GACCGAGCGC AGCGAGTCAG
TGAGCGAGGA AGCGGAAGAG CGCCCAATAC GCAAACCGCC TCTCCCCGCG
CGTTGGCCGA TTCATTAATG CAGCTGGCAC GACAGGTTTC CCGACTGGAA
AGCGGGCAGT GAGCGCAACG CAATTAATGT GAGTTAGCTC ACTCATTAGG
CACCCCAGGC TTTACACTTT ATGCTTCCGG CTCGTATGTT GTGTGGAATT
GTGAGCGGAT AACAATTTCA CACAGGAAAC AGCTATGACC ATGATTACGC
CAAGCGCGCA

REFERENCES

Ad'hiah A.H. and S.S. Papiha. (1996). Complement components C2,C3, And C4 (C4A and C4B) and BF polymorphisms in populations Of Indian subcontinent. Human Biology 68: 755-776.

AmosW and D.C. Rubinsztein. (1996). Microsatellites are subject to directional evolution.
Nature Genetic 12:13-14.

Bamshad M., A. Fraley, M.H. Crawford, R.L. Cann, B.R. Busi, J.M. Naidu, and L.B. Jorde. (1996). mtDNA variation in caste populations of Andhra Pradesh, India. Human Biology 68: 1-28.

Bamshad M.J., W.S. Watkins, M.E. Dixon, L.B. Jorde, B.B. Rao, J.M. Naidu, B.V. Ravi Prasad, A. Rasanayagam, and M.F. Hammer. (1998). Female gene flow stratifies hindu castes. Nature 395: 651.

Barnabas S., R.V. Apte, and C.G. Suresh. (1996). Ancestry and interrelationships of the Indians and their relationship with other World populations: A Study Based on mitochondrial DNA polymorphism. Annals of Human Genetics 60:409-422.

Barnabas S., B. Joshi, and C.G. Suresh. (2000). Indian – Asian relationships: mtDNA reveals more. Naturwissenschaften 87: 180-183.

Bassam B.J., G. Ceatano-Anollés, and P.M. Gresshoff. (1991). Fast and sensitive silver staining of DNA in polyacrylamide gels. Analytical Biochemistry 196: 80-83.

Bhattacharyya N.P., P. Basu, Das M., S. Pramanik, R. Banerjee, B. Roy, S. Roychoudhury, and P.P. Majumder. (1999). Negligible male gene flow across ethnic boundaries in India, revealed by analysis of Y-chromosomal DNA polymorphism. Genome Research 9: 711-719.

Bosch E., F. Calafell, F.R. Santos, A. Pérez-Lezaun, D. Comas, N. Benchemsi, C. Tyler-Smith, and J. Bertranpetit. (1999). Variation in short tandem repeats is deeply structured by genetics background on the human Y chromosome. American Journal of Human Genetics 65:1623-1638.

Bowcock A.M., A.L. Ruiz-Linares, J. Tomfohrde, E. Minch, J.R. Kidd, and L.L. Cavalli-

Sforza. (1994). High resolution of human evolutionary trees with polymorphic microsatellites. Nature 368: 455-457.

Cavalli-Sforza L.L., and A.W.F. Edwards. (1967). Phylogenetic analysis: models and estimation procedures. Evolution 32:550-570.

Cavalli-Sforza L.L, A. Piazza, P. Menozzi, and J. Mountain, (1988). Reconstruction of human evolution: Bringing together genetic, archaeological, and linguistic data. Proceedings of National Academy of Science 85:6002-6006.

Cavalli-Sforza L.L, P. Menozzi, and A. Piazza. (1993). Demic Expansion and human evolution. Science 259:639-646.

Cavalli-Sforza L.L., P. Menozzi, and A. Piazza. (1994). The history and geography of human genes. Princeton University Press, Princeton, New Jersey.

Chakraborty R. and L.Jin. (1993).A unified approach to study hypervariable polymorphisms: Statistical considerations of determining relatedness and population distances. DNA Fingerprinting: State of Science 153-175.

Deka R., M.D. Shriver, L.M. Yu, R.E. Ferrell, R. Chakraborty. (1995). Intra- and Inter-population diversity at short tandem repeat loci in diverse populations of the world. Electrophoresis 16: 1659-1664.

Excoffier L., P.E Smouse., and Quattro, (1992). Analysis of molecular variance inferred from metric distances among DNA haplotypes: Application to human mitochondrial DNA restriction data. Genetics 131:479-491.

Freije D., C. Helms, M.S. Watson, H. Donis-Keller. (1992). Identification of a second Psuudoautosomal region near the Xq and Yq telomers. Science 258: 1784-1787.

FAQs about Basques and the Basques.Website, https:// www.cogs.susx.ac.uk/users/larryt/basque.fags.html.

Ghosh A. (1989). An Encyclopaedia of Indian Archaeology, Volume 1, Munshiram Manoharlal PublishersPrivate Limited,New Delhi, India, 18-19.

Gibbons A. (1998). Anthropologists probe genes, Brains at Annual Meetings. Science 280: 380-381.

Gibbons A. (1997). Y Chromosome shows that Adams was an African. Science 278: 804-805.

Gokhale B.G. (1982). Bharatavasha: A politocal and cultural history of India. Sterling

Publishers Private Limited, New Delhi, 1-15.

Goldstein D.B., and D.D. Pollock. (1997). Launching microsatellites: A review of mutation processes and methods of phylogenetic inference. Journal of Heredity 88: 335-342.

Goldstein D.B., A.R. Linares, L.L. Cavalli-Sforza, and M.W. Feldman. (1995).Genetic absolute dating based on microsatellites and the origin of modern humans. Proceedings of National Academy of Science 92: 6723-6727.

Gupta, P.S. (ed) Census of India (1961). (1970) Capital Offset Printers, Delhi. pp330-335.

Guo S.W., and E.A. Thompson. (1992). Performing the exact test of Hardy-Weinberg proportion for multiple alleles. Biometrics 48:361-372.

Hammer M. F., (1994). A recent insertion of an Alu element on the Y chromosomes is a useful marker for human population studies. Molecular Biology and Evolution 11: 749-761.

Hammer M. F., A.B. Spurdle. T. Karafet, M.R. Bonner, E.T. Wood, A. Novelletto, P. Malaspina, R.J. Mitchell, S. Horai, T. Jenkins, and S.L. Zegura. (1997). The geographic distribution of human Y chromosome variation. Genetics 145: 787-805.

Hammer M.F. and S Horai. (1995). Chromosomal DNA variation and the peopling of Japan. American Journal of Human Genetics. 56: 951-962.

Hammer M.F., T. Karafet, A. Rasanayagam. E.T. Wood, T.K. Altheide, T. Jenkins, R.C. Griffiths, A.R. Templeton, and S.L. Zegura. (1998). Out of Africa and back again: nested cladistic analysis of human Y chromosome variation. Molecular Biology and Evolution 15: 427-441.

Henderson S.T. and T.D. Petes. (1992). Instability of simple sequence DNA in *Saccharomyces cerevisiae*. Molecular Cell Biology 12:2749-2757.

Introduction to Jainism. Website http://www.jainnet.com/intro.html

Jorde L.B., M.J. Bamshad, W.S. Watkins, R. Zenger, A.E Fraley, P.A. Krakowiak, K.D. Carpenter, H. Soodyall, T. Jenkins, and A.R. Rogers. (1995). Origins and affinities of modern humans: A comparison of mitochondrial and nuclear genetic data. American Journal of Human Genetics 57:523-538.

Jorde L.B., W.S. Watkins, M.J. Bamshad, M.E. Dixon, C.E. Ricker, M.T. Seielstad, and

M.A. Batzer. (2000). The distribution of human genetic diversity: A comparison of mitochondrial, autosomal, and Y chromosomal data. American Journal of Human Genetics 66: 979-988.

Kaessmann H., V. Wiebe, and S. Pääbo. (1999). Extensive nuclear DNA sequence diversity among chimpanzees. Science 286:1159-1162.

Kayser M., A. Caglia, D. Corach, N. Fretwell, C. Gehrig, G. Graziosi, F. Heidorn, S. Herrmann, B. Herzog, M. Hidding, K. Honda, M. Jobling, M. Krawczak, K. Leim, S. Meuser, E. Meyer, W. Oesterreich, A. Pandya, W. Parson, G. Penacino, A. Perez-Lezaun, A. Piccinini, M. Prinz, C. Schmitt, P.M. Schneider, R. Szibor, J. Teifel-Greding, G. Weichhold, P.d. Knijff, and L. Roewer. (1997). Evaluation of Y-chromosomal STRs: a multicenter study. International Journal of Legal Medicine 110:125-133.

Kennedy K A.R., A. Sonakia., J. Chiment, and K.K. Verma. (1991). Is the Narmada Hominid an Indian *Homo Erectus*. American Journal of Physical Anthropology 86:475-496.

Kennedy K.A.R. and S U. Deraniyagala. (1989). Fossil remains of 28,000- Year- old Hominids from Sri Lanka. Current Anthropology 30, 394-399.

Kennedy K.A.R., S U. Deraniyagala, W.J. Roertgen, J. Chiment, and T. Disotell. (1987). Upper Pleistocene fossil Hominids from Sri Lanka. American Journal of Physical Anthropology 72: 441-461.

Kittle R.A., M. Perola, L. Peltonen, A.W. Bergen, R.A. Aragon, M. Virkkunen, M. Linnoila, D. Goldman, and J.C. Long. (1998). Dual origins of Finns revealed by Y chromosome haplotype variation 62:1171-1179.

Kivisild T., M.J. Bamshad, K. Kaldma, M. Metspalu, E. Metspalu, M. Reidla, S Laos, J. Parik, W.S. Watkins, M.E. Dixon, S.S. Papiha, S.S. Mastana, M.R. Mir, V. Ferak, and R. Villems. (1999). Deep common ancestry of Indians And Western – Eurasians Mitochondrial DNA lineages. Current Biology 9: 1331-1334.

Knijff P.d., M. Kayser, A. Caglia, D. Corach, N. Fretwell, C. Gehrig, G. Graziosi, F. Heidorn, S. Herrmann, B. Herzog, M. Hidding, K. Honda, M. Jobling, M. Krawczak, K. Leim, S. Meuser, E. Meyer, W. Oesterreich, A. Pandya, W. Parson, G. Penacino, A. Perez-Lezaun, A. Piccinini, M. Prinz, C. Schmitt, P.M. Schneider, R. Szibor, J. Teifel-Greding, G. Weichhold, and L. Roewer. (1997). Chromosome Y microsatellites: population genetics and evolutionary aspects, International Journal of Legal Medicine 110: 134-140.

Lahermo P., A. Sajantila, P. Sistonen, M. Lukka, P. Aula, L. Peltonen, and M-

L. Savontaus. (1996). The genetic relationship between the Finns and the Finnish Saami (Lapps): analysis of nuclear DNA and mtDNA. American Journal of Human Genetics 58:1309-1322.

Lahr M.,and R. Foley. (1994). Multiple dispersals and modern Human Origins. Evol Anth,
3:48-60.

Lewontin R.C. (1972). The apportionment of human Diversity. Evolutionary Biology 6:381-398.

Li W. and L.A. Sadler. (1991). Low nucleotide diversity in Man. Genetics 129:513-523.

Lumley H.D., and A. Sonakia. (1985). Contexte Stratigraphique Et Archeologique De L'Homme De La Narmada, Hathnora, Madhya Pradesh, Inde. L'Anthropologie (Paris) 89: 3-12.

Majumder P.P. (1998). People of India: Biological diversity and affinities. Evolutionary Anthropology 6: 100-110.

Majumder P.P., B. Roy, S. Banerjee, M. Chakraborty, B. Dey, N. Mukherjee, M. Roy, P. Guha Thakurta, and S.K. Sil. (1999). Human-specific insertion/deletion polymorphisms in Indian populations and their possible evolutionary implications. European Journal of Human Genetics 7: 435-446.

Mastana S.S., J. Constans, and S. S. Papiha. (1996). New rare variants and interpretation of the genetic diversity at the group-specific component (GC) locus in Population of India. Human Biology 68:795-818.

Merril C.R., D. Goldman, S.A. Sedman, and M.H. Ebert, (1981). Ultrasensitive stain for proteins in polyacrylamide gels shows regional variation in cerebrospinal fluid proteins, Science 211: 1437-1438.

Merriwether D.A., A.G. Clark, S.W. Ballinger, T.G. Schurr, H. Soodyall, T. Jenkins, S.T. Sherry, and D.C. Wallace. (1991). The structure of human mitochondrial DNA variation. Journal of Molecular Evolution 33:543-555.

Moreland W.H., and A.C. Chatterjee, (1960). A short history of India. Fourth Edition. Longmans. 1-131.

Mountain J.L., J.M.H. Hebert, S Bhattacharyya, P.A. Underhill, C. Ottolenghi, M. Gadgil, and L.L. Cavalli- Sforza. (1995). Demographic history of India and mtDNA- sequence diversity. American Journal of Human Genetics 56: 979-992.

Neill S. (1970).The story of Christian churches in India and Pakistan. William B. Eermans Publishing Company Grand Rapids, Michigan.

Oota H., N. Saitou, T. Matsushita, and S. Ueda. (1999). Molecular genetic analysis of remains of a 2,000-Year –Old human population in China- and Its relevance for the Origin of the modern Japanese population. American Journal of Human Genetics 64:250-258.

Papiha S.S. (1996). Genetic variation in India. Human Biology 68: 607-628.

Papiha S.S., M.S. Schanfield., and R. Chakraborty. (1996). Immunoglobulin allotypes and estimation of genetic admixture among populations of Kinnaur district, Himachal Pradesh, India. Human Biology 68: 777-794.

Passarino G., O. Semino, L.F. Bernini, and A.S. Santachiara-Benerecetti. (1996). Pre-Caucasoid and Caucasoid genetic features of the Indian population, Revealed by mtDNA polymorphism. American Journal of Human Genetics 59: 927-934.

Passarino G., O. Semino, L. Quintana- Murci, L. Excoffier, M. Hammer, and A.S. Santachiara- Benerecetti. (1998). Different genetic components in the Ethiopian population, identified by mtDNA and Y–chromosomal polymorphism. American Journal of Human Genetics 62: 420-434.

Pérez-Lezaun A., F. Calafell, D. Comas, E. Mateu, E. Bosch, R. Martínez-Arias, J. Clarimón, G. Fiori, D. Luiselli, F. Facchini, D. Pettener, and J. Bertranpetit. (1999). Sex-Specific migration patterns in Central Asian populations, revealed by analysis of Y-chromosome short tandem repeats and mtDNA. American Journal of Human Genetics 65:208-219.

Pérez-Lezaun A., F. Calafell, M. Seielstad, D. Comas, E. Bosch, and J. Bertranpetit. (1997). Population genetics of Y-chromosome short tandem repeats in humans. Journal of Molecular Evolution 45:265-270.

Piazza A. 1998. Towards a genetic history of China. Nature 395:636-639.

Quintana-Murci L., C. Krausz, T. Zerjal, S.H. Sayar, M.F. Hammer, S.Q. Mehdi, Q. Ayub, R. Qamar, A. Mohyuddin, U. Radhakrishna, M.A. Jobling, C. Tyler-Smith, and K. McElreavey. (2001). Y-chromosome lineages trace diffusion of people and languages in Southeast Asia. American Journal of Human Genetics 68:537-542.

Quintana-Murci L, O. Semino, H.J. Bandelt, G. Passarino, K. McElreavey, and A.S. Santachiara-Benerecetti. (1999). Genetic evidence of an early exit of *Homo sapiens sapiens* from Africa through Eastern Africa Nature Genetics 23: 437-41.

Raymond M. and F. Rousset. (1994). An exact test For population differenciation. Evolution 49: 1280-1283.

Redd AJ, M. Stoneking. (1999). Peopling of Sahul: mtDNA variation in Aboriginal Australian and Papua New Guinean populations. American Journal of Human Genetics 65:808-828.

Registrar General and Census Commissioner Of India, (1999). The Census of India Website. http://www.censusindia.net

Relethford J.H. (1995). Genetics and modern human origins. Evolutionary Anthropology 95:53-63.

Relethford J.H., and H.C. Harpending. (1994). Craniometric variation, genetic theory and modern human origins. American Journal of Physical Anthropology 95:249-270.

Relethford J.H., and L.B. Jorde.(1999). Genetic evidence for Larger African population Size during recent human evolution 108:251-260.

Renfrew C., (1989). The origin of Indo-European languages. Scientific Americans 261: 82-90.

Rolf B., E. Meyer, B. Brinkmann, P.d. Knijff. (1998). Polymorphism at the tetranucleotide repeat locus DYS 389 in 10 populations reveals strong geographic clustering. European Journal of Human Genetics 6: 583-588.

Tambia S.J.(1973).From varna to caste through mixed unions. Googy J.,(ed). Character of Kinship. Cambridge University Press, Great Britian. 191-229.

Sankhyan A.R. (1997). Fossil clavicle of A middle Pleistocene Hominid from the central Narmada valley, India. Journal of Human Evolution 32: 3-16.

Seielstad M.T., J.M. Hebert, A.A. Lin, P.A Uderhill, M. Ibrahim, D. Vollrath, and L.L. Cavalli-Sforza. (1994). Construction of human Y- chromosomal haplotypes using a new polymorphic A to G transition. Human Molecular Genetics 3: 2159-2161.

Semino O., A. Torroni, R. Scozzari, A. Brega, and A.S.S.Benerecetti. (1991). Mitochondrial DNA polymorphisms among Hindus: A comparison with the Tharus of Nepal. Annual Human Genetics 55:123-136.

Sokal R.R., and F.J. Rohlf.(1995) Biometry. W.H.Freeman and Company, New York. 292.

Spurdle A.B., M.F. Hammer, and T. Jenkins. (1994). The Y Alu Polymorphism in

Southern African populations and its relationship to other Y – specific polymorphism. American Journal of Human Genetics 54: 319-330.

Stoneking M., J.J. Fontius, S.L.Clifford, H. Soodyall, S.S Arcot., N. Saha, T. Jenkins, M. Tahir, P.L. Deininger, and M.A. Batzer. (1997). *Alu* insertion polymorphisms and human evolution: Evidence for a larger population size in Africa. Genome Research 7:1061-1071.

Su B., J. Xiao, P. Underhill, R. Deka, W. Zhang, J. Akey, W. Huang, D. Shen, D. Lu, J. Luo, J. Chu, J. Tan, P. Shen, R. Davis, L. Cavalli-Sforza, R. Chakraborty, M. Xiong, R. Du, P. Oefner, Z. Chen, and L. Jin. (1999). Y-chromosome evidence for a northward migration of modern humans into Eastern Asia during the last Ice Age. American Journal of Human Genetics 65:1718-1724.

The Burusho. Sadie Anderson. Website, http://www.anthro.mnsu.edu/cultural/old world/asia/burusho.html.

The Major world Religions. Website http://www.omsakthi.org/religions.html.

Watkins W.S., M. Bamshad, M.E. Dixon, B. Bhaskara Rao, J.M. Naidu, P.G. Reddy, B.V.R Prasad, P.K. Das, P.C. Reddy, P.B. Gai, A. Bhanu, Y.S. Kusuma, J.K. Lum, P. Fischer, and L.B. Jorde. (1999). Multiple origins of the mtDNA 9-bp deletion in populations of South India. American Journal of Physical Anthropology 109:147-158.

Watson E., P. Forster, M. Richards, and H. Bandelt. (1997). Mitochondrial footprints of human expansions in Africa. 61: 691-704.

Weber J.L., and C. Wong. (1993). Mutation of human short tandem repeats. Human Molecular Genetics 2: 1123-1128.

Wolpert S., (1989). A new history of India, Third Edition, Oxford University Press, New York, Oxford, 1-134.

Wright S., (1951). The genetic Structure of Populations. Ann Eugen. 15:323-354.

Wright S., (1965). The interpretation of population structure by F –statistics with special regards to system of mating. Evolution 19: 395-420.

Zvelebil K.V., (1990). Dravidian linguistics An introduction. Pondicherry Institute of Linguistics and Culture.