THE GENETIC HISTORY OF THE SARASWAT BRAHMINS:
ORIGINS AND AFFINITIES WITH INDIAN POPULATIONS

By

Athma Anjali Pai

AN UNDERGRADUATE RESEARCH PAPER

In

Anthropology

Submitted to the
Department of Anthropology
University of Pennsylvania

Thesis Advisor: Dr. Theodore G. Schurr

2007

Table of Contents

Index of Supplementary Material

Abstract:

The Saraswat Brahmin community of western India is a highly endogamous group which has retained a distinct Indo-European language and culture through centuries of migration and displacement from various parts of India, and despite living in Dravidic dominated South-India for hundreds of years. This study represents the first analysis of genetic variation in this group using mitochondrial (mt) DNA markers. Mitochondrial markers were analyzed in 116 Saraswat Brahmin males residing in North America. Due to the effects from both high endogamy and complex migratory patterns, it previously was hypothesized that the haplotype frequencies in Saraswat Brahmins would distinguish them from other Indian language communities, although they might exhibit varying degrees of genetic affinity with geographically related language communities. The results have shown that the mtDNA haplotypes found in Saraswat Brahmins are predominantly South Asian-specific haploroups, though there are several unique variations on commonly found haplotypes. In general, there is some evidence for a higher maternal affinity with Indo-European North Indian groups. Analysis of male-specific Y-chromosome variation could enrich the data of distinct maternal lineages as presented in this thesis.

I. Background to Research Problem

I.1 Brief History of India

The South Asian subcontinent is an area of enormous linguistic, cultural, religious, historical, and biological diversity. The largest country on the subcontinent, India, particularly exemplifies this diversity with its 4,693 recognized ethnic communities and 325 functioning languages with 25 scripts (Singh, 2002). With more than a billion inhabitants, the country is the second most populated nation in the world. The extensive variation on all levels seems to have been generated over millennia of migrations, cultural development, and world interaction all happening in a topographically varied geographical crossroads for modern human migrations (TIGV, 2005). The initial inhabitants of India were possibly some of the earliest modern humans that migrated out of Africa in the Pleistocene (Kivisild et al., 1999a; Kivisild et al., 1999b; Kivisild et al., 2003). These settlers of the subcontinent are believed to be represented in the indigenous tribal populations that still inhabit India today. Later, it is known that pockets of extensive civilizations built up in the northern river regions of India with the Harappan culture around 2800 BCE, but then declined around the arrival of Indo-European speaking Aryan groups approximately around 1200 BCE (Snell, 1987). These Aryan nomads are credited by scholars for developing a religious caste hierarchy, spreading Indo-European language and culture, and pushing the Dravidian civilizations southward (Snell, 1987), though the details concerning their arrival and impact have long been debated.

Today, India has two dominant linguistic and cultural communities with anthropological distinctions that probably arose because of events related to the Aryan arrival (Rajkumar and Kashyap, 2004). Most northern communities are Indo-European speakers, while South Indians are generally Dravidian speakers. Both linguistic communities have had significant influences in shaping what is now thought of as general Indian culture (Basu et al., 2003). Overall, the history of India is one of a continuous inward flow of people either migrating or invading while leaving their marks on culture and the gene pool (Basu et al., 2003). There have been multiple communities leaving an indelible mark on India, including the Han Chinese, Greeks, Scytho-Iranians, Kushans, Jewish tribes, Zoroastrians, Moghuls, Muslims, Dutch, French, Portuguese, English, etc (Gaikwad and Kashyap, 2005; Kivisild et al., 2003; Rajkumar and Kashyap, 2004; Snell, 1987).

Several unique characteristics of Indian society need to be taken into account with any anthropological studies, including this study. Two of the major features are (1) the distinctions between non-Hindu tribal populations and the rest of Indian society and (2) the Hindu caste system that has embedded itself thoroughly into historical and contemporary Indian society. The non-Hindu tribal populations spread throughout India are thought to be the descendants of the original inhabitants who were barely impacted by the arrival of the Aryans – other than through geographical displacements either to the south or into remote forest regions (Metspalu et. al. 2004; Rajkumar et. al. 2005). Tribal populations make up a little less than 8% of the residents of India today (Census, 1991). They mostly follow ancestral worship, and speak languages in all of the four major linguistic groups found in India (Cordaux et. al. 2003). Due to their high isolation, they

are very interesting to study to elucidate elements of early Indian culture and genetic origins.

The second factor, the Hindu caste (or *varna*) system, is said to have originated with the arrival of the Aryans into India (Kivisild et. al. 2003; Bamshad et. al. 2001). Aryans incorporated themselves into the highest caste and the native populations into a spectrum of castes below their own. The system is made up of 5 predominant categories: Brahmin (priest), Kshatriya (warrior), Vaishya (trader/businessman), Sudra (laborer/farmer), and Panchama (the untouchable or absolute menial laborer). This system is mentioned in several Hindu texts and seen as a fundamental system of order for society that incorporates both the division of labor and the philosophical principles of rebirth and the attainment of higher goals (Snell, 1987). Due to strict cultural practices and high rates of endogamy, these castes seem to represent unique homogeneous communities that can be studied as particular units within the amalgam of Indian society. It is important to note, however, the inherent difference between a caste and a *jāti*. While castes can be thought of as a broader social grouping, *jātis* are more regional in nature. *Jātis* are particular communities that share a caste, language, culture, ancestral connection, and oftentimes a common occupation. Therefore, *jātis* are the actual units that can be considered to be homogeneous in nature, and can be considered as separate populations that might have more genetic similarity than a wider caste sampling.

I.2 Introduction to Biogenetic Studies in India

Anthropological genetic studies are often conducted (like the study in this thesis), to elucidate elements of historical, cultural, or evolutionary significance. Studies utilize

either classical markers (blood groups, etc), or molecular markers (DNA, genes, etc). The two most commonly used molecular markers are mitochondrial (mt) DNA and the Y-chromosome, both of which provide varying information about maternal and paternal histories of populations respectively. An individual's sequence of diagnostic DNA mutational markers is referred to as a haplotype. Haplotypes are then grouped into larger clusters called haplogroups (referred to by a letter nomenclature) that encompass lineages that share a discernable recent common ancestor.

The large diversity of biological variation, languages, religions, and cultural practices of people on the Indian subcontinent has led to an intensification of genetic diversity research on Indian communities (Baig et al., 2004; Bamshad et al., 2001). Remarkably, approximately one fifth of the human gene pool belongs primarily to people inhabiting the Indian subcontinent (Kivisild et al., 1999a). Early work focused on estimating the distinctions between groups in the well-known Indian caste systems, delineating proto-Asian versus West Eurasian origins of peoples, estimating molecular dates for waves of settlement on the subcontinent, and mapping genetic data onto language trees (Baig et al., 2004; Bamshad et al., 2001; Kivisild et al., 1999a; Kivisild et al., 2003; Palanichamy et al., 2004). Many studies investigating the spread of modern humans out of Africa have used genetic studies in India to demonstrate the rapid dispersal of humans along the Asian coasts after leaving Africa (Sun et al., 2006). Mitochondrial and Y-chromosome studies are especially relevant for Indian populations because initial results have shown that Indian lineages contain some very early branches of common haplotypes (Kivisild et al., 1999a; Palanichamy et al., 2004).

Many molecular genetics studies of populations from the South Asian continent have attempted to solve the problems of macrohaplogroup (M, N, and R) differentiation, clusters in the continent, patterns of regional variation, and caste differentiation. Due to the geographical placement of the subcontinent, it is easy to see how Indians are considered to be concurrently part of both the Eastern and Western Eurasian "metapopulations" (Kivisild et al., 1999b). Analytical work has demonstrated that about 60% of Indians today have maternal haplotypes belonging to one of the Indian-specific branches of haplogroup M (Bamshad et al., 2001; Kivisild et al., 2003). Such a finding is by no means surprising, as India has long been considered as an incubator for specific M haplotypes, leading to hypotheses that many Indian tribal and caste populations derive from extremely similar genetic heritages due to limited Pleistocene gene flow from regions outside of South Asia (Kivisild et al., 2003).

The deepest roots of the M lineage are found in India, showing signs that this macrohaplogroup most probably originated in South Asia (Palanichamy et al., 2004; Rajkumar et al., 2005; Sun et al., 2006). In addition, the frequency and diversity of the undifferentiated M haplotypes is found to be the highest in India (Kivisild et al., 2003). Many of the mutations that determine M lineages in India are rarely found in gene pools of populations residing in East Asia, Oceania, and Southeast Asia (Sun et al., 2006; Metspalu et. al. 2004). There are also occurrences of Indian specific U (N haplogroup) and R haplotypes, along with non-Asian haplogroups most commonly considered to be from West Eurasia (Bamshad et al., 2001). Three of these haplogroups (M2, U2i, and R5) are considered to be the oldest Indian lineages, and posited to be associated with the initial inhabitation of South Asia (Metspalu et al., 2004).

A number of studies have also investigated the frequencies of West Eurasian haplogroups in caste and tribal populations within India (Baig et al., 2004; Bamshad et al., 2001; Basu et al., 2003; Kivisild et al., 2003). Tribal and caste populations are highly differentiated based on genetic distance estimates (Basu et al., 2003). Studies have shown that affinities to European populations due to higher frequencies of West Eurasian haplogroups is highest in caste groups and proportionately increased by higher caste rank (Bamshad et al., 2001; Basu et al., 2003). These findings support the idea that the Indian caste system was introduced by invading Indo-Aryans and later Eurasian admixture through migrations and invasions (Baig et al., 2004; Bamshad et al., 2001; Basu et al., 2003; Kivisild et al., 2003). The linguistic similarities of the Indian branch of the Indo-European language tree and West European Indo-European languages most probably reflect contact on just a linguistic level rather than both linguistic and genetic levels at that early stage (Bamshad et al., 2001). Rather, the Indian lineages commonly thought to be common West Eurasian specific can probably be ascribed to later admixture which accounts for less than 10% of the current Indian gene pool and relatively minor impacts on the previously existing diversity of India (Kivisild et al., 1999a; Kivisild et al., 2000; Metspalu et al., 2004).

There are distinct differences in caste affiliation with respect to data from mtDNA and Y-chromosome markers. The maternal lineages present in Indian groups show a stronger East Asian similarity. In constrast, paternal lineages point to an increase in Eurasian haplogroups proportionate to caste rank, an idea supported by results showing that less than 10% of mitochondrial DNA in caste populations can be connected to non-Indian specific lineages (Baig et al., 2004; Bamshad et al., 2001; Kivisild et al., 2003).

Despite this general pattern, researchers have also noted the regional variation in haplogroup frequencies across India. There is a wide geographic dispersal of caste populations, leading to geographically dependent connections between caste and tribal populations (Bamshad et al., 2001). While one can argue for geographical or evolutionary connections between Indian caste and tribe populations, it is hard to ignore the possibility of West Eurasian admixture through various historical invasions that occurred throughout India's history (Kivisild et al., 2003).

I.3 Introduction to Saraswat Brahmins

Saraswat Brahmins are an Indian sub-caste, or *jāti*, distinguished from other Indian Brahmin communities by their history, traditions, and language (Bhatia et al., 1976; Conlon, 1977; Keni, 1998). The term "Saraswat Brahmin" is commonly used to refer to members of the wider caste-cluster known as the Gaud Saraswat Brahmin (GSB) cluster, and will be used in this study to refer to all members of this cluster (Conlon, 1977). The religious sub-structure of this caste-cluster will be discussed later. Saraswat Brahmins belong to the Gaud (northern) division of the 10 predominant Brahminical divisions in Hinduism (Conlon, 1977). They have maintained this high Brahmin identity through historical, though they have had to deal with criticism and distrust from fellow Brahmin groups because of the inclusion of fish in their diet.

The most notable characteristic of Saraswats is their use of Konkani as a native language. Konkani is a direct derivative of the Sanskrit language and is spoken by between 5 to 7 million speakers around the world today (Census, 1991; Keni, 1998). Although found sprinkled through Southern India, Konkani is the state language of Goa –

the most recent common homeland of the Saraswats. It seems that the membership criterion for the Saraswat Brahmin community is one that is common through most of Hindu India – i.e., they rely on a "natural identity," or an ancestral relationship to be considered part of the community (Conlon, 1977). However, their uniqueness as a community goes past their common use of Konkani, as their culture and history uncovers a complex pattern of social and cultural components.

Saraswats have an intricate past which has been explained through both oral history and scholarly investigations. Unfortunately, the historical documentation of Saraswat Brahmins in Goa is forever lost due to both natural forces and the Portuguese invasion (Conlon, 1977). Legend posits that they originated in Kashmir on the banks of the now dry Saraswati River, which gave its name to the community. The Hindu *Puranas* (religious legends built around historical events) tell a story of the Brahmin war-leader Parashurama leading the 'Aryans' (Brahmins) of the Saraswati River banks to his newly created land on the Konkan Coast (Keni, 1998). This legendary outline is not uncommon, as most dominant Brahmin castes along the western coast of India have similar origin myth linked to Parashurama (Conlon, 1977).

However, scholarly evidence however, follows a similar pattern of travel from Kashmir through Punjab, the Sind, Gujurat, Bengal, and Maharashtra before arriving in Goa, which is now considered the Saraswat homeland (Bhatia et al., 1976; Conlon, 1977). This supposed migratory route is shown by the black line in **Figure 1**. Their stay in Bengal on the eastern coast of India has been supported by the early appearance of fish into the diet of Bengali Brahmins and the parallels between Konkani and Bengali dialects

(Conlon, 1977; Keni, 1998). Religiously, Saraswats maintained many similar clan names and practices that clearly indicate their northern Gaud Brahmin origins (Conlon, 1977).

In Goa, they prospered and established themselves as the principal brahminical elite employed as land-owners, scribes, accountants, and master traders (Conlon, 1977). By the late 15th and early 16th centuries, Saraswats had began to move out of Goa – seeking more farmland, trade opportunities, or expanded employment potential (Conlon, 1977; Keni, 1998). When the Muslim Bijapur empire won control over Goa from the Hindu Vijayanagar dynasty, social order started to deteriorate significantly and fell apart after the Portuguese conquer of Goa started in 1510 (Conlon, 1977).

The Saraswats were subjected to persecution and forced to make religious conversions, resulting in many Konkani speaking Muslims and Catholics who can trace their ancestry back to Goan Saraswat Brahmins. Although Goan Hindus initially fought for religious freedom, large groups of Saraswats fled Goa to southern Maharashtra in the north and Karnataka and Kerala in the south to escape persecution. In 1559 CE, King Joao III demanded the expulsion of non-Christians from Goa, which prompted 12,000 Saraswat families to flee both to the north and south, as shown by the yellow lines in **Figure 1** (Keni, 1998).

As a result, Saraswats can now be found spread across the Konkan (western) coast of India (Bhatia et al., 1976; Conlon, 1977; Keni, 1998; Patil, 1970). However, the migrations clearly began before the Portuguese decrees and seem to have been instigated by Saraswat traders traveling as far south as Kerala. Those who moved north into Maratha-controlled Maharashtra were successful as Brahmin advisors in both politics and the economy (Conlon, 1977). Those moving south into Karnataka also prospered –

finding land to irrigate and employment in many port and trading towns, where they became known as "Konkanis," a moniker continuing throughout Karnataka even today (Conlon, 1977). Saraswats traveling even farther south into Kerala had to initially deal with the hostility of the Portuguese-fearing Muslim Zamorin of Kozikhode (Calicut) before they continued south to prosper as traders in Kochi (Cochin) (Keni, 1998).

Despite their geographic spread and relatively seamless into new areas, the Saraswat Brahmin community has maintained a surprising degree of "community consciousness" (Patil, 1970). Scholars attribute this to a continuance of the Konkani language, religious and cultural traditions, and high rates of endogamy (Conlon, 1977; Keni, 1998; Patil, 1970). Although the migration out of Goa broke the established geographically-based society, displaced Saraswats were still united by a remembrance and longing to go back to Goa (Conlon, 1977). However, divisions did eventually occur as at present the community is divided into four distinct religious communities (*mutts*), although association and intermarriage among members of these communities is common. The four communities – Kavle *Mutt* (~740 CE), Kashi *Mutt* (~1300 CE), Gokarn Parthagali *Mutt* (~1476 CE), and Chitrapur Saraswat *Mutt* (~1708 CE) – most probably developed out of geographic connections and various concurrent rising movements in Hinduism (Keni, 1998). Under strict definitions, these four religious groups would represent distinct Hindu subcastes. However, through their similar practices and tolerance of sub-caste intermarriage, they all fall under the common Gaud Saraswat Brahmin caste-cluster, as mentioned earlier.

The northern heritage of Saraswats living in South India is evident in their language (Konkani is one of the few Indo-European languages found in Dravidic

dominated South India), unique cuisine, and religious practices (Keni, 1998). Due to high education rates, Saraswats have become prominent within the professional and business communities in recent years. Their movement out of districts in Karnataka and Kerala has been a recent phenomenon. Early in the 20th century, they were widely distributed in villages along the Konkan coast, with very few living in major cities (Conlon, 1977). Following independence movement and post-independence trends, many Saraswats relocated to Mumbai (Bombay) (Conlon, 1977). Like many other Indians, Saraswat Brahmins have now immigrated outside of India, including numerous individuals who have settled in the United States and Canada (Keni, 1998).

Although well studied for their culture and religious practices, there has been very little genetic work done on the Saraswat Brahmin community. The complex migratory patterns and high degree of endogamy as a minority high-caste Indian group makes the community an interesting target for genetic work. The only ascertainable work that has been done on Saraswats was a classical genetic marker study by Bhatia et al. (1976) in which ABO and Rhesus blood marker variation was assessed in members of the community in various Indian cities (Bhatia et al., 1976). Because of the nature of these marker systems, this study provided limited information about the degree that geographic separation and endogamy has affected genetic variation in Saraswats. The increased resolution available through newer molecular methods and a growing knowledge of genetic ancestry markers will allow a more thorough investigation of the genetic make-up of this population.

An interesting addendum to the Bhatia et al. (1976) study was their conclusion that the Saraswats may have been genetically influenced by Greek invasions in Punjab

due to notable levels of Cooley's thalassemia found in both Saraswats and Greeks (Bhatia et al., 1976; Patil, 1970). The possible finding of Greek ancestry among this group would be quite remarkable. Overall, this research into the population history of Saraswat Brahmins will certainly contribute to our understanding of the genetic history of Indians and possibly provide an example to understand the nuances of admixture and migration on the Indian subcontinent.

The most comparable recent molecular research has been on other Brahmin populations from the Konkan coast. Gaikwad and Kashyap (2005) looked at mtDNA and Y-chromosomal variation in four Maharasthrian populations in this region. The population that they looked at with the closest cultural and migratory history to Saraswats was the Chitpavan Brahmins. They hypothesized that the continuous trade and cultural contact between populations on the Konkan coast and non-Indians allowed for extensive genetic diversity, especially after historical evidence that some invading populations were incorporated into the caste system (Gaikwad and Kashyap, 2005). The study pointed to the uniqueness of the Konkan coast as the residence of a wide variety of cultural groups, such as the Indian Jews, Parsis (Zorastrians), Anglo-Indias, and the Indo-Portuguese (Gaikwad and Kashyap, 2005). Their results indicated that the Chitpavan Brahmins are genetically associated with Western-Eurasian genotypes through both non-recombining uniparental (especially paternal Y-chromosome indications) and biparental variation, thus leading to the idea that they are of "Scytho-Iranian" ancestry (Gaikwad and Kashyap, 2005).

I.4 Specific Aims of Study

India is a unique region of the world with regard to genetic diversity and population history studies. Both classical and molecular marker studies have shown that India has the greatest amont of genetic diversity among comparable regions outside of Africa, making it the perfect "natural genetic laboratory" (Majumder, 1998; Rajkumar and Kashyap, 2004; TIGV, 2005). Indian populations present opportunities to conduct a variety of studies due to their many ascertainable sub-structures, the geographic importance in the spread of modern humans, and the relatively low genetic diversity values useful for various disease studies (Kivisild et al., 2000; Rajkumar et al., 2005; TIGV, 2005). It will also be important to study as many Indian populations as possible since the genomic unity of these diverse populations allows the ability to test the generality of specific results (Bamshad et al., 2001; TIGV, 2005). It will be especially interesting to study the southwest regions of India due to extensive past colonization by European traders, large trade networks, and a diversity of linguistic and cultural practices spread over a small geographically isolated region (Gaikwad and Kashyap, 2005; Rajkumar and Kashyap, 2004; Roy et al., 2003).

This particular study seeks to address many related questions about the history of Indian genetic diversity by using Saraswat Brahmins as a case example. As a high caste population with potentially significant degrees of admixture, Saraswat Brahmins represent a unique group from which there are few genetic data. This study clarified these elements of genetic population history by (1) examining the haplogroup diversity of the population, (2) measuring degrees of admixture with related groups, (3) investigating possible migration routes seen through genetic mixing and (4) comparing results with

frequencies among non-Indian populations which might have some relation to the Saraswats. Ultimately, the Saraswat Brahmin genetic diversity results present a model for the uniqueness and variation within the larger arena of Indian genetic diversity.

II Research Design and Methods

II.1 Sample Collection

To elucidate the genetic diversity of the Saraswat Brahmins, I planned to analyze mtDNA and NRY variation. This study is part of a larger collaboration with Dr. T.D. Dogra, Department of Forensic Medicine and Toxicology, All Indian Institute of Medical Sciences, New Delhi, India, who is undertaking a similar analysis of genetic variation in Saraswat Brahmins there. Under the terms of the collaboration, I will receive and analyze some of the DNAs obtained by Dr. Dogra and exchange data to facilitate further statistical and phylogenetic analyses to determine the relative biological relationships of the different Saraswat populations.

After receiving University of Pennsylvania Institutional Review Board approval for this proposed study and monetary support from a College Alumni Award for Undergraduate Research, I prepared for sample collection. Knowing that samples would be collected during the course of the summer of 2006 and in a variety of ways, I put together a kit that would simplify the collection process. The kit contained an Informed Consent form (Appendix I) that all participants were required to sign for use of their biological samples, a Genealogical form (Appendix II) which asked genealogical questions about themselves and their families for 2-3 generations, two Omni Swab applicators with two small cryovials with stabilizing TE buffer to collect two buccal cell samples from each participant, and an Information sheet (Appendix III) to provide basic information about the project and instructions on how to donate samples using the applicators. Through interviews and the genealogical data, I was able not to collect

samples from individuals who were closely related on either the maternal or paternal sides of their families. However, given that the Saraswat Brahmins are part of a highly endogamous community with documented cases of close familial marriage, extra caution had to be taken to ensure that individuals were not related over the past 2-3 generations.

As a member of the Saraswat Brahmin community myself, I had easier access to leaders and organizations of the non-resident Indian (NRI) community settled in the North America. This community is mostly comprised of families that left India over the past 50 years and resettled in the United States and Canada. I recruited participants for this study through my personal contacts within the regional communities. The recruitment of study participants and the collection of genealogical data and buccal cell samples from these participants were conducted in two ways. First, I contacted potential participants through e-mail and telephone. To those who agreed to participate, I mailed a kit (described above) with information about the project and the means to provide their sample. The second of these two collection methods involved traveling to Saraswat gatherings to talk to potential participants and either hand out kits or take samples on site.

The largest of these gatherings happened from July 1-4, 2006, in Hamilton, Ontario, Canada. Called the North American Konkani *Sammelan*, this gathering was a community convention that occurs every two years to bring together Konkani speakers who have settled throughout North America and other places. In reality, the event is a four day festival of performances, speeches, workshops, exhibitions, and food. The main purpose seems to be to unite the NRI Saraswat Brahmin community and allow members a chance to enjoy with family and long-time friends in one location.

Knowing that this would be an ideal place to recruit large numbers of un-related participants, I contacted the organizers of this convention. Like other members of the community, the organizers showed considerable interest in the study since Saraswat Brahmins – like many other minority communities – are very interested in elucidating their population history in any way. Many members of the North American community are doctors, so the organizers of the *Sammelan* referred me to the head of the concurrently organized Saraswat Brahmin medical symposium. Following the assumption that mentioning this study to doctors would obtain scientific validity throughout the community, the head of the medical symposium, Dr. Mohan Pai, agreed to talk about and encourage participation in this study during the symposium. Through the connections of Dr. Pai, my father, and my uncle, I was able to hand out 100 kits to people recruited at the Sammelan, bringing the total number of kits mailed and handed out throughout the summer up to 150 kits. At the end, I received a sample from 77% of these kits handed out, thus resulting in 116 total number of participants.

The process of recruiting participants in person was a very interesting experience. Many people were extremely curious about either the study design, the use of the samples, or the particular molecular markers being studied. Others wanted to know more about the hypotheses of the study, the dissemination of results, or the range of molecular population history studies being conducted throughout the world. I will never forget when one elderly gentleman pulled me aside from talking to a group of potential participants and told me something to the effect of "Young lady, I've seen the growth of this science business and all the buildup around genes and DNA. But to finally tell us that we 'Konkanis' are the best community – that's worth reading about!" This comment

reflected the generally enthusiastic sentiment expressed by most community membersto whom I talked. There were a few people who were too afraid of misuse of the samples to donate to the study. I suspect that my status as a younger female member of the community who had not yet completed an undergraduate degree did not help to reassure the cynics, many of whom were primarily concerned with the general ethics of DNA testing and the possibility of disease susceptibility results.

Overall, the experience was reassuring and exciting. The participants' continued interest in my work and the ability to explain the process erased any doubts that I might have had. I left this small field experience of collecting samples with about 40 samples in hand, promises of many more in the mail, and the ethical reassurance that the community was behind the project.

II.2 Laboratory Analysis

Sample Preparation

When the samples arrived back at the lab, they were extracted with the procedure described below and marked with labels that reflected only the population to which they belong and a numerical indicator. It is imperative to respect the confidentiality of participant donation and molecular results, therefore the key showing the link between the participants' personal histories and the sample number is accessible only to myself and my advisor Dr. Schurr. In addition, any connections between participants' individual histories and results will never be divulged or published. Results will soon be sent to each participant, along with a review of the overall conclusions of the study.

The samples obtained allowed us to extract DNA and subsequently analyze variation in the mtDNA and NRY of the Saraswat Brahmin populations. These two genetic systems differ in their inheritance patterns (maternal vs. paternal), mutation rates (fast vs. slow), and levels of diversity (high vs. moderate). This thesis reflects the results of the mtDNA studies, since the Y-chromosome studies are in the process of being completed.

Extraction of DNA

Once labeled and organized, the buccal samples in cryovials were briefly vortexed to resuspend the cells in solution. Then the buffer solution with cell material was transferred to clean 1.5 mL microfuge tubes, whereupon the cells were pelleted through high speed centrifugation (15 min, 14K rpm). The supernatants were discarded and extracted with phenol-chloroform (Maniatis et al., 1982). After purification with ethanol washes, the dry DNA samples were re-hydrated in water to create stock solutions. These stocks were used to create 1:10 DNA dilutions which were then used for all further molecular analyses.

Mitochondrial DNA Analysis

The mtDNA work involved the initial sequencing of a total of 400 nucleotide base pairs each from hypervariable region I (HVR-I) (nucleotide position 16000-16400) of the coding region, using the primers and PCR conditions reported in Schurr et al. (1999) (Schurr et al., 1999). This HVR-I sequencing analysis allowed a fine-grained evaluation of mutational changes that have occurred within individual mtDNAs while also

identifying nucleotide changes which define mtDNA haplogroups. All sequencing reactions were read on an ABI 3130xl DNA Analyzer in the Laboratory of Molecular Anthropology, and the resulting sequences were aligned and compared to the mitochondrial Cambridge Reference Sequence using the Sequencer 3.1 software tool (Gene Codes Corporation).

The sequence of HVR-I single nucleotide polymorphisms (SNPs) defined the haplotypes for each individual, while subsequent polymerase chain reaction-restriction fragment length polymorphism (PCR-RFLP) analysis confirms the mtDNA haplogroup identity by looking at mutational changes that are conserved through certain haplogroup families. This analysis involves the discrimination of conserved base substitutions by restriction endonucleases, which are enzymes that cut DNA at known 4-6 base pair motif locations. The fragments produced by the digestion of amplified DNA by restriction enzymes are then analyzed using an agarose gel electrophoresis method. The PCR-RFLP analysis includes screening for common South Asian haplotype branches of undifferentiated M types (M2 – M6, M18, M25, etc.), undifferentiated R types (R2, R5, R6, etc.), and specific N types (U2, U7, W, etc.) (Metspalu et al., 2004; Palanichamy et al., 2004; Sun et al., 2006). The diagnostic HVR-I and/or PCR-RFLP SNPs that were used to assign haplogroup placement are listed in **Table 1**.

Table 1. Haplogroup Diagnostic Molecular Markers

Haplogroup	HVR-I Motif (+16,000)	Coding Region SNP	Restriction Enzyme
M	223	+10389/+10400	DdeI/AluI
M2	223, 274, 319	"	"
M3	126, 223	"	"
M4	223, 311	"	"
M5	129, 223	"	"
M6	223, 231, 362	"	"
M25	223, 304	"	"
N	223	-10389/-10400	DdeI/AluI

N5	111, 223, 311	"	"
R		+12705	MboII
R5	304	"	"
U		+12308	HinfI
U1	183C, 189, 249	"	"
U2	51	"	"
U2a	51, 206C	"	"
U2b	51	"	"
U2c	51, 234	"	"
U7	318T	"	"
J	069, 126	-13704	BstNI

II.3 Statistical and Phylogenetic Methods

Statistical Analysis

To determine the placement of the Saraswat Brahmin mtDNA diversity within other Indian populations, I conducted a statistical analysis of the various aspects of Saraswat population structure, history, and genetic relationships. A primary component of this work involved comparing the Saraswat data set with comparable data sets from other Indian populations found in published literature (Cordaux et al., 2003; Gaikwad and Kashyap, 2005; Kivisild et al., 1999a; Metspalu et al., 2004; Mountain et al., 1995; Thangaraj et al., 2003) (See **Table 7** for chart of comparison populations used). This was done by using a number of statistical tools available in the Arlequin 3.0 software package (Excoffier et al., 2005) to assess Saraswat intra- and inter-population variation. Intra-group genetic distances were analyzed individually for each data set using the unbiased gene diversity estimate (Nei, 1972). This statistic determines the probability that two alleles drawn at random from a population will be different from each other (Jobling et al., 2004). The sequence data were also used to calculate mean pairwise differences (a simple description of overall diversity in the population which can be represented in a graphical form) and nucleotide diversity estimates (a measure similar to gene diversity that is independent of sample size) estimates in Arlequin (Excoffier et al., 2005). I also

used Arlequin to calculate the pairwise F_{ST} values, which measure the proportion of total variance in allele frequency that is seen between subpopulations. To examine the inter-population diversity, I used the AMOVA method (Excoffier et al., 1992), which estimates the genetic structure indices of populations using the allelic context of haplotypes and their frequencies. Both the F_{ST} and AMOVA methods assess the proportion of genetic variance due to population subdivision. The significance of the distances between subdivisions was analyzed with non-parametric permutation tests.

Phylogenetic Analysis

For the phylogenetic analysis of this data, I constructed networks using two methods. With the median-joining network method (Bandelt et al., 1999; Bandelt et al., 1995), I generated a network of mtDNA haplotypes to better understand the phylogenetic connections between the haplotypes present in Saraswats. Such networks construct trees from ancestral nodes while allowing for maximum parsimony in the branching relationships (Jobling et al., 2004). In addition, a neighbor-joining (N-J) tree (Saitou and Nei, 1987) was constructed with PHYLIP 3.572c (Felsenstein, 1998) from the F_{ST} genetic distances estimated from haplotype frequencies and DNA sequence data. The N-J method is based on the 'minimum evolution' method to create the best tree possible while attempting to have the shortest sum of branch lengths (Jobling et al., 2004).

To investigate the migration patterns of Saraswat Brahmins from north to south India, I estimated the age of the genetic lineages using diagnostic polymorphisms for these lineages that signal their emergence as distinct clusters in the overall phylogeny. With the median network tree, mtDNA HVR-I data were used to estimate the coalescence

times for each haplogroup branch, using the mutation rate of Forster and Ingman that estimates 1 mutation per 20,180 years (Forster et al., 1996; Ingman et al., 2000). While this method of estimating coalescence dates from HVR-I data is commonly used, it may not represent the best method for dating, since it uses the highly variable HVR control region, which contains variable amounts of mutational hotspots and conserved regions. In addition, the coalescence times for these branches can only indicate the date of emergence of this branch in the particular population being analyzed, not the age of the entire haplogroup.

III Analysis and Interpretation of the Data

III.1 Sample Background

To understand exactly where these individual samples came from and the demographic history for the group of participants, I asked all participants to fill out a genealogical form (Appendix II) which asked for information about their birthplace, their ethnicity, their parents' names and ethnicities and their grandparents' names and ethnicities. In addition, it asked for the respective religious community affiliation (*mutt* affiliation) of their mother's and father's families. The purpose of this information was two-fold. Firstly, it was a method to ensure that none of the participants were directly related (brothers or first cousins). Secondly, this information allowed me to test some ideas that I hypothesized about the community demography through several generations, namely that (1) the trend towards urban areas is a recent phenomenon, and (2) the *mutt* distinctions and affiliations did not form a barrier to genetic admixture between the communities. I will touch on the first hypothesis in this section, and consider the genetic evidence for the second hypothesis a little bit later. Unfortunately, this demographic data are slightly incomplete. Many participants told me that they either could not remember or did not know their parents' birthplaces or grandparents' names and/or birthplaces. **Table 2** shows the statistics of the responses that I received on these genealogical forms.

Table 2: Response to Birth Location on Genealogy Form

	Expected	Actual	Yield
Participants' birth locations	116	96	83%
Parents' birth locations	232	174	75%
Grandparents' birth locations	464	196	42%

As seen in **Table 2**, the participants' knowledge of their family's birth locations decreased with each preceding generation. All of the data analysis that considered these birth locations took into account the discrepancies in the numbers of reported locations for each group, and therefore, compared the percentages rather than the actual numbers.

Using these data, I located each birthplace on a map and categorized it as either a village, town (> 35,000 inhabitants), or a city (> 1 million inhabitants), as defined by the guidelines of the Indian 2001 Census (Census, 1991). **Figure 2** shows a map with the distribution of the birth locations for three generations of Saraswat Brahmins across the South India. It is important to note that the birthplaces were distributed predominantly down the southwest coast of India, with very few falling outside of that region. Interestingly, the few that are out of that region tend to be in urban areas and found in either the participant or parental generation – in Mumbai to the north, Bangalore in the center of the map, Chennai on the eastern coast, and the few international locations. These findings correlate with the fact that the trend towards urban areas seems to be a more recent phenomenon, as seen in **Table 3** below.

Table 3: Distribution of Birth Locations

	Village	Town	City
Participants	20%	29%	51%
Parents of Participants	45%	30%	25%
Grandparents of Participants	46.50%	30%	23.50%

These results show that almost 25% of Saraswats seemed to have migrated towards the city in the last generation. Therefore, the current image of Saraswats as being city-dwelling successful businessmen is probably a recent creation, rather than an image perpetuated for generations. Additionally, if I were to correlate genetic data with

participant birth location, the resulting distribution would not be indicative of the initial spread of Saraswats out of Goa. Finally, all of the participants now live in North America, meaning a correlation to their current residence location would be even more uninformative.

III.2 Saraswat Brahmin Population Results

Of the 116 samples received from participants, 110 were successfully extracted and amplified to provide sufficient data sets with both HVR-I and PCR-RFLP analyses. The samples were divided into 43 different haplotypes, which were then placed into various haplogroups in the M, N and R macrohaplogroup families. The distribution of haplotypes is shown in **Figure 3**. The yellow circles at the middle or ends of each branch indicate a particular haplotype, and the numbers along the branches are mitochondrial mutational changes found in each haplotype. By haplotype numbers alone, these results indicate low maternal diversity, with a perpetuation of similar haplotypes and haplogroups through the community. Comparisons with other India populations (see **Table 7**) indicate that these particular haplotypes occur with the highest frequencies in northern Indo-European speaking populations, supporting historical and linguistic data showing Saraswat Brahmin origins in North India. The most important thing to note about this particular phylogenetic network is the very long branches connecting haplotypes with very diverse SNPs. These long branches could indicate the presence of old haplotypes which have undergone little recent diversification (Schurr and Wallace, 2002), and would support theories of Saraswat Brahmin cultural isolation by showing

evidence that they have been carrying the same haplotypes for generations with little outside maternal genetic influence.

There are a few particular haplotypes and haplogroup clusters that deserve notice on this network. First, the diversity of M (both subhaplogroup and undifferentiated) types is extremely large, as is common with Indian populations (Kivisild et al., 2003; Metspalu et al., 2004). There are many common Indian M subgroups represented within Saraswats, as well as several undifferentiated types that could not be associated to any other population found in published data sets. Second, the haplogroup J branch is quite unique. Haplogroup J is a common West Eurasian lineage that typically has much shorter sequence motifs than the six mutational changes in the HVR-I sequences seen in these J samples. Phylogeographically, this case is also interesting because the only other Indian populations with similar (but not exact) J haplotypes are the Kashmiri caste population and the Gujurati caste population – both living along the hypothesized migration route of Saraswat Brahmins. Another intriguing haplotype is an undifferentiated U haplotype with HVR-I mutations 16172 and 16311. There is a possibility that this haplotype could fall into the U6 haplogroup, which is predominantly found in Northern Africa (Maca-Meyer et al., 2003). The samples within these J and putative U6 haplotypes will be further analyzed to determine what can be elucidated from extended mitochondrial sequencing and coding region analysis.

Haplogroup Diversity

Haplogroup diversity is used to ascertain the extent of sequence variability of mtDNAs in a population. **Figure 4a** shows the overall diversity in Saraswats as represented by the three main macrohaplogroup clusters: M, N, and R. As expected for

Indian populations, these percentages show a predominance of M types and lesser frequencies of generally West Eurasian R types and N types. The percentage of M types found in Saraswats is about 71%, which is slightly higher than the estimated national average of about 60% (Kivisild et al., 2003), and actually much closer to the expected M percentage in tribal populations (Metspalu et al., 2004). Haplogroup M encompasses 78 Saraswat Brahmin samples, which are further examined in **Figure 4b**. This figure shows the percentages of various M subhaplogroups in the Saraswat sample set. As seen here, undifferentiated M haplotypes comprise the largest group of this pie, represented as 41% of the samples. This is expected since India is known to have the largest number of undifferentiated M lineages (Kivisild et al., 2003). The high percentage (27%) of M4 types wa unusual, since M4 – though an autochthonous type – is thought to be one of the youngest lineages introduced to India (Metspalu et al., 2004). These haplotypes will require additional SNP typing to ascertain a more complete categorization of these samples as part of the M4 haplogroup.

Figure 4a shows that about 23% of the samples belonged to macrohaplogroup R. Most of these types are Indian specific haplotypes. The J haplotypes discussed above belong to this macrohaplogroup as well. The major R lineage that appears in Saraswat Brahmin samples was haplogroup U. Thirteen out of the 25 R samples belonged to haplogroup U in various subhaplogroups as shown in **Figure 4c**. U2, a haplogroup found throughout South Asian and very rarely elsewhere, was the predominant subhaplotype in the Saraswat samples.

Saraswat Brahmin Statistical Analysis

Statistical analysis of the Saraswat Brahmin haplotypes was performed using the software program Arlequin 2.0 to yield the population specific results found in **Table 4**.

Table 4: Saraswat Brahmin Statistical Summary

Gene diversity	0.958*
Tajima's D	-1.797*
Fu's F	-25.081*

*statistically significant (< 0.05)

Nei's gene diversity test is a measure of the probability of obtaining two different alleles from a population when selecting alleles at random (Jobling et al., 2004; Nei, 1972). Ranging from 0 to 1, higher gene diversity estimates indicate an increased frequency of mutational changes and therefore different haplotypes. The Saraswat Brahmin gene diversity estimate of 95.8% falls within a standard range for human populations, and will be later compared to the gene diversities of other Indian populations later. Tajima's D is a test comparing the statistical parameters accounting for the number of segregating sites (differentiating mutational sites) and nucleotide diversity to estimate the genetic neutrality of the population. Positive values indicate a population subdivision or balancing selection, while negative values indicate population growth or positive selection (Jobling et al., 2004). The negative Tajima's D value found in Saraswat Brahmin populations most probably indicates a population expansion, indicating that the Saraswats split off from some larger group rather than the other way around. Fu's F is an additional neutrality test that analyzes the haplotype frequency distribution and usually falls near the value of -25.00.

Another common statistical method used to look at population expansion and neutrality is the analysis of pairwise mismatch distributions. This distributional

histogram approximates the diversity present within a sample set by looking at the number of distinct differences between all pairs of alleles/haplotypes in a population (Jobling et al., 2004). The curve of the distribution is often used as a sign of population history, with smooth, bell-shaped curves indicating a rapid population expansion from a solitary haplotype, and ragged, multi-peaked curves indicating a population with either no great expansions or a population bottleneck. (Jobling et al., 2004).

Figure 5a shows the overall pairwise mistmatch distribution for the Saraswat Brahmins. This figure shows a definite bell-shaped curve with an ostensibly smooth fit line. The smoothness or raggedness of the distribution is measured by Harpending's raggedness index, which is essentially a statistical measure comparing neighboring peaks in a distribution (Jobling et al., 2004). Values less than 0.03 indicate a smooth distribution – and, therefore, a previous population expansion. The age of this expansion is usually measured by the mean of the pairwise differences, with a larger number indicating an older expansion based on sequence diversity.

In order to tease out the details of macrohaplogroup expansions in the Saraswat population, I performed the same mismatch analysis sequentially on just the M and then R haplotypes in the data set (N samples were excluded since there are only seven of them). The mismatch distributions for M and R are shown in **Figures 5b and 5c** respectively. As expected, due to the large predominance of M types in the data set, most of the overall mismatch curve and all of the later (left) part of the curve came from the individual M distribution, which is very smooth. However, the ragged R mismatch distribution indicates a much more ragged shape, with an older possible expansion time, although it must be noted that the lower number of R samples presents a confounding

factor when comparing this distribution to the M distribution. Based on these curves and the mean pairwise differences, this evidence points to the possibility that the M types are somewhat younger than the R types in the Saraswat Brahmin set.

One way to test the haplogroup age estimates is by using the ρ dating statistic (Morral et al., 1994). This method uses a previously constructed phylogeny or network to examine intra-allelic diversity by analyzing the average number of mutational changes between the specified ancestral node and the descendant nodes (Jobling et al., 2004). The estimate is then multiplied by a default mitochondrial HVR-I mutational rate of 1 mutation per 20,180 years to produce an estimated age of coalescence for all of the specified nodes in the branch (Forster et al., 1996; Ingman et al., 2000). Likewise, the σ statistic, or standard deviation of the ρ calculation, gives the range of error for this coalescence date (Saillard et al., 2000). These calculations were carried out with the Network program using the Bandelt median-joining network as shown in **Figure 3**. **Table 5** shows the estimated ranges of divergence for the major branches found in the Saraswat Brahmin data set.

Table 5: Branch-specific Age Estimates

	ρ	Σ	Age	±
M	3.04	0.85	61,316	17,171
M2	1.41	0.48	28,435	9,750
M3	0.33	0.29	6,726	5,945
M4	3.10	1.51	62,461	30,735
M5	1.08	0.46	21,732	9,183
M25	0.16	0.12	3,186	2,374
N	4.71	1.65	95,134	33,246
R	7.64	1.75	154,175	35,360
R5	6.33	2.08	127,806	42,008
U	5.00	1.39	100,900	28,070
U1	6.00	2.45	121,080	49,430
U2	4.75	1.44	95,855	28,981
U7	2.00	1.41	40,360	28,538
J	6.00	2.24	121,080	45,123

The estimates in **Table 5** support the previous conclusion that the R and N branches in this data set possibly diverged earlier than the M types. Especially interesting among these dating estimates are those for the M4 and J lineages, which are much older than expected from previous calculations with other data sets (Metspalu et al., 2004). However, there are some caveats with these age estimates. Firstly, these calculations only take HVR-I data into account, which is problematic due to the exclusion of prominent diagnostic PCR-RFLP markers found in the coding region, which is known to mutate at slower rates than the control region HVR-I. Based on the phylogenetic network data, the coalescence calculations also take into account ubiquitous HVR-I mutations that occur in different mtDNA backgrounds (recurrent). Additionally, as shown above, standard deviation ranges for each of these age estimates are extremely broad, and frequently overlap. More accurate age estimates might obtained by acquiring more SNP changes on a whole mt genome scale and use additional dating methods that take both non-coding and coding region mutational rates into account.

Religious Mutt *Differentiation*

As mentioned earlier, one of the primary reasons that I asked participants for their religious community or *mutt* affiliation was to compare the genetic differences between each of the four mutts. I asked for both their mother's and father's affiliation, and, in many cases, received information that they were of a mixed marriage between two different *mutts*. While recruiting participants for the study, I tried to acquire an equal number of samples per *mutt*. Out of the 110 samples analyzed, 74 (67%) belonged to Kashi *Mutt,* which is the largest one (Conlon, 1977). The Chitrapur Saraswat *Mutt* had

18 samples; Gokarn Partagali *Mutt* had 16 samples; and Kavle *Mutt* had 2 samples. This largely skewed sample distribution means that all results of *mutt* comparisons need to be taken with a grain of salt. If this study is to be expanded in the future, then I would be sure to design a sampling strategy that aims to even the sample ratios.

To examine the maternal genetic differentiation between *mutts*, I divided each haplotype by the number of samples that it contained from each *mutt*. **Figure 6** shows the spread of the four *mutts* across the Saraswat Brahmin haplotype network. Yellow indicates Kashi *Mutt*; red indicates Chitrapur Saraswat *Mutt*; blue indicates Gokarn Parthagali *Mutt*; and green indicates Kavle *Mutt*. The wide spread of colors on this network indicates low phylogenetic differentiation between the religious communities. Some haplotypes do seem limited to particular *mutts*, although the same branch usually contains samples from other *mutts*.

To further investigate the possibility of these *mutts* underlying a genetic structure, I conducted an Analysis of Molecular Variance (AMOVA), which examines the allelic molecular relationships between varied tiers of population structures (Jobling et al., 2004). To conduct this analysis, it is necessary to group the populations into a hierarchical structure. I chose Tier I as Kavle *Mutt*, Kashi *Mutt*, and Gokarn Partagali *Mutt* – the three older communities. I placed the Chitrapur Saraswat *Mutt* samples into a separate tier, since they split off later due to religious philosophical differences and desire to maintain a separate cultural distinction, although they still frequently intermingle and intermarry with the greater Gaud Saraswat Brahmin caste-cluster (Keni, 1998). **Table 6** shows the results of this AMOVA analysis.

Table 6: *Mutt* Structure AMOVA estimates

among populations	1.01%*
among populations within group	0.61%*
within populations	98.38%*

*not significant (> 0.05)

As expected the from close cultural and marital ties between the four populations, the AMOVA analysis indicated close genetic ties as well. Only a total of 1.52% of the genetic variation could be attributed to any differences between the four populations, with the remaining 98.38% of variation coming from within the sample sets themselves. These results indicate that, based on my unequal sampling of the four *mutt* communities, they are genetically very similar, and, therefore, can be grouped together as one population for the remaining analysis.

III.3 Saraswats in a broader context: Comparisons with other Indian Populations

To obtain a better understanding of the broader spectrum of Indian diversity, I compared the Saraswat Brahmin mtDNA data with those of other Indian populations. In this analysis, I chose 23 other populations – 17 caste populations and 6 tribal groups. Their names, states of residence, language group, sample size, and reference study are shown in **Table 7**. In addition, **Table 7** shows the three letter code by which each individual population will be referenced in all of the proceeding figures and tables. I tried to pick a range of populations that were from areas through which the Saraswat Brahmins passed on their migration to Goa. The populations from Karnataka, Kerala, and Tamil Nadu are the South Indian Dravidic speakers. Two non-Hindu populations were also incorporated into the caste hierarchy – the Cochin Jews and the Maharashtrian Zoroastrian Parsis. It is important to note that since there seemed to be a lack of

Table 7: Index of Comparative Populations

Population Name	Code	State	Grouping	Sample Size	Reference
Saraswat Brahmins	SB	Goa/Karnataka/Kerala/Maharashtra	Caste - Brahmin	110	this study
Bengali Brahmins	WBB	West Bengal	Caste - Brahmin	40	Metspalu et. al. 2004
Bengali Sudras	WBS	West Bengal	Caste - Sudra	11	Metspalu et. al. 2004
Gujurati Caste	GGU	Gujurat	Caste	58	Metspalu et. al. 2004/Kivisild et. al. 1999
Kashmiri Caste	KKA	Kashmir	Caste	19	Kivisild et. al. 1999
Havik Brahmins	KBH	Karnataka	Caste - Brahmin	47	Mountain et. al. 1995
Cochin Caste	KCC	Kerala	Caste	55	Metspalu et. al. 2004
Cochin Jews	KCJ	Kerala	Caste - Non Hindu	45	Metspalu et. al. 2004
Chitpavan Brahmins	MCB	Maharashtra	Caste - Brahmin	19	Gaikwad and Kashyap 2005
Desasth Brahmins	MDB	Maharashtra	Caste - Brahmin	19	Gaikwad and Kashyap 2005
Dhangars	MDH	Maharashtra	Caste - Sudra	19	Gaikwad and Kashyap 2005
Konkanastha Brahmins	MKB	Maharashtra	Caste - Brahmin	58	Metspalu et. al. 2004
Marathas	MMA	Maharashtra	Caste - Middle	19	Gaikwad and Kashyap 2005
Parsis	MPA	Maharashtra	Caste - Non Hindu	55	Metspalu et. al. 2004
Punjabi Brahmins	PBP	Punjab	Caste - Brahmin	26	Metspalu et. al. 2004
Punjabi Kshatriyas	PKP	Punjab	Caste - Kshatriya	34	Metspalu et. al. 2004
Rajputs	RKR	Rajasthan	Caste - Kshatriya	36	Metspalu et. al. 2004
Tamil Caste	TNA	Tamil Nadu	Caste	51	Cordaux et. al. 2003
Great Andamanese	AGA	Andaman	Tribe	20	Thangaraj et. al. 2003
Onge	AON	Andaman	Tribe	62	Thangaraj et. al. 2003
Kurumban Mullu	KKM	Karnataka	Tribe	44	Cordaux et. al. 2003
Koragas	KKC	Kerala	Tribe	53	Cordaux et. al. 2003
Kuruchian	KKL	Karnataka	Tribe	46	Cordaux et. al. 2003
Yoruva	KYE	Kerala	Tribe	53	Cordaux et. al. 2003

published data on South Indian caste populations, I included more Dravidic tribal groups. The Andamanese were included as geographic outliers, since they are isolated tribal populations living on the Indian-controlled Andaman Islands. The sample size variation through this group of comparison populations was also quite large. While there are 999 total samples represented overall, they are not evenly distributed through the populations.

Figure 7 indicates the geographic distribution of these comparison populations distributed by political state boundaries. Each pie chart is proportional in size to the number of samples analyzed from that particular state. Pie charts outlined in black indicate caste populations, while those outlined in red indicate tribal populations. Each individual pie chart indicates the percentages of the three macrohaplogroups (M, N, and R) that are found in the state's populations. These relationships will be examined in-depth in the next section.

Haplogroup Variance

The distribution of mitochondrial haplogroup variation over wide regions has been of primary interest to researchers studying the migrations and population histories of different communities. The Indian haplogroup distribution is especially intriguing due to the large amount of inherent diversity in culture, religion, and phenotypic characteristics. **Figure 8a** shows the percentages of macrohaplogroups across all the comparative populations, which are arranged in a north to south orientation from left to right. The six tribal populations to the extreme right are separated from the caste populations by their noticeably lower haplogroup diversity. The caste and tribal populations can be clearly differentiated by the lack of macrohaplogroup N mtDNAs in tribal populations and the

rather high percentages of M haplotypes. Although not particularly significant, it seems as though there is a north to south trend in increasing frequencies of M mtDNAs, perhaps pointing to the ancestral Indian migrations of these M haplogroups through South India before Indo-Aryan groups arrived in North India. Interestingly, the Saraswat Brahmins have a similar M frequency to that of the Rajasthani Rajputs (RKR) and the Konkanastha Brahmins (MKB) from the western coast of India.

Figure 8b shows the distribution of varying M subhaplogroups, again arranged from north to south, with the tribal groups being separated on the extreme right. This figure shows the same trend of greater diversity in caste populations as compared to tribal populations. The large percentages of undifferentiated M haplotypes is astounding, and is the primary evidence used by other scholars to point to the ancestral origins of many basal M types on the Indian subcontinent (Bamshad et al., 2001; Kivisild et al., 1999a; Metspalu et al., 2004). Also prominent is the distribution of M2 (an ancestral Indian subhaplogroup believed to have arrived with the original inhabitants of subcontinent) with these populations (Metspalu et al., 2004). The high occurrence of this subhaplogroup is in the tribal populations. Its frequency decreases through the southern caste populations, and it is nearly absent in the extreme northern populations (Kashmiris, Punjabis, Rajasthanis, and Bengalis). Both Saraswat Brahmins (SB) and the Konkanastha Brahmins (MKB) have low percentages of this subhaplogroup (6.4% and 2.5%), possibly indicating a low degree of early tribal or Dravidic admixture.

The second most prominent haplogroup was U from the macrohaplogroup R family. U haplotypes are generally considered to be West Eurasian rather than East Eurasian in origin, but both U2 and U7 have a significant presence in India and are

thought of as Indian-specific types (Palanichamy et al., 2004). **Figure 8c** shows the percentages of U subhaplogroups in the comparative populations, being arranged north to south from left to right with the tribal populations at the extreme right. As seen before, there is a difference between most of the tribal groups and the caste populations. Half of the six tribal groups lacked any U types at all (including the two Andamanese populations), while a third had primarily U1 haplotypes rather than the more common U2 haplotype seen in all of the caste populations. The Saraswat Brahmins also had 3 U haplotypes that are undefined (as mentioned earlier, they could be U6 types), and need to be further analyzed to place them into a subhaplogroup.

Overall, the haplogroup diversity in all the comparative populations indicated a strong caste versus tribal distinction. Many of the very common Indian-specific haplotypes such as M2-M6, U1, U2, and U7 were found distributed through all geographic areas, supporting hypotheses that they are basal Indian branches (Metspalu et al., 2004). However, there are two important caveats to note about this data. First, the sample sizes for these populations are not equal, and therefore might bias the assessment of diversity. Secondly, one must not to base all estimate of genetic relationships on the comparative haplogroup variation through populations, as these correlations can often be insufficient to determine actual diversity at the individual level (Kivisild et al., 1999b).

Summary Statistical Analysis

With the data available, the easiest method to overcome the insufficiency of haplogroup comparisons between populations is to utilize statistical tools that will compare relative gene diversity, neutrality estimates, molecular variance, population structures, and genetic distances. Using the Arlequin 3.0 program, it was possible to

Table 8: Summary Statistics for Comparative Populations

CASTE POPULATIONS	KKA	PBP	PKP	RKR	WBB	WBS	GGU	MPA	MCB	MKB	MDB	MDH	MMA	SB	KBH	KCO	KCJ	TNA
gene diversity	0.953	0.991	0.991	0.989	0.991	1.000	0.995	0.953	0.988	0.985	1.000	1.000	0.994	0.958	0.984	0.987	0.878	0.996
mean pairwise differences	5.139	5.705	5.360	5.060	6.227	5.745	6.220	4.506	6.837	6.025	6.316	6.339	7.128	5.572	5.552	6.493	3.721	6.155
Harpendings raggedness	0.087	0.013	0.028	0.017	0.011	0.061	0.011	0.017	0.024	0.010	0.017	0.030	0.008	0.007	0.012	0.015	0.158	0.011
Tajima's D	-1.658	-1.825	-1.522	-1.978	-1.822	-1.072	-1.936	1.847	-1.459	-2.051	-1.550	-1.908	-1.808	-1.797	-1.617	-1.550	-0.605	-1.906
Fu's Fs	-5.357	-17.278	-24.705	-25.322	-25.214	-6.366	-25.218	-17.208	-9.032	-25.260	-15.440	-15.407	-11.062	-25.081	-23.911	-25.158	-2.383	-25.237

TRIBE POPULATIONS	AGA	AON	KYE	KKO	KKM	KKU
gene diversity	0.447	0.713	0.775	0.746	0.621	0.787
mean pairwise differences	2.153	2.545	2.642	3.003	1.783	5.200
Harpendings raggedness	0.337	0.101	0.079	0.238	0.101	0.099
Tajima's D	-1.084	2.422	-0.231	0.688	0.766	-0.776
Fu's F	-0.083	4.906	-2.120	1.979	-0.231	1.110

NOTE: red indicates significace (< 0.05)

NOTE: mean pairwise differences have no valid significance estimate

analyze all of the 23 comparative populations and the Saraswat Brahmins obtain a better understanding of overall diversity through India (Schneider et al., 2000).

Table 8 presents the summary statistics for all of the 24 populations analyzed, with gene diversity, mean pairwise differences, raggedness indices, Tajima's D and Fu's F neutrality values. While comparing values in this table, it is important to note that the WBS, KKA, MCB, MDB, MDH, and MMA represent data sets with less than 20 samples each. This certainly affected the statistical values for many of the indices, as the haplotypic variation was not sufficient to be indicative of the whole population. The gene diversity values again presented an interesting picture of low diversity within tribal populations when compared to caste populations. Also interesting were the few positive Tajima's D indices. Although only the MPA value was significant, these positive Tajima's D trends indicate some degree of population subdivision, or evidence of genetically significant divisions in the population for which these values were obtained.

Population Structure

While looking at the overall diversity of Indian populations as a whole can be very informative and intriguing, it is also necessary to explore various elements of sub-structure that might exist to detect patterns within the fabric of the genetic quilt evident throughout India. **Table 9** shows the results of three different AMOVA calculations that examined diversity based on varying elements that might subdivide the Indian population: (1) geographical boundaries, (2) tribal vs. caste elements, and (3) linguistic affiliation(s).

Table 9: Indian Population Structure AMOVA Estimates

	Geographic	Tribal	Linguistic
among populations	4.30%	2.38%*	5.10%
among populations within group	9.51%	10.84%	8.64%
within population	86.19%	86.78%	86.26%

* not significant (> 0.05)

The first analysis involved a simple geographic division in which populations were separated into northern or southern residences throughout India. Group I (North India) included the Kashmiri, Punjabi, Rajasthani, Bengali, Gujurati, and Maharashtrian populations, Group II (South India) included the Karnataka, Kerala, and Tamil populations with the Saraswat Brahmins, and Group III (islands) comprised of the 2 Andamanese populations. The AMOVA analysis indicates that 4.3% of the variation was due to differences between these geographic groups, while 9.51% of the variation can be explained by differences between populations within the groups.

The second analysis examined tribal versus caste distinctions within the comparison populations. Since this is based on an aspect of religious subdivision in Indian society, I placed populations into three groups. Group I (Hindu caste populations) included the Bengali, Gujurati, Kashmiri, Punjabi, Rajasthani, Tamil, Maharashtrian caste, Kerala caste, and Karnataka caste populations with the Saraswat Brahmins, Group II (tribal populations) included the Andamanese, Karnataka tribal, and Kerala tribal populations, and Group III (non-Hindu upper-class populations) included the Cochin Jews and the Parsis. The historical and legendary origins of both the Cochin Jews and Parsis are thought to be based in the Middle East and Persia, respectively, although both were thriving successful communities in India at one time. The AMOVA analysis

indicates that only 2.38% of variation resulting from separations between the three groups, while 10.84% of variation was due to population differences within the groups.

The last AMOVA analysis involved the linguistic separations determined by the major language families found in India. Group I (Indo-European speakers) included the Kashmiri, Punjabi, Rajasthani, Bengali, Gujurati, and Maharashtrian populations with the Saraswat Brahmins, Group II (caste Dravidic speakers) included the Karnataka, Kerala, and Tamil caste populations (including the Cochin Jews), Group III (tribal proto-Dravidic speakers) included the Karnataka and Kerala tribal populations, and Group IV (Andamanese speakers) included the two Andamanese populations. The AMOVA analysis showed that 5.10% of the variation was explained by linguistic differences, while 8.64% of the variation was due to differences between the populations within the group.

The common trend in these three AMOVA analyses was that 86-87% of the genetic variation of this full data set can be explained by diversity within distinct populations, while about 13-14% can be explained by geographic, cultural, or linguistic substructure inherent in the gene pool. Overall, these findings indicate population structure must be taken into account when looking at Indian genetic diversity.

Genetic Distances

The Arlequin 3.0 program also has algorithms that directly measure genetic distances between populations. One set of values, called F_{ST}s, describe the variation between subdivisions of a total data set. Thus, in this case, the total data set would be the 999 individual results with the subdivisions being their ethnic community affiliations.

This method examines the variation in individual mutational positions between each subdivision sequentially. Large F_{ST} values indicate high amounts of continued admixture between the two populations, while small F_{ST} values indicate highly distinct populations (Jobling et al., 2004). **Table 10** shows the F_{ST} matrix for all comparative populations used for this study. As seen, the tribal populations have significantly large F_{ST} values when compared to other populations, again indicating their noticeable separation from caste groups throughout India.

This matrix of F_{ST} values was used to generate graphical representations of the genetic distances between populations. The first such representation is a Multidimensional Scaling (MDS) plot of the F_{ST} values between all of the populations, as seen in **Figure 9a**. MDS plots use a multivariate analysis to reduce a data set with greater than three dimensions down to a comprehensible two- or three-dimensional plot, thereby minimizing the possibilities of information loss (Jobling et al., 2004). In **Figure 9a**, all of the caste populations were clustered together in the middle of the plot, with red squares indicating their placement, while the tribal populations widely surrounded this cluster, with blue squares indicating their placement. In addition, geography seems have played a factor in the genetic distances between tribal populations, with the Andaman Islands populations clustering together in the upper right quadrant, the Karnataka tribe populations together in the lower right quadrant, and the Kerala tribe populations together in the lower left quadrant.

To get a better estimate of the placement of Saraswat Brahmins based on genetic distance, I increased the resolution on the middle caste cluster by removing the

Table 9: F_{ST} Genetic Distance Matrix for Comparative Populations

	SB	AGA	AON	WBB	WBS	GGU	KKA	KBH	KKM	KKO	KKU	KYE	KCO	KCJ	MCB	MDB	MDH	MKB	MMA	MPA	PBP	PKP	RKR	TNA
SB	0.000	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
AGA	0.082	0.000	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
AON	0.239	0.304	0.000	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
WBB	0.046	0.150	0.288	0.000	+	+	-	-	+	+	+	+	-	+	-	+	-	-	+	+	-	-	-	-
WBS	0.062	0.247	0.392	0.048	0.000	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
GGU	0.041	0.146	0.286	0.015	0.047	0.000	-	+	+	+	+	+	+	+	+	+	-	+	+	+	-	-	+	+
KKA	0.068	0.239	0.387	0.012	0.060	0.003	0.000	+	+	+	+	+	-	+	+	+	-	+	+	+	-	-	+	+
KBH	0.032	0.143	0.286	0.016	0.057	0.012	0.032	0.000	+	+	+	+	+	+	+	+	+	+	+	+	+	-	-	+
KKM	0.191	0.443	0.414	0.204	0.351	0.187	0.327	0.197	0.000	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
KKO	0.240	0.435	0.521	0.247	0.312	0.219	0.279	0.238	0.474	0.000	+	+	+	+	+	+	+	+	+	+	+	+	+	+
KKU	0.157	0.288	0.359	0.193	0.194	0.180	0.236	0.198	0.304	0.364	0.000	+	+	+	+	+	+	+	+	+	+	+	+	+
KYE	0.081	0.300	0.372	0.101	0.219	0.094	0.205	0.089	0.297	0.365	0.253	0.000	+	+	+	+	+	+	+	+	+	+	+	+
KCO	0.046	0.134	0.272	0.012	0.040	0.012	0.007	0.024	0.169	0.231	0.179	0.112	0.000	+	+	+	-	+	+	+	-	+	+	+
KCJ	0.059	0.205	0.360	0.037	0.098	0.036	0.065	0.040	0.256	0.290	0.239	0.161	0.041	0.000	+	+	+	+	+	+	+	+	+	+
MCB	0.057	0.151	0.319	0.008	0.056	0.031	0.025	0.027	0.253	0.297	0.210	0.165	0.019	0.056	0.000	+	-	+	-	+	-	+	-	+
MDB	0.056	0.176	0.345	0.055	0.085	0.019	0.058	0.042	0.286	0.300	0.214	0.160	0.049	0.096	0.063	0.000	+	+	+	+	+	+	+	+
MDH	0.033	0.125	0.322	0.012	0.037	0.012	0.012	0.021	0.295	0.268	0.200	0.180	0.014	0.044	0.022	0.028	0.000	+	-	+	-	-	-	+
MKB	0.040	0.120	0.257	0.011	0.058	0.030	0.045	0.027	0.149	0.256	0.182	0.092	0.024	0.047	0.017	0.056	0.037	0.000	-	+	+	+	+	+
MMA	0.030	0.113	0.277	0.018	0.033	0.023	0.037	0.019	0.203	0.259	0.187	0.101	0.017	0.062	0.023	0.032	0.010	0.013	0.000	+	-	+	-	-
MPA	0.029	0.114	0.296	0.025	0.081	0.028	0.057	0.030	0.252	0.272	0.224	0.073	0.039	0.068	0.054	0.064	0.038	0.035	0.024	0.000	+	+	-	+
PBP	0.029	0.126	0.327	0.015	0.040	0.007	0.001	0.023	0.277	0.251	0.207	0.160	0.013	0.049	0.019	0.032	-0.011	0.029	0.007	0.024	0.000	-	-	-
PKP	0.032	0.159	0.314	0.009	0.047	-0.001	0.009	0.013	0.244	0.254	0.198	0.107	0.018	0.035	0.031	0.029	0.010	0.023	0.021	0.020	0.003	0.000	-	-
RKR	0.022	0.125	0.298	0.009	0.058	0.011	0.033	0.010	0.218	0.262	0.202	0.083	0.012	0.027	0.019	0.035	0.015	0.011	0.011	0.013	0.009	0.007	0.000	-
TNA	0.030	0.120	0.259	0.007	0.040	0.011	0.024	0.019	0.168	0.237	0.158	0.099	0.019	0.036	0.028	0.041	0.015	0.015	0.015	0.025	0.010	0.006	0.007	0.000

red and + indicates significance (< 0.05)

tribal populations from the analysis. This new MDS plot is shown in **Figure 9b**. Assuming that geography might play some role in the distribution of genetic distance in the caste populations, I colored the circles by geographical affinity, with red squares indicating North Indian populations, blue squares indicating populations on the upper to middle western coast of India, purple squares indicating populations on the eastern coast of India, and green squares indicating South Indian populations. The Saraswat Brahmins were grouped with the populations on the western coast due to their historical longtime settlement in Goa and current placement in both the middle and lower areas of the western coast. However, the plot did not show any geographic grouping of these populations. Instead, there was no discernable pattern based on any previous population structure evident in this plot.

There were five outliers in this plot including the Cochin Jews of Kerala in the lower left quadrant, the Saraswat Brahmins and the Maharashtrian Parsis in the upper left quadrant, and the Maharashtrian Desasth Brahmins and Bengali Sudras in the upper right quadrant. Of these, the Bengali Sudras had a sample size of 11 individuals – all with distinct haplotypes, thereby appearing as highly diverse even though the low sample size indicates that it is not accurate to make this assumption. It is also interesting that the non-Hindu, non-tribal groups represented in the larger data set were both outliers from the rest of the caste populations.

Of primary importance, is the separation of Saraswat Brahmins from the rest of the groups, with the exception of Parsis. This pattern will be further analyzed in **Figure 10** by looking at a different representation of the genetic distances between populations. Overall, this MDS plot indicates that Indian caste diversity does not fall into an expected

pattern of geographical differences. Not all of these groups clustered very closely together when compared to Indian tribal populations. Thus, the tribe versus caste differentiation is clearly a more significant factor than any internal caste population differences.

The other way to represent this matrix of F_{ST} values is by using it to generate neighbor-joining (N-J) trees. N-J trees use the principle of "minimum evolution," which was suggested as a method to use distance matrices to make phylogenetic trees that have the shortest sum of branch lengths (Cavalli-Sforza and Edwards, 1967; Jobling et al., 2004). The N-J method uses an iterative formula to find trees with the minimum total branch length that is the closest to the optimal minimum evolution tree (Jobling et al., 2004; Saitou and Nei, 1987). For the trees shown in **Figure 10**, I analyzed only the caste populations since analyses with the tribal populations did not give high resolution of the caste population relationships.

Figure 10a shows a cladogram made with the N-J method in which the branching order of all 24 populations from the most recent common ancestor (represented by the root) are outlined. It revealed that both Saraswat Brahmins and Maharashtrian Parsis (MPA) branched off and differentiated themselves much earlier than the rest of the populations represented in it. It is possible that this represents the high levels of endogamous marriage practices evident through both Saraswat Brahmin and Maharashtrian population history – especially in recent times. Another relevant relationship seen in this cladogram is the clustering of the Bengali Brahmins (WBS) and Maharashtrian Marathas (MMA), which are two lower caste groups represented in this sample set.

While the cladograms indicated the divergence from a rooted ancestor, the same N-J method and F_{ST} matrix can be used to make an unrooted N-J tree, as seen in **Figure 10b**. This tree shows shows the two-dimensional relationships of populations based on genetic distances. The important things to note about this tree are the large genetic distances of the Bengali Sudras (WBS), the Cochin Jews (KCJ), and the Maharashtrian Desasth Brahmins (MDB). The length of their branches indicates that they are relatively distinct from the other comparative populations.

IV. Discussion and Conclusions

This study was designed with the intent of examining the genetic variation in a unique Indian community, the Saraswat Brahmins, to elucidate their population history and their place within the broader spectrum of Indian genetic diversity. The mtDNA results from 110 individuals were been compared to those for 23 other Indian populations with a wide range of affiliations, geographic locations, and backgrounds. The hypothesis of this study was that the Saraswat Brahmins would show genetic signatures of a highly endogamous group, and have different affinities to other Indian groups based on possible admixture during their migratory routes and similar origins.

As the Analysis and Interpretation Section of this thesis shows, there are many interesting conclusions that can be drawn from careful inspection of the data resulting. The Saraswat Brahmin community is certainly a unique world population. Elements of its haplotype and haplogroup variation show evidence of low maternal diversity with a high predominance of basal M types. In addition, the Indian-specific West Eurasian types (N and R) found within the Saraswat Brahmin data set are more indicative of ties to Indo-European Aryan communities rather than from tribal populations, establishing the Saraswat Brahmins as a caste community. Overall, the mtDNAs observed in Saraswat Brahmins are primarily Indian-specific haplotypes that are closely related to many of those appearing in Northern Indian populations. The results – in concert with legendary, historical, linguistic, and cultural elements – point to the northern Indian origins of the Saraswat Brahmin community. Statistical measures indicate that the community experienced a population expansion some time in the past, rather than undergoing recent subdivisions that reduced diversity in the population. This observation supports the

AMOVA results on the religious sub-structure within the community, which indicated that there are no significant genetic differences between the four religious sub-castes, most likely because of the freedom of intermarriage propagated between these sub-castes. While this analysis did not incorporate a similar sample size in each set, it could indicate trends that might arise even when all of the sub-communities are equally sampled.

When compared to other Indian populations, Saraswat Brahmins exhibit values well within the ranges of general Indian diversity, although they appeared as outliers in MDS plots of F_{ST} genetic values. This pattern could indicate their highly endogamous nature with low admixture, since their early branching from other Indian populations could indicate that they experienced little admixture with these groups for some time. It is hard to ascertain the degree of admixture with other Indian populations based on maternal evidence alone, but a more complete picture using Y-chromosome data would probably fill out the preliminary indications of very low direct and recent admixture. Rather, it is probably more likely that the Saraswat Brahmin population arose from the same ancestors as many of the other populations. The same can be said for their migratory route, although it seems clear that they migrated from the north and moved into the south at later dates.

The complexity of interpreting the results from the Saraswat Brahmin data set also aligns them with other Indian populations. The mitochondrial gene pool of India seems to be an intricate web of complex population structures. The overall analysis of this study shows that there are clear distinctions between the tribal and caste populations present on the subcontinent. Results of haplogroup comparisons show that although caste populations have many Indian-specific types, they appear at a very high frequency in the

tribal populations, with no evidence of West Eurasian admixture. As with the Saraswat Brahmin data, all Indian populations have many undifferentiated M types. The diversity of macrohaplogroup M in India indicates that it has had a long time to diversify and establish itself as the major lineage throughout India. However, these data also indicate that more work needs to be done to differentiate these haplotypes so as to better understand the origins of this important macrohaplogroup before it spread to the rest of the world. The same is true with respect to the N and R types that are autochthonous to India. These types are interesting remnants of either the original out-of-Africa haplogroups or more recent Indo-Aryan arrivals to the subcontinent.

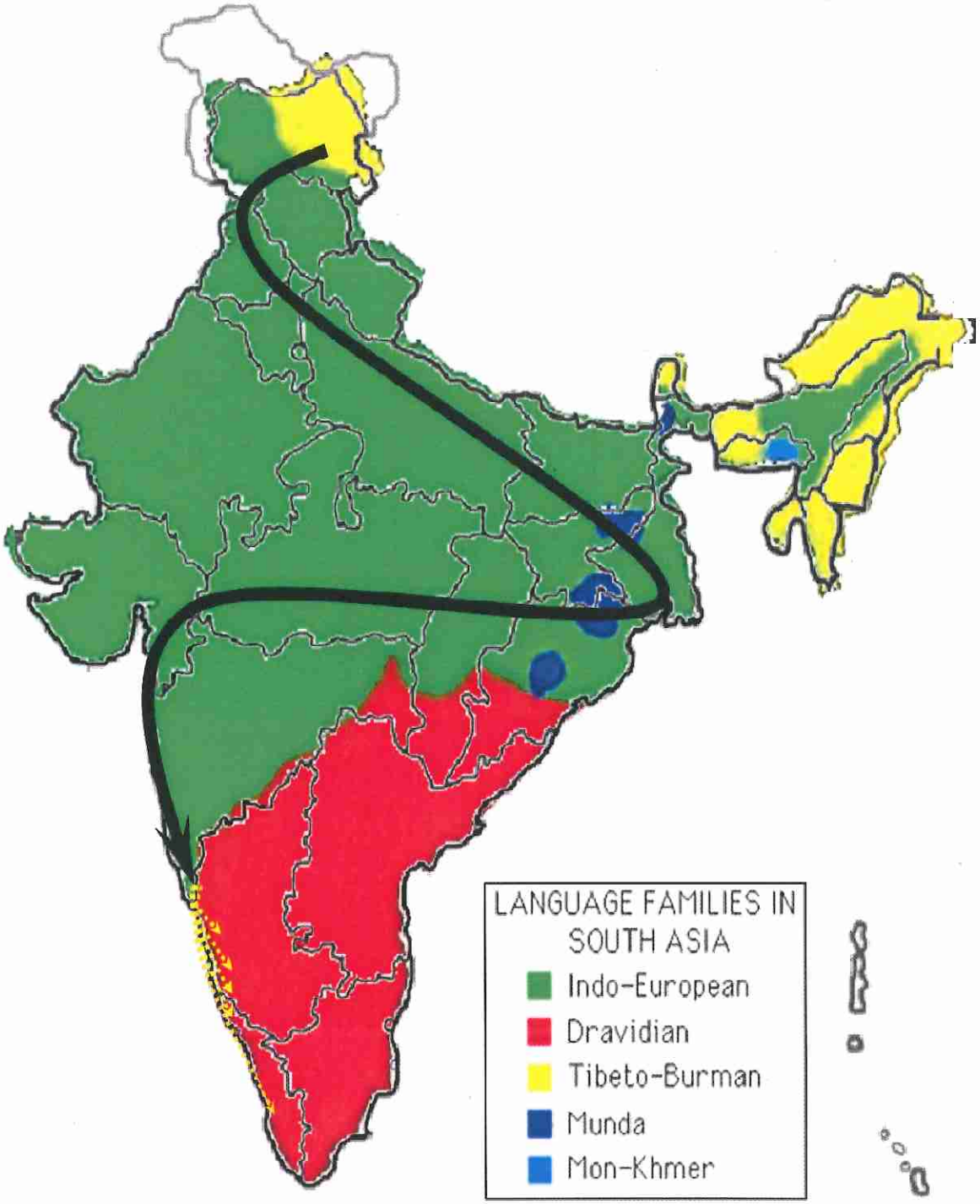
When focusing on the caste populations of India, it is clear that there are multiple factors which play a role in determining the genetic affinities, relationships, and branches of these comparison populations. From the summary statistics, AMOVA estimates, and genetic distance representations, there is no one element among geographic, linguistic, and religious affiliations that creates absolute divisions within the society. This observation is quite interesting, since many other areas of the world have principal factors that enable a classification of subpopulations. Genetics can be added to the growing list of factors such as culture, religion, and language that are too complex to describe as singular features. Instead, it is necessary to take all of them into account when putting together the population history of the subcontinent.

Although this study points towards many interesting conclusions and factors that are useful in elucidating Saraswat Brahmin and general Indian population history, it also strongly points to further research that will be necessary to obtain a more complete picture. Many specific samples in the Saraswat Brahmin data set need to be further typed

for intriguing markers. To continue the mtDNA work, I would like to focus on (1) the Indian-specific N and R types found in the sample set, (2) the further analysis of the subhaplogroup age estimates with more in-depth typing, and (3) comparisons with other highly endogamous communities. Additionally, the Y-chromosome data from the Saraswat Brahmins will help to provide a clearer picture of the migratory patterns, variable admixture rates, and possible influence from non-Indian populations because of differential male and female gene flow. The soon-to-arrive samples from the Dogra lab in India will also increase the data set, which in turn will help us to make a stronger statistical case for any conclusions raised here.

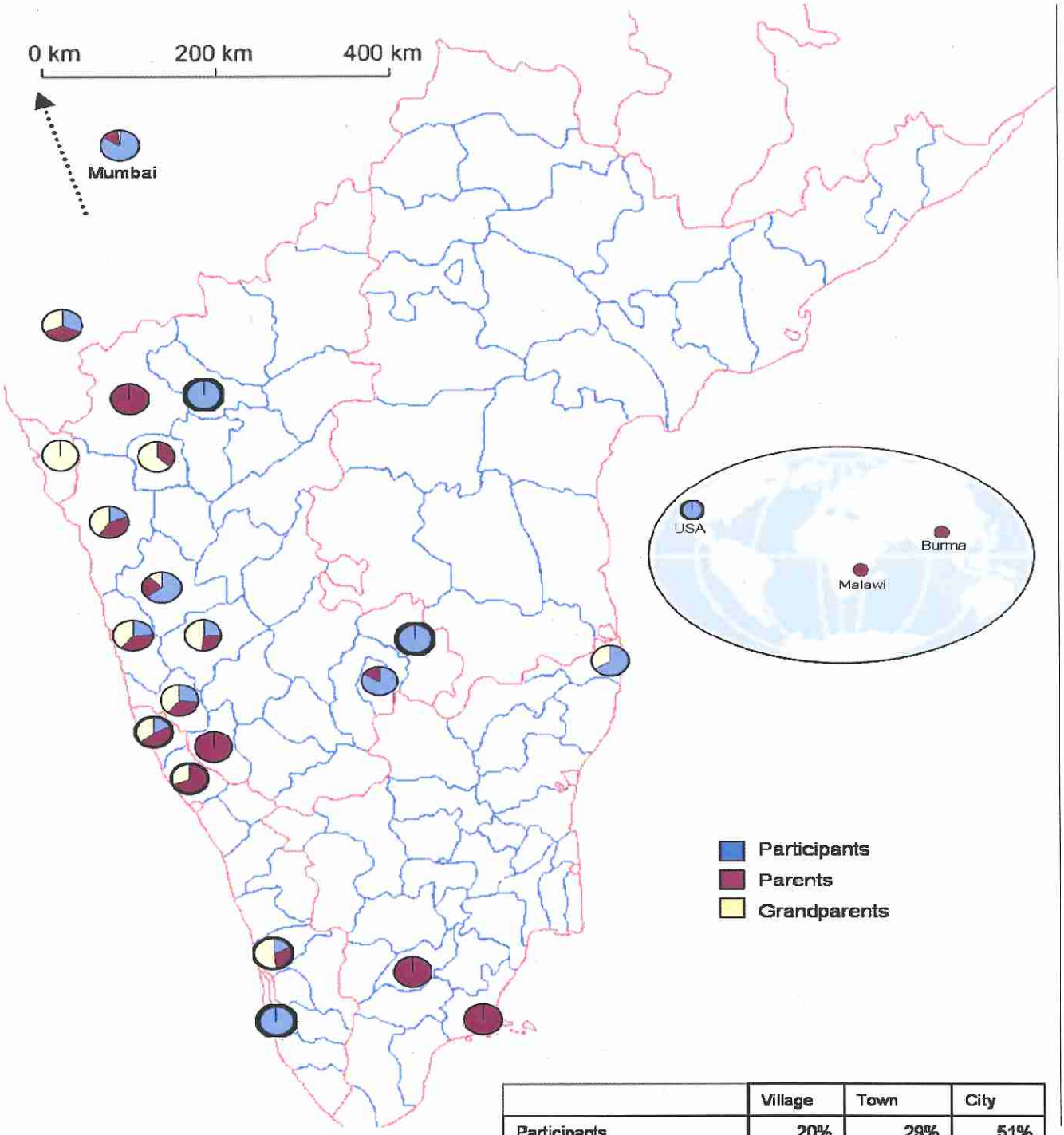
As has been stated many times before, the diversity of Indian populations represents an important area that has been underrepresented in genetics research. It is clear from this study of Saraswat Brahmins as a group and as part of the broader Indian gene pool that there are many complex patterns and trends that can be elucidated with more samples from more populations. The uniqueness of India lies not in its multi-faceted population structure, but rather in the ability for those multiple facets to mix with each other, influence each other, and still find a way to maintain distinctions while together into one nation. It is this intermingled nature of the discernable homogeneous units of Indian society that must be tapped by geneticists to help study the complexity of human genetics.

Figure 1. The Proposed Migration Route of Saraswat Brahmins



The black line indicates the legendary and scholarly assumed migratory pattern of Saraswat Brahmins from North India down to Goa. The yellow dotted lines indicate the outward migrations from Goa into surrounding areas of Saraswat Brahmin settlement down the Western 'Konkan' coast of India.

Figure 2: Generational Birth Locations of Saraswat Brahmins



	Village	Town	City
Participants	20%	29%	51%
Parents of Participants	45%	30%	25%
Grandparents of Participants	46.50%	30%	23.50%

Each pie chart indicates the percentage of each generation that was born in a particular locale. This map indicates the distribution of Saraswat Brahmins across the western coast of India after migration out of Goa.

Figure 3: Saraswat Brahmin Network

This network indicates the wide distribution of Saraswat Brahmin haplotypes. The majority of the haplotypes are Indo-European Indian-specific types, but each branch is unusually diverse with a large number of mutational positions per lineage.

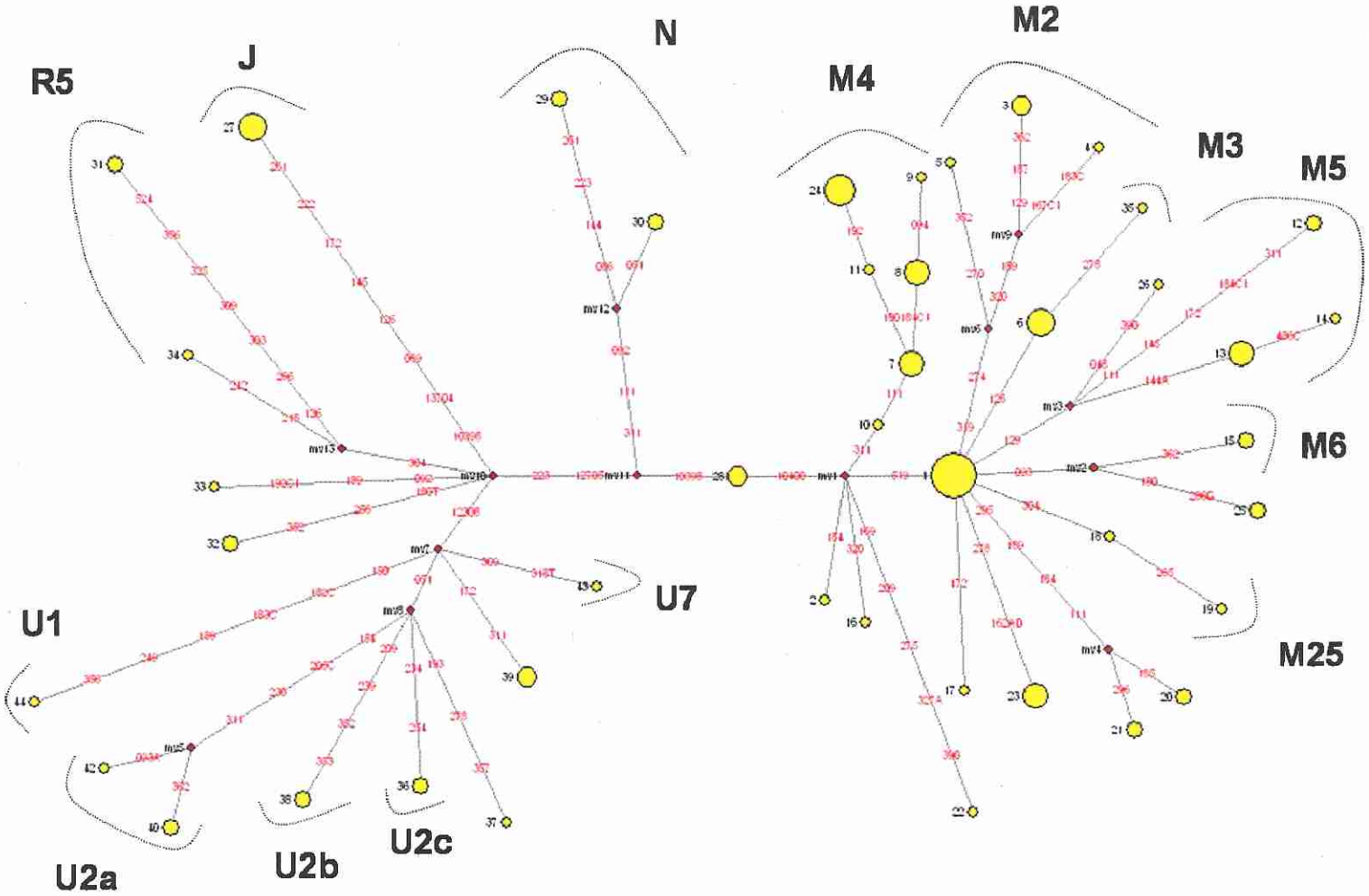
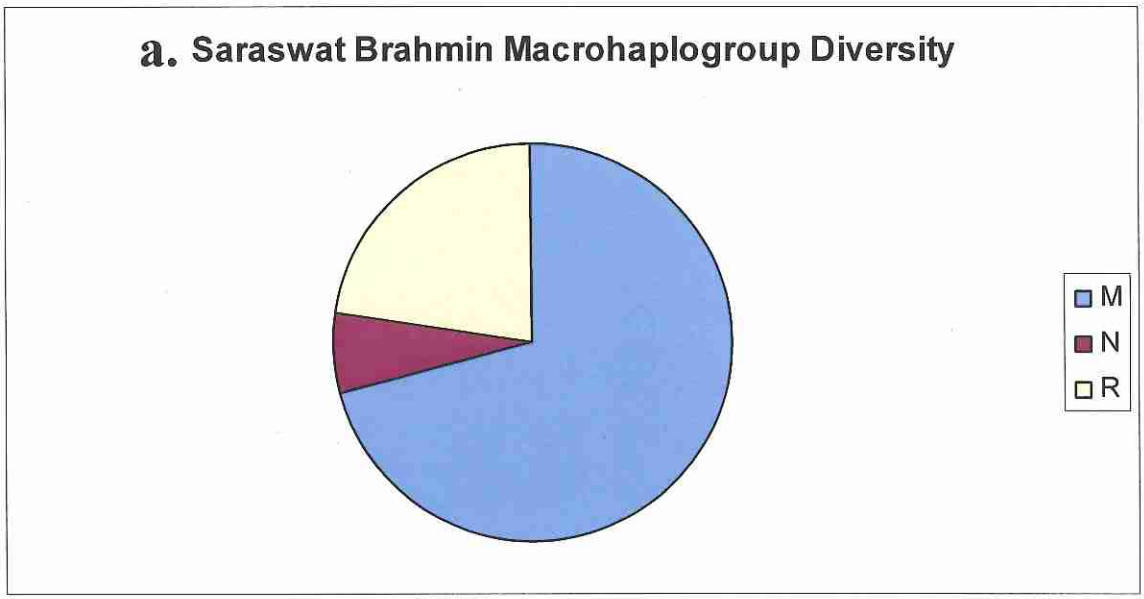
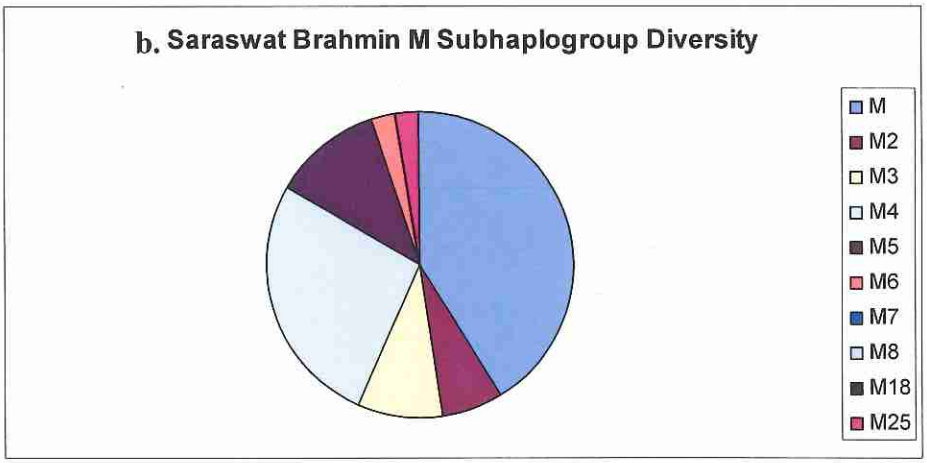


Figure 4: Percentages of Haplogroups in Saraswat Brahmins





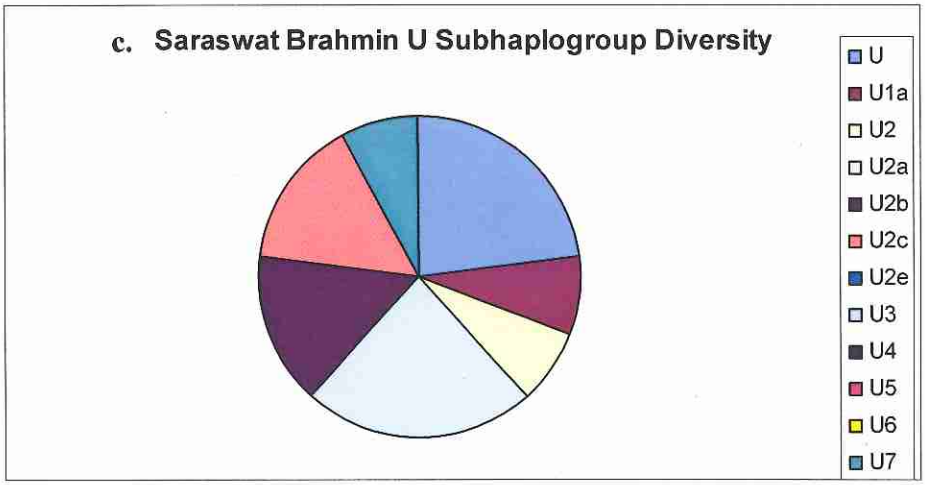
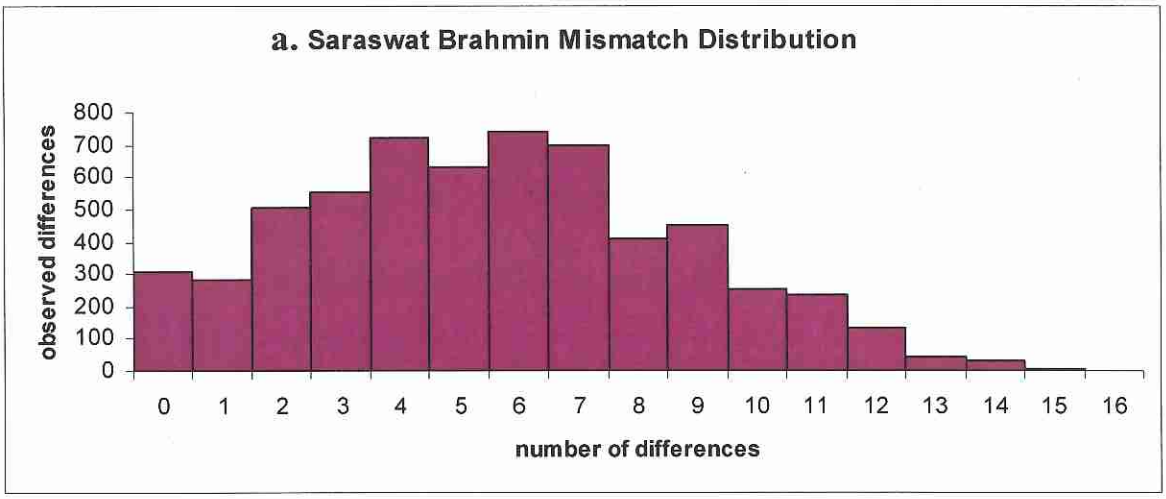
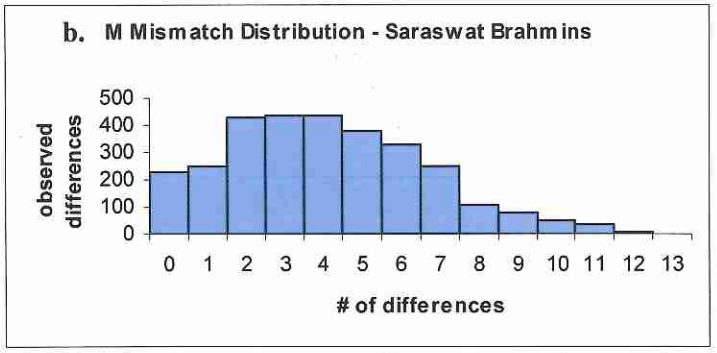


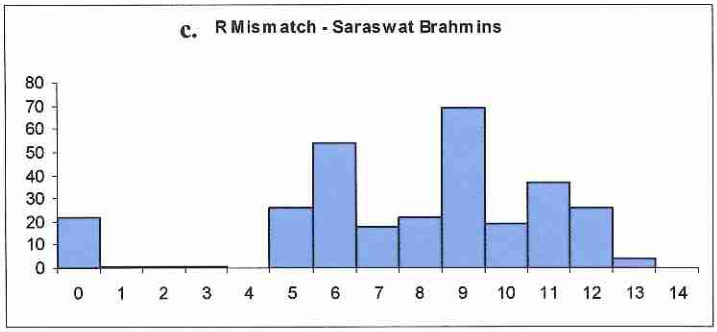
Figure 5: Saraswat Brahmin Pairwise Mismatch Distributions



Mean pairwise differences = 5.56
Raggedness index = 0.0064
N = 110

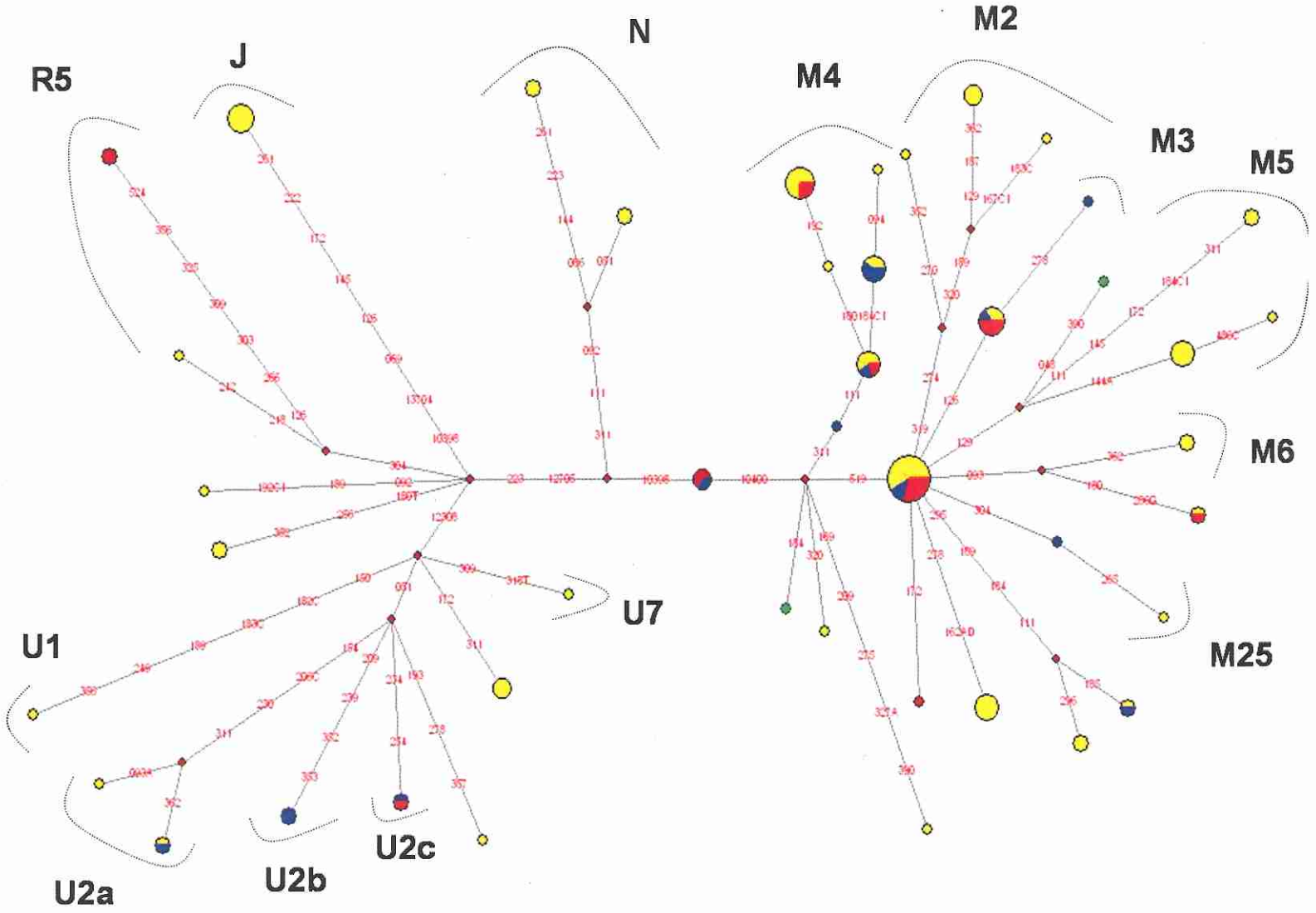


Mean pairwise differences = 4.08
Raggedness index = 0.008
N = 78



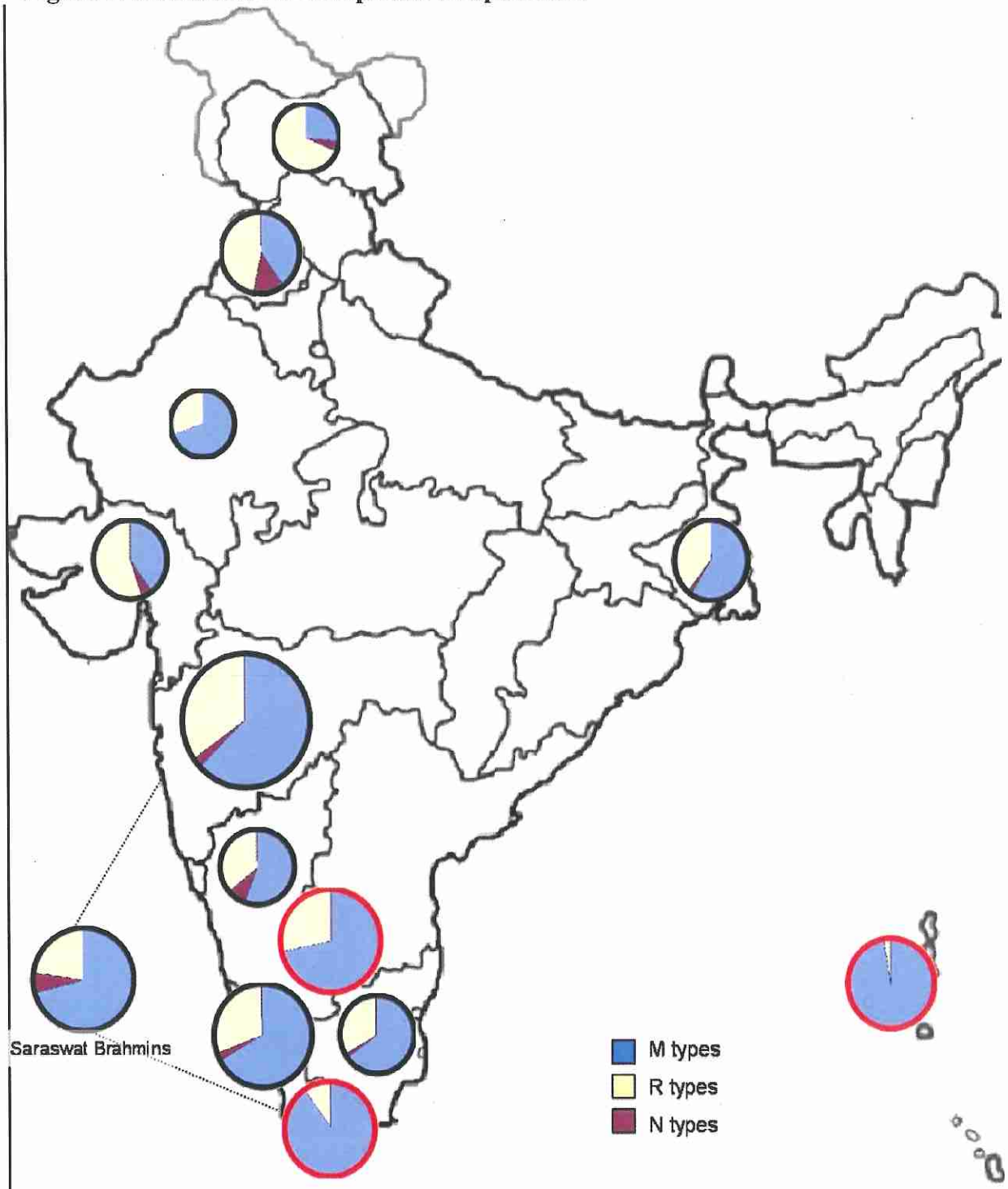
Mean pairwise differences = 7.81
Raggedness index = 0.1
N = 25

Figure 6: *Mutt* Distribution Network



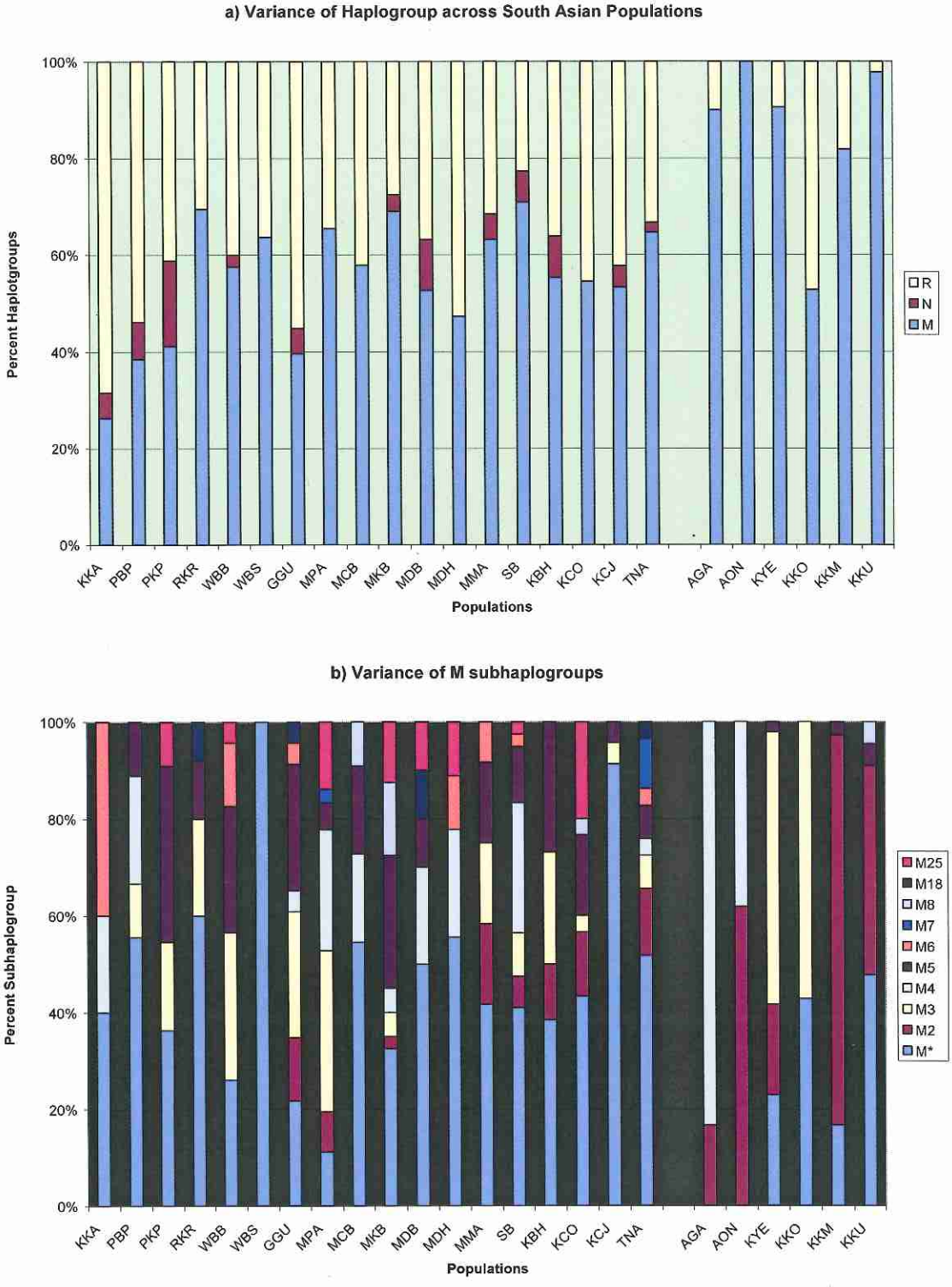
This network indicates the unspecific distribution of various Mutt *haplotypes. The yellow nodes indicate samples belonging to Kashi* Mutt, *the red nodes indicate Chitrapur Saraswat* Mutt, *the nodes circles indicate Gokarn Parthagali* Mutt, *and the green nodes indicate Kavle* Mutt.

Figure 7: Distribution of Comparative Populations



Each piechart on the map is proportional to the number of samples analyzed from the state the pie chart is placed on. Circles outlined in black indicate caste populations, while those outlined in red indicate tribal populations. The individual pie charts represent the macrohaplogroup percentages found in that state's overall populations used for this comparative study.

Figure 8: Variance of Haplogroups



c) Variance of U subhaplogroups

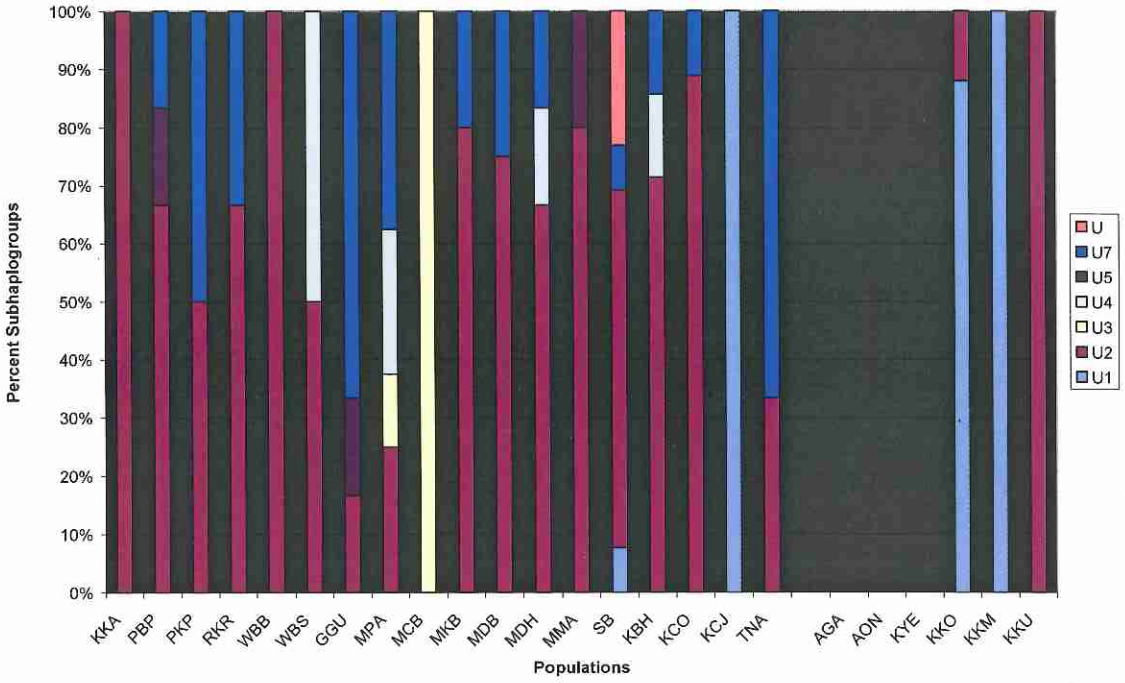
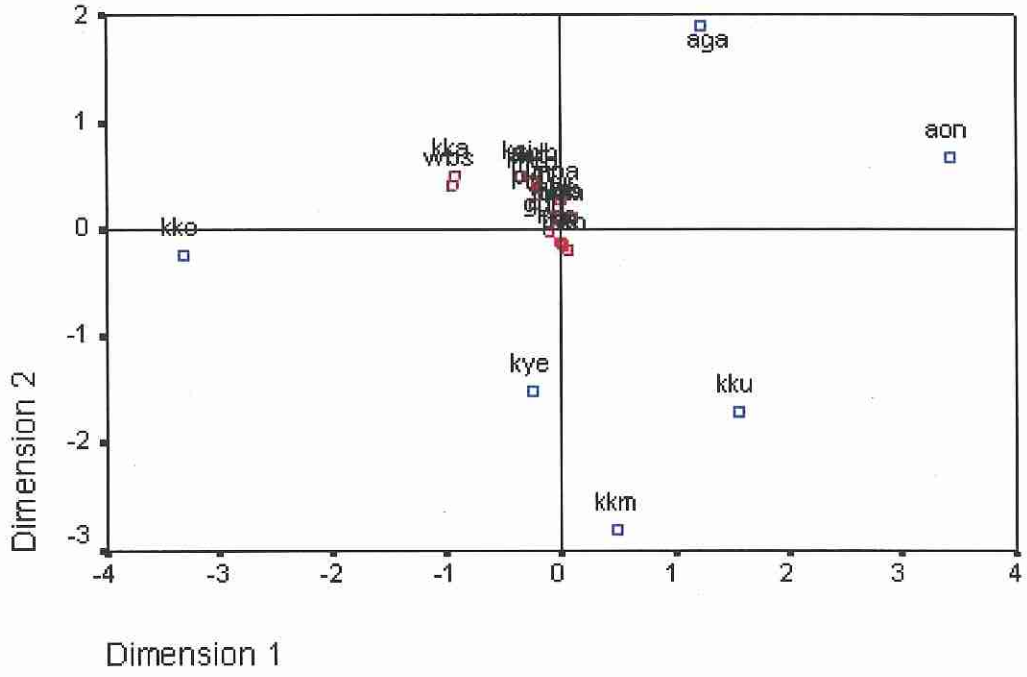


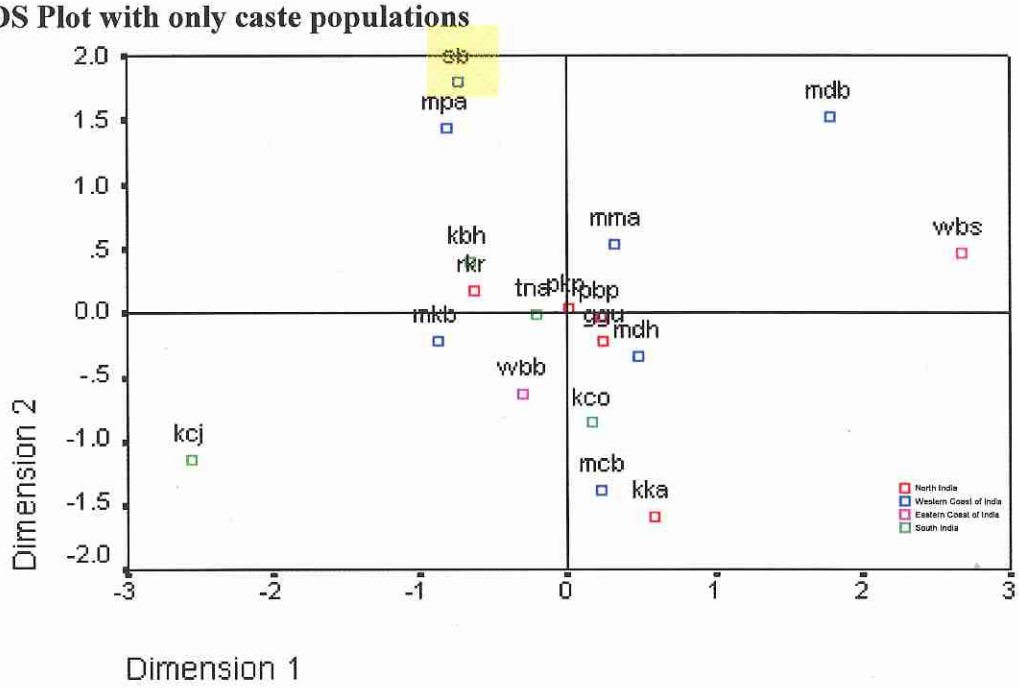
Figure 9: Multidimensional Scaling Plots

a. MDS Plot for All Comparison Populations



Tribal populations are indicated with blue squares, while caste populations – indicated by red squares – are all clustered in the middle of the plot.

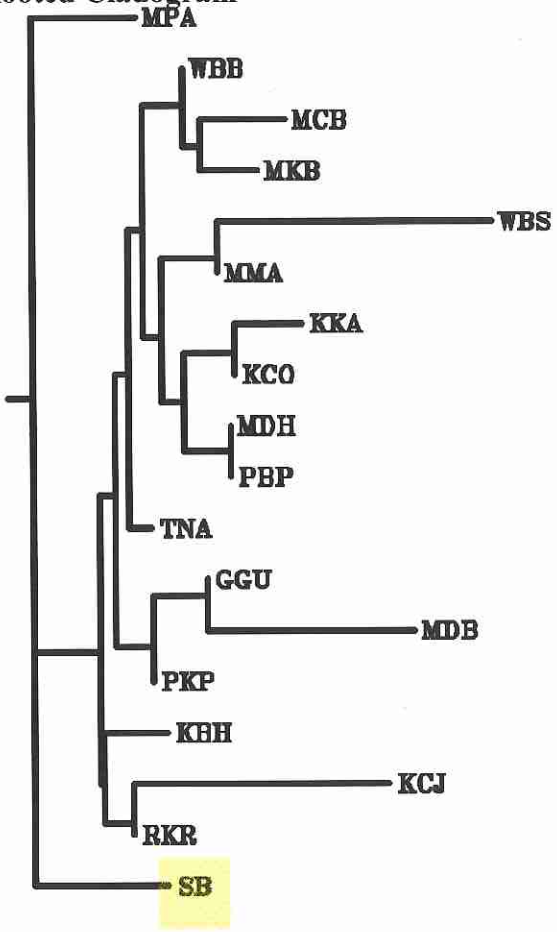
b. MDS Plot with only caste populations



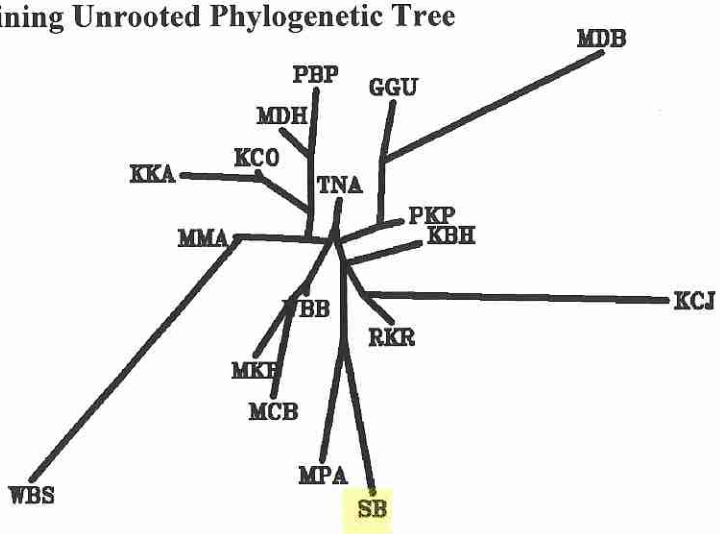
The higher resolution on the center caste cluster shows the distribution of caste populations (red: North India, blue: Western coast, purple: Eastern coast, green: South India).

Figure 10: Graphical Representations of the Neighbor-Joining Phylogenetic Method

a. Neighbor-Joining Rooted Cladogram



b. Neighbor-Joining Unrooted Phylogenetic Tree



Appendix I (2 pages)

INFORMED CONSENT FORM
The Genetic History of the Saraswat Brahmins: Origins and Affinities with Indian Populations

Principal Investigator: Theodore G. Schurr, Ph.D., Assistant Professor
Department of Anthropology, University of Pennsylvania,
325 University Museum, 3260 South Street, Philadelphia, PA. 19104-6398
Tel#: 215-573-7632; Email: tgschurr@sas.upenn.edu

Collaborators: Ms. Athma A. Pai
63 Westover Lane, Stamford CT 06902
Tel#: (203) 912-0533; Email: aapai@sas.upenn.edu

Invitation to Participate: You are invited to participate in a study of genetic variation of the Saraswat Brahmins to investigate their history and migrations in South Asia.

Purpose: In this research project, we will survey your mitochondrial DNA (mtDNA) and Y chromosome genes for sequence variation. Small mutations in these portions of your genome create genetic markers that can be used to characterize human populations. These changes identify specific genetic lineages within human populations that can be traced through maternal (mtDNA) and paternal (Y chromosome) lineages. For these reasons, we can trace these mutations through human families from the present to the distant past with considerable accuracy. We can also reconstruct patterns of human movement through geographic areas by tracking the spread of these lineages in different human groups. We have included a packet with further material explaining mtDNA and Y chromosome variation in general and specifically in India.

Procedures: If you decide to participate, we will ask you to provide us with a buccal (cheek) cell sample. The materials included in this package contain the items that are needed to obtain this sample. They include two cheek swab kits in individual wrappers and two vials of buffer. Open one of the swab kits, rub the serrated part of the head vigorously (without drawing blood) up and down on the inside of your mouth about 10 times on both sides of your mouth, then eject the head into a vial by pushing down on the narrow end of the shaft. Seal the cap of the vial securely. Repeat this procedure with the other swab kit and vial. This procedure requires only 5-10 minutes, and will not subject you to any significant health risks. We will also ask you to complete a short questionnaire about your family history, which should require no more than a few minutes to complete. This information will help us to interpret the results of your genetic tests, and will not be shared with anyone not associated with this project at Penn.

We have enclosed a pre-addressed envelope for you to use when mailing the buccal cell samples to Dr. Schurr's laboratory at the University of Pennsylvania. DNA will be extracted from the buccal cell samples, and then analyzed for mtDNA and Y-chromosome variation using molecular biology methods. We will report the results of the genetic analyses to you by mail or email. Upon the completion of the study, the DNA samples will be destroyed to prevent further use of them outside of the parameters of this project.

Risks or Discomfort to Participant: The buccal cell collection procedure requires only a few minutes to complete, and may result in very minor discomfort. Beyond the temporary discomfort, this procedure will subject you to no other significant health risks.
Benefits to Participant: There are no direct benefits to you as a result of your participation in this study. However, it is our hope that, by analyzing the sequence variation in the DNA from your buccal cells, we will be able to more fully elucidate the history of the Saraswat Brahmins.

Confidentiality: During the research project, all genetic and genealogical data will be kept in computerized form in the PI's laboratory, and accessed by only those persons involved in the project (Pai and Schurr). Any data under the researchers' control will be disclosed in scientific reports or public presentations in a manner that does not reveal your identity. The data obtained by the analysis of your DNA will also be made available to you upon your request.

In addition, your genetic data may be shown to Indian scientists who are collaborating with Dr. Schurr on the analysis of Saraswat Brahmin populations from North America and India. They are working together to better understand the biological history of the Sarawsats. In this case, no identifying information will be associated with your genetic data; only the number associated with your DNA sample will be provided to the Indian researchers to maintain your privacy. If you don't wish for your data to be shared, then you can simply indicate this preference to Dr. Schurr and Ms. Pai, and they will not include your data in the information to be shared with their Indian colleagues.

Compensation and Medical Treatment: While there is minimal risk to you in taking buccal cell samples, should any discomfort arise during this process, please do not proceed with the procedure. There will also be no financial compensation for your participation.

Research Contacts: Should you have any questions about this project or your rights as a research participant, please contact the Dr. Schurr at the mailing address, telephone number, or email address provided above.

Terms of Participation: Your participation in this research project is completely voluntary, and you may withdraw from it at any time.

Conclusion: You have read and understand the consent form. You agree to participate in this research study.

Name of Participant	Signature of Participant	Date

Appendix II (1 page)

GENEALOGICAL INFORMATION

The Genetic History of the Saraswat Brahmins: Origins and Affinities with Indian Populations

Principal Investigator: Theodore G. Schurr, Ph.D., Assistant Professor
Department of Anthropology, University of Pennsylvania
325 University Museum, 3260 South Street, Philadelphia, PA 19104-6398
Tel#: 215-573-7632; Email: tgschurr@sas.upenn.edu

Collaborator: Ms. Athma A. Pai
63 Westover Lane, Stamford, CT 06902
Tel#: (203) 912-0533; Email: aapai@sas.upenn.edu

Sample #_____ Date Collected_____

Name_____

Place of Birth_____ Ethnicity_____

Mother's Name_____

Place of Birth_____ Ethnicity_____

Family Religious Math: *Kashi*_____ *Gokarn Partagali*_____ *Kavle*_____ *Chitrapur* _____ *Other*:_____

Father's Name_____

Place of Birth_____ Ethnicity_____

Family Religious Math: *Kashi*_____ *Gokarn Partagali*_____ *Kavle*_____ *Chitrapur* _____ *Other*:_____

Maternal Grandmother_____

Place of Birth_____ Ethnicity_____

Maternal Grandfather_____

Place of Birth_____ Ethnicity_____

Paternal Grandmother_____

Place of Birth_____ Ethnicity_____

Paternal Grandfather_____

Place of Birth_____ Ethnicity_____

Additional Notes:

Appendix III (3 pages)

Indian Genetic Diversity: A Study of Saraswat Brahmins

Principal Investigator: Theodore G. Schurr, Ph.D., Assistant Professor
Department of Anthropology, University of Pennsylvania
325 University Museum, 3260 South Street, Philadelphia, PA 19104-6398
Tel#: 215-573-7632; Email: tgschurr@sas.upenn.edu

Collaborator: Ms. Athma A. Pai
63 Westover Lane, Stamford, CT 06902
Tel#: (203) 912-0533; Email: aapai@sas.upenn.edu

Molecular Anthropology is a field that combines the power of genetics and anthropological techniques to delve into the past. Using specific autosomal, mitochondrial, and Y-chromosome markers, molecular anthropologists can work to answer questions of origins, migrations, evolution patterns, and cultural practices. Past work has included the use of DNA mutation rates to calibrate a molecular evolution clock, tracing of populations back to ancestral homelands, and linking a group of people to a notable ancestor. This particular study aims to address questions of Indian Genetic Diversity by studying Saraswat Brahmin mitochondrial DNA (mtDNA) and Y-chromosome marker patterns. We hope to achieve the following:

1) Examine the general genetic diversity present within the community
2) Measure the degrees of genetic mixing with other Indian communities
3) Investigate possible migration routes into South India
4) Compare Saraswats with non-Indian populations which might have come into contact with Saraswats.

Ultimately, we hope to use Saraswat Brahmin genetic diversity results to propose a model for the uniqueness and variation within the larger arena of Indian genetic diversity.

Please note that this study will be looking at Saraswat Brahmins at a population level. No samples will be assessed on an individual basis. Due to the nature of recruitment for this study (ie. no two samples from the same immediate family), paternity or maternity issues cannot and will not come into question. Also, there is no such thing as proving that a person is not a Saraswat Brahmin through genetics. All populations have genetic variation, and that is exactly what we are trying to study.

Molecular Markers: mitochondrial and Y-chromosome DNA

Both mitochondrial and Y-chromosome DNA studies are commonly used to study population studies. This two types of DNA are particularly useful since they are indicative of maternal and paternal lineages respectively.

Mitochondrial DNA is a circular DNA fragment which is about 16,000 base pairs long. Each person (both male and female) inherits their mtDNA from their mother, so looking at mtDNA variation is a study of a maternal lineage. There is a coding region (genes which are translated into proteins) and a non-coding region on mtDNA. Both of

these regions have characteristic mutations that can be assessed by molecular biology techniques to assign a maternal haplogroup, or genetic lineage, to the person. These haplogroups have been well studied for their place of origin, evolution time, and the frequencies in various world populations. mtDNA haplogroup frequencies are often used to assess group or female migration patterns, cultural patterns, population variance, and species variation among all primate species.

The Y-chromosome works a little differently than mtDNA, since it is comprised of nuclear DNA. Since it is a sex chromosome, the Y-chromosome undergoes very little recombination with other chromosomes, and is purely paternally inherited. Given its nature as a sex chromosome defining the male gender, only males can be studied for Y-chromosome variation. Two regions on the Y-chromosome, the Non-Recombining Region (NRY) and the Sex Determining Region (SRY), have characteristic mutations that again can be assessed to assign a paternal haplogroup. Y-chromosome variation is often used to study male migration patterns, cultural patterns, and population variance.

Overall Indian Genetic Diversity

The large diversity of traditions and cultural practices of people on the Indian subcontinent has lead to an intensification of genetic diversity research of Indian communities (Baig et al. 2004, Bamshad et al. 2001). Remarkably, approximately one fifth of the human gene pool belongs primarily to people inhabiting the Indian subcontinent (Kivisild et al. 1999). Early work has been focused upon estimating the distinctions between groups in the well-known Indian caste system, delineating proto-Asian versus West Eurasian origins of peoples, estimating molecular dates for waves of settlement of the subcontinent, and mapping genetic data onto language trees (Baig et al. 2004, Bamshad et al. 2001, Kivisild et al. 1999, Kivisild et al. 2003, Palanichamy et al. 2004). Mitochondrial and Y chromosome work is especially relevant for Indian populations because initial results have shown that Indian lineages contain some very early branches of common haplotypes (Kivisild et al. 1999, Palanichamy et al. 2004). Much of the population genetics work focused upon the South Asian continent has been centered around solving problems of major haplogroup (M, N, and R) differentiation and clustering, caste differentiation, and regional variation.

Previous Saraswat Brahmin Genetic Studies

Though well-studied for their culture and religious practices, there has been little to no genetic work done on the Saraswat Brahmin community. The complex migratory patterns and high degree of endogamy as a minority high-caste Indian group makes the community an interesting target for genetic work. The only ascertainable work that has been done on Saraswats was a 1976 blood study by Bhatia et. al. which looked at ABO and Rhesus blood markers in members of the community in various Indian cities (Bhatia et al. 1976). This study was able to conclude limited information about high degrees of geographic separation and endogamy affecting genetic variation, however, increased technology and knowledge of genetic ancestry markers will be able to give more precise results. An interesting addendum to the Bhatia et. al. study was the conclusion that Saraswats might have been genetically influenced by Greek invasions in Punjab due to

notable levels of a particular thalessimia found in both Greeks and Saraswats (Bhatia et al. 1976, Patil 1970). A possible finding of Greek ancestry among the group would be quite remarkable. Work on Saraswat Brahmins can be contributed to this effort of categorizing the genetic history of Indians and possibly provide an example to understand the nuances of admixture and migration on the Indian subcontinent.

LABORATORY OF MOLECULAR ANTHROPOLOGY
Department of Anthropology, University of Pennsylvania

The Laboratory of Molecular Anthropology at UPenn is run by Dr. Theodore Schurr. The lab has several projects being conducted at the moment, which range from studies of molecular genetic variation in Altaian ethnic groups to biomedical studies of mtDNA diseases in Slavic populations to studies of Holocene archeological populations from the Lake Baikal region. There are quite a few other specific population diversity projects as well, including Afghani, Mongolian, and Caucasus regional genetic diversity studies. For more information about the lab and projects, please visit

http://www.sas.upenn.edu/~tgschurr

The lab recently received the honor of being chosen as the North American center for the prestigious worldwide Genographic Project launched by the National Geographic Society in order to study human evolution, history, and migrations through genetic studies. This 5 year project with 10 regional bases around the world will help to better understand global genetic diversity in a way that has never been considered before. The results from this study on Saraswat Brahmin diversity could possibly be integrated into the Genographic Project database in the future. For more information about the Genographic Project (including more background information about population genetic markers), please visit:

http://www.nationalgeographic.com/genographic

---

[1] Baig MM, Khan AA, and Kulkarni KM. (2004) Mitochondrial DNA Diversity in Tribal and Caste Groups of Maharastra (India) and its Implications on Their Genetic Origins. *Annals of Human Genetics* 68:453-60.

[2] Bamshad M, Kivisild T, Watkins WS, et al. (2001) Genetic Evidence on the Origins of Indian Caste Populations. *Genome Research* 11:994-1004.

[3] Bhatia HM, Shanbhag SR, Baxi AJ, et al. Genetic Studies among Endogamous Groups of Saraswats in Western India. *Hum. Hered.* 26:458-67.

[4] Jobling MA, Hurles ME, and Tyler-Smith C. (2004) Human Evolutionary Genetics: Origins, Peoples, and Disease. *Garland Science*: New York.

[5] Kivisild T, Bamshad MJ, Kaldma K, et al. (1999) Deep common ancestry of Indian and western-Eurasian mitochondrial DNA lineages. *Current Biology* 9:1-5.

[6] Kivisild T, Rootsi S, Metspalu M, et al. (2003) The Genetic Heritage of the Earliest Settlers Persists Both in Indian Tribal and Caste Populations. *Am. J. Hum. Genet.* 72:313-332.

[7] The Genographic Project. (2006) *National Geographic Society.* http://www.nationalgeographic.com/genographic

[8] Palanichamy MG, Sun C, Agrawal S et al. (2004) Phylogeny of Mitochondrial DNA Macrohaplogroup N in India, Based on Complete Sequencing: Implications for the Peopling of South Asia. *Am. J. Hum. Genet.* 75:966-978.

[9] Patil V. (1970) Saraswats. *The Illustrated Weekly of India* 26:6-19.

References Cited:

Baig M, Khan A, and Kulkarni K (2004) Mitochondrial DNA diversity in tribal and caste groups of Maharashtra (India) and its implications on their genetic origins. Annals of Human Genetics *68:*453-460.

Bamshad M, Kivisild T, Watkins W, Dixon M, Ricker C, Rao B, Naidu J, Prasad B, Reddy P, Rasanayagam A, Papiha S, Villems R, Redd A, Hammer M, Nguyen S, Carroll M, Batzer M, and Jorde L (2001) Genetic Evidence on the Origins of Indian Caste Populations. Genome Research *11:*994-1004.

Bandelt H-J, Forster P, and Rohl A (1999) Median-joining networks for inferring intraspecific phylogenies. Molecular Biology and Evolution *16:*37-48.

Bandelt H-J, Forster P, Sykes B, and Richards M (1995) Mitochondrial portraits of human populations using median networks. Genetics *141:*743-753.

Basu A, Mukherjee N, Roy S, Sengupta S, Banerjee S, Chakraborty M, Dey B, Roy M, Roy B, Bhattacharyya N, Roychoudhury S, and Majumder P (2003) Ethnic India: A Genomic View, With Special Reference to Peopling and Structure. Genome Research *13:*2277-2290.

Bhatia H, Shanbhag S, Baxi A, Bapat J, Sathe M, Sharma R, Kabeer H, Bharucha Z, and Surlacar L (1976) Genetic Studies among Endogamous Groups of Saraswats in Western India. Human Heredity *26:*458-467.

Cavalli-Sforza L, and Edwards A (1967) Phylogenetic analysis: models and estimation procedures. Evolution *21:*550-570.

Census (1991) Census of India: Government of India.

Conlon F (1977) A Caste in a Changing World: The Chitrapur Saraswat Brahmans, 1700-1935. Berkeley: University of California Press.

Cordaux R, Saha N, Bentley G, Aunger R, Sirajuddin S, and Stoneking M (2003) Mitochondrial DNA analysis reveals diverse histories of tribal populations from India. European Journal of Human Genetics *11:*253-264.

Excoffier L, Laval G, and Schneider S (2005) Arlequin (version 3.0): An integrated software package for population genetics data analysis. Evolutionary Bioinformatics Online *1:*47-50.

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

Felsenstein J (1998) PHYLIP (Phylogeny Interference Package), version 3.572c. Seattle, University of Washington: Distributed by author.

Forster P, Harding R, Torroni A, and Bandelt H-J (1996) Origin and evolution of Native American mtDNA variation: A reappraisal. American Journal of Human Genetics *59:*935-45.

Gaikwad S, and Kashyap V (2005) Molecular Insight into the genesis of ranked caste populations of western India based upon polymorphisms across non-recombinant and recombinant regions in genome. Genome Biology *6:*10.

Ingman M, Kaessmann H, Paabo S, and Gyllensten U (2000) Mitochondrial genome variation and the origin of modern humans. Nature *408:*708-713.

Jobling M, Hurles M, and Tyler-Smith C (2004) Human Evolutionary Genetics: Origins, Peoples & Disease. New York, NY: Garland Science.

Keni C, ed. (1998) Saraswats: in Goa and Beyond. Vasco-da-Gama, Goa, India: Murgaon Mutt Sandul Samiti.

Kivisild T, Bamshad M, Kaldma K, Metspalu M, Metspalu E, Reidla M, Laos S, Parik J, Watkins W, Dixon M, Papiha S, Mastana S, Mir M, Ferak V, and Villems R (1999a) Deep common ancestry of Indian and Western-Eurasian mitochondrial DNA lineages. Current Biology *9:*1-5.

Kivisild T, Kaldma K, Metspalu M, Parik J, Papiha S, and Villems R (1999b) The Place of the Indian mtDNA Variants in the Global Network of Maternal Lineages and the Peopling of the Old World. In R Deka, S Papiha and R Chakraborty (eds.): Genomic Diversity: Applications in Human Population Genetics: Klumer Academic Plenum Publishers, pp. 135-152.

Kivisild T, Papiha S, Rootsi S, Parik J, Kaldma K, Reidla M, Laos S, Metspalu M, Pielberg G, Adojaan M, Metspalu E, Mastana S, Wang Y, Golge M, Demirtas H, Schnakenberg E, Stefano Gd, Geberhiwot T, Claustres M, and Villems R (2000) An Indian Ancestry: a Key for Understanding Human Diversity in Europe and Beyond. In C Renfrew and K Boyle (eds.): Archaeogenetics: DNA and the population prehistory of Europe. Cambridge, United Kingdom: McDonald Institute for Archaeological Research, pp. 267-275.

Kivisild T, Rootsi S, Metspalu M, Mastana S, Kaldma K, Parik J, Metspalu E, Adojaan M, Tolk H-V, Stepanov V, Golge M, Usanga E, Papiha S, Cinnioglu C, King R, Cavalli-Sforza L, Underhill P, and Villems R (2003) The Genetic Heritage of the Earliest Settlers Persists Both in Indian Tribal and Caste Populations. American Journal of Human Genetics *72:*313-332.

Maca-Meyer N, Gonzalez A, Pestano J, Flores C, Larruga J, and Cabrera V (2003) Mitochondrial DNA transit between West Asia and North Africa inferred from U6 phylogeography. BMC Genetics *4:*15.

Majumder P (1998) People of India: biological diveristy and affinities. Evolutionary Anthropology *6:*100-110.

Maniatis T, Fritsch E, and Sambrook J (1982) Molecular Cloning: A Laboratory Manual. New York: Cold Spring Harbor Laboratory Press.

Metspalu M, Kivisild T, Metspalu E, Parik J, Hudjashov G, Kaldma K, Serk P, Karmin M, Behar D, Gilbert M, Endicott P, Mastana S, Papiha S, Skorecki K, Torroni A, and Villems R (2004) Most of the extant mtDNA boundaries in South and Southwest Asia were likely shaped during the initial settlement of Eurasia by anatomically modern humans. BMC Genetics *5:*26.

Morral N, Bertranpetit J, Estivill X, Nunes V, Casals T, Gimenez J, Reis A, Varon-Mateeva R, Macek M, Kalaydijieva L, Angelicheva D, Dancheva R, Romeo G, Russo M, Garnerone G, Perrari M, Magnani C, Claustres M, Desgeorges M, Schwartz M, Schwartz M, Dallapiccola B, Novelli G, Ferec C, Arce Md, Nemeti M, Kere J, Anvret M, Dahl N, and Kadasi L (1994) The origin of the major cystic fibrosis mutation (delta F508) in European populations. Nature Genetics *7:*169-175.

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

Nei M (1972) Genetic Distances Between Populations. American Naturalist *106:*283-292.

Palanichamy M, Sun C, Agrawal S, Bandelt H-J, Kong Q-P, Khan F, Wang C-Y, Chaudhur TK, Palla V, and Zhang Y-P (2004) Phylogeny of Mitochondrial DNA Macrohaplogroup N in India, Based on Complete Sequencing: Implications for the Peopling of South Asia. American Journal of Human Genetics *75:*966-978.

Patil V (1970) Saraswats. The Illustrated Weekly of India *26:*6-19.

Rajkumar R, Banerjee J, Gunturi H, Trivedi R, and Kashyap V (2005) Phylogeny and antiquity of M macrohaplogroup inferred from complete mt DNA sequence of Indian specific lineages. BMC Evolutionary Biology *5:*26.

Rajkumar R, and Kashyap V (2004) Genetic structure of four socio-culturally diverse caste populations of southwest India and their affinity with related Indian and global groups. BMC Genetics *5:*23.

Roy S, Thakur C, and Majumder P (2003) Mitochondrial DNA variation in ranked caste groups of Maharashtra (India) and its implications on genetic relationships and origins. Annals of Human Biology *30:*443-454.

Saillard J, Forster P, Lynnerup N, Bandelt H-J, and Norby S (2000) mtDNA variation among Greenland Eskimos: the edge of the Beringian expansion. American Journal of Human Genetics *67:*718-726.

Saitou N, and Nei M (1987) The neighbor-joining method: A new method for recontructing phylogenetic trees. Molecular Biology and Evolution *4:*406-425.

Schneider S, Roessli D, and Excoffier L (2000) ARLEQUIN, version 2.0: A software for population genetic data analysis. Geneva, Switzerland: University of Geneva.

Schurr T, Sukervik R, Starikovskaya Y, and Wallace D (1999) Mitochondrial DNA diversity in Koryaks and Itel'men: Population replacement in the Okhotsk-Bering Sea region during the Neolithic. American Journal of Human Genetics *101:*1-40.

Schurr T, and Wallace D (2002) Mitochondrial DNA Diversity in Southeast Asian Populations. Human Biology *74:*431-452.

Singh K (2002) People of India: introduction national series. Dehli, India: Anthropological Survey of India-Oxford University Press.

Snell R, ed. (1987) Aspects of the Culture of South Asia. London, UK: School of Oriental and African Languages, University of London.

Sun C, Kong Q-P, Palanichamy M, Agrawal S, Bandelt H-J, Yao Y-G, Khan F, Zhu C-L, Chaudhuri T, and Zhang Y-P (2006) The Dazzling Array of Basal Branches in the mtDNA Macrohaplogroup M from India as Inferred from Complete Genomes. Molecular Biology and Evolution *23:*683-690.

Thangaraj K, Singh L, Reddy A, Rao V, Sehgal S, Underhill P, Pierson M, Frame I, and Hagelberg E (2003) Genetic Affinities of the Andaman Islanders, A Vanishing Human Population. Current Biology *13:*86-93.

TIGV (2005) The Indian Genome Variation database (IGVdb): a project overview. Human Genetics *118:*1-11.