

7-5-2011

Analysis of Y-chromosome Diversity in Lingayat and Vokkaliga Populations of Southern India

Shilpa Chennakrishnaiah

Florida International University, schen011@fiu.edu

DOI: 10.25148/etd.FI11072506

Follow this and additional works at: <https://digitalcommons.fiu.edu/etd>

Recommended Citation

Chennakrishnaiah, Shilpa, "Analysis of Y-chromosome Diversity in Lingayat and Vokkaliga Populations of Southern India" (2011).
FIU Electronic Theses and Dissertations. 423.
<https://digitalcommons.fiu.edu/etd/423>

This work is brought to you for free and open access by the University Graduate School at FIU Digital Commons. It has been accepted for inclusion in FIU Electronic Theses and Dissertations by an authorized administrator of FIU Digital Commons. For more information, please contact dcc@fiu.edu.

FLORIDA INTERNATIONAL UNIVERSITY

Miami, Florida

ANALYSIS OF Y-CHROMOSOME DIVERSITY IN LINGAYAT AND VOKKALIGA
POPULATIONS OF SOUTHERN INDIA

A thesis submitted in partial fulfillment of the

requirements for the degree of

MASTER OF SCIENCE

in

FORENSIC SCIENCE

by

Shilpa Chennakrishnaiah

2011

To: Dean Kenneth G. Furton
College of Arts and Sciences

This thesis, written by Shilpa Chennakrishnaiah, and entitled Analysis of Y-chromosome Diversity in Lingayat and Vokkaliga Populations of Southern India, having been approved in respect to style and intellectual content, is referred to you for judgment.

We have read this thesis and recommend that it be approved.

DeEtta K. Mills

Bruce R. McCord

Rene J. Herrera, Major Professor

Date of Defense: July 5, 2011

The thesis of Shilpa Chennakrishnaiah is approved.

Dean Kenneth G. Furton
College of Arts and Sciences

Interim Dean Kevin O'Shea
University Graduate School

Florida International University, 2011

DEDICATION

I dedicated this thesis to my parents, brother and husband for their unconditional love, unwavering support and encouragement throughout the course of this study.

ACKNOWLEDGMENTS

I would like to thank my major professor Dr. Rene J. Herrera, and the members of my committee, Dr. Bruce McCord and Dr. DeEtta Mills for their constant support and guidance during this work. I would like to specifically thank David Perez, who helped me to finish the project in a timely fashion. I also would like to extend my gratitude to Tenzin Gayden, Luis Rivera, Harlette Lacau, Dr. Maria Regueiro, Kristian Herrera, Areej Bukhari and Dr. Tanya Simms for their help and constructive criticisms.

ABSTRACT OF THE THESIS

ANALYSIS OF Y-CHROMOSOME DIVERSITY IN LINGAYAT AND VOKKALIGA
POPULATIONS OF SOUTHERN INDIA

by

Shilpa Chennakrishnaiah

Florida International University, 2011

Miami, Florida

Professor Rene J. Herrera, Major Professor

Archaeological and genetic evidence have long supported the notion that the Indian subcontinent played an important role in the dispersal of modern humans out of Africa. In the present study, two Dravidian populations, namely Lingayat (N=101) and Vokkaliga (N=102) were examined using high-resolution analyses of Y-chromosome single nucleotide polymorphism (Y-SNP) and their associated seventeen short tandem repeat (STR) loci. The results revealed a prevalence of the major indigenous Indian Y-haplogroups (H, L, F* and R2), which collectively accounted for three-fourths of the Lingayat and Vokkaliga paternal gene pool. In addition, the presence of ancient lineages such as F*-M213, H*-M69 and C*-M216 suggested that modern humans reached India very early after their migration out of Africa. Finally, high haplotype diversity values at 17 Y-STR loci for Lingayat (0.9981) and Vokkaliga (0.9901) populations as well as the absence of shared haplotypes between them emphasized the importance of independent databases for forensic casework.

TABLE OF CONTENTS

CHAPTER	PAGE
I. INTRODUCTION.....	1
II. MATERIALS AND METHODS.....	5
Sample collection and DNA extraction.....	5
Y-SNP and Y-STR genotyping.....	5
Statistical and Phylogenetic Analyses.....	7
III. RESULTS.....	10
Y-Haplogroup Diversity.....	10
Genetic structure.....	12
Haplotype diversity and Coalescence Times.....	13
Network Analysis.....	15
IV. DISCUSSION.....	17
Haplogroups R1a1a-M198 and R2a-M124 are pan-Indian markers.....	17
Neolithic signals in the south Indian peninsula.....	20
Ancient South Asian lineages.....	22
Conclusion.....	24
REFERENCES.....	26

LIST OF FIGURES

FIGURE	PAGE
Figure 1. Y-haplogroup phylogenetic tree	33
Figure 2. Correspondence Analysis	34
Figure 3. Maximum Likelihood.....	35
Figure 4. A Median-Joining network of haplogroup R1a1a*-M198	36
Figure 5. A Median-Joining network of haplogroup R2a-M124.....	37
Figure 6. A Median-Joining network of haplogroup H1-M52	38
Figure 7. A Median-Joining network of haplogroup L1-M76.....	39

LIST OF TABLES

TABLE	PAGE
Table 1. Populations analyzed	40
Table 2. AMOVA	41
Table 3. Forensic parameters using the 9-loci and the 17-loci Y-STR haplotypes	42
Table 4. Y-chromosome haplogroup variance and time estimates	43
Supplementary Table 1. Y-STR haplotype data	44

CHAPTER I

INTRODUCTION

India's unique geostrategic position in South Asia has served as a major corridor for the dispersal of modern humans out of Africa (Cann, 2001; Basu et al., 2003; Cordaux et al., 2004; Thangaraj et al., 2005; Sengupta et al., 2006; Majumder, 2008; Chandrashekar et al., 2009; Petraglia et al., 2010). The people of contemporary India represent a mosaic of ethnic, linguistic, cultural and genetic diversity (Basu et al., 2003; Sahoo et al., 2006; Terreros et al., 2007). The vast majority of the Indian populations (80.5%) (Census of India, 2001) belong to the Hindu religion, the members of which are socially stratified into castes on the basis of their occupation. The four main castes are Brahmin (scholars and priests), Kshatriyas (kings and warriors), Vaishyas (agriculturists and traders), and Shudras (menial laborers) (Basu et al., 2003; Terreros et al., 2007).

Linguistically, Indian populations display strong regional patterns with Austro-Asiatic and Tibeto-Burman speaking groups prevalent in East and Northeast India, Indo-European speakers distributed in the north and central parts of the country, while Dravidian language is spoken predominantly in the southern peninsula. The Indo-European speakers are the largest speaking group followed by Dravidians, Austro-Asiatic and Tibeto-Burman language families (Basu et al., 2003; Sahoo et al., 2006). With the exception of the Austro-Asiatic speakers, who are exclusively tribal, the other three language families are spoken by members of both caste and tribal groups (Sahoo et al., 2006; Majumder, 2008).

Dravidians are believed to be the indigenous inhabitants of India prior to the arrival of Indo-European speakers approximately 3500 year ago (Sarkar 1958; Majumder 2008). These immigrants were nomadic people from Central Asia who conquered the Indus Valley following the decline of the Harappan (Indus Valley) civilization. The Indo-European speakers are thought to have entered India from the northwest and then spread southward displacing the Dravidians to the southern peninsula. They also introduced the highly rigid caste system that led to the current endogamous groups in the Indian population (Sarkar 1958; Watkins et al., 2008; Majumder 2008).

In this study, two Dravidian Kannada-speaking populations namely Lingayat and Vokkaliga from the Karnataka state were examined. Karnataka is located in the southern part of India and shares its borders with the states of Maharashtra in the north, Andhra Pradesh in the east, Tamil Nadu in the southeast, Kerala in the southwest, Goa in the northwest and Arabian Sea to the west. The region covers an area of approximately 191,976 square kilometres or 5.83% of the total geographical area of India. It is ranked as the eighth and ninth largest Indian state by area and population size, respectively, and consists of 30 districts. The official language of Karnataka is Kannada and the capital city is Bangalore. The Lingayat and Vokkaliga populations represent the two major populations of Karnataka state (Rajkumar and Kashyap, 2004).

The Lingayat population was founded by Basavanna, a social reformer of the 12th century from Karnataka, India. He started this new group to protest against the rigid rituals and discrimination of people based on the caste system that was deeply rooted within the Hindu religion. Basavanna allowed people from all castes and creeds to join the Lingayat group to avoid discrimination based on superstitious beliefs and rigid social

hierarchy that was inherent in Hinduism. However, the Lingayat population itself gradually became another Indian caste that followed strict endogamy (i.e., the practice of marriage within an ethnic group). Vokkaligas, on the other hand, are individuals who depend on agriculture and farming for their living. Vokkaligas are also referred to as Gowdas, a title usually given to the head of the family or of a village, in some cases.

Previous genetic studies on the origin of Indian populations presented conflicting results, at least with respect to the Y-chromosome distributions in the country. Several studies (Bamshad et al., 2001; Basu et al., 2003; Cordaux et al., 2004) reported genetic affinities between Central Asian/European populations and Indian upper castes groups, supporting the migration of Indo-European speakers into the subcontinent. In contrast, Kivisild et al. (2003) suggested that the Indian tribal and caste populations share similar genetic origins as the Pleistocene southern and western Asian populations, with limited gene flow from external regions since the Holocene. Furthermore, a comprehensive study of Y-chromosome diversity within the sub-continent (Sengupta et al., 2006) found that Central Asian influence in India is more cultural than genetic. Recently, a genome-wide survey of Indian populations (Reich et al., 2010) revealed that the ancestry of all Indians can be traced to two ancient populations, namely 1) ‘Ancestral North Indian’ (ANI) and 2) ‘Ancestral South Indian’ (ASI). As a result of these conflicting conclusions, it is apparent that in order to understand the complex genetic history of Indian populations, more studies using different marker systems as well as representative populations across India are required.

The human Y-chromosome provides a powerful molecular tool for the investigation of human evolutionary history from a paternal perspective given that the Y-

chromosome does not undergo meiotic recombination with the exception of the pseudoautosomal regions (Jobling and Tyler-Smith, 2003). As a result, any mutational events that occur along the non-recombining region of the Y-chromosome (NRY) are retained and transferred paternally from one generation to another. By studying the worldwide distributions of these mutations on the NRY, one can identify signals of various waves of human migrations and reconstruct the patterns of their respective contributions to the contemporary genetic landscape of modern populations (Jobling and Tyler-Smith, 2003).

Although a previous study (Rajkumar and Kashyap, 2004) reported autosomal STR diversity in the Lingayat and Vokkaliga populations, the current project is the first to genetically characterize these groups using high-resolution Y-single nucleotide polymorphism (Y-SNP) analyses and their associated short tandem repeats (STRs). In addition to investigating the paternal diversity in the Lingayat and Vokkaliga populations, the present study seeks 1) to estimate the time of most common recent ancestor (TMCRA) for select Y-haplogroups using Y-STR data to determine the initial human settlement in southern India; 2) to shed light on the origin of the Indian population with respect to Indo-European vs. Dravidian groups by comparing the data obtained from this study with the previously published work from India and other global populations; and 3) to establish a Y-STR database of the Lingayat and Vokkaliga collections for forensic and population genetic studies.

CHAPTER II

MATERIALS AND METHODS

Sample collection and DNA extraction

Blood samples or buccal swabs were collected after receipt of informed consent from 203 males belonging to two Dravidian populations, namely Lingayat (101) and Vokkaliga (102). Both of these populations are Kannada-speaking groups from Karnataka state of southern India. The paternal ancestry of every donor was recorded for a minimum of two generations. The DNA was isolated from buccal swabs using the Genra Buccal Cell Kit (Puregene, Genra Systems, Minneapolis, MN), whereas the phenol-chloroform method was employed for extraction from the blood samples as described elsewhere (Antunez de Mayolo et al., 2002). All the ethical guidelines were followed as specified by the Florida International University Institutional Review Board (IRB).

Y-SNP and Y-STR genotyping

Sixty six biallelic markers were genotyped following the hierarchy of the human Y-chromosome tree using standard methods, including PCR-RFLP, allele-specific PCR (Gayden et al., 2008) or direct sequencing. Allele-specific PCR was done for those binary markers where the SNP did not generate a restriction site. The technique involves designing primers that are complementary to one of the two allelic variants at the 3' end of the oligonucleotide such that amplification will take place only when there is a perfect

match. The nomenclature followed for the Y-SNP haplogroups is as recommended by Karafet et al. (2008) and a recent update by Myres et al. (2011).

The PCR amplification for 17 Y-STR loci (DYS19, DYS385a/b, DYS389I, DYS389II, DYS390, DYS391, DYS392, DYS393, DYS437, DYS438, DYS439, DYS448, DYS456, DYS458, DYS635 and GATA H4) was performed in a multiplex reaction using the AmpFISTR[®] Y-filer kit (Applied Biosystems, Foster City, CA) according to the manufacturer's instructions in an Eppendorf Mastercycler[®] gradient thermal cycler (Eppendorf AG, Hamburg, Germany). The amplified fragments were separated by multi-capillary electrophoresis on an ABI Prism 3130xl Genetic Analyzer and the ABI Genescan 500 LIZ[®] was used as an internal size standard. The software GeneMapper[®] v3.1 (Applied Biosystems, Foster City, CA) was employed to determine fragment sizes, while alleles were designated by comparison to an allelic ladder supplied by the manufacturer.

Since DYS458 null samples are frequently associated with the deletion of Y-homolog of amelogenin locus, a primer pair of AMEL-2F/3R (Chang et al., 2003; Cadenas et al., 2006) was used to test the amelogenin locus for four Lingayat samples that showed a null allele at the DYS458 locus. On the other hand, the DYS389 locus yields two PCR fragments (DYS389I and DYS389II) that overlap as a result of a duplicated priming site for the forward primer (Kayser et al., 1997). As a result, DYS389II produces a longer fragment that contains the DYS389I allele. Therefore, the allele for DYS389II is obtained by subtracting the size of the corresponding DYS389I allele. In addition, DYS385 was not considered for any of the Y-STR based analyses given the duplicated nature of the marker.

Statistical and Phylogenetic Analyses

A total of thirty-three reference populations (Table 1) from the published literature were included for comparison at the major haplogroup level (A to T) of the Y-chromosome tree. A correspondence analysis (CA) was carried out using the NTSYSpc-2.02i software by Rohlf (2002). The CA provides a two dimensional plot of genetic relationships between populations on the basis of their Y-chromosome haplogroup distributions. Thus, the populations that map closer together in the CA graph are more genetically similar than those that plot further apart. The two components that constitute the axes of the plot represent the two vectors that account for the greatest proportion of the total variance detected within the given data (represented by respective eigen-values for each dimension). A maximum likelihood (ML) tree was constructed using the CONTML program in the Phylip package v3.6 (Felsenstein, 2002). The statistical robustness of the phylogenetic relationships within the ML phylogeny was evaluated with bootstrap analysis (SEQBOOT option) involving 1,000 replications, while the CONSENSE program determined the best-fit tree. In order to evaluate any correlations between genetic diversity and language or geography, an analysis of molecular variance (AMOVA) was performed using ARLEQUIN software version 3.1 (Excoffier et al., 2005). Populations were partitioned into groups on the basis of their geographic location and linguistic affiliation as listed in Table 1, with the significance of the test assessed at $\alpha = 0.05$.

Haplotype diversity (HD) was calculated using the software package Arlequin v. 3.1 (Excoffier et al., 2005). Discrimination capacity (DC) was estimated by dividing the number of different haplotypes by the total number of individuals in a given population.

Chromosomes carrying null alleles were excluded from the haplotype calculation at the 17 Y-STR loci level. For the purpose of comparison, HD and DC values were also computed for the minimal 9-loci haplotype (DYS19, DYS385a/b, DYS389I, DYS389II, DYS390, DYS391, DYS392 and DYS393).

The age of STR variation within the Y-chromosome haplogroup was estimated as the average squared difference in allele size between all chromosomes and the founder haplotype, averaged over microsatellite loci and divided by the evolutionary effective mutation rate of 6.9×10^{-4} per 25 years (Zhivotovsky et al., 2004). As recommended by Sengupta et al. (2006), a founder haplotype was estimated by generating a haplotype from the median repeat number values at each Y-STR locus within a given haplogroup (Sengupta et al., 2006). Microsatellite variances for major binary haplogroups were computed using the V_p (intrapopulation genetic variance) equation of Kayser et al. (2001), which sets the overall variance as equal to the mean of the variances found at each locus.

Median joining networks were constructed using the program NETWORK 4.6.0.0 and Network Publisher version 1.2.0.0 (www.fluxus-engineering.com) in order to assess the genetic relationships between the haplotypes within haplogroups R1a1a*-M198, R2a-M124, H1a-M52 and L1-M76. Networks were generated from the eight Y-STR loci (DYS19, DYS389I, DYS389II, DYS390, DYS391, DYS392, DYS393 and DYS439) common among all relevant reference populations (Table 1) included in the analyses. The Indian populations used in the network analyses were taken from Sengupta et al. (2006). The weight of each locus was assigned to be inversely proportional to its variance (Martinez et al., 2007). To simplify the network projections, singletons were excluded

from the analyses involving multiple populations, with the exception of L1-M76 which consists of a few populations. In addition, all MJ networks were post-processed using the Maximum Parsimony (MP) algorithm.

CHAPTER III

RESULTS

Y-Haplogroup Diversity

A total of 66 binary markers typed in this study define 16 paternal lineages each in both Lingayat and the Vokkaliga populations. The phylogenetic relationships of the total 21 Y-chromosome haplogroups and their observed frequencies in the two Dravidian collections are shown in Figure 1. Overall, three major clades (H, L and R), commonly found in the Indian subcontinent, account for more than three-fourths of the samples analyzed in the present study.

Haplogroup R-M207 is the most common marker in Lingayat's paternal gene pool (35.64%), while the L-M20 clade predominates in the Vokkaliga population (32.35%). Within haplogroup R, the sub-clade R1a1a*-M198, which has been associated with the spread of the Indo-European language (Wells et al., 2001), is found in Lingayat (19.8%) at a frequency twice that of the Vokkaligas (6.86%), while R2a-M124 is present at equivalent proportions in both collections (15.84% and 13.73%, respectively). Interestingly, only one R1a1a-M198 sample from Vokkaliga is polymorphic for the downstream mutation (R1a1a1e-PK5). On the other hand, with the exception of four Vokkaliga samples that belong to L3*-M357, all of the L-M20 chromosomes are L1-M76 derivatives, an observation congruent with the results previously reported in the subcontinent (Sengupta et al., 2006).

The second-most frequent major clade in both groups is an Indian-specific lineage, H-M69, which shows comparable distributions in Lingayat (24.75%) and

Vokkaliga (23.53%) populations, with the majority of the samples exhibiting the H1a-M82 mutation (17.82% and 11.76%, respectively). In addition, low levels of subtype H1a1-M197 (0.99% and 0.98%, respectively) were detected in the Indian peninsula. It should be noted that the sub-clade H2-Apt is restricted to the Vokkaliga population (7.84%).

The frequency of paraphyletic F*-M213 in Vokkaliga (13.73%) is comparable with the value reported in the Dravidian population included in Sengupta et al. (2006) but it is relatively less polymorphic in Lingayat (4.95%). In contrast, J-M304 is found in higher proportion in Lingayat (11.88%) than in Vokkaligas (3.92%). Within this clade, the majority of the samples are characterized by the J2b2*-M241 branch (3.96% and 2.94% in Lingayat and Vokkaliga, respectively). Although the sub-lineage J2a*-M410 is observed exclusively in the Lingayat gene pool (3.96%), its derivative J2a3-M68 is present in both of the south Indian collections, albeit at low frequencies (<2%). In addition, one Lingayat male is derived at J2a13-P279.

Other informative major haplogroups common between the two Dravidian populations include C-M216 and G-M201. With the exceptions of an undifferentiated C*-M216 chromosome in Vokkaliga and one G1*-M285 derived individual in Lingayat, sub-clades C5a-P92 (2.97% and 1.96% in Lingayat and Vokkaliga, respectively) and G2a*-P15 (0.99% and 0.98% in Lingayat and Vokkaliga, respectively) are equally represented in both the populations. Interestingly, an individual carrying O2a*-M95 is found only in the Vokkaliga collection. This East Asian specific marker was previously reported in Andhra Pradesh (Ramana et al., 2001) and Tamil Nadu populations (Wells et al., 2001).

Genetic structure

Figure 2 displays a Correspondence Analysis (CA) plot generated from the haplogroup frequencies of the major Y-clades (A-T) from 35 populations (Table 1), including the two Dravidian groups examined in this study. The Central Asian and South Central Asian populations cluster loosely in the lower left quadrant of the graph, with populations from southern India exhibiting some degree of partitioning along axis I. Vokkaliga, Orissa and Kyrgyzstan are outliers on the periphery of the group, whereas Lingayat plots closely with two neighboring Dravidian populations from Tamil Nadu and Andhra Pradesh states. As expected, Pakistan maps close to Punjab while maintaining equidistant position from Uzbekistan and two Afghan collections. Interestingly, Sri Lanka lies in the middle of the cluster flanked by Uttar Pradesh on the left and Gujarat to the distant right.

The European populations, on the other hand, segregate in the upper left portion with the exception of Kosovo. The positioning of Kosovo Roma close to south Indian populations, including the two Dravidian collections, may be influenced by the extensive sharing of haplogroup H-M69 among them (Regueiro et al., 2011). On the other side of the graph, the populations from the Caucasus/Anatolia segregate near the center of the plot while the Middle Eastern group, excluding north and south Iran, lies at the extreme right of the chart. The genetic association between two Iranian collections and the Caucasus/Anatolia assemblage is consistent with previous observations (Shepard and Herrera, 2006; Regueiro et al., 2006; Terreros et al., 2011).

The topology of the ML tree (Figure 3) roughly corroborates the distribution of the populations in the CA plot. The dendrogram displays three distinct groups, namely

South Central Asian, European, and a cluster consisting of populations from the Middle East and Anatolia/Caucasus regions. However, there are several notable exceptions to these groupings. For instance, Orissa, West Bengal and two Nepalese collections (Newar and Kathmandu) separate from their South Central Asian family and instead branch within the European cluster. In addition, Kyrgyzstan and Uzbekistan also display genetic similarity to the European group as opposed to north and south Afghanistan. The two Afghan collections, along with Pakistan and North Iran, mirror their geographic locations by displaying an intermediate positioning between South Central Asia and Anatolia/Caucasus groups in the ML phylogeny.

As expected, the AMOVA results demonstrate that most of the genetic variation is found within the population (86%) in both the linguistic and geographic classifications (Table 2). When populations are grouped according to language family (Table 1), a similar variance is observed between among-groups ($\Phi_{CT} = 0.07$) and among-populations-within-groups ($\Phi_{SC} = 0.07$), indicating no preferential impact of language on either diversity component. However, the geographic grouping shows a higher proportion of among-group variance ($\Phi_{CT} = 0.09$) compared to among populations residing within the same geographical region ($\Phi_{SC} = 0.05$) (Table 2).

Haplotype diversity and Coalescence Times

The overall haplotype diversity for both Dravidian populations, collectively, at the 17 Y-STR loci level is 0.9970, while the corresponding value for the minimal 9-loci haplotype is 0.9915. Lingayat displays higher haplotype diversity (0.9981, 0.9917) than the Vokkaligas (0.9901, 0.9814) at both haplotype resolutions (Table 3). There are no

profiles shared between the two Dravidian collections at the 17-loci Yfiler haplotype data set, indicating an absence of recent gene flow between the two populations. However, haplotype sharing is observed within the individual population, particularly in the Vokkaligas, where the first and second most frequent haplotypes are represented by eight and five samples, respectively. In addition, eight other shared haplotypes, consisting of less than three chromosomes each, are noted (Supplementary Table 1). On the other hand, Lingayats are relatively more diverse, with 82 of the 89 haplotypes occurring only once in the population. Of the seven shared haplotypes, one haplotype is composed of three individuals while the remaining six are represented by two samples each. It is notable that four Lingayat males displayed a null allele at locus DYS458. Interestingly, these four individuals failed to amplify at J2b-M12 as well as for the Y homolog of the amelogenin locus (AMGY), congruent with a previous report that showed a deletion in the short arm of the Y-chromosome encompassing these two sites (Cadenas et al., 2006).

Table 4 presents the time estimates calculated according to the method described in Zhivotovsky et al. (2004) and as modified by Sengupta et al. (2006). Of all the lineages, haplogroup R2a-M124 from the Lingayat population yields the oldest time estimate (31.25 ± 6.41 kya), although a younger date is obtained for the same marker in Vokkaligas (14.92 ± 3.57 kya). A similar trend is also observed for the paragroup F*-M213, with a coalescent time twice older in Lingayat (24.65 ± 6.36 kya) compared to the Vokkaliga population (11.91 ± 4.26 kya). These findings are consistent with the high microsatellite variance accumulated within the R2a-M124 (0.82) and F*-M213 (0.693) chromosomes of the Lingayat population. However, the divergence times achieved for these two haplogroups in Lingayat are most likely inflated since this ethnic group was

historically formed from multiple source populations, thereby violating the assumption of a single founder in the coalescent model proposed by Zhivotovsky et al. (2004).

In contrast, time estimates for haplogroups H-M69 and L-M20 are older in Vokkaligas (26.97 ± 5.34 kya and 15.52 ± 4.61 kya, respectively) relative to the Lingayat population (22.03 ± 4.24 kya and 11.14 ± 3.14 kya, respectively). In addition, both of these haplogroups display higher genetic differentiation in the Vokkaliga collection, harboring additional sub-clades H2*-Apt and L3*-M357, respectively, which are completely absent in the Lingayat population. It is noteworthy that the ages for sub-haplogroups H1a*-M82 and R1a1a*-M198 are equivalent in both Dravidian collections (Table 3).

Network Analysis

Median joining networks based on the eight Y-STR loci haplotypes within the haplogroups R1a1a*-M198 and R2a-M124 are presented in Figures 4 and 5. It is notable that while there are no common haplotypes between the Lingayat and Vokkaliga collections in the M198 network, they share certain profiles between themselves and with other reference populations in the M124 network. In addition, Afghan M198 samples display extensive sharing among themselves as well as with other populations. It is also interesting to note that the most common M198 haplotype differs from the second most frequent haplotype by only one repeat at locus DYS390 (24/25). Furthermore, Newar shows limited to no haplotype sharing in both the network projections.

A Network analysis of the H1-M52 samples (Figure 6) exhibits a certain degree of partitioning between South Asian populations and Roma samples from Kosovo,

despite some haplotype sharing between these two groups. Of the populations analyzed, south Indian samples display the highest microsatellite diversity with widespread distributions throughout the network. The most frequent haplotype that represents the South Asian cluster comprises of one Roma, two Pakistani and 14 Indian samples, including four Lingayat males. Interestingly, a haplotype consisting of one individual each from Lingayat, Vokkaliga and north India is also observed. The remaining Lingayat and Vokkaliga haplotypes do not display any specific pattern in their distributions.

In contrast, with the exception of the west Indian samples, the haplogroup L1-M76 network (Figure 7) demonstrates high levels of Y-STR haplotype sharing between Lingayat and Vokkaliga samples as well as with other reference populations examined. In addition, the three major shared haplotypes in the network are separated from each other by one mutational step at locus DYS439 (11/12/13), suggesting a recent gene flow and/or a shared common ancestry for this marker. Therefore, with the exclusion of the DYS439 marker, these three core haplotypes share a 7-loci modal haplotype for L1-M76 (DYS19=14, DYS389I=12, DYS389II=16, DYS390=22, DYS391=10, DYS392=14 and DYS393=11), consisting of 53 individuals. In addition, one common haplotype exclusive to the Lingayat and Vokkaliga populations and several unique haplotypes belonging to the two Dravidian collections are present in the network diagram.

CHAPTER IV

DISCUSSION

Archeological and genetic studies support the notion that India's vast coastal area played an important role in the early southern dispersal of modern humans out of Africa (Kennedy et al., 1987; Quintana-Murci et al., 1999; Stringer, 2000; Cann, 2001; Macaulay et al., 2005; Petraglia et al., 2010; Morlote et al., 2011). The south Indian peninsula encompasses four Dravidian-speaking states, namely Andhra Pradesh, Karnataka, Kerala and Tamil Nadu. While the populations from Andhra Pradesh and Tamil Nadu have been widely studied (Ramana et al., 2001; Wells et al., 2001; Kivisild et al., 2003; Sengupta et al., 2006; Watkins et al., 2008), the genetic information on the other two southern states from the west coast is limited. The present study address this deficiency by examining the Y-chromosome diversity in the Lingayat and Vokkaliga populations, the two major communities of the Karnataka state, as well as evaluate their genetic relationships to other Indian and global populations.

Haplogroups R1a1a-M198 and R2a-M124 are pan-Indian markers

The geographic origin of haplogroup R1a1a-M198 remains contentious because of its widespread distributions across Eurasia (except for East Asia) as well as the current absence of molecular resolution within this marker (Wells et al., 2001; Zerjal et al., 2002; Kivisild et al., 2003; Sengutpa et al., 2006). The occurrence of haplogroup R1a1a-M198 in India has previously been associated with the Indo-European expansion from Central Asia into northern India about 3,500 years ago (Wells et al., 2001). This theory was

supported by subsequent genetic studies (Basu et al., 2003; Cordaux et al., 2004) which showed that upper caste north Indians are genetically more closely related to Central Asians than to south Indian populations. However, a recent report (Sengupta et al., 2006) involving comprehensive sampling across the subcontinent coupled with high resolution Y-chromosome analyses demonstrated that the Central Asian influence on the Indian gene pool, if any, is limited. Furthermore, Sengupta and colleagues (2006) proposed that an early Holocene expansion in northwestern India contributed R1a1a-M198 chromosomes in both South Asian and Central Asian populations.

The present study revealed appreciable frequencies of R1a1a-M198 in the two Dravidian caste populations examined, particularly in Lingayat (19.8%). While this marker reaches a frequency of 40% in northern India, it was also reported in south Indian Dravidian-speaking tribal groups of Chenchus (27%) and Valmiki (4.2%) from Andhra Pradesh and the Kallar (4.0%) from Tamil Nadu (Kivisild et al., 2003; Ramana et al., 2001; Wells et al., 2001). The presence of R1a1a-M198 in both southern Indian caste and tribal groups is inconsistent with the argument that this mutation is specific to Indo-European speakers in the north (Kivisild et al., 2003). In addition, the estimated ages of microsatellite variance within this chromosome in Lingayat (12.80 kya) and Vokkaliga (11.73 kya) are comparable to the average age reported for the Indian population (14.0 kya) (Sengupta et al., 2006), attesting to the antiquity of this marker in the subcontinent. This observation is further supported by the R1a1a-M198 network (Figure 4) which illustrated a high haplotype diversity among the South Asians, while the Afghani populations displayed limited variation, suggesting a founder effect and/or genetic isolation in the latter. Furthermore, AMOVA results (Table 2) also support genetic

differentiation of populations based on their geographical distance as opposed to the linguistic classification. Therefore, the data from the present study do not support the Indo-European contribution of R1a1a-M198 chromosomes to the Indian gene pool, given its occurrence in both language groups and the fact that the estimated age of this marker substantially predates the Indo-European expansion. Instead, the distribution of R1a1a-M198 in India is more consistent with Renfrew's (1992) elite dominance model in which the indigenous Dravidian speakers in northern India embraced the Indo-European language either by force or as a result of the social status associated with the linguistic affiliation (Basu et al., 2003).

Haplogroup R2a-M124, on the other hand, is geographically more confined relative to R1a1a-M198 but is present at informative levels in Iran (3.03%), Central Asia (8.0%), Pakistan (7.39%), Nepal (26%) and India (9.34%) (Wells et al., 2001; Regueiro et al., 2006; Sengupta et al., 2006; Gayden et al., 2007). The relatively high frequency of R2a-M124 in the Lingayat (15.84%) and Vokkaliga (13.73%) populations, combined with their high microsatellite variance (Figure 5 and Table 4) and deep coalescence times (24.64 kya and 11.91 kya, respectively) support the indigenous origin of this haplogroups in India (Kivisild et al., 2003; Sahoo et al., 2006; Sengupta et al., 2006). However, the age for R2a-M124 in Lingayat (24.64 kya) should be interpreted with caution given the history of the population. Lingayat was initially started as a religious group that consisted of people from a diverse set of castes and geographic regions, which may have contributed to an inflated age for R2a-M124. On the other hand, the high percentage of R2a-M124 in the Newar population from Nepal and the distinct partitioning of their

haplotypes in the network (Figure 5) are indicative of pronounced genetic drift in this ethnic group as described previously (Gayden et al., 2007; 2009).

Neolithic signals in the south Indian peninsula

At the basal haplogroup level, the L-M20 clade is distributed across Anatolia, the Middle East, Central Asia, India and Pakistan (Wells et al., 2001; Semino et al., 2000; Qamar et al., 2002; Cinnioglu et al., 2004). It is present at highest frequency near the Indus Valley (Sengupta et al., 2006), suggesting a potential role for this ancient marker in the expansion of farming in the subcontinent (Cordaux et al., 2004). Following the molecular resolution achieved within the haplogroup L (Sengupta et al., 2006), the majority of the M20 derived Indian Y-chromosomes possess the L1-M76 mutation, while the L3-M357 sub-clade is found mainly in Pakistan (Sengupta et al., 2006) and Afghanistan (Lacau, unpublished data). Accordingly, in the current study, with the exception of four Vokkaliga males that belong to L3-M357 lineage, all of the L-M20 samples in both Dravidian populations are L1-M76 derivative, with Vokkaligas (28.43%) displaying a higher frequency compared to the Lingayats (17.82%).

In contrast to the high Y-STR diversity reported for L1-M76 in South Asian tribes (Sengupta et al., 2006), the data from the present study show rather limited microsatellite variation (Figure 6), although a relatively older age is obtained for this marker (Table 4). Unlike R1a1a and R2 lineages, the L1-M76 haplotypes display extensive sharing between Lingayat and Vokkaliga collections as well as with reference populations examined (Figure 7). In addition, the 7-loci modal haplotype (DYS19=14, DYS389I=12, DYS389II=16, DYS390=22, DYS391=10, DYS392=14 and DYS393=11) corresponds to

five of the six markers common between the current study and the widespread modal haplotype shared among Chenchus, Lambadis, Punjabis and Iranians (Kivisild et al., 2003), suggesting a recent expansion of this marker from the southern peninsula (Sengupta et al., 2006). The latter inference is congruent with the low average mean variance obtained for L1-M76 (0.25) at the 15 Y-STR loci level (Table 4) is consistent with a previous report (Thanseem et al., 2006). Thus, it is possible that farmers from the fertile Indus Valley carrying the M20 lineage may have moved southward along the Indian coast and developed the L1-M76 mutation upon reaching the southwestern coast of the peninsula. The high occurrence of L1-M76 in Vokkaligas (28.43%), who consists primarily of agriculturalists, while L-M20 or its representatives are nearly absent in the populations on the east coast, namely Orissa and West Bengal (Sahoo et al., 2006) further supports that L1-M76 may be a neolithic marker. Alternatively, the elevated level of L1-M76 in Vokkaliga may be a result of a founder effect and/or genetic drift, given its low STR variance compared to other Y-lineages (Table 4). On the basis of the high frequency of L1-M76 and microsatellite variance among Dravidian speakers in southern India when compared to Pakistan, Sengupta et al. (2006) proposed an indigenous Indian origin for this language group.

Another haplogroup that is associated with the spread of agriculture from the Fertile Crescent and Anatolia regions is J2-M172 (Cinnioglu et al., 2004; Semino et al., 2004). Its sub-clade J2a-M410 is believed to have expanded eastward to Iraq, Iran and Central Asia (Sengupta et al., 2006) in parallel with the findings of painted pottery and ceramics attributed to agriculturalists (Cauvin, 2000). Similar Neolithic artifacts are also found near the Indus Valley thereby linking them with the possible spread of J2a-M410

in Pakistan. In India, haplogroup J2a-M410 is restricted to upper caste Dravidian and Indo-European speakers, with limited presence in the middle and lower castes (Sengupta et al., 2006). This sub-clade is present exclusively in Lingayat (6.93%), except for one individual from Vokkaligas, whereas its sister clade J2b-M12 occurs at equivalent proportions in both the Dravidian populations. Currently, the geographic source(s) and migration scenario for these two sub-clades remain unknown (Sengupta et al., 2006).

In addition, four J2b2-M241 Lingayat males displayed a null allele at DYS458 and failed to produce the AMGY PCR amplicon while their X homolog (AMGX) amplified successfully. Comparison of Y-STR haplotypes of the affected males from the present study with those from the literature showed high level of allele sharing (Cadenas et al., 2006), implying shared paternal lineages or a recent common ancestry for these groups of individuals. The elevated incidence of male amelogenin dropout in the Indian subcontinent (Cadenas et al., 2006; Kashyap et al., 2006) presents significant implications for forensic case work and paternal testing.

Ancient South Asian lineages

Besides its widespread distributions in South Asian populations, including India, Pakistan, Sri Lanka and Nepal (Wells et al., 2001; Kivisild et al., 2003; Sahoo et al., 2006; Sengupta et al., 2006; Thanseem et al., 2006; Gayden et al., 2007), haplogroup H-M69 is also present at high frequency among Romani groups residing in Europe (Gresham et al., 2001; Pericic et al., 2005; Gusmao et al., 2008; Klaric et al., 2009; Regueiro et al., 2011). Therefore, it is not surprising that the Kosovo Roma from Serbia maps with Indian populations in the CA plot (Figure 2) and ML (Figure 3) phylogeny. In

India, the majority of H-M69 chromosomes belong to sub-clade H1-M52 and its downstream mutation H1a-M82, which is reported at high proportions in Koya (61.0%) and the Chenchu (37.0%) tribal groups from southern India (Kivisild et al., 2003). The H1-M52 percentages in the Lingayat (20.79%) and Vokkaliga (11.76%) collections are relatively low but are equivalent to other Dravidian caste populations (Thanseem et al., 2006; Sengupta et al., 2006). It is interesting to note that Vokkaliga shows H2-Apt derived samples (7.84%), which is previously observed in Koya (9.8%) and Tharus (11.7%) from Nepal (Kivisild et al., 2003; Fornarino et al., 2009).

It is noteworthy that the most common haplotype representing the Roma cluster in the H1-M52 network (Figure 6) matches the 7 Y-STR loci Romani Modal Haplotype (DYS19=15, DYS389I=14, DYS389II=16, DYS390=22, DYS391=10, DYS392=11, DYS393=12) (Gusmao et al., 2008; Rugeiro et al., 2011). Interestingly, this haplotype differs from the other two South Asian representative haplotypes in the network (Figure 6) by only one mutational step at DYS389I (13), suggesting their shared common ancestry (Gusmao et al., 2008; Rugeiro et al., 2011) and recent dispersal of Roma from the Indian subcontinent (Fraser 1992). The considerable degree of haplotype diversity presented by south Indian H1-M52 chromosomes confirms their antiquity (Table 4) and possibly reflects an ancestral homeland for this marker (Sengupta et al., 2006).

In addition to H*-M69, haplogroup F*-M213 is also thought to have originated in India and together account for ~25.0%-31.0% of the Indian Y-chromosomes (Cordaux et al., 2004; Sengupta et al., 2006). The paraphyletic F* is more frequently observed in tribal populations than in the caste groups (Cordaux et al., 2004) but in the present study it represents 13.73% of the Vokkaliga paternal gene pool. The age of five F* Lingayat

males (24.64 kya) is twice that of the Vokkaligas which is most likely inflated given the limited sample size. The paragroup F* shares a deep phylogenetic node with C-M216, whose lineages are also present in two Dravidian populations. The presence of one C*-M216 individual in Vokkaliga may represent the genetic signature left by the original out of Africa migrants via the southern coastal route some 60,000 to 75,000 years ago.

Conclusion

In summary, the data from the present study revealed the prevalence of indigenous Indian Y-haplogroups (H, L, F* and R2) that constitute three-fourths of the Lingayat and Vokkaliga paternal gene pool. Although, there is no distinction in the presence or absence of these major clades, the frequency varies between the two Dravidian populations. For example, haplogroup L-M20 predominates in Vokkaliga (32.35%) whereas R-M207 is more common in Lingayat (35.64%), reflecting the differential genetic impact these two populations experienced. The occurrence of relatively high frequency of R1a1a-M198 in both the south Indian collections along with the ancient average age of 13.15 kya is incompatible with the theory that the Indo-European speakers contributed this marker to the Indian gene pool. The genetic data support the historical fact that Lingayat was formed from multiple source populations. The presence of ancient lineages such as F*-M213, H*-M69 and C*-M216 suggest that modern humans reached the Indian subcontinent very early after their migration out of Africa.

The overall haplotype diversity at both the minimal 9-loci (0.9915) and 17-loci (0.9970) Y-STR systems indicate that the data from this study provide a high power of

discrimination that is useful for forensic applications. In addition, there are no haplotypes shared between Lingayat and Vokkaliga populations at the 17 Y-STR loci level, underscoring the genetic uniqueness of these two Dravidian collections as well as the need for independent databases for forensic analysis and paternity testing. The presence of four Lingayat males with the deletion of amelogenin Y-chromosome locus highlights the importance of typing alternative and/or additional loci for gender determination in forensic casework, especially in the Indian subcontinent where such cases are prevalent.

REFERENCES

- Abu-Amero KK, Hellani A, González AM, Larruga JM, Cabrera VM, Underhill PA (2009) Saudi Arabian Y-chromosome diversity and its relationship with nearby regions. *BMC Genet* 10:59.
- Al-Zahery N, Semino O, Benuzzi G, Magri C, Passarino G, Torroni A, Santachiara-Benerecetti AS (2003) Y-chromosome and mtDNA polymorphisms in Iraq, a crossroad of the early human dispersal and of post-Neolithic migrations. *Mol Phylogenet Evol* 28:458-472
- Antunez-de-Mayolo G, Antunez-de-Mayolo A, Antunez-de-Mayolo P, Papiha SS, Hammer M, Yunis JJ, Yunis EJ, Damodaran C, Martinez de Pancorbo M, Caeiro JL, Puzyrev VP, Herrera RJ (2002) Phylogenetics of worldwide human populations as determined by polymorphic Alu insertions. *Electrophoresis* 19:3346-3356
- Applied Biosystems (2006) AmpFISTR Yfiler PCR amplification kit user's manual. Applied Biosystems, Foster City
- Bamshad M, Kivisild T, Watkins WS, Dixon ME, Ricker CE, Rao BB (2001) Genetic evidence on the origins of Indian caste populations. *Genome Res* 6:994-1004
- Basu A, Mukherjee N, Roy S, Sengupta S, Banerjee S, Chakraborty M, Dey B, Roy M, Roy B, Bhattacharyya NP, Roychoudhury S, Majumder PP (2003) Ethnic India: a genomic view, with special reference to peopling and structure. *Genome Res* 13:2277-2290
- Battaglia V, Fornarino S, Al-Zahery N, Olivieri A, Pala M, Myres NM, King RJ, Rootsi S, Marjanovic D, Primorac D, Hadziselimovic R, Vidovic S, Drobic K, Durmishi N, Torroni A, Santachiara-Benerecetti AS, Underhill PA, Semino O (2008) Y-chromosomal evidence of the cultural diffusion of agriculture in Southeast Europe. *Eur J Hum Genet* 17:820-830
- Cadenas AM, Regueiro M, Gayden T, Singh N, Zhivotovsky LA, Underhill PA, Herrera RJ (2006) Male amelogenin dropouts: phylogenetic context, origins and implications. *Forensic Sci Int* 166:155-163
- Cadenas AM, Zhivotovsky LA, Cavalli-Sforza LL, Underhill PA, Herrera RJ (2007) Y-chromosome diversity characterizes the Gulf of Oman. *Eur J Hum Genet* 16:374-386
- Cann RL (2001) Genetic clues to dispersal in human populations: retracing the past from the present. *Science* 291:1742-1748

- Cauvin J (2000) *The birth of the gods and origins of agriculture*. Cambridge University Press, Cambridge, United Kingdom
- Chandrasekar A, Kumar S, Sreenath J, Sarkar BN, Urade BP, Mallick S, Bandopadhyay SS, Barua P, Barik SS, Basu D, Kiran U, Gangopadhyay P, Sahani R, Prasad BV, Gangopadhyay S, Lakshmi GR, Ravuri RR, Padmaja K, Venugopal PN, Sharma MB,
- Rao VR (2009) Updating phylogeny of mitochondrial DNA macrohaplogroup m in India: dispersal of modern human in South Asian corridor. *PLoS One* 10:e7447.
- Chang YM, Burgoyne LA, Both K (2003) Higher failures of amelogenin sex test in an Indian population group. *J Forensic Sci* 48:1309-1313
- Cinnioglu C, King R, Kivisild T, Kalfoglu E, Atasoy S, Cavalleri GL, Lillie AS, Roseman CC, Lin AA, Prince K, Oefner PJ, Shen P, Semino O, Cavalli-Sforza LL, Underhill PA (2004) Excavating Y-chromosome haplotype strata in Anatolia. *Hum Genet* 114:127-148
- Cordaux R, Aunger R, Bentley G, Nasidze I, Sirajuddin SM, Stoneking M (2004) Independent origins of Indian caste and tribal paternal lineages. *Curr Biol* 14:231-235
- Census of India (2001) Government of India. Retrieved 24 September 2010 from http://censusindia.gov.in/Census_Data_2001/India_at_glance/religion.aspx
- Excoffier, Laval LG, Schneider S (2005) Arlequin ver. 3.0: An integrated software package for population genetics data analysis. *Evol Bioinform Online* 1:47-50
- Fechner A, Quinque D, Rychkov S, Morozowa I, Naumova O, Schneider Y, Willuweit S, Zhukova O, Roewer L, Stoneking M, Nasidze I (2008) Boundaries and clines in the West Eurasian Y-chromosome landscape: insights from the European part of Russia. *Am J Phys Anthropol* 137:41-47
- Felsenstein J (2002) *Phylogeny inference package (PHYLIP)*, version 3.6a3. Seattle: Department of Genetics, University of Washington.
- Fornarino S, Pala M, Battaglia V, Maranta R, Achilli A, Modiano G, Torroni A, Semino O, Santachiara-Benerecetti SA (2009) Mitochondrial and Y-chromosome diversity of the Tharus (Nepal): a reservoir of genetic variation. *BMC Evol Biol* 9:154
- Fraser A (1992) *The Gypsies*. Oxford: Blackwell Publishers
- Gayden T, Cadenas AM, Regueiro M, Singh NB, Zhivotovsky LA, Underhill PA, Cavalli-Sforza LL, Herrera RJ (2007) The Himalayas as a directional barrier to gene flow. *Am J Hum Genet* 80:884-894

- Gayden T, Regueiro M, Martinez L, Cadenas AM, Herrera RJ (2008) Human Y-chromosome haplotyping by allele-specific polymerase chain reaction. *Electrophoresis* 29:2419-2423
- Gayden T, Mirabal S, Cadenas AM, Lacau H, Simms TM, Morlote D, Chennakrishnaiah S, Herrera RJ (2009) Genetic insights into the origins of Tibeto-Burman populations in the Himalayas. *J Hum Genet* 54:216-223
- Gresham D, Morar B, Underhill PA, Passarino G, Lin AA, Wise C, Angelicheva D, Calafell F, Oefner PJ, Shen P, Tournev I, de Pablo R, Kucinskaskas V, Perez-Lezaun A, Marushiakova E, Popov V, Kalaydjieva L (2001) Origins and divergence of the Roma (Gypsies). *Am J Hum Genet* 69:1314-1331
- Gusmao A, Gusmao L, Gomes V, Alves C, Calafell F, Amorim A, Prata MJ (2008) A perspective on the history of the Iberian Gypsies provided by phylogeographic analysis of Y-chromosome lineages. *Ann Hum Genet* 72:215-227
- Jobling MA, Tyler-Smith C (2003) The human Y chromosome: an evolutionary marker comes of age. *Nat Rev Genet* 4:598-612
- Karafet TM, Mendez FL, Meilerman MB, Underhill PA, Zegura SL and Hammer MF (2008) New binary polymorphisms reshape and increase resolution of the human Y chromosomal haplogroup tree. *Genome Res* 18:830-838
- Kashyap VK, Sahoo S, Sitalaximi T, Trivedi R (2006) Deletions in the Y-derived amelogenin gene fragment in the Indian population. *BMC Med Genet* 7:37
- Kayser M, Caglia A, Corach D, Fretwell N, Gehrig C, Graziosi G, et al., (1997) Evaluation of Y-chromosomal STRs: a multicenter study. *Int J Legal Med* 110:125-133
- Kennedy KA, Deraniyagala SU, Roertgen WJ, Chiment J, Disotell T (1987) Upper pleistocene fossil hominids from Sri Lanka. *Am J Phys Anthropol* 72:441-461
- Kivisild T, Rootsi S, Metspalu M, Mastana S, Kaldma K, Parik J, Metspalu E, Adojaan M, Tolk HV, Stepanov V, Gölge M, Usanga E, Papiha SS, Cinnioglu C, King R, Cavalli-Sforza L, Underhill PA, Villems R (2003) The genetic heritage of the earliest settlers persists both in Indian tribal and caste populations. *Am J Hum Genet* 72:313-32
- Klaric IM, Salihovic MP, Lauc LB, Zhivotovsky LA, Rootsi S, Janic ijevic B (2009) Dissecting the molecular architecture and origin of Bayash Romani patrilineages: genetic influences from South-Asia and the Balkans. *Am J Phys Anthropol* 138:333-342
- Luis JR, Rowold DJ, Regueiro M, Caeiro B, Cinnioglu C, Roseman C, Underhill PA, Cavalli-Sforza LL, Herrera RJ (2004) The Levant versus the Horn of Africa: evidence for bidirectional corridors of human migrations. *Am J Hum Genet* 74:532-544

Macaulay V, Hill C, Achilli A, Rengo C, Clarke D, Meehan W, Blackburn J, Semino O, Scozzari R, Cruciani F, Taha A, Shaari NK, Raja JM, Ismail P, Zainuddin Z, Goodwin W, Bulbeck D, Bandelt HJ, Oppenheimer S, Torroni A, Richards M (2005) Single, rapid coastal settlement of Asia revealed by analysis of complete mitochondrial genomes. *Science* 308:1034-1036.

Majumder PP (2008) Genomic inferences on peopling of south Asia. *Curr Opin Genet Dev.* 18:280-284

Martinez L, Underhill PA, Zhivotovsky LA, Gayden T, Moschonas NK, Chow CE, Conti S, Mamolini E, Cavalli-Sforza LL, Herrera RJ (2007) Paleolithic Y-haplogroup heritage predominates in a Cretan highland plateau. *Eur J Hum Genet* 15:485-493

Mirabal S, Regueiro M, Cadenas AM, Cavalli-Sforza LL, Underhill PA, Verbenko DA, Limborska SA, Herrera RJ (2009) Y-chromosome distribution within the geo-linguistic landscape of northwestern Russia. *Eur J Hum Genet* 17:1260-1273

Morlote DM, Gayden T, Arvind P, Babu A, Herrera RJ (2011) The Soliga, an isolated tribe from Southern India: genetic diversity and phylogenetic affinities. *J Hum Genet* 56:258-269

Myres NM, Rootsi S, Lin AA, Järve M, King RJ, Kutuev I, Cabrera VM, Khusnutdinova EK, Pshenichnov A, Yunusbayev B, Balanovsky O, Balanovska E, Rudan P, Baldovic M, Herrera RJ, Chiaroni J, Di Cristofaro J, Villems R, Kivisild T, Underhill PA (2011) A major Y-chromosome haplogroup R1b Holocene era founder effect in Central and Western Europe. *Eur J Hum Genet* 1:95-101

Nasidze I, Sarkisian T, Kerimov A, Stoneking M (2003) Testing hypotheses of language replacement in the Caucasus: evidence from the Y-chromosome. *Hum Genet* 112:255-261

Petraglia MD, Haslam M, Fuller DQ, Boivin N, Clarkson C (2010) Out of Africa: new hypotheses and evidence for the dispersal of *Homo sapiens* along the Indian Ocean rim. *Ann Hum Biol* 37:288-311

Qamar R, Ayub Q, Mohyuddin A, Helgason A, Mazhar K, Mansoor A, Zerjal T, Tyler-Smith C, Mehdi SQ (2002) Y-chromosomal DNA variation in Pakistan. *Am J Hum Genet* 70:1107-1124

Quintana-Murci L, Semino O, Bandelt HJ, Passarino G, McElreavey K, Santachiara-Benerecetti AS (1999) Genetic evidence of an early exit of *Homo sapiens sapiens* from Africa through eastern Africa. *Nat Genet* 23:437-441

- Ramana GV, Su B, Jin L, Singh L, Wang N, Underhill P, Chakraborty R (2001) Y-chromosome SNP haplotypes suggest evidence of gene flow among caste, tribe, and the migrant Siddi populations of Andhra Pradesh, South India. *Eur J Hum Genet* 9:695-700
- Regueiro M, Cadenas AM, Gayden T, Underhill PA, Herrera RJ (2006) Iran: tricontinental nexus for Y-chromosome driven migration. *Hum Hered* 61:132-143
- Regueiro M, Stanojevic A, Chennakrishnaiah S, Rivera L, Varljen T, Alempijevic D, Stojkovic O, Simms T, Gayden T, Herrera RJ (2011) Divergent patrilineal signals in three Roma populations. *Am J Phys Anthropol* 144:80-91
- Reich D, Thangaraj K, Patterson N, Price AL, Singh L (2010) Reconstructing Indian population history. *Nature* 461:489-94
- Renfrew C (1992) Archaeology, genetics and linguistic diversity. *Man* 27:445-478
- Rajkumar R, Kashyap VK (2004) Genetic structure of four socio-culturally diversified caste populations of southwest India and their affinity with related Indian and global groups. *BMC Genet* 5:23
- Rohlf F (2002) NTSTSSpc. Exter Publishing, Setauket, NY
- Sahoo S and Kashyap VK (2006a) Phylogeography of mitochondrial DNA and Y-chromosome haplogroups reveal asymmetric gene flow in populations of Eastern India. *Am J Phys Anthropol* 131:84-97
- Sahoo S, Singh A, Himabindu G et al., (2006b) A prehistory of Indian Y chromosomes: evaluating demic diffusion scenarios. *Proc Natl Acad Sci* 103:843-848
- Sarkar SS (1958) Race and race movements in India, In: The Cultural Heritage of India. S.K.Chatterjee (Ed.) Ramakrishna Mission Institute of Culture, Calcutta.
- Semino O, Passarino G, Oefner PJ, Lin AA, Arbuzova S, Beckman LE, De Benedictis G, Francalacci P, Kouvatsi A, Limborska S, Marcikiae M, Mika A, Mika B, Primorac D, Santachiara-Benerecetti AS, Cavalli-Sforza LL, Underhill PA (2000) The genetic legacy of Paleolithic *Homo sapiens sapiens* in extant Europeans: a Y chromosome perspective. *Science* 290:1155-1159
- Semino O, Magri C, Benuzzi G, Lin AA, Al-Zahery N, Battaglia V, Maccioni L, Triantaphyllidis C, Shen P, Oefner PJ, Zhivotovsky LA, King R, Torroni A, Cavalli-Sforza LL, Underhill PA, Santachiara-Benerecetti AS (2003) Origin, diffusion, and differentiation of Y-chromosome haplogroups E and J: inferences on the neolithization of Europe and later migratory events in the Mediterranean area. *Am J Hum Genet* 74:1023-1034

Sengupta S, Zhivotovsky LA, King R, Mehdi SQ, Edmonds CA, Chow CE, Lin AA, Mitra M, Sil SK, Ramesh A, Usha Rani MV, Thakur CM, Cavalli-Sforza LL, Majumder PP, Underhill PA (2006) Polarity and temporality of high-resolution y-chromosome distributions in India identify both indigenous and exogenous expansions and reveal minor genetic influence of Central Asian pastoralists. *Am J Hum Genet* 78:202-221

Shepard EM and Herrera RJ (2006) Iranian STR variation at the fringes of biogeographical demarcation. *Forensic Sci Int* 158:140-148

Stringer C (2000) Palaeoanthropology: coasting out of Africa. *Nature* 405:24-27

Thangaraj K, Chaubey G, Kivisild T, Reddy AG, Singh VK, Rasalkar AA, Singh L (2005) Reconstructing the origin of Andaman Islanders. *Science* 308:996-996

Thangaraj K, Naidu BP, Crivellaro F, Tamang R, Upadhyay S, Sharma VK, Reddy AG, Walimbe SR, Chaubey G, Kivisild T, Singh L (2010) The influence of natural barriers in shaping the genetic structure of Maharashtra populations. *PLoS One* 5:e15283

Thanseem I, Thangaraj K, Chaubey G, Singh VK, Bhaskar LV, Reddy BM, Reddy AG, Singh L (2006) Genetic affinities among the lower castes and tribal groups of India: inference from Y chromosome and mitochondrial DNA. *BMC Genet* 7:42

Terreros MC, Rowold D, Luis JR, Khan F, Agrawal S, Herrera RJ (2007) North Indian Muslims: enclaves of foreign DNA or Hindu converts? *Am J Phys Anthropol* 133:1004-1012

Terreros MC, Rowold DJ, Mirabal S, Herrera RJ (2011) Mitochondrial DNA and Y-chromosomal stratification in Iran: relationship between Iran and the Arabian Peninsula. *J Hum Genet* 56:235-246

Watkins WS, Thara R, Mowry BJ, Zhang Y, Witherspoon DJ, Tolpinrud W, Bamshad MJ, Tirupati S, Padmavati R, Smith H, Nancarrow D, Filippich C, Jorde LB. (2008) Genetic variation in South Indian castes: evidence from Y-chromosome, mitochondrial, and autosomal polymorphisms. *BMC Genet* 9:86

Wells RS, Yuldasheva N, Ruzibakiev R, Underhill PA, Evseeva I, Blue-Smith J, Jin L, Su B, Pitchappan R, Shanmugalakshmi S, Balakrishnan K, Read M, Pearson NM, Zerjal T, Webster MT, Zholoshvili I, Jamarjashvili E, Gambarov S, Nikbin B, Dostiev A, Aknazarov O, Zalloua P, Tsoy I, Kitaev M, Mirrakhimov M, Chariev A, Bodmer WF (2001) The Eurasian heartland: a continental perspective on Y-chromosome diversity. *Proc Natl Acad Sci U SA* 98:10244-10249.

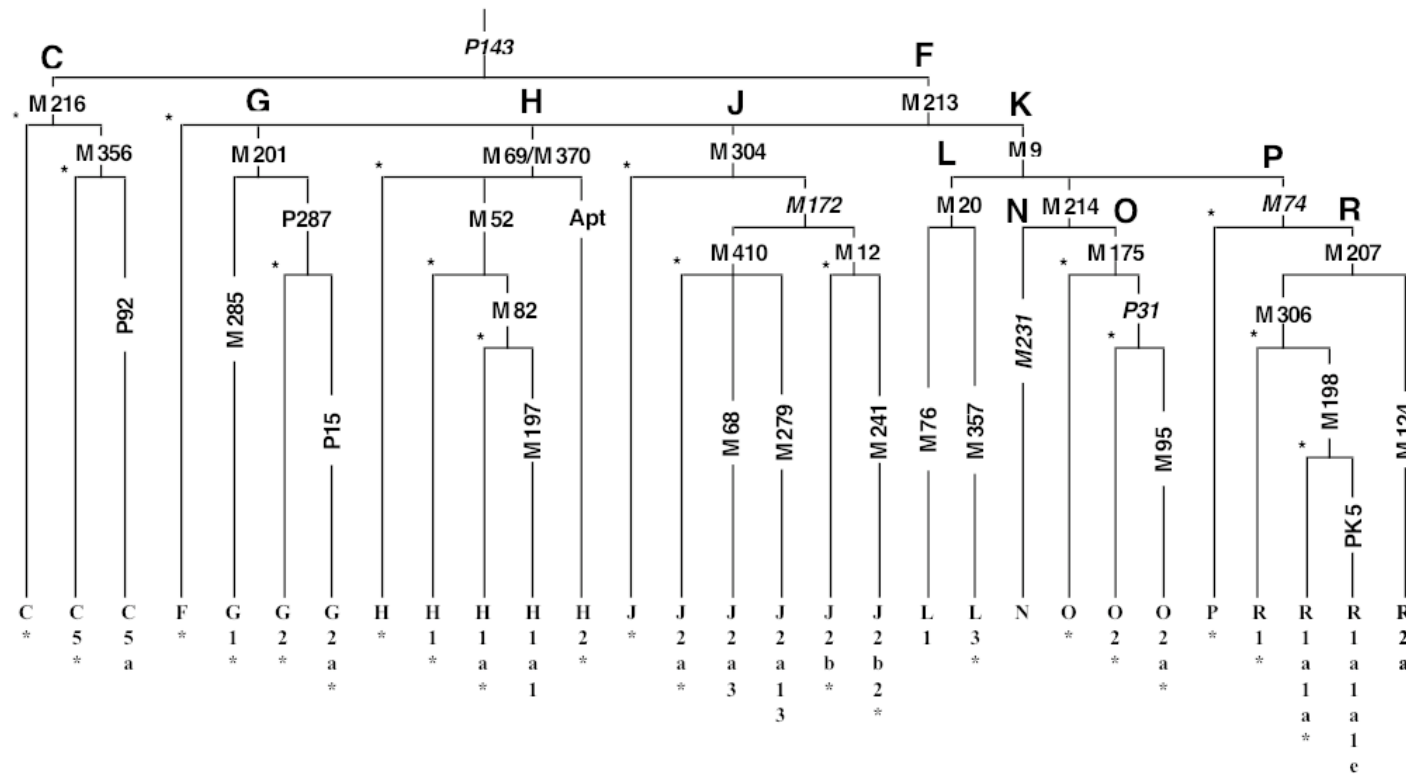
Y Chromosome Consortium (2002) A nomenclature system for the tree of human Y chromosomal binary haplogroups. *Genome Research* 12:339-348

Zerjal T, Wells RS, Yuldasheva N, Ruzibakiev R, Tyler-Smith C (2002) A genetic landscape reshaped by recent events: Y-chromosomal insights into central Asia. *Am J Hum Genet* 71:466-482

Zerjal T, Pandya A, Thangaraj K, Ling EY, Kearley J, Bertoneri S, Paracchini S, Singh L, Tyler-Smith C (2007) Y-chromosomal insights into the genetic impact of the caste system in India. *Hum Genet* 121:137-144

Zhivotovsky LA, Underhill PA, Cinnioğlu C, Kayser M, Morar B, Kivisild T, Scozzari R, Cruciani F, Destro-Bisol G, Spedini G, Chambers GK, Herrera RJ, Yong KK, Gresham D, Tournev I, Feldman MW, Kalaydjieva L. (2004) The effective mutation rate at Y chromosome short tandem repeats, with application to human population-divergence time. *Am J Hum Genet* 74:50-61.

Figure 1. The phylogenetic relationships of the total 21 Y-chromosome haplogroups and their observed frequencies in the Lingayat and Vokkaliga populations.



LINGAYAT		3	5	1	1	3	3	18	1		4	2	1	1	4	18					20		16
N = 101		0.03	0.05	0.01	0.01	0.03	0.03	0.18	0.01		0.04	0.02	0.01	0.01	0.04	0.18					0.20		0.16
VOKKALIGA	1	2	14		1	3		12	1	8		1			3	29	4			1	7	1	14
N = 102	0.01	0.02	0.14		0.01	0.03		0.12	0.01	0.08		0.01			0.03	0.28	0.04			0.01	0.07	0.01	0.14

Figure 2. Correspondence Analysis based on the Y-SNP haplogroups frequency data of 35 populations listed in Table 1.

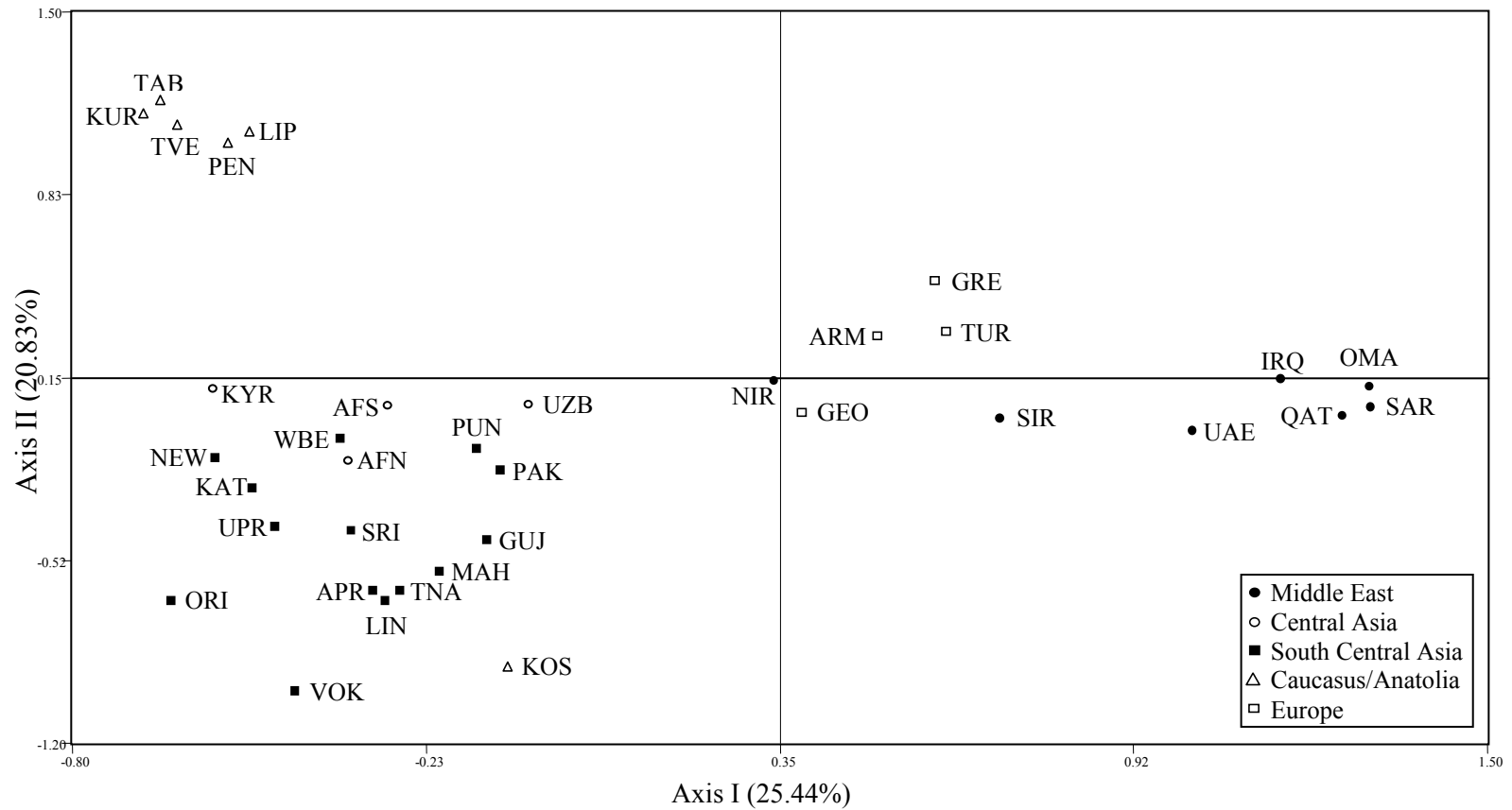


Figure 3. Maximum Likelihood Tree based on the Y-SNP haplogroups frequency data of 35 populations listed in Table 1.

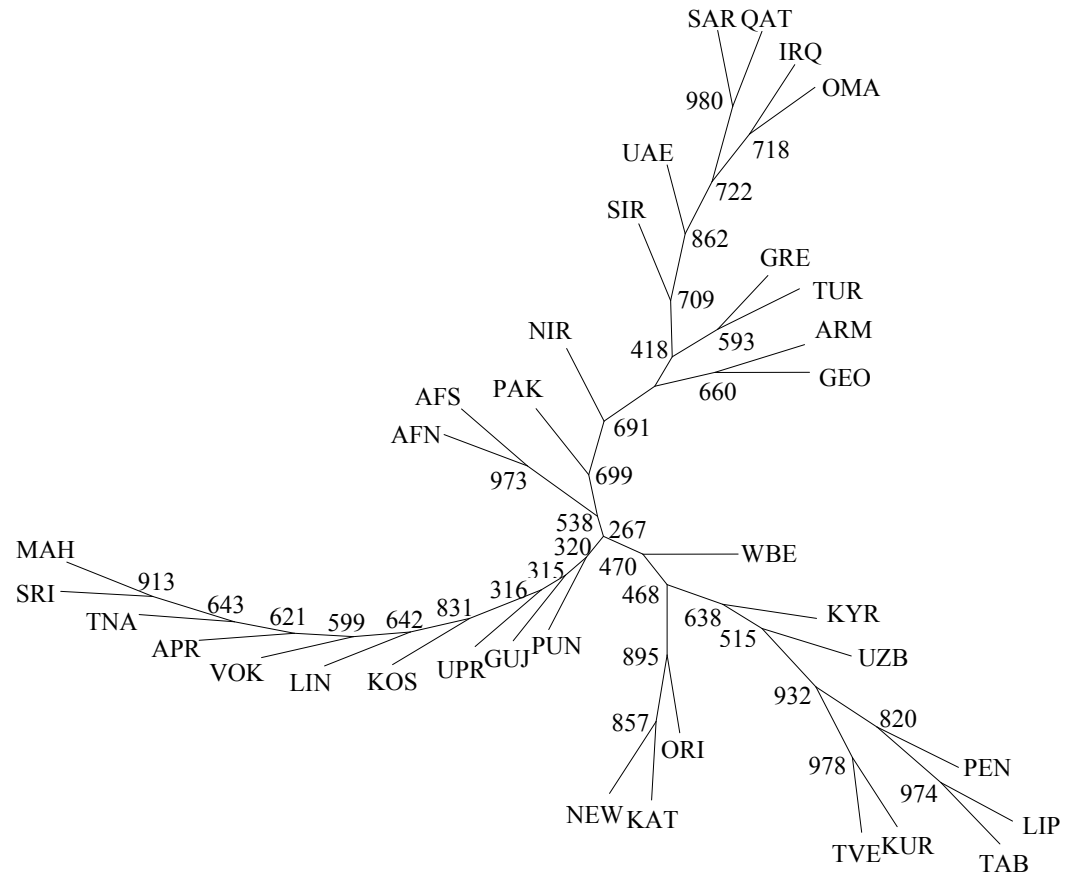


Figure 4. A Median-Joining Network based on 8 Y-STR loci within the R1a1a-M198 haplogroup. The area of the circles is proportional to the haplotype frequency and the smallest circle corresponds to two individuals.

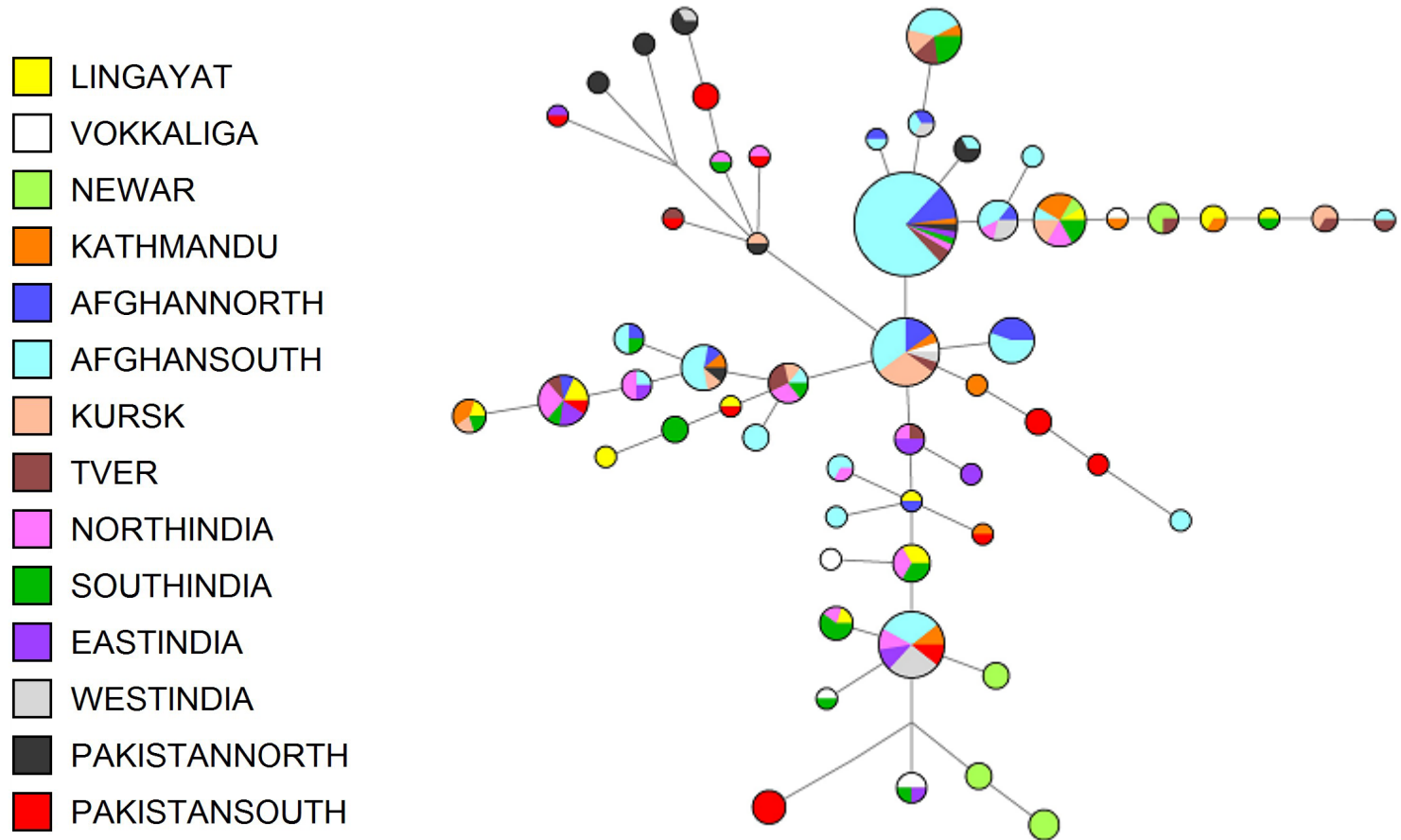


Figure 5. A Median-Joining Network based on 8 Y-STR loci within the R2a-M124 haplogroup. The area of the circles is proportional to the haplotype frequency and the smallest circle corresponds to two individuals.

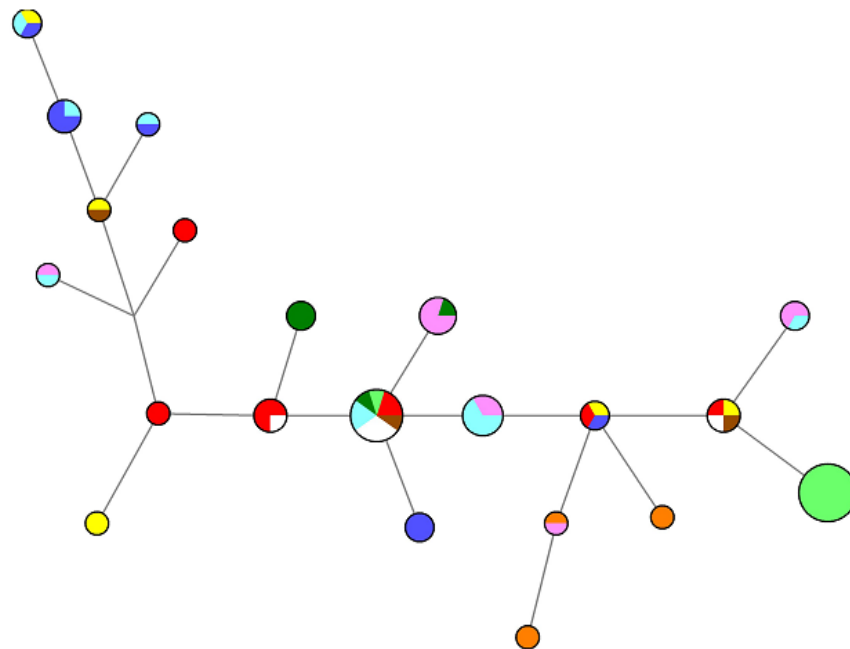


Figure 6. A Median-Joining Network based on 8 Y-STR loci within the H1-M52 haplogroup. The area of the circles is proportional to the haplotype frequency and the smallest circle corresponds to two individuals.

- LINGAYAT
- VOKKALIGA
- NEWAR
- KATHMANDU
- ROMA
- AFGHANSOUTH
- NORTHINDIA
- SOUTHINDIA
- EASTINDIA
- WESTINDIA
- PAKISTANNORTH

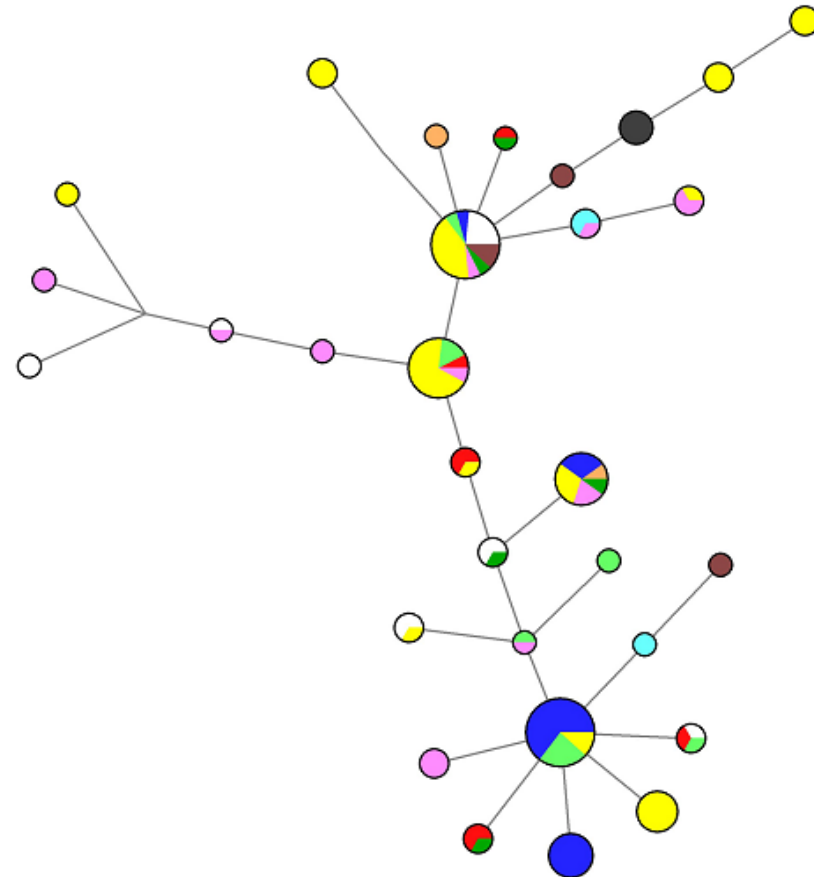


Figure 7. A Median-Joining Network based on 8 Y-STR loci within the L1-M76 haplogroup. The area of the circles is proportional to the haplotype frequency and the smallest circle corresponds to one individual.

- LINGAYAT
- VOKKALIGA
- SOUTHINDIA
- WESTINDIA
- PAKISTANSOUTH

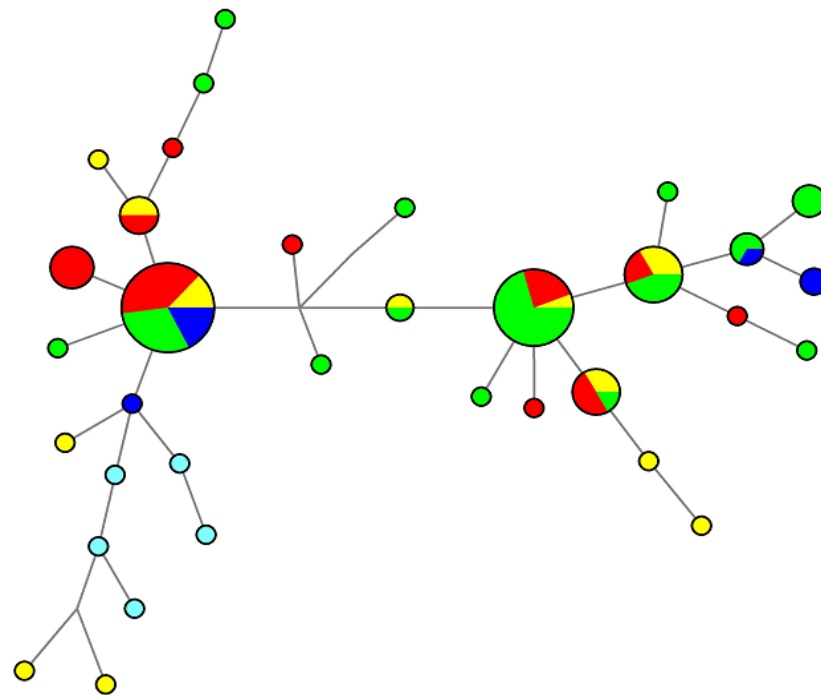


Table 1. Populations analyzed in this study.

Population	Abbreviation	Linguistic Affiliation	N	References
MIDDLE EAST				
Northern Iran	NIR	Indo-European	33	Regueiro et al., 2006
Southern Iran	SIR	Indo-European	117	Regueiro et al., 2006
Iraq	IRQ	Afro-Asiatic	139	Al-Zahery et al 2003
Oman	OMA	Afro-Asiatic	121	Luis et al., 2004
Qatar	QAT	Afro-Asiatic	72	Cadenas et al., 2008
Saudi Arabia	SAR	Afro-Asiatic	157	Abu-Amerno et al., 2009
UAE	UAE	Afro-Asiatic	164	Cadenas et al., 2008
CENTRAL ASIA				
Afghan north	AFN	Indo-European	44	<i>Lacau et al., (unpublished)</i>
Afghan south	AFS	Indo-European	146	<i>Lacau et al., (unpublished)</i>
Kyrgyztan	KYR	Altaic	52	Wells et al., 2001
Uzbekistan	UZB	Altaic	70	Wells et al., 2001
SOUTH CENTRAL ASIA				
Lingayat	LIN	Dravidian	101	<i>Present Study</i>
Vokkaliga	VOK	Dravidian	102	<i>Present Study</i>
Andra Pradesh	APR	Dravidian	167	Watkins et al., 2009
Gujarat	GUJ	Indo-European	29	Kivisild et al., 2003
Maharastra	MAH	Indo-European	50	Thangaraj et al., 2010
Orrisa	ORI	Indo-European	23	Sahoo and Kashyap, 2006a
Punjab	PUN	Indo-European	66	Kivisild et al., 2003
Tamil Nadu	TNA	Dravidian	118	Watkins et al., 2008
Uttar Pradeah	UPR	Indo-European	141	Zerjal et al., 2007
West Bengal	WBE	Indo-European	31	Kivisild et al., 2003
Katmandu	KAT	Indo-European	77	Gayden et al., 2007
Newar	NEW	Sino-Tibetan	66	Gayden et al., 2007
Pakistan	PAK	Indo-European	176	Sengupta et al., 2006
Sri Lanka	SRI	Indo-European	39	Kivisild et al., 2003
CAUCASUS/ANATOLIA				
Armenia	ARM	Indo-European	413	<i>Herrera et al., (unpublished)</i>
Georgia	GEO	Kartvelian	77	Nasidze et al., 2003
Turkey	TUR	Altaic	523	Cinnioglu et al., 2004
Greece	GRE	Indo-European	92	Battaglia et al., 2009
EUROPE				
Lipezkaja (Russia)	LIP	Indo-European	47	Fechner et al., 2008
Penzenskaja (Russia)	PEN	Indo-European	81	Fechner et al., 2008
Tambovskaya	TAB	Indo-European	48	Fechner et al., 2008
Kursk (Russia)	KUR	Indo-European	40	Mirabal et al., 2009
Tver (Russia)	TVE	Indo-European	38	Mirabal et al., 2009
Kosovo (Roma)	KOS	Indo-European	42	Regueiro et al., 2010

Table 2. Analysis of molecular variance (AMOVA) results based on major Y-haplogroup frequencies.

Group	Number of groups	Within populations		Among populations within groups		Among groups	
		Variation (%)	Φ_{ST}	Variation (%)	Φ_{SC}	Variation (%)	Φ_{CT}
Language ^a	6	86.47	0.14*	6.39	0.07*	7.14	0.07*
Geography ^b	5	86.35	0.14*	4.24	0.05*	9.41	0.09*

* P -value <0.00001.

^a Linguistic partitioning (6 groups): 1-Afro-Asiatic (IRQ, OMA, QAT, SAR, UAE); 2-Altaic (KYR, UZB, TUR); 3-Dravidian (LIN, VOK, APR, TNA); 4-Indo-European (NIR, SIR, AFN, AFS, GUJ, MAH, ORI, PUN, UPR, WBE, KAT, PAK, SRI, ARM, GRE, LIP, PEN, TAB, KUR, TVE, KOS); 5-Kartvelian (GEO); 6-Sino-Tibetan (NEW).

^b Geographical partitioning (5 groups): 1-Middle East (NIR, SIR, IRQ, OMA, QAT, SAR, UAE); 2-Central Asia (AFN, AFS, KYR, UZB); 3-South Central Asia (LIN, VOK, APR, GUJ, MAH, ORI, PUN, TNA, UPR, WBE, KAT, NEW, PAK, SRI); 4-Caucasus/Anatolia (ARM, GEO, TUR, GRE); 5-Europe (LIP, PEN, TAB, KUR, TVE, KOS).

Table 3. Forensic parameters of the Lingayat and Vokkaliga populations using the minimal 9-loci and the 17-loci Y-STR haplotypes.

Haplotypes	Lingayat	Vokkaliga	All Populations
Minimal 9 loci Y-STR Haplotype			
Sample Size	101	102	203
Number of different Haplotypes	81	71	147
Number of Unique Haplotypes	68	59	120
Discrimination Capacity	0.8020	0.6961	0.7241
Haplotype Diversity \pm SD	0.9917 \pm 0.0041	0.9814 \pm 0.0064	0.9915 \pm 0.0024
Y-filer 17 loci Y-STR Haplotype			
Sample Size	97*	102	199
Number of different Haplotypes	89	80	169
Number of Unique Haplotypes	82	69	151
Discrimination Capacity	0.9175	0.7843	0.8492
Haplotype Diversity \pm SD	0.9981 \pm 0.0019	0.9901 \pm 0.0043	0.9970 \pm 0.0013

*Excludes 4 Lingayat samples that carry null allele at DYS458

Table 4. Y-chromosome haplogroup variance and time estimates based on 15 Y-STR loci using Zhivotovsky et al. (2004) method.

Haplogroups	N	Evolutionary mutation rate (kya \pm sd)	Mean variance
L1-M76			
Lingayat	18	11.137 \pm 3.142	0.269
Vokkaliga	29	9.578 \pm 2.605	0.232
All Population	47	10.484 \pm 2.804	0.250
R1a1a*-M198			
Lingayat	20	12.801 \pm 2.752	0.289
Vokkaliga	7	11.732 \pm 3.092	0.295
All Population	27	13.151 \pm 2.593	0.287
R2-M124			
Lingayat	16	31.250 \pm 6.409	0.820
Vokkaliga	14	14.924 \pm 3.574	0.387
All Population	30	24.960 \pm 4.554	0.614
H1-M52			
Lingayat	22	16.689 \pm 3.597	0.406
Vokkaliga	14	17.512 \pm 4.430	0.489
All Population	36	17.244 \pm 3.268	0.441
F*- M213			
Lingayat	5	24.638 \pm 6.357	0.693
Vokkaliga	14	11.905 \pm 4.255	0.299
All Population	19	23.773 \pm 8.318	0.521

Supplementary Table 1. The 17 loci Y-STR haplotype data of the Lingayat (N=101) and Vokkaliga (N=102) populations of southern India.

Sample	DYS456	DYS389I	DYS390	DYS389II	DYS458	DYS19	DYS385a/b	DYS393	DYS391	DYS439	DYS635	DYS392	GATA_H4	DYS437	DYS438	DYS448
Lin1	15	14	25	31	19	15	11-13	13	10	10	24	11	13	14	11	20
Lin2	16	13	22	29	16	17	14-14	12	10	10	21	11	11	14	10	19
Lin3	15	14	25	31	16	17	11-14	13	11	10	23	11	13	14	11	20
Lin4	14	13	24	30	19	14	12-15	12	10	12	21	11	11	14	9	19
Lin5	15	15	23	31	17	15	16-17	12	10	11	20	11	11	14	9	19
Lin6	15	14	23	30	17	15	15-19	13	10	11	27	10	11	16	11	19
Lin7	16	14	23	31	17	15	13-19	12	11	10	21	11	11	14	10	19
Lin9	16	13	23	29	17	14	11-18	14	11	10	24	10	13	16	9	19
Lin10	14	14	23	31	16	10	14-19	13	10	11	21	11	11	14	9	20
Lin11	14	14	22	30	14	16	15-18	13	10	13	21	11	13	14	10	17
Lin12	16	14	23	30	18	14	13-19	14	10	11	25	10	13	16	11	19
Lin13	17	12	22	30	14	14	13-13	11	10	14	23	14	12	15	10	19
Lin14	16	12	22	29	14	14	13-16	11	10	12	23	14	12	15	10	19
Lin15	17	13	22	29	18	15	15-18	12	10	11	20	11	12	14	9	18
Lin16	15	14	22	30	14	17	14-16	13	11	11	23	11	12	14	10	18
Lin17	15	13	22	29	17	15	16-17	12	10	13	20	11	11	15	9	19
Lin18	15	14	22	30	17	16	15-17	12	10	11	20	11	12	14	9	19
Lin19	15	13	23	28	18	14	12-17	14	10	11	24	10	12	15	11	18
Lin20	16	13	23	29	17	15	16-17	12	10	12	20	11	12	15	9	19
Lin21	15	14	20	31	15	14	14-17	14	10	13	23	11	13	14	10	19
Lin22	15	13	22	28	17	17	15-19	12	10	12	20	11	13	15	10	19
Lin23	14	14	23	29	19	14	14-19	12	10	12	23	11	11	14	9	21
Lin24	10	13	24	29	16	14	13-19	8	10	10	24	10	12	16	11	19
Lin25	14	12	22	27	19	14	14-21	14	11	12	24	10	12	16	11	18
Lin26	15	14	23	30	16	14	13-16	13	10	11	24	11	11	15	9	21
Lin27	15	14	20	31	15	14	14-17	14	10	12	23	11	12	14	10	19
Lin28	15	14	23	30	15	14	14-19	13	10	11	22	11	11	14	9	20
Lin29	15	13	24	29	17	14	12-18	15	10	10	27	10	12	16	11	20
Lin30	16	12	22	28	14	14	13-16	11	10	12	24	14	12	15	10	19
Lin31	15	14	25	31	16	15	11-14	13	10	10	24	11	12	14	11	20

Sample	DYS456	DYS389I	DYS390	DYS389II	DYS458	DYS19	DYS385a/b	DYS393	DYS391	DYS439	DYS635	DYS392	GATA_H4	DYS437	DYS438	DYS448
Lin32	16	13	23	29	17	14	12-18	14	10	11	25	10	11	16	11	19
Lin33	15	12	22	28	15	15	13-17	11	10	12	23	14	12	15	10	19
Lin34	13	13	24	30	16	16	13-19	12	10	12	21	11	11	15	9	18
Lin35	16	14	22	30	15	17	16-19	12	10	11	20	11	11	14	9	18
Lin36	15	15	23	31	15	16	15-17	13	10	11	20	11	11	14	9	19
Lin37	16	13	22	29	15	15	15-17	12	10	11	22	11	11	14	9	19
Lin38	15	14	23	30	16	14	13-16	13	10	11	24	11	11	15	9	21
Lin39	14	13	24	30	17	16	12-17	12	10	12	22	11	12	14	10	19
Lin40	15	12	24	29	17	17	11-13	13	11	12	23	11	13	14	11	20
Lin41	15	12	24	28	17	14	13-20	13	10	12	24	10	12	16	11	18
Lin42	16	13	22	29	17	15	15-15	12	11	11	19	11	12	14	9	19
Lin44	15	12	22	31	16	15	15-15	14	11	11	23	11	12	16	10	21
Lin45	15	13	22	30	14	14	13-17	11	10	13	25	14	11	15	11	19
Lin46	15	14	23	30	19	15	13-21	14	10	10	24	11	10	16	11	19
Lin47	15	13	23	29	17	14	13-17	12	10	12	22	11	12	15	9	21
Lin48	15	14	24	32	16	17	12-14	13	11	10	23	11	12	14	12	20
Lin49	14	10	25	27	18	15	12-14	13	10	12	24	11	12	14	9	20
Lin50	15	12	22	28	14	14	13-17	11	10	14	24	14	12	16	10	19
Lin51	16	13	25	32	15	15	12-14	13	11	11	23	11	13	14	11	20
Lin52	12	12	24	28	Null	16	13-16	12	10	12	21	11	11	15	9	18
Lin146	14	13	23	31	17	15	13-17	15	10	12	25	10	12	16	11	19
Lin147	16	13	22	29	19	15	14-16	12	10	11	19	11	12	14	9	19
Lin148	15	13	25	31	15	15	11-14	13	11	10	23	11	13	14	11	20
Lin149	16	13	25	31	15	15	13-14	13	10	12	23	11	13	14	11	19
Lin150	14	12	23	28	17	16	14-15	13	10	13	21	12	11	16	10	20
Lin151	15	13	25	31	16	15	11-14	13	10	11	23	11	12	14	11	20
Lin152	15	14	25	31	16	15	11-14	13	11	10	23	11	13	14	11	20
Lin153	15	13	25	31	15	15	12-14	13	10	11	23	11	13	14	11	20
Lin154	15	14	22	31	17	15	16-19	12	10	13	20	11	12	14	9	19
Lin155	15	13	26	31	16	15	11-12	13	11	11	24	11	12	14	11	20

Sample	DYS456	DYS389I	DYS390	DYS389II	DYS458	DYS19	DYS385a/b	DYS393	DYS391	DYS439	DYS635	DYS392	GATA_H4	DYS437	DYS438	DYS448
Lin156	16	13	25	31	15	16	11-13	13	11	10	23	11	13	14	11	20
Lin157	15	14	23	30	17	14	13-16	15	11	10	24	12	12	16	11	16
Lin158	15	14	24	32	16	15	11-15	13	10	10	23	11	13	14	11	20
Lin159	16	13	25	30	14	15	11-14	13	10	10	23	11	13	14	11	20
Lin160	14	14	22	31	15	16	15-15	12	10	11	21	11	12	14	10	19
Lin161	15	13	25	31	16	15	11-14	13	11	10	23	11	13	14	11	20
Lin162	15	12	21	28	15	14	14-17	11	9	12	24	14	12	15	11	19
Lin163	15	13	22	28	19	14	13-17	14	10	10	25	10	12	16	11	19
Lin164	15	14	23	29	17	13	13-22	15	10	12	27	10	11	14	11	20
Lin165	15	14	23	30	19	15	13-20	14	10	10	25	11	11	16	11	19
Lin166	15	14	23	30	19	16	15-18	14	10	10	24	12	11	16	11	19
Lin167	15	12	22	28	16	14	13-17	11	10	13	24	14	12	15	10	19
Lin168	14	10	25	27	18	15	12-14	13	10	12	24	11	12	14	9	20
Lin169	16	14	22	30	18	16	15-18	12	9	12	22	11	12	15	10	18
Lin170	17	12	22	28	14	14	13-17	11	10	13	22	14	12	15	10	19
Lin171	15	13	22	28	17	15	15-17	12	10	12	21	11	12	14	9	19
Lin172	16	13	25	30	16	15	11-14	13	10	11	23	11	12	14	11	20
Lin173	19	13	22	29	17	15	15-18	12	10	11	22	11	13	14	9	19
Lin174	13	13	24	29	Null	15	13-18	12	10	12	21	11	12	15	9	17
Lin175	15	13	23	30	15	14	13-19	11	10	12	24	15	12	15	11	19
Lin176	14	14	22	31	17	15	15-17	12	10	12	20	11	12	14	9	20
Lin177	16	14	22	31	17	16	15-17	12	10	11	21	11	12	14	9	19
Lin178	15	12	24	27	16	15	12-17	13	10	8	21	11	12	14	10	18
Lin179	15	12	22	28	15	14	13-17	11	10	11	24	14	12	15	10	19
Lin180	16	13	23	30	16	15	11-14	13	10	10	23	11	12	14	11	20
Lin181	15	12	22	28	15	14	13-17	11	10	11	24	14	12	15	10	19
Lin182	16	14	22	31	16	15	16-18	12	11	11	20	11	12	14	9	18
Lin183	15	12	24	27	16	15	12-17	13	10	8	21	11	12	14	10	18
Lin184	16	14	22	31	16	15	16-18	12	11	11	20	11	12	14	9	18
Lin185	15	13	23	31	14	14	14-16	14	10	11	18	11	11	14	10	20

Sample	DYS456	DYS389I	DYS390	DYS389II	DYS458	DYS19	DYS385a/b	DYS393	DYS391	DYS439	DYS635	DYS392	GATA_H4	DYS437	DYS438	DYS448
Lin186	15	14	23	31	15	14	13-17	11	10	11	25	14	12	15	10	19
Lin187	13	13	24	29	Null	15	13-18	12	10	12	21	11	12	15	9	17
Lin188	13	13	24	29	Null	15	13-18	12	10	12	21	11	12	15	9	17
Lin189	14	14	22	31	17	15	15-17	12	10	12	20	11	12	14	9	20
Lin190	15	12	22	28	14	14	13-17	11	10	14	24	14	12	16	10	19
Lin191	16	13	23	30	16	15	11-14	13	10	10	23	11	13	14	11	20
Lin192	17	12	22	29	15	14	13-21	11	10	12	24	14	12	15	10	19
Lin193	15	13	25	31	15	15	11-14	13	10	10	24	11	12	14	12	19
Lin194	16	12	22	28	15	14	13-17	11	10	13	24	14	12	15	10	19
Lin195	15	12	22	28	15	14	13-17	11	10	11	24	14	12	15	10	19
Lin196	15	12	22	29	16	14	13-16	11	9	12	23	14	12	15	10	19
Vok197	15	14	22	29	19	17	16-19	12	10	12	20	11	13	15	10	19
Vok198	13	12	25	28	16	15	13-18	12	10	13	21	12	11	15	9	19
Vok199	16	12	23	28	15	14	14-17	11	10	13	24	14	12	15	10	19
Vok200	15	14	25	31	17	15	11-15	13	10	11	23	11	12	14	11	20
Vok201	15	13	22	29	19	15	7-16	12	11	14	22	14	10	16	10	19
Vok202	16	14	23	31	19	15	14-19	14	10	12	24	10	12	16	11	20
Vok203	16	14	23	30	18	14	12-18	14	11	10	24	10	13	16	11	19
Vok204	15	13	23	29	18	15	13-16	13	10	12	21	11	11	16	10	19
Vok205	15	13	23	29	18	15	15-19	14	10	10	23	10	11	16	11	20
Vok206	16	12	23	28	15	14	14-17	11	10	13	24	14	12	15	10	19
Vok207	15	12	22	28	16	16	13-17	11	10	13	23	14	12	15	10	19
Vok208	15	12	22	28	15	14	13-16	11	10	13	24	14	13	15	10	18
Vok209	15	13	23	29	18	15	15-19	14	10	10	23	10	11	16	11	20
Vok210	15	13	22	29	17	15	17-17	12	10	12	21	11	13	14	9	20
Vok211	15	12	22	28	14	14	13-16	11	10	14	23	14	12	15	8	19
Vok212	15	14	25	31	16	15	11-14	13	10	10	23	12	13	14	11	20
Vok213	15	14	23	30	18	15	13-16	13	10	12	21	11	11	16	10	19
Vok214	15	12	22	28	15	14	13-16	11	10	13	24	14	13	15	10	18
Vok215	16	12	23	28	15	14	14-17	11	10	13	24	14	12	15	10	19

Sample	DYS456	DYS389I	DYS390	DYS389II	DYS458	DYS19	DYS385a/b	DYS393	DYS391	DYS439	DYS635	DYS392	GATA_H4	DYS437	DYS438	DYS448
Vok216	15	14	21	30	18	15	14-19	13	10	12	21	11	11	15	10	18
Vok217	15	12	22	28	14	14	13-16	11	10	14	23	14	12	15	8	19
Vok218	15	12	22	28	14	14	13-17	11	10	12	25	14	12	15	10	19
Vok219	16	12	23	28	15	14	14-17	11	10	13	24	14	12	15	10	19
Vok220	15	12	22	27	15	14	14-17	11	10	12	25	14	12	15	10	19
Vok221	16	13	25	31	16	16	11-14	13	11	11	23	11	12	14	11	20
Vok222	16	13	22	29	17	16	15-17	12	10	12	20	11	12	14	9	19
Vok223	15	12	22	28	15	14	13-16	11	10	13	24	14	13	15	10	18
Vok224	16	12	23	28	15	14	14-17	11	10	13	24	14	12	15	10	19
Vok225	15	12	22	28	14	14	13-18	11	10	13	23	14	11	15	10	19
Vok226	15	13	23	29	19	15	13-16	13	10	12	21	11	11	16	10	19
Vok227	15	14	21	31	17	15	12-16	13	11	12	21	11	11	15	10	18
Vok228	15	13	23	29	18	15	13-16	13	10	12	21	11	11	16	10	19
Vok229	15	12	22	28	15	14	13-17	11	10	12	24	14	12	15	10	19
Vok230	16	14	23	30	18	14	12-18	14	11	10	24	10	13	16	11	19
Vok231	15	12	22	28	15	14	13-16	11	10	13	24	14	13	15	10	19
Vok232	15	13	21	30	16	17	18-18	13	10	13	20	11	11	15	10	20
Vok233	14	14	23	30	19	14	13-18	14	11	12	26	10	11	16	11	18
Vok234	15	12	22	28	14	14	13-16	11	11	11	24	14	12	15	10	19
Vok235	17	12	22	29	16	14	14-17	11	10	12	23	14	12	15	11	19
Vok236	15	14	25	31	16	15	11-14	13	10	10	23	12	12	14	11	20
Vok237	16	13	23	29	17	14	12-18	14	10	10	23	10	12	16	11	19
Vok238	15	14	25	31	17	15	11-14	13	10	11	23	11	12	14	11	20
Vok239	16	14	22	30	17	15	15-17	12	11	11	19	11	12	14	9	19
Vok240	15	13	22	29	19	15	7-16	12	11	14	22	14	10	16	10	19
Vok241	15	12	22	28	14	14	13-16	11	10	13	23	14	12	15	8	19
Vok242	15	13	23	29	18	14	12-16	14	10	10	24	10	12	16	12	19
Vok243	16	13	22	30	18	15	15-17	12	10	12	20	11	11	14	9	19
Vok244	16	12	22	29	15	14	14-18	11	10	12	23	14	12	15	10	19
Vok245	13	12	22	29	20	14	12-13	13	11	12	26	12	11	14	11	19

Sample	DYS456	DYS389I	DYS390	DYS389II	DYS458	DYS19	DYS385a/b	DYS393	DYS391	DYS439	DYS635	DYS392	GATA_H4	DYS437	DYS438	DYS448
Vok246	15	12	22	28	14	14	13-16	11	10	13	23	14	12	15	8	19
Vok247	15	13	21	30	16	17	18-18	13	10	13	20	11	11	15	10	20
Vok248	15	12	22	28	14	14	13-16	11	10	13	23	14	12	15	8	19
Vok249	15	12	24	28	18	15	14-16	13	10	13	24	11	11	15	10	19
Vok250	15	14	23	30	20	16	13-15	13	10	12	21	11	11	16	10	19
Vok251	16	13	22	30	18	15	15-17	12	10	12	20	11	11	14	9	19
Vok252	16	12	22	28	15	14	14-17	11	10	13	24	14	12	15	10	19
Vok253	17	12	22	28	14	14	14-16	11	10	12	24	14	12	15	10	19
Vok254	15	12	22	28	15	14	13-17	11	10	12	25	14	11	15	10	19
Vok255	15	13	23	29	17	15	12-18	14	10	10	25	10	12	15	12	19
Vok256	15	15	24	32	16	14	13-16	13	10	11	24	11	11	15	9	21
Vok257	14	12	22	28	14	14	13-16	12	10	14	24	14	12	15	10	19
Vok258	14	14	23	30	19	15	13-22	14	10	10	25	10	12	16	11	18
Vok259	15	14	24	31	17	15	15-16	12	10	11	20	11	12	14	9	19
Vok260	13	14	23	30	18	15	13-19	14	10	10	26	10	12	16	11	19
Vok261	16	14	25	31	16	17	11-14	13	10	11	23	11	13	14	11	20
Vok262	13	12	24	28	16	15	11-16	13	10	12	20	11	11	16	9	19
Vok263	14	13	24	29	16	14	15-15	13	10	11	20	11	11	14	9	18
Vok264	15	13	21	30	18	17	15-20	14	11	12	23	11	12	15	10	19
Vok265	15	14	22	30	15	15	14-16	13	10	13	21	11	14	14	11	18
Vok266	16	15	23	31	19	14	12-17	14	10	12	24	10	12	16	11	19
Vok267	16	14	22	29	17	15	14-17	12	10	11	20	11	11	14	9	19
Vok269	15	13	21	30	18	17	15-20	14	11	12	23	11	12	15	10	19
Vok271	14	10	24	27	18	15	12-14	13	10	12	24	11	12	14	9	20
Vok272	15	12	22	28	19	16	12-13	14	10	12	21	11	12	16	10	21
Vok273	13	14	25	33	16	15	11-15	13	10	10	23	11	14	14	11	20
Vok274	15	13	21	30	18	17	15-20	14	11	12	23	11	12	15	10	19
Vok275	15	13	21	30	18	17	15-20	14	11	12	23	11	12	15	10	19
Vok276	15	13	21	30	18	17	15-20	14	11	12	23	11	12	15	10	19
Vok277	15	13	21	30	18	17	15-20	14	11	12	23	11	12	15	10	19

Sample	DYS456	DYS389I	DYS390	DYS389II	DYS458	DYS19	DYS385a/b	DYS393	DYS391	DYS439	DYS635	DYS392	GATA_H4	DYS437	DYS438	DYS448
Vok278	17	14	23	32	18	14	15-16	14	10	11	18	11	13	14	10	19
Vok279	13	11	24	27	16	15	13-18	12	12	12	22	11	11	15	9	20
Vok280	15	13	23	28	16	14	13-18	14	10	11	25	10	11	16	11	19
Vok281	15	13	21	31	18	17	15-20	14	11	12	23	11	12	15	10	19
Vok282	15	13	21	29	15	15	15-16	12	10	11	20	11	11	14	8	19
Vok283	15	12	22	29	14	14	13-16	11	10	12	24	14	12	15	10	19
Vok284	17	14	22	30	17	16	18-18	12	10	11	20	11	12	14	9	19
Vok285	15	13	21	30	18	17	15-20	14	11	12	23	11	12	15	10	19
Vok286	15	12	22	28	15	14	14-17	11	10	11	24	14	11	15	10	19
Vok287	15	13	21	30	18	17	15-20	14	11	12	23	11	12	15	10	19
Vok288	15	13	21	30	18	17	15-20	14	10	12	23	11	12	15	10	19
Vok289	18	13	22	30	16	15	15-16	14	10	12	20	11	12	14	9	18
Vok290	17	13	24	30	15	15	15-18	14	10	14	23	13	11	14	10	18
Vok291	15	13	23	29	16	14	14-19	14	10	11	25	10	12	16	12	18
Vok292	15	13	22	29	17	15	15-16	12	10	12	21	14	11	16	10	19
Vok293	15	13	21	30	18	17	15-19	14	12	12	23	11	12	15	10	19
Vok295	16	13	22	25	17	15	15-18	12	10	11	20	11	11	14	9	19
Vok296	15	13	22	29	18	15	7-16	12	10	13	22	14	11	16	10	19
Vok298	15	12	22	28	14	14	13-16	11	10	11	22	14	12	15	10	19
Vok299	16	14	22	29	17	15	16-17	12	10	11	20	11	11	14	9	19
Vok300	15	14	22	30	17	16	14-14	13	10	13	20	11	11	15	10	18
Vok301	15	13	25	29	17	16	11-15	13	10	11	22	11	13	14	11	19
Vok302	16	13	25	30	15	16	11-14	13	11	10	24	11	13	14	11	20