# GENETIC POLYMORPHISMS ON Y-CHROMOSOMAL HAPLOTYPE STR LOCI IN GARO AND SANTAL POPULATION IN BANGLADESH

A DISSERTATION SUBMITTED TO THE UNIVERSITY OF DHAKA IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF MASTER OF PHILOSOPHY IN BIOCHEMISTRY AND MOLECULAR BIOLOGY


SUBMITTED BY

MD. MAHAMUD HASAN
EXAMINATION ROLL NO.: 01
REGISTRATION NO.: 346
SESSION: 2008-2009

SUPERVISED BY

PROFESSOR DR. SHARIF AKHTERUZZAMAN
DEPARTMENT OF BIOCHEMISTRY AND MOLECULAR BIOLOGY
FACULTY OF BIOLOGICAL SCIENCES
UNIVERSITY OF DHAKA
BANGLADESH

JULY 21, 2013

***DEDICATED TO MY BELOVED PARENTS***

# ACKNOWLEDGEMENTS

First of all I would like to say all the findings of this study in only the result of the mercy of Allah.

It is great pleasure to express my enormous gratefulness and sincere appreciation to my supervisor Professor Dr. Sharif Akhteruzzaman, Department of Biochemistry and Molecular Biology, University of Dhaka, for his scholastic guidance, valuable advice, continuous cooperation and endless inspiration throughout the period of this research work and emendation of the manuscript.

I wish to record my sincere thanks to the M.Phil committee members of University of Dhaka.

I would like to thank from the bottom of my heart to Pilu Momtaz , Scientific Officer and Md. Mizanur Rahman, Lab Technologist , National Forensic DNA Profiling Laboratory, Dhaka Medical College for their assistance in the lab work.

I am also grateful acknowledge to Corpus Christy High School, Jalchatra, Madhupur, Tangail and Saint Joseph High School, Dinajpur for helping to collect the blood and buccal samples from Garo and Santal males, respectively for my research work.

Many thanks also go to all the Garo and Santal individuals who so generously donated biological materials for the study.

Finally, I like to express utmost gratitude to my parents and all other family members and friends for their enthusiastic support, constant inspiration and blessings during my study and thesis work.

# ABSTRACT

Allele frequencies and haplotype or gene diversity distribution of seventeen Y-chromosomal short tandem repeats (Y-STRs) such as, DYS456, DYS389I, DYS390, DYS389II, DYS458, DYS19, DYS385a/b, DYS393, DYS391, DYS439, DYS635, DYS392, Y GATA H4, DYS437, DYS438 and DYS448 were examined in 120 Garo and 139 unrelated male individuals of Santal community from two different places of Bangladesh. From 120 Garo individuals, a total of 99 different haplotypes were identified with a discrimination capacity (DC) of 0.825 and from 139 Santal individuals, a total of 129 different haplotypes were observed which corresponds to a discrimination capacity (DC) of 0.928. The highest allele frequency (0.8250) was observed in locus DYS391 for Garo, whereas for Santal it was 0.8993 for locus DYS437. The gene diversity values ranged from 0.2939 (DYS391) to 0.9505 (DYS385a/b) in Garo, whereas in Santal it was observed from 0.1849 (DYS437) to 0.9212 (DYS385a/b). To portray the genetic association among the populations, Neighbour-Joining (NJ) tree were constructed based on the $F_{st}$ value using the PHYLIP software. A comparison of the studied data with the published data from YHRD (Y Chromosome Haplotype Reference Database ) based on AMOVA analysis showed that, Garo population is closely related to Khalkh population from Ulaanbaatar, Mongolia and Santal population is closely related to Rajbanshi population from West Bengal, India. The high haplotype diversity obtained from the two studied population indicated the high potential for differentiating between male individuals in this two populations and indicated their usefulness in criminal investigation, parentage testing and in determining paternal lineage.

# TABLE OF CONTENTS


| | Title | Page No. |
|---|---|---|
| **CHAPTER ONE: INTRODUCTION** | | **01-15** |
| 1.1 | General overview……. | 01 |
| 1.2 | Population Genetics……. | 02 |
| 1.3 | Development of Y-STR Database……... | 03 |
| 1.4 | Chromosomal Locations of Markers……... | 04 |
| 1.5 | Polymorphisms: The Basis of DNA typing……. | 05 |
| 1.6 | Polymorphisms of the Human Y-Chromosome…….. | 05 |
| 1.7 | Y-STR Haplotypes……... | 05 |
| 1.8 | Genetic Diversity……. | 06 |
| 1.9 | Y chromosome STR Studied on Various Populations…… | 07 |
| 1.10 | Objectives of the present study…… | 14 |
| 1..11 | Importance of the present study…….. | 15 |
| **CHAPTER TWO: MATERIALS AND METHODS** | | **17-27** |
| 2.1 | Study Population……..... | 17 |
| 2.2 | Ethical Consideration…… | 17 |
| 2.3 | Materials…….. | 17 |
| 2.3.1 | Laboratory Instruments…… | 17 |
| 2.3.2 | Chemicals and Reagents……..... | 18 |
| 2.4 | Methods…….. | 19 |
| 2.4.1 | Sample Collection…… | 19 |
| 2.4.2 | Chelex DNA Extraction……..... | 19 |
| 2.4.3 | Reagents to be Prepared for Extraction……..... | 19 |
| 2.5 | DNA Extraction Procedure by Chelex Method……... | 20 |
| 2.5.1 | Extraction of DNA from FTA™ Card……. | 20 |
| 2.5.2 | Extraction of DNA from Buccal Swab Samples……..... | 21 |
| 2.6 | DNA Quality and quantification Determination……..... | 21 |
| 2.7 | PCR Amplification of STR Loci……. | 22 |
| 2.8 | Materials for PCR of Y-STR Loci Amplification……... | 22 |

| | Title | Page No. |
|---|---|---|
| 2.9 | Reagents Required for Fragment Analysis……..... | 24 |
| 2.10 | Analysis of Amplified Fragments using Capillary Electrophoresis | 24 |
| 2.10.1 | Samples Preparation for Electrophoresis……. | 25 |
| 2.10.2 | Running the Plate on Genetic Analyzer……... | 26 |
| 2.10.3 | Data Collection and Analysis…….. | 26 |
| 2.10.4 | Quality Control for Genotyping…….. | 26 |
| 2.10.5 | Statically Analysis of Data……. | 27 |
| **CHAPTER THREE: RESULTS** | | **28-37** |
| 3.1 | Population Genetic Data……..... | 28 |
| 3.1.1 | Y-STR Genotyping of DNA Samples……..... | 28 |
| 3.2 | Statistical Analysis of Genetic Data…… | 29 |
| 3.2.1 | Allele Frequencies……... | 29 |
| 3.2.2 | Gene or Haplotype Diversity……... | 31 |
| 3.2.3 | Genetic Distance with Pairwise Fst using AMOVA……... | 33 |
| 3.2.4 | Neighbour-Joining (NJ) Tree Generating…… | 36 |
| 3.2.5 | Multidimensional Scaling (MDS) Population Plot Designing…… | 37 |
| **CHAPTER FOUR: DISCUSSION** | | **48-57** |
| **REFERENCES** | | **58-64** |
| **APPENDICES** | | **65-75** |
| Appendix 1 | A list of Y Chromosome STR haplotypes detected in 120 unrelated males in Garo population……. | 65-67 |
| Appendix 2 | Allele frequencies and gene diversity for 17 Y-STR markers in 120 unrelated Garo males …… | 68-69 |
| Appendix 3 | A list of Y Chromosome STR haplotypes detected in 139 unrelated males in Santal population……... | 70-73 |
| Appendix 4 | Allele frequencies and gene diversity for 17 Y-STR markers in 139 unrelated Santal males…….. | 74-75 |

# LIST OF TABLES


| Table No. | Title | Page No. |
|---|---|---|
| Table 1.1 | Major equipments used in the present study…… | 18 |
| Table 1.2 | Chemicals and reagents used in the present study…….. | 18 |
| Table 1.3 | Master Mix Preparation…… | 23 |
| Table 1.4 | Thermal Cycling Profile……... | 23 |
| Table 1.5 | Formamide:Size standard cocktail preparation…… | 25 |
| Table 1.6 | Φst genetic distances (below diagonal) and Φst P-values (above diagonal) for pair-wise comparisons of Garo and other 9 Asian populations……. | 35 |
| Table 1.7 | Φst genetic distances (below diagonal) and Φst P-values (above diagonal) for pair-wise comparisons of Santal and other 8 Asian and 1 African populations…….. | 35 |

# LIST OF FIGURES


| Figure No. | Title | Page No. |
|---|---|---|
| Figure 1.1 | Chromosomal locations for commonly used and new Y-STR markers…… | 04 |
| Figure 1.2 | Allele frequency distribution of ten Y-STR loci. The figures compared Bangladeshi Bengali, Garo and Santal genetic data | 30 |
| Figure 1.3 | Comparison of Gene or Haplotype diversity among three population group of Bangladesh (Bangladeshi, Garo and Santal…… | 32 |
| Figure 1.4 | Neighbor-Joining tree structure from pairwise genetic distance (Φst value) between Garo population (Bangladesh) and Nine Asian population…….. | 36 |
| Figure 1.5 | Neighbor-Joining tree structure from pairwise genetic distance (Φst value) between Santal (Bangladesh) and Eight Asian and One African population……. | 37 |
| Figure 1.6 | MDS plot based on pairwise Φst values calculated for the Garo (colored) and nine reference populations of Asia……. | 38 |
| Figure 1.7 | MDS plot based on pairwise Φst values calculated for the Santal (colored) and eight reference populations of Asia and one African population…… | 39 |
| Figure 1.8 | Matrix of pairwise FST values indicates short-term genetic distances between populations……. | 40 |
| Figure 1.9 | Slatkins Linearized Fst value indicates short-term genetic distances between populations……. | 41 |
| Figure 1.10 | Matrix of co-ancestry coefficients indicates the relationship between different populations…….. | 42 |
| Figure 1.11 | Average number of pairwise differences indicates the relationship between different populations…….. | 43 |
| Figure 1.12 | Number of alleles at different loci shows the relationship between different populations……. | 44 |

**Title** **Page No.**

Figure 1.13 Expected heterozygosity indicates the closely relationship between different populations…….. 45

Figure 1.14 Molecular distance at different loci ($\square_H$) indicates the molecular distance among various populations based on expected heterozygosity…….. 46

Figure 1.15 Modified Garza-Williamson index at different loci indicates the molecular distance among different populations…… 47

# ABBREVIATIONS

| | |
|---|---|
| A/N | Accession Number |
| ABI | Applied Biosystems Incorporation. |
| AMEL | Amelogenin |
| AMOVA | Analysis of Molecular Variance |
| ASO | Allele-specific oligonucleotide |
| CE | Capillary Electrophoresis |
| CODIS | Combined DNA Index System |
| DC | Discrimination Capacity |
| EDTA | Ethylene diamine tetraacetic acid |
| FAM | 5-Carboxy Fluorescein |
| Fst | F-statistics |
| GD | Gene Diversity |
| GDB | Genome Database |
| HD | Haplotype Diversity |
| Hi-Di | Highly Deionized |
| HVRs | Hyper Variable Regions |
| ISFG | International Society of Forensic Genetics |
| LINEs | Long Interspersed Nuclear Elements |
| MDS | Multidimensional scaling |
| MLP | Multi-Locus Probe |
| MRCA | Most Recent Common Ancestor |
| MSY | Male-specific Y-chromosome Region |
| NCBI | National Centre for Biotechnology Information |
| NED | (Proprietary to Applied Biosystems) |
| NIST | National Institute of Standards and Technology |
| NJ | Neighbour-Joining |
| NRY | Non-recombining Y-chromosome Region |
| OD | Optical Density |

| | |
|---|---|
| PAR | Pseudo-Autosomal Region |
| PCR | Polymerase Chain Reaction |
| PHYLIP | PHYLogeny Inference Package |
| POP | Performance Optimized Polymer |
| RFLP | Restriction Fragment Length Polymorphism |
| SINEs | Short Interspersed Nuclear Elements |
| SLP | Single Locus Polymorphism |
| SNPs | Single Nucleotide Polymorphisms |
| SSRs | Simple Sequence Repeats |
| STR | Short Tandem Repeat |
| SWGDAM | Scientific Working Group on DNA Analysis Methods |
| TE | Tris-EDTA |
| VNTR | Variable Number of Tandem Repeats |
| YHRD | Y Chromosome Haplotype Reference Database |
| Y-STR | Short Tandem Repeat (STR) on the Y-chromosome |

**Chapter One**

# INTRODUCTION

## 1.1 General Overview

DNA typing has become one of the most definitive tools for human identification in forensic science, paternity testing, missing person identification, mass disaster investigations as well as genealogical mapping. The majority portion of a DNA molecule (99.9%) is the same between individuals. Only a small portion of DNA (0.1% or one million nucleotide) differs between people and makes us unique individuals. In these small portions of DNA, some hypervariable polymorphic regions provide the capability of using DNA information for human identity as well as genealogical purposes (Butler J.M., 2001).

Recently, geneticists often used additional markers as the need arises to obtain further information about a particular sample. Sex-typing is performed in conjunction with available STR markers to provide the gender of the individual who is the source of the DNA sample in question. When DNA samples are highly degraded and fail to provide useful information with conventional STR typing, mitochondrial DNA. Additionally, Y chromosome systems are becoming more popular as a means to extract information from the male portion of a sample mixture (e.g., evidence in rape cases).

The Y chromosome and mitochondrial DNA markers are known as 'lineage markers'. They are passed down from generation-to-generation without changing (except for mutational events). Maternal lineages can be traced with mitochondrial DNA sequence information while paternal lineages can be followed with Y chromosome markers. With lineage markers, the genetic information from each marker is referred to as a haplotype rather than a genotype because there is only a single allele per individual. Because Y chromosome markers are linked on the same chromosome and are not shuffled with each generation. Therefore, haplotype obtained from lineage markers can never be as effective in differentiating between two individuals as genotypes from autosomal markers that are unlinked and segregate separately from generation to generation.

The genetic markers forming a DNA profile are found at different frequencies and gene or haplotype diversity in a population. This collection of genetic markers is known as a population genetic database. If the frequencies of the genetic markers making up a DNA profile are very high and found in half of the population, the statistical strength or discrimination power of the particular DNA profile is not strong. On the contrary, if the genetic markers of the DNA profile are found only in one in a million individuals, then the statistical discrimination power of the DNA profile is very strong. It is therefore, important that the data on the genetic markers of a population genetic database are statistically tested before the data are used to determine the occurrence rate of a particular DNA profile in criminal caseworks. Neglecting to do so can exaggerate the strength of the evidence against the defendant (Balding *et al.* 1996).

Due to the fact that the Y chromosome is passed down unchanged (except for mutations) from father to son, the observation of a match with Y-STRs does not carry the power of discrimination and weight into court. The lack of recombination between Y chromosome markers means that Y-STR results have to be combined into a haplotype for searching available databases as well as estimating the rarity of a particular haplotype. Population DNA databases are important for comparison purposes to understand how frequent or how rare a crimescene DNA profile may be in a particular population. Within the last several years, a number of Y-STR haplotype databases are developed by different organization. The largest and most widely used Y-STR database was created by Lutz Roewer and colleagues at Humbolt University in Berlin, Germany and has been available online since 2000. As of May 2004, more than 24000 samples from greater than 224 populations around the world can be searched via the Internet at the following web sites: http://www.yhrd.org.

**1.2 Population Genetics**

Population genetics is a field of biology that studies the genetic composition of biological populations and the changes in genetic composition that result from the operation of various factors, including natural selection. Population geneticists pursue their goals by developing abstract mathematical models of gene frequency dynamics, trying to extract conclusions from those models about the likely patterns of genetic variation in actual populations, and testing the conclusions against empirical data.

Population genetics is intimately bound up with the study of evolution and natural selection, and is often regarded as the theoretical cornerstone of modern Darwinism. This is because natural selection is one of the most important factors that can affect a population's genetic composition. Natural selection occurs when some variants in a population out-reproduce other variants as a result of being better adapted to the environment. Presuming the fitness differences are at least partly due to genetic differences, this will cause the population's genetic makeup to be altered over time. By studying formal models of gene frequency change, population geneticists therefore hope to shed light on the evolutionary process, and to permit the consequences of different evolutionary hypotheses to be explored in a quantitatively precise way. The field of population genetics came into being in the 1920s and 1930s, thanks to the work of R.A. Fisher, J.B.S. Haldane and Sewall Wright. It is the study of the frequency and interaction of alleles and gene in population [Postlethwalt J., 2009]. A sexual population is a set of organisms in which any pair of members can breed together. This implies that all members belong to the same species and live near each other [Hartl D., 2007].

### 1.3 Development of Y-STR database

The use of Y-STRs as inclusion evidence involves the definition of population genetics profiles, with the elaboration of a large number of population databases. The construction of Y-specific STR databases is more complex than those for unlinked autosomal markers because the informative content of Y-specific STRs results from the possibility of constructing highly discriminative haplotypes. The suitability of the Y-STR databases for practical use in the forensics field will be greatly increased with the typing of each individual to as many loci as possible instead of typing a great number of individuals for a small number of Y-STRs. The Y-STR markers with the most use in sample testing to date are the “minimal haplotype” loci. These 9 markers (if one counts the two DYS385 alleles as separate “loci”) have been used to generate more than 16,000 profiles in the Y-STR Haplotype Reference database across approximately 100 European, U.S., and Asian populations (http://www.ystr.org). Within the past several years, studies with additional Y-STR loci beyond the minimal haplotype loci have been conducted.

### 1.4 Chromosomal Locations of Markers

The efforts of the Human Genome Project have generated a publicly available human Y-chromosome sequence that is approximately 51 megabases (Mb) in size. However, a "heterochromatin" region around 20 Mb in size toward the end of the long arm of the Y-chromosome may never be completely deciphered [Tilford *et al.* 2001]. The availability of a human reference sequence now permits location of the various Y-STR markers along the Y-chromosome **(Figure 1.1)**.

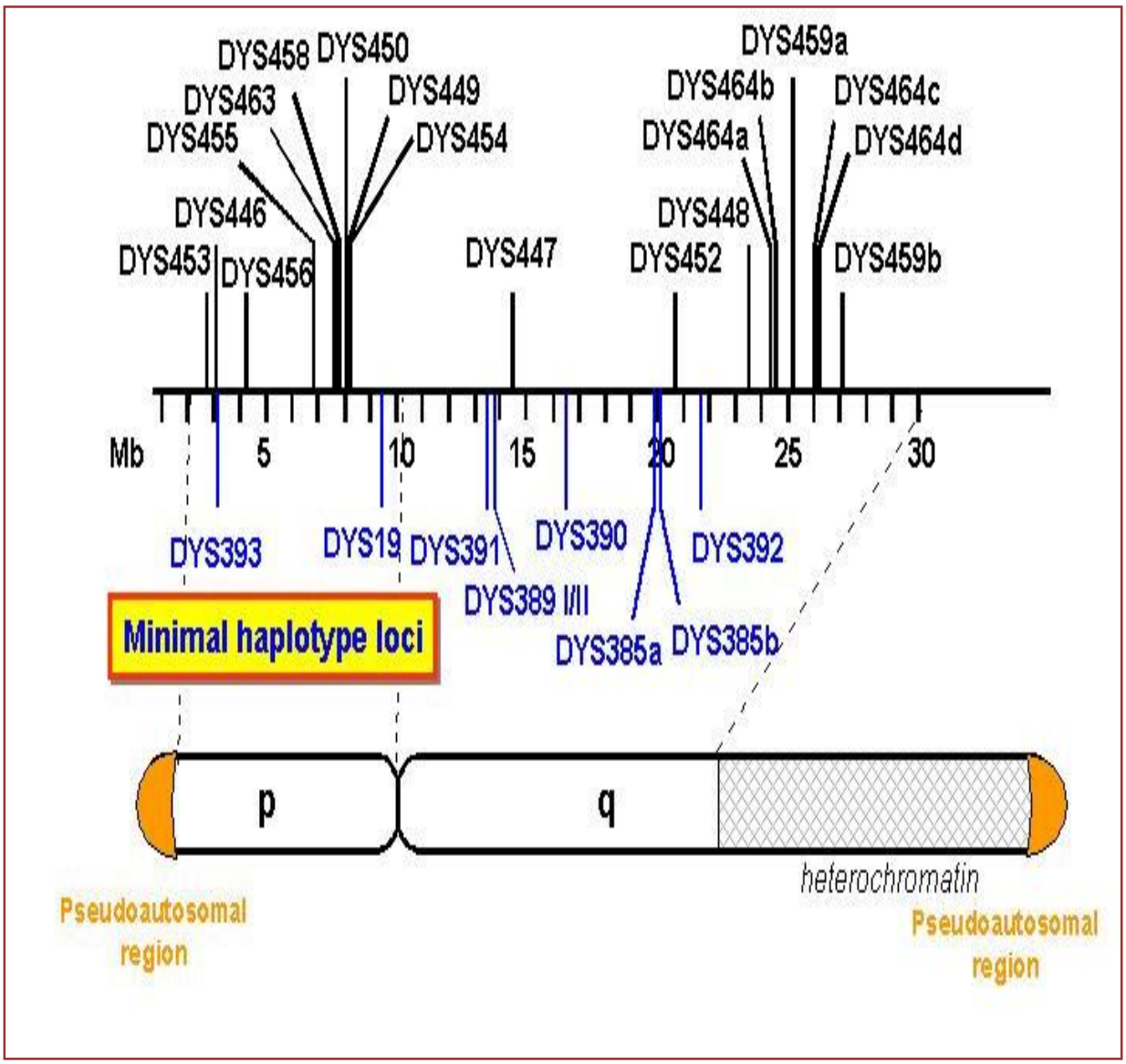


**Figure 1.1:** Chromosomal locations for commonly used and new Y-STR markers (Michael Hammer, 2002).

### 1.5 Polymorphisms: The Basis of DNA typing

All DNA testing is based on the observation that the genome of each person or animal is unique (except of course identical twins). The myriad of small and large differences in nucleotide sequence among individuals are known as DNA polymorphisms. Two fundamentally different types of polymorphisms have been widely exploited for DNA typing: tandem repeats and restriction fragment length polymorphisms.

Polymorphism is the co-occurrence of at least two different alleles in a population. Most of the DNA is identical to DNA of others. There are inherited regions of DNA that can vary from person to person. DNA variation is exhibited in the form of different alleles, or various possibilities at a particular locus. So variations in DNA Sequence between individuals are termed polymorphisms. The polymorphisms are also seen in short tandem repeat sequences due to the different number of copies of the repeat element that can occur in a population of individuals. This high degree of length polymorphism in the human population owing to frequents slippage by DNA polymerase during replication.

### 1.6 Polymorphisms of the Human Y-Chromosome

Mutations create DNA polymorphisms and these may serve as genetic markers.
These are below:

- Single-Copy DNA – e.g., Y-SNPs (about 200 binary polymorphisms characterized)
- Repetitive DNA – e.g.,Y-STRs (above 300 microsatellites characterized)
- Minisatellite – e.g., MSY1 (only 1 minisatellite characterized)

### 1.7 Y-STR haplotypes

A haplotype is a combination of alleles at adjacent locations (loci) on a chromosome that are transmitted together. A haplotype may be one locus, several loci, or an entire chromosome depending on the number of recombination events that have occurred between a given set of loci. Y-STR haplotypes means the set of results from the Y-STR markers tested. A Y-STR haplotype is the unique set of values obtained when a defined set of Y-DNA STR markers are tested. A minimum haplotype refers to the selected set of Y-DNA STR markers chosen to facilitate comparison between different individuals and different populations. Genetic distance

is a value which represents the number of non-matches between tested haplotype and another haplotype. The smaller the value of genetic distance, the more recent that two individuals shared a common ancestor.

**1.8 Genetic Diversity**

Genetic diversity refers to the total number of genetic characteristics in the genetic makeup of a species. It is distinguished from genetic variability, which describes the tendency of genetic characteristics to vary. It serves as a way for populations to adapt to changing environments. With more variation, it is more likely that some individuals in a population will possess variations of all alleles that are suited for the environment. Those individuals are more likely to survive to produce offspring bearing that allele. The population will continue for more generations because of the success of these individuals [Groom *et al.,* 2006].

The academic field of population genetics includes several hypotheses and theories regarding genetic diversity. The neutral theory of evolution proposes that diversity is the result of the accumulation of neutral substitutions. Diversifying selection is the hypothesis that two subpopulations of a species live in different environments that select for different alleles at a particular locus. This may occur, for instance, if a species has a large range relative to the mobility of individuals within it. Frequency dependent selection is the hypothesis that as alleles become more common, they become more vulnerable.

There are many different way to measure genetic diversity. The modern causes for the loss of animal genetic diversity have also been studied and identified [Tisdell C., 2003]. In 2007, a study conducted by the National Science Foundation and found that genetic diversity and biodiversity (biodiversity is the degree of variation of life forms within a given ecosystem) are dependent upon each other –that diversity within a species is necessary to maintain diversity among species, and vice versa. Genotypic and phenotypic diversity has been found in all species at the protein, DNA, and organism levels. Genome-phenome organization in nature is nonrandom, heavily structured, and correlated with abiotic and environmental diversity and stress [Frankham R, 2005].

The interdependence between genetic and biological diversity is delicate. Changes in biological diversity lead to changes in the environment, leading to adaptation of the remaining species. Changes in genetic diversity, such as in loss of species, leads to a loss of biological diversity. Genetic diversity plays an important role in the survival and adaptability of a species [Pullin A.S., 2002]. When a population habited changes, the population may have to adapt to survive. Variation in the population's gene pool provides variable traits among the individuals of that population. The variable traits can be selected for, via natural selection; ultimately leading to an adaptive change in the population, allowing it to survive in the changed environment. If a population of a species has a very diverse gene pool then there will be more variability in the traits of individuals of that population and consequently more traits for natural selection to act upon to select the fittest individuals to survive.

Genetic diversity is also essential for a species to evolve. Species that have less genetic variation are at a greater risk. With very little gene variation within the species, healthy reproduction becomes increasingly difficult, and offspring are more likely to deal with problems such as inbreeding. The vulnerability of a population to certain types of diseases can also increase with reduction in genetic diversity.

### 1.9 Review of Literature

Y-STR loci are important genetic markers for forensic investigation. However, paternal inheritance, reduced effective population size, and lack of independence between loci can reduce Y-STR diversity and may yield greater population substructure effects on a locus-by-locus basis compared with the autosomal STR loci. Population studies are necessary to assess the genetic variation of forensically relevant markers so that proper inferences can be made about the rarity of DNA profiles. The 8 Y-STR polymorphic regions (DYS19, DYS388, DYS389I, DYS389II, DYS390, DYS391, DYS392 and DYS393) were typed in four populations from northwest (NW) Africa (Moroccan Arabs, southern Moroccan Berbers, Saharawis and Mozabites). Allele frequency distributions showed statistically significant differences for all loci among all the populations except for DYS19. The haplotype diversity ranged from 0.856 (Mozabites) to 0.967 (southern Moroccan Berbers). For some loci, allele frequencies in NW Africans were clearly different from those in Europeans. The most common NW African haplotype was found only in one individual out of a total of 494 Europeans typed

for the whole STR set. Thus, NW African and European Y chromosomes are clearly differentiated [Bosch *et al.* 2000].

Another population studies were carried out on the Y-specific STR polymorphisms DYS19, DYS385 I+II, DYS389 I+II, DYS390, DYS391, DYS392, and DYS393 in population samples from North India, Turkey, and Germany. In all three populations the vast majority of haplotypes was observed only once, especially in the Turkish group [Henke *et al.* 2001]. In 2001, Aisha and co-workers studied on 16 Y-specific STR loci in 711 males from 12 populations in Pakistan. Individual loci showed between 4 and 10 alleles, and diversities ranged from 0.07 to 0.77. A total of 527 different haplotypes were found and the haplotype diversity ranged from 0.92 to 0.99 for the different populations. 446 haplotypes occurred in single individuals, and only 19 haplotypes were present in more than three males, but two striking examples of haplotype sharing were found, one involving 13 individuals, and the other 17. The 13 individuals were all Parsis, and 16 of the 17 were Brahuis, providing evidence for population sub structuring.

In 2001, Chinese geneticist Hou and his colleagues found that there is a need to develop more Y-specific STR markers, because the discriminating power of each STR locus is limited. Their study focused on the analysis of the DNA sequence for each allele at all six Y-specific STR loci in order to understand their structures in the human genome and to construct human allelic ladders, which are necessary for forensic DNA typing. In addition, the haplotype distribution for all six analyzed loci was studied in a Chinese Han population sample. The results indicate that DYS434, DYS435, DYS436, DYS437, DYS438 and DYS439 are useful Y-specific STR markers for forensic sciences.

In one study, the genetic variance at 7 Y-STRs was studied among 986 male individuals from 20 globally dispersed human populations. A total of 598 different haplotypes were observed, of which 437 (73.1%) were each found in a single male only. Population-specific haplotype-diversity values were 0.86–0.99. Analyses of haplotype diversity and population-specific haplotypes revealed marked population structure differences between more-isolated indigenous populations (e.g., Central African Pygmies or Greenland Inuit) and more-admixed populations

(e.g., Europeans or Surinamese). Furthermore, male individuals from isolated indigenous populations shared haplotypes mainly with male individuals from their own population. By analysis of molecular variance, we found that 76.8% of the total genetic variance present among these male individuals could be attributed to genetic differences between male individuals who were members of the same population. Haplotype sharing between populations, $F_{ST}$ statistics, and phylogenetic analysis identified close genetic affinities among European populations and among New Guinean populations. Our data illustrate that Y-chromosomal STR haplotypes are an ideal tool for the study of the genetic affinities between groups of male subjects and for detection of population structure [Manfred *et al.* 2001].

In 2003, scientist investigated geographic structure within U.S. ethnic populations. They analyzed 1705 haplotypes on the basis of 9 Y-STR from 9–11 groups each of African-Americans, European-Americans, and Hispanics. They found that there were no significant differences in the distribution of Y-STR haplotypes among African-American groups, whereas European-American and Hispanic groups did exhibit significant geographic heterogeneity. However, the significant heterogeneity resulted from one sample; removal of that sample in each case eliminated the significant heterogeneity. Multidimensional scaling analysis of $R_{ST}$ values indicated that African-American groups formed a distinct cluster, whereas there was some intermingling of European-American and Hispanic groups. MtDNA data exist for many of these same groups; estimates of the European-American genetic contribution to the African-American gene pool were 27.5%–33.6% for the Y-STR haplotypes and 9%–15.4% for the mtDNA types. The lack of significant geographic heterogeneity among Y-STR and mtDNA haplotypes in U.S ethnic groups means that forensic DNA databases do not need to be constructed for separate geographic regions of the U.S. Moreover, absence of significant geographic heterogeneity for these two loci means that regional variation in disease susceptibility within ethnic groups is more likely to reflect cultural/environmental factors, rather than any underlying genetic heterogeneity [Kayser *et al.* 2003].

In 2005, Korean biologist Park and his co-workers studied on 19 Y-specific STR loci and were analyzed in 301 unrelated Korean males by three multiplex PCR systems. A total of 297 different haplotypes were identified using the 19 Y-STR markers, of which 293 were unique and 4 were found twice. The overall haplotype diversity was 0.9999. The evaluation of the

information of selected markers by combination of each marker with the minimal haplotype showed that DYS434, DYS435, DYS436, DYS437, and DYS438 do not significantly contribute to increment of haplotype diversity. However, respective conjunction of DYS464, DYS449, and DYS446 with the minimal haplotype considerably increased the haplotype diversity. Especially, DYS464 is expected to be the most useful marker that can be included in the expanded minimal haplotype. These results including the haplotype data at 19 Y-STR loci in the present study would provide useful information in forensic practice in a Korean population.

A total of 2443 male individuals distributed across the five North American population groups African American, Asian, Caucasian, Hispanic, and Native American were typed for the 12 Y-STR loci. Results showed that all population samples were highly polymorphic with the marker DYS385a/b being the most polymorphic across all sample populations. The Native American population groups demonstrated the lowest genetic diversity, most notably at the DYS393 and DYS437 loci. Haplotype diversities were greater than 99.6% for the African Americans, Caucasians, Hispanics, and Asians. The Native Americans had the lowest haplotype diversities (Apaches, 97.0%; Navajo, 98.1%). Population substructure effects were greater for Y-haplotypes, compared with that for the autosomal loci. For the apportionment of variance for the 12 Y-STRs, the within sample population variation was the largest component (>98% for each major population group and approximately 97% in Native Americans), and the variance component contributed by the major population groups was less than the individual component, but much greater than among sample populations within a major group (11.79% versus 1.02% for African Americans/Caucasians/Hispanics and 15.35% versus 1.25% for all five major populations). When each major population is analyzed individually, the $R_{ST}$ values were low but showed significant among group heterogeneity [Bruce *et al.* 2005].

Twelve Y-STR loci were genotyped in the three major ethnic populations in Singapore, namely the Chinese, Malay and Indian. Allele frequency distribution, locus diversity, haplotype diversity and discrimination capacity were estimated. Analysis of molecular variance between the three ethnic populations indicated that 87.71% of the haplotypic variation is found within population and 12.29% is between populations (Fixation Index $F_{ST}$ = 0.123, p = 0.000). Population pairwise comparisons showed significant Fst values between all population pairs,

with the lowest ($R_{ST}$ = 0.05) for Chinese–Malay and the highest ($R_{ST}$ = 0.19) for Chinese–Indian [Yong *et al.* 2006].

In 2006, twenty six Y-STR loci were amplified in a sample of 856 unrelated males from Bhutan, using two multiplex polymerase chain reaction (PCR) assays.. The 26-loci give a discriminating power of 0.9957, though even at this resolution one haplotype occurs 24 times. We identify novel alleles at five loci and microvariants at a further three, which were characterized by sequencing. Extended (11- locus) haplotypes for these samples have been submitted to the YHRD [Parkin *et al.* 2006].

Allele frequencies and haplotypes of the 17 Y-chromosome STRs loci were determined in a sample of 131 healthy unrelated males from the Lassa area of Tibet Autonomy Region of China (SW China). The overall haplotype diversity was 0.9998. The results demonstrate that these loci will be very useful for human identification in forensic cases and paternity tests in the Lassa region [Kang Long Li *et al.* 2007]. In the same year 2007, Parkin and his colleagues Y-STR studied on 769 unrelated males from Nepal, using two multiplex polymerase chain reaction (PCR) assays. They found discriminating power of 0.997, with 59% unique haplotypes, and the highest frequency haplotype occurring 12 times. They also identified novel alleles at four loci, microvariants at a further two, and nine examples of amelogenin-Y deletions (1.2%). Comparison with a similarly sized Bhutanese sample typed with the same markers suggested histories of isolation and drift, with drift having a greater effect in Bhutan.

The observed 17-loci haplotypes and the individual allele frequencies for each locus were estimated, whilst the locus diversity, haplotype diversity and discrimination capacity were calculated in the three ethnic populations. Analysis of molecular variance indicated that 88.7% of the haplotypic variation is found within population and 11.3% is between populations (fixation index $F_{ST}$ = 0.113, p = 0.000). This study has revealed Y-chromosomes with null alleles at several Y-loci, namely DYS458, DYS392, DYS389I, DYS389II, DYS439, DYS448 and YGATA H4; and several occurrences of duplications at the highly polymorphic DYS385 loci. Some of these deleted loci were in regions of the Yq arm that have been implicated in the occurrence of male infertility [Chang *et al.* 2007].

Haplotype frequencies for 11 Y-STR markers in a Romani population (n = 63) from Slovakia, Jats of Haryana (n = 84) and Jat Sikhs (n = 80) from India were determined. AMOVA revealed non-significant differences between the Romanies and significant differences with non-Romani populations. The Macedonian Romani population differed from all Romani populations examined. Frequent haplotypes observed in Romani populations were sporadic in northwest Indian populations. Thirteen out of 316 populations worldwide were found to share the six most frequent haplotypes of the Slovakian Romanies when the screening conditions were narrowed based on the population size to be over 40, the occurrence of the haplotypes was more than one and the sum frequencies of the most frequent haplotypes was at least 0.02. The most common haplotypes were also observed in other Romani groups. When searching with two Indian (Malbar and Malaysian Indian) most frequent haplotypes under the same conditions matches could be detected in all Romani populations except for the Macedonian Romanies. The search with the Jat Sikhs and Jats of Haryana most frequent haplotypes resulted no matches in Romani populations [Nagy *et al.* 2007].

According to Sa´nchez's reports, and those of other participants (2007), they were determined 17 Y-STR in 247 unrelated healthy individuals from the Barcelona metropolitan area (Catalonia, NE Spain). In this sample, all haplotypes were unique. The aim of this study was to acquire haplotype frequencies for mathematic processing of the forensic diagnosis in Barcelona metropolitan area. From the forensic point of view, the combined polymorphisms of the Y-Filer™ kit provide a high diagnostic efficiency. In the same year 2007, Portuguese Scientist Pontes and his co-workers genotyped 17 Y-STR loci for determining allele frequencies from 175 healthy unrelated males and 45 father–son pairs from the North of Portugal. A total of 171 haplotypes were identified, of which 167 were unique and 4 were found in 2 individuals. The haplotype diversity (99.97%) and discrimination capacity (95.43%) were calculated. They reported some non-standard situations, such as allele duplications and mutations. They also reported a case of disputed paternity in which duplicated alleles plus an inconsistency of the transmitted alleles appeared.

China biologist Zhu and colleagues have co-amplified and analyzed 17 T-STRs loci from 167 healthy unrelated autochthonous male individuals of Chinese Tibetan ethnic minority group residing in Qinghai province of China. A total of 163 haplotypes were identified in the set of Y-STR loci, of which 159 were unique and 4 found in two individuals. They observed that the gene diversity values for the Y-STRs loci ranged from a minimum 0.3581 for DYS391 locus to a maximum of 0.8702 for DYS385a/b loci in Tibetan population. The overall haplotype diversity for the set of Y-STRs loci was 0.9998, and the discrimination capacity was 0.9938. Research results will be valuable for human identification and paternity tests in the region and for Chinese population genetic study in the future [Zhu *et al.* 2008].

According to Mizuno's reports and those of other participants (2008), a total of 1079 Japanese males were genotyped. A total of 950 haplotypes for the 16 Y-STR markers were detected and, of these, 886 haplotypes were unique. The most frequent haplotype was found in 22 Japanese males. They observed that the haplotype diversity was 0.9992, indicating a high potential for differentiating between male individuals. They also found that 10 haplotypes with no allele detected at the DYS448 marker. They recommended, the presence of such a typical haplotypes should be noted, when DNA typing results obtained from degraded DNA samples and/or DNA mixture samples from more than one male individual are being interpreted.

In 2008, Taiwanese scientist Huang and co-workers studied on 17 Y-STRs on 200 unrelated Taiwanese males. A total of 192 different haplotypes were found from 200 unrelated individuals, of which 176 were unique, and 8 were found in 2 individuals. The found the gene diversity values of ranged from 0.409 (DYS438) to 0.957 (DYS385a/b). DYS438 is one of the least polymorphic with a GD value of 0.409. The most polymorphic marker in this study was the multi-copy loci DYS385a/b with a GD value of 0.957 [Huang *et al.* 2008]. Another study revealed that diversity values correspond to expected heterozygosity values for autosomal STR loci and provide the probability that two randomly chosen haplotypes will be different. Average autosomal STR gene diversity values were calculated at 0.811 and 0.804 for Caucasian and African American populations. Southern Indian population data collected in this study identified an average Y-STR gene diversity value of 0.657. The haplotype diversity value identified for the Southern Indian population in this study was 0.9989 with 118 unique

haplotypes identified in a population of 123. One gene duplication event was identified at the DYS385 locus in this population. Haplotype data generated in this study has been included in the YHRD Database and is presented in the same format [William *et al.* 2008].

In 2010, Bangladeshi Scientist Alam and his co-workers 17 Y-STR loci genotyped from 216 unrelated Bangladeshi males. A total of 211 haplotypes for the 17 Y-STR markers were detected and, of these, 206 haplotypes were unique. They observed that the haplotype diversity was 0.9998 and indicating a high potential for differentiating between male individuals in this population. They also constructed of Neighbor-Joining (N-J) Tree with the help of AMOVA tool and revealed a close association of Bangladeshi population with Indian Gaddi and Southern Indian populations [Alam *et al.* 2010].

### 1.10 Objectives of the Present Study

In this context, efficient but also affordable Y-STR genotyping systems are needed for positive identification of criminals in incidences of sexual violence. In order to evaluate the genetic markers used for a forensic DNA databases and the compatibility between the manual DNA typing system and the automatic DNA typing system, a testing DNA database should be constructed. These Y-STR systems are powerful forensic tools especially for the area of sexual assault evidence and are superior to autosomal systems. Considering the important issue, the purpose of the study was:

- To determine the Y-Chromosome genetic structure of Garo and Santal ethnic populations in Bangladesh using 17 Y-STR loci namely DYS456, DYS389I, DYS390, DYS389II, DYS458, DYS19, DYS385a/b, DYS393, DYS391, DYS439, DYS635, DYS392, Y_GATA_H4, DYS437, DYS438 and DYS448.
- To analyze 17 Y-STR loci in the Garo and Santal population in order to investigate the genetic relationship between the tribes and the neighboring populations. The results obtained from this study could provide molecular genetic evidence for human settlement of the Pacific.
- To generating population data for the different subpopulations and ethnic groups in the country for different loci.
- To create a database for calculating forensic efficiency parameters to utilize judiciary systems

- To robust assay for accurate characterization of Y markers
- To Y-STR testing for inferring genealogical relationships particularly for surname testing
- To complement other types of DNA ancestral systems such as haplogroup determination.

### 1.11 Importance of the Present Study

Bangladesh is inhabited by many different ethnic groups. Historically, there are mainly four ethnic groups in Bangladesh: Dravidian, proto-Australian, Mongolian and Bengalis. The Bengalis are by far the largest group of all, constituting about 98% of the total population. The Dravidian element of population is represented mainly the Oraons, a tribe of central Indian in origin. The Proto-Australian group includes Khasia and Santals, mainly at Sylhet district and North-West region of Bangladesh. The Mongolian group is confined to the hills along the southern spur of the Shilong plateau, in Chittagong hill tracts and Madhupur forest. Chakama, Tripura, Garo, Murang and Moghs are the major Mongolid tribes in Bangladesh.

Among them Garo is one of the largest indigenous group of Bangladesh. The Garos are mainly distributed over the Kamrup, Goalpara and Karbi Anglong districts of Assam, Garo Hills in Meghalaya, and substantial numbers, about 200,000 are found in greater Mymensingh (Tangail, Jamalpur, Sherpore, Netrakona) and Gazipur, Rangpur, Sunamgonj, Sylhet, Moulovibazar district of Bangladesh.

The Santal is another largest indigenous group of Bangladesh. Santals are the descendants of Austric-speaking Proto-Australoid race. They are mainly living in the Himalayan sub-mountain region in different districts of Rajshahi and Rangpur division. Their principal home is in West Bengal, the forests of adjacent Bihar (Jhadkhand) and Orissa, and Chhota Nagpur. According to Bangladesh Population Census of 1991, the total Santal populace in Bangladesh is almost 261,746 (UNESCO, 2005).

Until today, the autosomal STR and Y-STR DNA data for ethnic and indigenous populations of Bangladesh is limited. The available data is only for general Bangladeshi Bengalis population [Ahmed et al. 2006; Alam et al. 2010], which is inadequate for forensic DNA examinations as

well as population studies. Therefore, to fill the void, for the first time here we report the Y-STR data of two different ethnic populations of Bangladesh based on forensically and genealogical research important of 17 Y-STR markers. This data will contribute to establishment of Y-STR DNA database that may improve the judiciary system in Garo and Santal ethnic populations in Bangladesh. However, a Y-STR haplotype database for these tribes could be of great utility in determining the geographic origin of male individuals. In addition, because the Y chromosome haplotype has remained generally unchanged over the generations, this database can be used for determining the origin of different closely-related populations and contribute to better understanding of geographic effects and population migration.

**Chapter Two**

# MATERIALS AND METHODS

## 2.1 Study Population

In this study, blood samples and buccal swab samples were collected from two different ethnic groups, Garo and Santal, from different places of Bangladesh. Garo is one of the largest indigenous groups of Bangladesh. The Garos are mainly distributed over the Garo Hills in Meghalaya, and substantial numbers, about 200,000 are found in greater Mymensingh (Tangail, Jamalpur, Sherpore, Netrakona) and Gazipur, Sunamgonj, Sylhet, Moulovibazar district of Bangladesh. For study purpose, biological samples were collected of 120 healthy unrelated **Garo male individuals** from Madhupur, Tangail District.

The Santal is another largest indigenous group of Bangladesh. They are mainly living in the Himalayan sub-mountain region in different districts of Rajshahi and Rangpur division. For study purpose, biological samples were collected of 139 healthy unrelated **Santal male individuals** from Dinajpur District.

## 2.2 Ethical Consideration

Blood samples and buccal swab samples were collected from all participants with written informed consent.

## 2.3 Materials

### 2.3.1 Laboratory Instruments

The major equipments that were used to genotyping from blood and buccal samples of Garo and Santal male individuals, respectively, are listed in **Table 1.1**

**Table 1.1:** Major equipments used in the present study

| Instrument Name | Model | Manufacturer |
|---|---|---|
| Genetic Analyzer including Data Collection Sofwere v2.0 GeneMapper ID Software v3.2 | 3100-*Avant* | Applied Biosystems, Foster City, USA |
| Thermal Cycler System | GeneAmp 2700 | Applied Biosystems, USA |
| Real Time PCR System | 7500 | Applied Biosystems, USA |
| Spectrophotometer | NanoDrop-1000 | Thermo Scientific, USA |
| Microcentrifuge Machine | Biofuge Primo | Heraeus, Germany |
| Heat Block | QBT2 | Grant, England |
| Magnetic Stirrer | MR 3001 | Heidolph, Germany |
| Vortex | REAX Control | Heidolph, Germany |
| Micropipettes | Eppendorf Research | Eppendorf, Germany |
| Deionized Water System | Milli-Q | Millipore, Franch |
| pH Meter | PHM210 | Radiometer Analytical, Franch |

### 2.3.2 Chemicals and Reagents

Chemical and reagents used in this study are shown in **Table 1.2.** All chemicals were of analytical and molecular grade.

**Table 1.2:** Chemicals and reagents used in the present study

| Chemicals and Reagents | Company Name |
|---|---|
| TE-Buffer | Sigma-Aldrich, USA |
| Chelex-100 Resin | Bio-Rad, USA |
| Mineral Oil | Applied Biosystems, USA |
| Quantifiler® Y Human Male DNA Quantification Kit | Applied Biosystems, USA |
| AmpFlSTR® Yfiler™ PCR Amplification Kit | Applied Biosystems, USA |
| Matrix Standard | Applied Biosystems, USA |
| LIZ™-500 Size Standard | Applied Biosystems, USA |
| AmpFlSTR®Yfiler Allelic Ladder | Applied Biosystems, USA |
| 3100 POP-4 Polymer | Applied Biosystems, USA |
| Hi-Di Formamide | Applied Biosystems, USA |
| 10X Buffer | Applied Biosystems, USA |

## 2.4 Methods

### 2.4.1 Sample Collection

Blood samples and buccal swab samples were collected from 139 randomly selected unrelated Santal males and from 120 Garo males with written informed consent. Blood samples were collectedin EDTA tube (Yaohua, China), on FTA™ Card (Whatman, UK) and buccal samples were collected in Swab Stick (Shenzhen Cleanmo, China).

### 2.4.2 Chelex DNA Extraction

The DNA from Garo and Santal inhabitants in Bangladesh was extracted from blood and buccal swabs with the Chelex protocol [Singer-Sam *et al.* 1989; Walsh, P.S., 1991].Chelex®100 is a chelating resin that has a high affinity for polyvalent metal ions. It has a particularly high selectivity for divalent ions and differs from ordinary ion exchangers because of its high selectivity for metal ions and its higher bond strength [Lincoln *et al.* 1998].

### 2.4.3 Reagents to be prepared for extraction

- Sterile deionized water (dd water)
- 20% Chelex
- TE buffer

**Reagent Preparation**

**20% and 5% (w/v) Chelex Solutions**

Chelex resin (Chelex 100, sodium form, biotech grade) is available from BioRad (Hercules, CA). Store the Chelex resin at room temperature protected from light. To prepare the Chelex dilutions:

- Add 20 g of Chelex resin to 100 mL of water, and shake moderately (approximately 30 seconds) to mix.
- Allow the Chelex resin to settle (should take no longer than approximately two minutes), then remove the water supernatant and discard.
- Repeat this Chelex resin wash step for a total of three washes.
- Add autoclaved, distilled water to a final volume of 100 mL to make a 20% Chelex solution.

- To prepare a 5% Chelex solution, add 25 mL of the washed 20% Chelex solution to 75 mL of autoclaved, distilled water.
- Measure the pH of the 20% and 5% Chelex solutions; the pH value should be in the range of pH 9-11. Store at 2-6°C, and protected from light.

**TE Buffer (10 mM Tris-HCl, 0.1 mM EDTA, pH 8.0) preparation**

- 10 mL of 1 M Tris-HCl, pH 8.0 and 0.2 mL of 0.5 M EDTA was mixed together
- 990 mL deionized water was added to solution and mixed properly
- 100mL solution was aliquot
- Then solution was autoclaved and stored at room temperature

### 2.5 DNA Extraction Procedure by Chelex Method

#### 2.5.1 Extraction of DNA from FTA™ Card

At first1 ml of TE buffer was taken into an autoclaved 1.5mL microcentrifuge tube. The FTA paper was cut approximately 5X5mm, added to tube and mixed gently. It was incubated at room temperature for 30 minutes and mixed occasionally by inversion or gentle vortexing. The samples were centrifuged for 3 minutes at 10,000–15,000 × g (maximum speed) at room temperature. The supernatant was removed as much as possible. 50 µL of residual supernatant was left in the tube.

Then the FTA™ Paper was washed with dd water and spinned the sample for 2 minutes at 10,000–15,000 × g (maximum speed) at room temperature. The supernatant was removed as much as possible. 50 µL of residual supernatant was left in the tube. Then 20% Chelex (50µL) was added to reach a final volume of 200 µL. It was taken for vortex at high speed for 5 sec. The sample was incubated at 56 °C for 30 minutes.

After complete incubation, the sample was taken for vortexing at high speed for 5–10 seconds. Then the sample was incubated at 100 °C in a heat block for 8 minutes. The sample was taken for vortexing at high speed for 5–10 seconds and spinned in a microcentrifuge for 2–3 minutes at 10,000–15,000 × g (maximum speed) at room temperature. The sample was then ready for DNA quantitation and the PCR amplification process.

### 2.5.2 Extraction of DNA from Buccal Swab Samples

The buccal swabstick was poured into an autoclaved 1.5mL microcentrifuge tube. Then it was incubated at room temperature for 30-60 minutes and mixed occasionally by gentle vortexing. Then the substrate was removed from the microcentrifuge tube. After removing substrate, the sample tube was spinned for 3 minutes at 10,000–15,000 × g (maximum speed) at room temperature. The supernatant was removed as much as possible. 50 µL of residual supernatant was left in the tube.

Then 20% Chelex (50 µL) was added to reach a final volume of 200 µL and 2-4µL Proteinase K was added into the sample tube. It was taken for vortex at high speed for 5 sec. The sample tube was incubated at 56 °C for 30 minutes.

After complete incubation, the sample was taken for vortexing at high speed for 5–10 seconds. Then the sample was incubated at 100 °C in a heat block for 8 minutes. The sample was taken for vortexing at high speed for 5–10 seconds and spinned in a microcentrifuge for 2–3 minutes at 10,000–15,000 × g (maximum speed) at room temperature. The sample was then ready for DNA quantitation and the PCR amplification process.

### 2.6 DNA Quality and Quantification Determination

The efficiency of PCR amplification is influenced by the quality (degree of degradation), purity, and total quantity of DNA in a sample. Lack of amplification is usually due to highly degraded DNA, the presence of PCR inhibitors, insufficient DNA quantity, or any combination of these factors. Nucleic acid samples can be readily checked for concentration and quality using the NanoDrop™ 1000 Spectrophotometer. About 1ul of samples are sufficient to ensure accurate and reproducible results when measuring aqueous nucleic acid samples.

The Optical Density (OD) readings were recorded at 260 and 280 nm along with the ratio between these readings. The DNA concentration was found at a time with OD readings from NanoDrop measurement. Our interested data were DNA quantity (ng/µL), 260/280 ratio (check for protein contamination; ~1.8 value is accepted) and 260/230 ratio (check for other contamination; must be greater than 260/280 value). From NanoDrop measurement, the concentration of all study DNA samples was more than 2ng/µL concentration.

Real Time PCR was also done with a small portion of samples, in order to check the presence of any PCR inhibitor and actual human genome copy number in extracted DNA samples. DNA concentrations were determined using the 7500 Real Time PCR System and the Quantifiler® Y Human Male DNA Quantification Kit.

Since the extracted genomic DNA was in higher concentration for PCR, a portion of the extracted sample was diluted with TE buffer so that only 2.5 ng of DNA is in a volume of 20 µl (final sample concentration 0.125ng/µl).

### 2.7 PCR Amplification of STR Loci

The polymerase chain reaction (PCR) is a technique widely used in molecular biology. Its key components, a DNA polymerase, such as *Taq* polymerase an enzyme used to amplify a piece of DNA by *in vitro* enzymatic replication[Karry Mullis, 1993]. As PCR progresses, the DNA thus generated is itself used as a template for replication. This sets in motion a chain reaction in which the DNA template is exponentially amplified. With PCR it is possible to amplify a single or few copies of a piece of DNA across several orders of magnitude, generating millions or more copies of the DNA piece [Joseph Sambrook *et al.* 2001]. PCR can be extensively modified to perform a wide array of genetic manipulations [Bartlett *et al.* 2003].

### 2.8 Materials for PCR of Y-STR Loci Amplification

The AmpF*l*STR® Yfiler™ PCR Amplification Kit is a short tandem repeat (STR) multiplex assay that amplifies 17 Y-STR loci (DYS456, DYS389I, DYS390, DYS389II, DYS458, DYS19, DYS385a/b, DYS393, DYS391, DYS439, DYS635, DYS392, Y_GATA_H4, DYS437, DYS438, DYS448) in a single PCR reaction [Mulero *et al.* 2006]. This Yfiler kit contains the following materials:

- AmpF*l*STR Yfiler PCR Reaction Mix
- AmpF*l*STR Yfiler Primer Set
- Ampli*Taq* Gold® DNA Polymerase
- AmpF*l*STR Control DNA 007
- AmpFlSTR Yfiler Allelic Ladder

**Preparing the Reactions**

AmpFlSTR®Yfiler Master Mix was prepared by adding the following volumes of reagents to a 1.5mL micro centrifuge tube.

1. Calculate the volume of each component needed to prepare the reactions, using the table below-

**Table 1.3:** Master Mix preparation

| Component | Volume Per Reaction (µL) |
|---|---|
| AmpF*l*STR®Yfiler Kit PCR Reaction Mix | 4.6 |
| AmpF*l*STR®Yfiler Kit Primer Set | 2.5 |
| Ampli*Taq* Gold®DNA Polymerase | 0.4 |
| **Total volume** | **7.5** |

2. Thaw the AmpF*l*STR®Yfiler Kit PCR Reaction Mix, AmpF*l*STR®Yfiler Kit Primer Set and Ampli*Taq* Gold®DNA Polymerase, then vortex 3 to 5 seconds and centrifuge briefly before opening the tubes.
3. Pipette the required volumes of components into an appropriately sized polypropylene tube.
4. Vortex the PCR master mix for 3 to 5 seconds, then centrifuge briefly.
5. Dispense 7.5µL of the PCR master mix into each reaction tube.
6. Add 2µL of sample or control (AmpF*l*STR®Yfiler Control DNA) (0.1ng/µL) and 3 µL of dd water to the appropriate tubes. The final reaction volume should be 12.5µL.
7. Vortex the PCR tubes for 3 to 5 seconds, then centrifuge briefly
8. Program the thermal cycling conditions according to AmpFlSTR®Yfiler Kit PCR Amplification kit use manual 2004.

**Table 1.4:** Thermal Cycling Profile

| Initial Incubation Step | Cycle (30 cycles) | | | Final Extension | Final Hold |
|---|---|---|---|---|---|
| | Denaturation | Annealing | Extension | | |
| 95°C | 94°C | 61°C | 72°C | 60°C | 4°C |
| 11 minute | 1 minute | 1 minute | 1 minute | 80 minutes | Infinite |

9. Load the tube into the thermal cycler.
10. Close the heated cover and Start the run.

### 2.9 Reagents Required for Fragment Analysis

**Hi-Di Formamide**

The highly deionized (Hi-Di) Formamide is used to resuspend samples before electrokinetic injection in capillary electrophoresis systems. Hi-Di Formamide chemically denatures the DNA strands prior to capillary electrophoresis, eliminating the need for heating.

**Size Standard**

During the time of separation of amplicons by capillary electrophoresis, the amplicons are subject to run along with internal size standards, sets of DNA fragments with known size, so that the precise size of the amplicons can be computed relative to the internal size standard. GeneScan™ -500 LIZ® is an internal lane size standard used for sizing Y-STR alleles. The use of an internal lane size standard enables automated data analysis and is also essential for achieving high run to run precision in sizing DNA fragments by electrophoresis. GeneScan™ -500 LIZ Size Standard is designed for sizing DNA fragments in the 35-500 bp range and provides 16 single-stranded labeled fragments of 35, 50, 75, 100, 139, 150, 160, 200, 250, 300, 340, 350, 400, 450, 490, and 500 bases. Each of the DNA fragments is labeled with LIZ, a proprietary fluorophore, which results in a single peak when run under denaturing or native conditions.

**Allelic Ladders**

An allelic ladder is an artificial mixture of the common alleles present in the human population for a particular STR marker. They are generated with the same primers as tested samples and thus provide a reference DNA size for each allele included in the ladder. These allelic ladders serve as a standard like a measuring stick for each STR locus. STR genotyping was performed by comparison of the size of a sample's alleles to size of alleles in allelic ladders for the same loci being tested in the sample. The Y-STR allele designation was based on the number of repeat motifs according to the guidelines set by International Society of Forensic Genetics.

### 2.10 Analysis of Amplified Fragments using Capillary Electrophoresis

The detection analysis used for PCR amplified product was optimized on an ABI PRISM® 3100-*Avant* Genetic Analyzer. Capillary Electrophoresis (CE) can be used to separate ionic

species by their charge and frictional forces. This capillary electrophoresis equipment provides high resolution for short fragments and uses a minimal amount (1.0µl) of PCR product (AmpF*l*STR®Yfiler™ Kit PCR Amplification Kit User's Manual 2004). Before electrophoresis the ABI 3100-*Avant* Genetic Analyzer was spectral calibrated that separates the four different fluorescent dye colors into distinct spectral components. After proper installation of 36 cm capillary array spatial and spectral calibrations were done and passed with Q value 0.9998 and condition 6.100 for spectral calibration using Matrix Standard Set DS-32 (Dye Set F) (User Bulletin, 2005).

### 2.10.1 Samples Preparation for Electrophoresis

Allele sizing of amplified PCR product was performed using the ABI 3100-*Avant* Genetic Analyzer according to the AmpF*l*STR®Yfiler ™ Kit PCR Amplification Kit User's Manual protocol. Prior loading the samples onto the Genetic Analyzer, the loading cocktail was prepared by combining and mixing the Hi-Di Formamide and GeneScan™ 500 LIZ® internal size standard as follows:

**Table 1.5:** Formamide: Size-standard cocktail preparation

| Reagent | Volume (µL) per reaction |
|---|---|
| Hi-Di™ Formamide | 8.7 |
| GeneScan™ 500 LIZ® Size Standard | 0.3 |
| **Total volume** | **9.0** |

1. Pipette the required volumes of components into an appropriately sized polypropylene tube. Vortex the tube, then centrifuge briefly.
2. Into each well of a MicroAmp® 96-Well reaction plate, add
    a. 9 µL of the formamide: size standard mixture
    b. 1 µL of PCR product or Allelic Ladder or 10 µL of Hi-Di formamide for blank
3. Seal the reaction plate with appropriate septa, and then briefly centrifuge the plate to ensure that the contents of each well are mixed and collected at the bottom.
4. Heat the reaction plate in a thermal cycler for 3 min at 95°C.
5. Immediately place the plate on ice for 3 min.
6. Place the plate into the ABI Prism 3100-*Avant* Genetic Analyzer

Sample Analysis for ABI 3100-Avant Genetic Analyzer:

I. Set-up G5 matrix file

II. Use POP-4™ (Performance Optimized Polymer-4)

III. BioLims Project: 3100-Avant_Project

IV. Dye Filter Set: G5

V. Run Module: HIDFragmentAnalysis36_POP4

VI. Analysis Module: GS500Analysis

### 2.10.2 Running the Plate on Genetic Analyzer

To run the plate on the 3100-Avant Genetic Analyzer, at first we select the record of the plate. The plate position indicator changed from yellow to green when linked and the green run button was activated. The active spectral calibration matches verified the dye set and capillary array length. It was verified that the Autoanalysis Manager is running. Then we clicked the green run button and then we clicked the OK button in the Processing Plates Dialog Box.

### 2.10.3 Data Collection and Analysis

The sample run data were collected using Data Collection Software V 2.0 together with an Y-Filer® allelic ladder and positive and negative controls using GeneMapper ID™ Software V3.2.

### 2.10.4 Quality Control for Genotyping

Negative and positive controls were run concurrently with each sample batch.

**Positive control**

Control DNA K01 (0.1 ng/µl) was used. The genotype of K01 DNA (positive control) those were found after capillary electrophoresis. If genotype was not found in this case then it must be confirmatory for PCR component contamination.

**Negative Control**

Deionized water instead of template DNA was used as negative control to check for any DNA contamination. Here no genotype was found because no PCR amplification was occurred as there was no template DNA and this ensures there was no contamination in the PCR component and appropriate for the study. If any genotype was found in this case then it must be confirmatory for PCR component contamination.

**Re-genotyping**

5-10% of the total samples were re-genotyped to confirm reproducibility of the DNA profile.

Our laboratory has participated in the Y-STR haplotype reference database (YHRD) quality assurance exercise in 2008 typing the YHRD core loci as well as additional loci DYS437, DYS448, DYS456, DYS458, DYS635 and Y-GATA-H4 (date: September 11, 2008).

### 2.10.5 Statically Analysis of Data

Different sophisticated software's were used for statistical analysis such as:

1. Allele frequencies were estimated by direct counting. Gene or haplotype diversity (D) was calculated as D = n/ (n-1) (1-∑Pi2), where Pi is the frequency of the ith allele or haplotype and n indicates the number of samples.

2. Discrimination capacity was determined by dividing the number of different haplotypes (not unique haplotype) seen in a given population by the total number of samples.

3. Pair wise values of Φst were calculated to measure the genetic distance corresponding to 17 Y-STR loci of our population as well as others 18 published data from Y-STR Haplotype Reference Database (YHRD) submitted Asian populations of whose most of them are neighbour to Bangladesh geographically, using online AMOVA tool available at http://www.yhrd.org/Population+Analysis/Online+AMOVA.

4. To portray the genetic association among the populations, a Neighbour Joining (NJ) tree was constructed based on the Φst value using the PHYLIP package version 3.6 [ Felsenstein, J. 2005].

5. The tree was visualized with the TreeViewX Software which is freely available at http://darwin.zoology.gla.ac.uk/~rpage/treeviewx/index.html.

6. Matrix of pairwise Fst, co-ancestry coefficient, pairewise differences and measurement of ancestral population sizes and expected heterozygosity were generated by Arlequin Software V3.5 program available at http://cmpg.unibe.ch/software/arlequin3/.

## Chapter Three

# RESULTS

### 3.1 Population Genetic Data

#### 3.1.1 Y-STR Genotyping of DNA Samples

After performing samples collection, DNA extraction, PCR amplification, and capillary electrophoresis of 17 Y-STR loci (DYS456, DYS389I, DYS390, DYS389II, DYS458, DYS19, DYS385a/b, DYS393, DYS391, DYS439, DYS635, DYS392, Y GATA H4, DYS437, DYS438 and DYS448) of Garo and Santal population were genotyped. The present study analyzed Y-STR polymorphisms in 120 male Garo individuals and 139 male Santal individuals. These data can be of considerable utility in ethnological studies and geographical genetics. The genetic data on Y-STRs of Garo and Santal populations are shown in **Appendix 1** and **Appendix 3**, respectively.

Seventeen Y-STRs were analyzed in 120 unrelated Garo males residing in Tangail region of Bangladesh. A total of 99 different haplotypes were identified of which 82 were unique, 14 were shared in two individuals, 2 were shared in three individuals and 1 was shared in four individuals. In total, 99 different haplotypes among 120 individuals which correspond to a discrimination capacity (DC) was 0.825 (99/120). The most frequent haplotype occurs 21 times (17.5%). The most common alleles of the 17 Y-chromosomal STR loci were DYS19*15, DYS389I*12, DYS389II*28, DYS390*23, DYS391*10, DYS392*14, DYS393*12, DYS438*11, DYS439*12, DYS437*14, DYS448*20, DYS456*15, DYS458*17, DYS635*21, Y_GATA_H4*12 and DYS385a/b*13/18.

Another, 17 Y-STR loci were analyzed in 139 unrelated Santal males residing in Dinajpur region of Bangladesh. A total of 129 different haplotypes were identified, of which 120 were unique, 8 were shared in two individuals and 1 was shared in three individuals. Total of 129 different haplotypes among 139 individuals which correspond to a discrimination capacity (DC) was 0.928 (129/139). The most frequent haplotype occurs 10 times (7.19%). The most common alleles of the 17 Y-chromosomal STR loci were DYS19*15, DYS389I*13, DYS389II*29, DYS390*25, DYS391*10, DYS392*13, DYS393*14, DYS438*10,

DYS439*12, DYS437*14, DYS448*18, DYS456*15, DYS458*16, DYS635*21, Y_GATA_H4*11 and DYS385a/b*15/20.

### 3.2 Statistical Analysis of Genetic Data

#### 3.2.1 Allele Frequencies

The allele frequency represents the incidence of a gene variant in a population. Alleles are variant forms of a gene that are located at the same position, or genetic locus, on a chromosome. An allele frequency is calculated by dividing the number of times the allele of interest is observed in a population by the total number of copies of all the alleles at that particular genetic locus in the population. Allele frequencies can be represented as a decimal, a percentage, or a fraction. In a population, allele frequencies are a reflection of genetic diversity. Changes in allele frequencies over time can indicate that genetic drift is occurring or that new mutations have been introduced into the population.

The genotyping information was then converted into allele frequencies by counting the number of times each allele was observed. The distributions of allele frequencies of Garo and Santal population for the 17 Y-STR loci examined have been shown in **Appendix 2** and **Appendix 4**, respectively.

For Garo males, allele frequencies at DYS loci varied from 0.0083 to 0.8250. The highest allele frequency (0.8250) observed in locus DYS391. On the contrary, for Santal males, allele frequencies at DYS loci varied from 0.0072 to 0.8993. The highest allele frequency (0.8993) observed in locus DYS437.

Comparison of Y-STR allele frequencies among Bangladeshi Bengali, Garo and Santal populations shown in **Figure 1.2.**

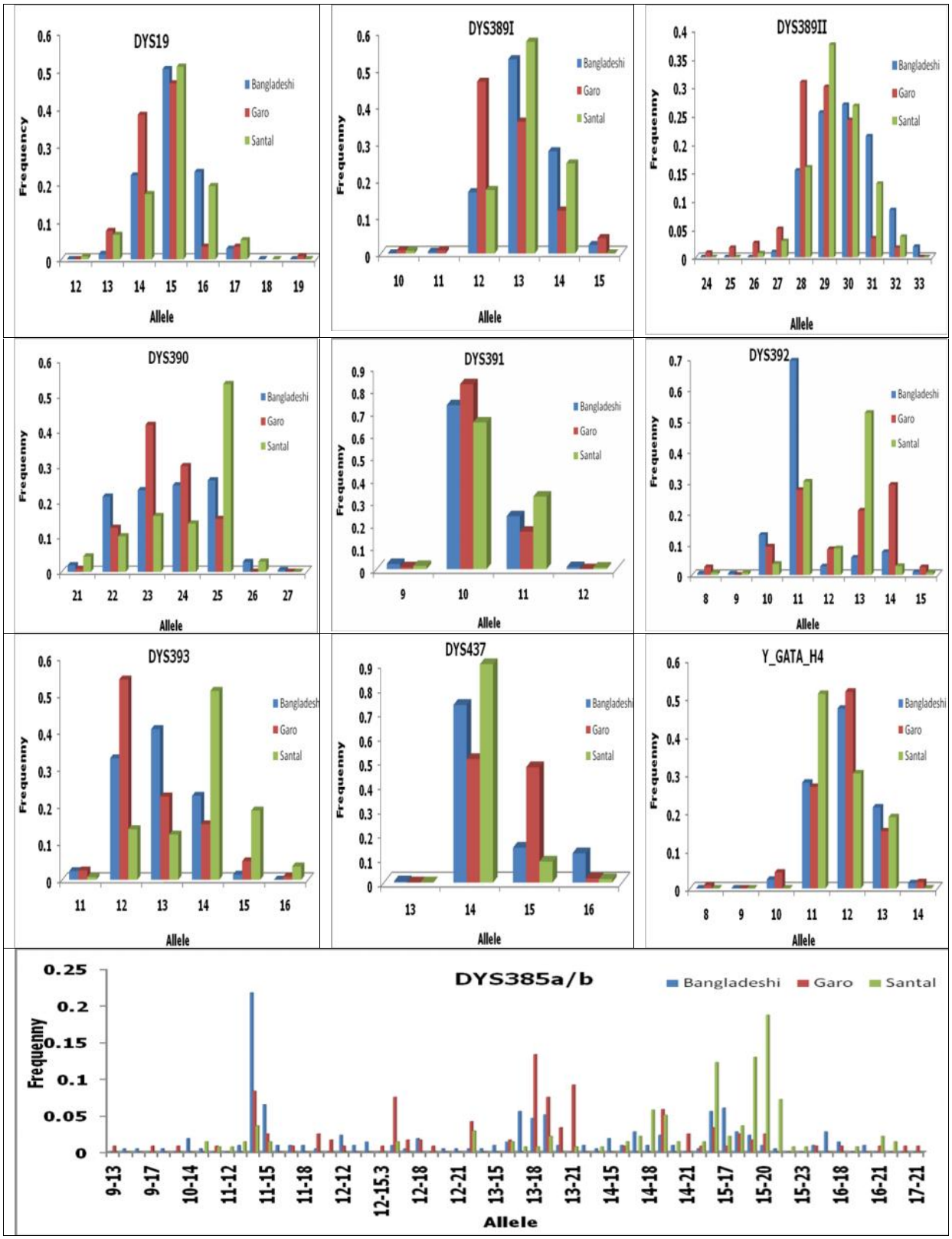


**Figure 1.2:** Allele frequency distribution of ten Y-STR loci. The figures compared Bangladeshi Bengali, Garo and Santal genetic data.

### 3.2.2 Gene or Haplotype Diversity

Gene diversity means the genetic variation between and within species, which is measured by determining the proportion of polymorphic loci across the genome, or by the number of heterozygous individuals in a population. The importance of genetic diversity is evident in terms of survival and adaptability of a species. For instance, a species with high genetic diversity will tend to produce a wider variety of offspring, where some of them may become the most fit variants. In contrast, a species that has little or no genetic diversity will produce offspring that are genetically alike and therefore will likely be susceptible to diseases or problems similar to those of their parent. Hence, little or lack of genetic diversity reduces biological fitness and increases the chances of species extinction.

Allele frequencies and haplotype frequencies were determined using the gene count method. To obtain an estimate of variation in the Garo and Santal population, Y chromosome microsatellite gene diversity and haplotype diversity (D) were calculated based on the following equation for unbiased estimates , $D = n/(n-1)\ (1-\sum P_i 2)$, where $P_i$ is the frequency of the $i$th allele or haplotype and n indicates the number of samples [Nei M. 1987]. The haplotype or gene diversity of these tested population are shown in **Appendix 2** and **Appendix 4**, respectively.

For Garo tribes, DYS385a/b and DYS391 were calculated to be the highest (0.9505) and lowest (0.2939) values for gene or haplotype diversity, respectively. On the other hand, for Santal tribes, DYS385a/b and DYS437 were calculated to be the highest (0.9212) and lowest (0.1849) values for gene or haplotype diversity, respectively. High degree of haplotype diversity means high intra-individual variation of a population.

Comparison of gene or haplotype diversity among Bangladeshi Bengali, Garo and Santal populations represented in the **Figure 1.3.**

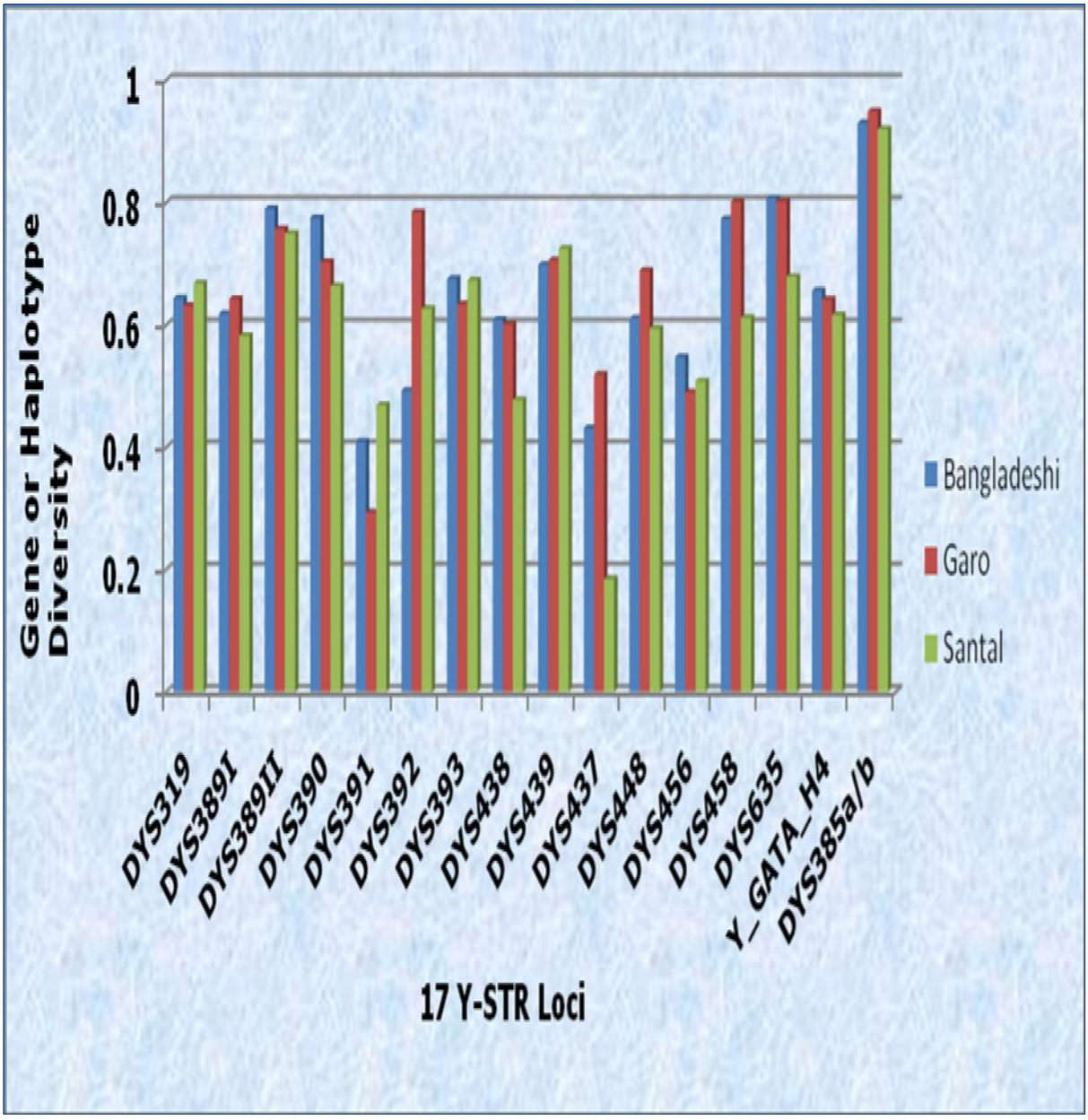


**Figure 1.3:** Comparison of Gene or Haplotype diversity among three population groups of Bangladesh (Bangladeshi, Garo and Santal)

### 3.2.3 Genetic Distance with Pairwise Fst using AMOVA

Fixation Index (Fst) value is standard values for comparison of populations. It is a measure of population differentiation, genetic distance based on genetic polymorphism data, such as microsatallites. It is often expressed as the proportion of genetic diversity due allele frequencies differences among populations. This comparison of genetic variability within and between populations is frequently used in the field of population genetics. The values range from 0 to 1. A zero (0) value implies complete panmixis that the two populations are interbreeding freely. A value of one (1) would imply the two populations are completely separate. The concept was first developed by Sewall Wright in 1920s. Matrix of pairwise Fst values provides little information on ancient demography. With the arrival of molecular data, a number of Fst-like statistics have been proposed to use molecular information (i.e. molecular distances between alleles).

Pair wise values of Φst were calculated to measure the genetic distance corresponding to 17 Y-STR loci of Garo and Santal population as well as from other 17 submitted Asian populations published in Y-STR Haplotype Reference Database (YHRD) whose most of them are neighbor to Bangladesh geographically, using online AMOVA tool (Fst Statistics taking the single step-mutation model into account) including significance tests for all pairwise comparisons [Willuweit S, 2007], which available at http://www.yhrd.org/Population+Analysis/Online+AMOVA.

For having extensive illustration of the genetic relation, 120 Garo individuals from Bangladesh were compared via AMOVA tool using online YHRD with 400 Han individuals from Beijing, China [Kim *et al.* 2001, Accession Number: YA003197], 103 Thai individuals from Chiang Mai, Thailand [Bhoopat *et al.* 2003, A/N: YA002969], 216 Bangladeshi individuals from Dhaka, Bangladesh [Alam *et al.* 2010, A/N: YA003445], 48 Vietnamese individuals from Hanoi, Vietnam [Kwak *et al.* 2005, A/N: YA003201], 77 Nepalese individuals from Kathmandu, Nepal [Gayden T. *et al.* 2009, A/N: YA003656], 573 Korean individuals from Seoul, South Korea [Park *et al.* 2005, A/N: YA003406], 97 Japanese individuals from Tokyo, Japan [Mizuno *et al.* 2008, A/N: YA003459], 65 Tripuri individuals from Tripura, India [Ghosh *et al.* 2011, A/N: YA003627] and 61 Khalkh individuals from Ulaanbaatar, Mongolia [Kwak K. *et al.* 2006, A/N: YA003733].

Nevertheless, a reduction to the core haplotype consisting of 17-YSTR loci mainly returns matches from the neighboring populations in Asia indicating the historical relatedness of populations in this Asian region. In order to analyze the relationship of Garo population living in Bangladesh as well as with other population from the neighboring countries including China, Thailand, Vietnam, Nepal, South Korea, Japan, India and Mongolia. In this study compared 17 Y-STR loci haplotypes via AMOVA with published data from the YHRD. AMOVA was used to establish the total variance among and within groups. AMOVA was done using fixation index analysis and with 10000 permutations. From the AMOVA calculation, we obtained pairwise difference and genetic distance between populations (multidimensional scaling analysis).

In total, 1,705 haplotypes from Asian nine populations were compared for Garo population. All pairwise comparisons between the ten populations for Garo (nine populations published in YHRD, http://www.yhrd.org) analyzed here and the significant at the 5% level but closer vicinity was found to the Garo population (Bangladesh) and Khalkh population (Mongolia). For having extensive illustration of the genetic relation, 139 Santal individuals from Bangladesh were compared via AMOVA tool using online YHRD with 216 Bangladeshi individuals from Dhaka, Bangladesh [Alam *et al.* 2010, A/N: YA003445], 194 African individuals from Guiné-Bissau [Rosa *et al.* 2007, A/N: YA003667], 180 Bhil individuals from Gujarat, India [Sharma *et al.* 2012. A/N: YA003762], 68 Munda individuals from Jharkhand, India [Ghosh *et al.* 2011, A/N: YA003617], 103 Brahmin individuals from Karnataka, India [Zerjal *et al.* A/N: YA003688], 77 Nepalese individuals from Kathmandu, Nepal [Gayden *et al.* 2009, A/N: YA003656], 53 Paliya individuals from West Bengal, India [Roy *et al.* 2010, A/N: YA003629], 22 Rabha individuals from West Bengal, India [Sharma et a. 2011, A/N: YA003679] and 39 Rajbanshi individuals from West Bengal, India [Roy et al. 2010, A/N: YA003630].

Nevertheless, a reduction to the core haplotype consisting of 17-YSTR loci mainly returns matches from the neighboring populations in Asia and one in Africa indicating the historical relatedness of populations in this Asian region. In order to analyzed the relationship of Santal populations living in Bangladesh as well as with other population from the neighboring countries including Nepal, India, Guiné-Bissau (Africa). In this study compared 17 Y-STR loci

haplotypes via AMOVA with published data from the YHRD. In total, 1,091 haplotypes were compared for Santal population. All pairwise comparisons between the ten populations for Santal (nine populations published in YHRD) analyzed here and the other groups were significant at the 5% level but closer vicinity was found to the Santal population (Bangladesh) and Rajbanshi (West Bengal, India). The Φst genetic distances and Φst P-values for pair-wise comparisons of populations shown in **Table 1.6** and **Table 1.7**, respectively.

**Table 1.6:** Φst genetic distances (below diagonal) and Φst P-values (above diagonal) for pair-wise comparisons of Garo and other 9 Asian populations

| | CHI | THA | BAN | VIE | NEP | KOR | JAP | TRI | MON | GAR |
|---|---|---|---|---|---|---|---|---|---|---|
| CHI | | 0.1599 | 0.0001 | 0.0000 | 0.0000 | 0.0000 | 0.0001 | 0.0000 | 0.6335 | 0.0000 |
| THA | 0.0001 | | 0.2793 | 0.1519 | 0.1272 | 0.1726 | 0.2834 | 0.2073 | 1.0000 | 0.0729 |
| BAN | 0.0002 | 0.0001 | | 0.0003 | 0.0000 | 0.0014 | 0.0003 | 0.0006 | 0.7587 | 0.0000 |
| VIE | 0.0030 | 0.0029 | 0.0030 | | 0.0010 | 0.0004 | 0.0024 | 0.0013 | 0.6937 | 0.0005 |
| NEP | 0.0014 | 0.0014 | 0.0015 | 0.0043 | | 0.0004 | 0.0001 | 0.0003 | 0.4744 | 0.0000 |
| KOR | 0.0004 | 0.0003 | 0.0004 | 0.0029 | 0.0016 | | 0.0041 | 0.0011 | 0.5782 | 0.0000 |
| JAP | 0.0007 | 0.0006 | 0.0007 | 0.0034 | 0.0019 | 0.0007 | | 0.0004 | 0.6953 | 0.0000 |
| TRI | 0.0013 | 0.0012 | 0.0012 | 0.0041 | 0.0026 | 0.0014 | 0.0018 | | 0.4706 | 0.0004 |
| MON | 0.0002 | 0.0000 | 0.0002 | 0.0034 | 0.0016 | 0.0003 | 0.0007 | 0.0014 | | 0.1847 |
| GAR | 0.0023 | 0.0022 | 0.0023 | 0.0051 | 0.0035 | 0.0024 | 0.0027 | 0.0034 | 0.0026 | |

Note: CHI=Beijing, China [Han], THA=Chiang Mai, Thailand [Thai], BAN=Dhaka, Bangladesh [Bangladeshi], VIE=Hanoi, Vietnam [Vietnamese], NEP=Kathmandu, Nepal [Nepalese], KOR=Seoul, South Korea [Korean], JAP=Tokyo, Japan [Japanese], TRI=Tripura, India [Tripuri], MON=Ulaanbaatar, Mongolia [Khalkh], GAR= Garo, Bangladesh

**Table 1.7:** Φst genetic distances (below diagonal) and Φst P-values (above diagonal) for pair-wise comparisons of Santal and other 8 Asian and 1 African population

| | BAN | AFR | GUJ | JHA | KAR | KAT | PAL | RAB | RAJ | SAN |
|---|---|---|---|---|---|---|---|---|---|---|
| BAN | | 0.2928 | 0.0000 | 0.0000 | 0.0000 | 0.0001 | 0.0000 | 0.0188 | 0.0000 | 0.0000 |
| AFR | 0.0001 | | 0.1148 | 0.0232 | 0.2074 | 0.1373 | 0.0342 | 0.1830 | 0.0811 | 0.1878 |
| GUJ | 0.0014 | 0.0013 | | 0.0000 | 0.0000 | 0.0001 | 0.0000 | 0.0345 | 0.0007 | 0.0000 |
| JHA | 0.0058 | 0.0058 | 0.0069 | | 0.0000 | 0.0000 | 0.0000 | 0.0231 | 0.0007 | 0.0000 |
| KAR | 0.0009 | 0.0008 | 0.0020 | 0.0065 | | 0.0000 | 0.0000 | 0.0365 | 0.0001 | 0.0000 |
| KAT | 0.0015 | 0.0014 | 0.0025 | 0.0071 | 0.0021 | | 0.0000 | 0.0457 | 0.0003 | 0.0002 |
| PAL | 0.0048 | 0.0048 | 0.0059 | 0.0104 | 0.0055 | 0.0061 | | 0.0217 | 0.0334 | 0.0000 |
| RAB | 0.0022 | 0.0021 | 0.0034 | 0.0080 | 0.0029 | 0.0035 | 0.0070 | | 0.0363 | 0.0294 |
| RAJ | 0.0047 | 0.0048 | 0.0059 | 0.0104 | 0.0054 | 0.0060 | 0.0056 | 0.0069 | | 0.0000 |
| SAN | 0.0008 | 0.0007 | 0.0019 | 0.0055 | 0.0015 | 0.0020 | 0.0053 | 0.0028 | 0.0053 | |

Note: BAN=Dhaka, Bangladesh [Bangladeshi], AFR=Guiné-Bissau [African], GUJ=Gujarat, India [Bhil], JHA=Jharkhand, India [Munda], KAR=Karnataka, India, KAT=Kathmandu, Nepal [Nepalese], PAL=West Bengal, India [Paliya], RAB=West Bengal, India [Rabha], RAJ=West Bengal, India [Rajbanshi], SAN=Santal, Bangladesh

### 3.2.4 Neighbour-Joining (NJ) Tree Generating

To portray the association among the populations, a Neighbour Joining (NJ) tree was constructed based on the $\Phi_{st}$ value using the PHYLIP package V3.6 [ Felsenstein, J. 2005]. The tree was visualized with the TreeView software [ Page, R. D. M. 1996]. The genealogical trees of these populations are shown in Figure 1.4 and Figure 1.5, respectively. In this population study this figure indicated that Garo and Mongolian population have been evolved from the same Most Recent Common Ancestor (MRCA), while among the ten (10) population the initial oldest MRCA was split into three population: Chinese (Han), Thailand (Thai) and another population which is the MRCA for rest eight population.

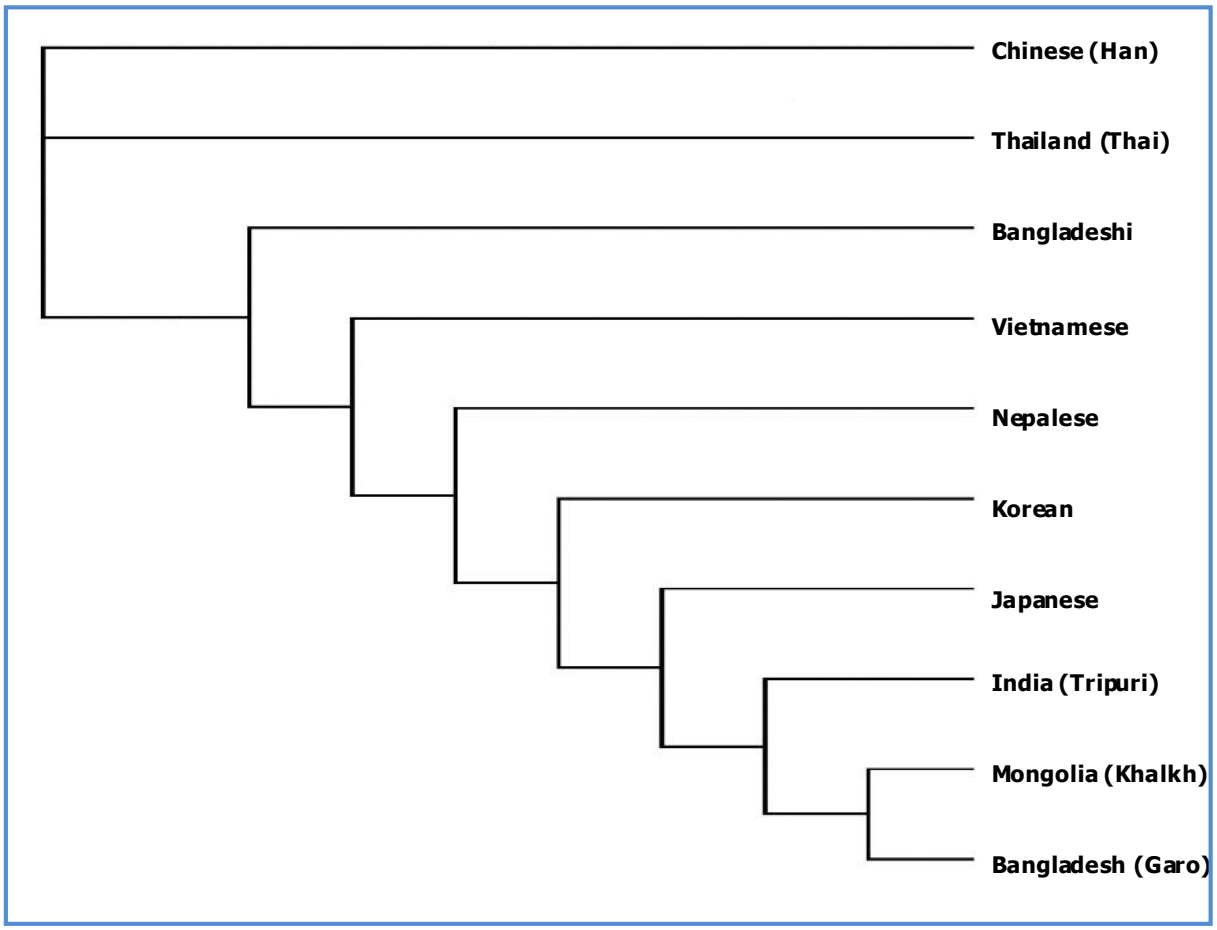


**Figure 1.4**: Neighbor-Joining tree structure from pairwise genetic distance (Φst value) between Garo population (Bangladesh) and Nine Asian population

This Figure 1.5 indicated that Santal (Bangladesh) and Rajbanshi (West Bengal, India) population have been evolved from the same (Most Recent Common Ancestor) while among this ten population, the initial oldest MRCA was split into three population: Guine-Bissau (African), Bangladeshi population and another population which is the MRCA for rest eight.

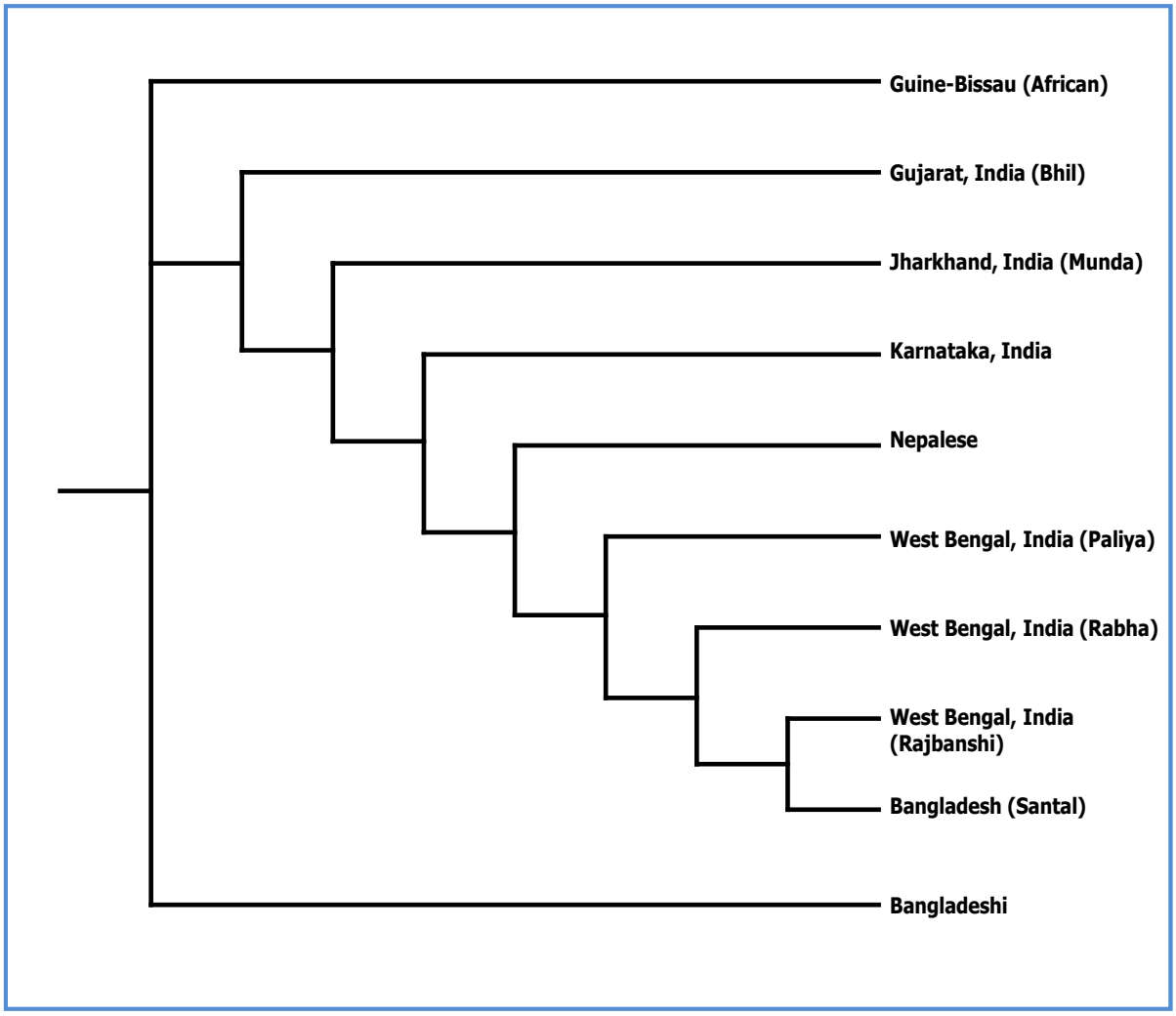


**Figure 1.5:** Neighbor-Joining tree structure from pairwise genetic distance (Φst value) between Santal (Bangladesh) and Eight Asian and One African population

### 3.2.5 Multidimensional Scaling (MDS) Population Plot Designing

Multidimensional scaling (MDS) refers to a broad class of procedures that scale objects based on a reduced set of new variables derived from the original variables [Cox & Cox, 1994]. As the name suggests, MDS is specifically designed to graphically represent relationships between

objects in multidimensional space. The objects are represented on a plot with the new variables as axes and the relationship between the objects on the plot should represent their underlying dissimilarity. All pairwise comparisons between the indigenous Garo and Santal and the other populations show large $\Phi_{st}$ values and are significant at the 5% level. As shown by AMOVA and illustrated in the MDS plot (**Fig. 1.6** and **Fig. 1.7**), the Garo and Santal indigenous tribes show little genetic relationship to other Asian populations tested.

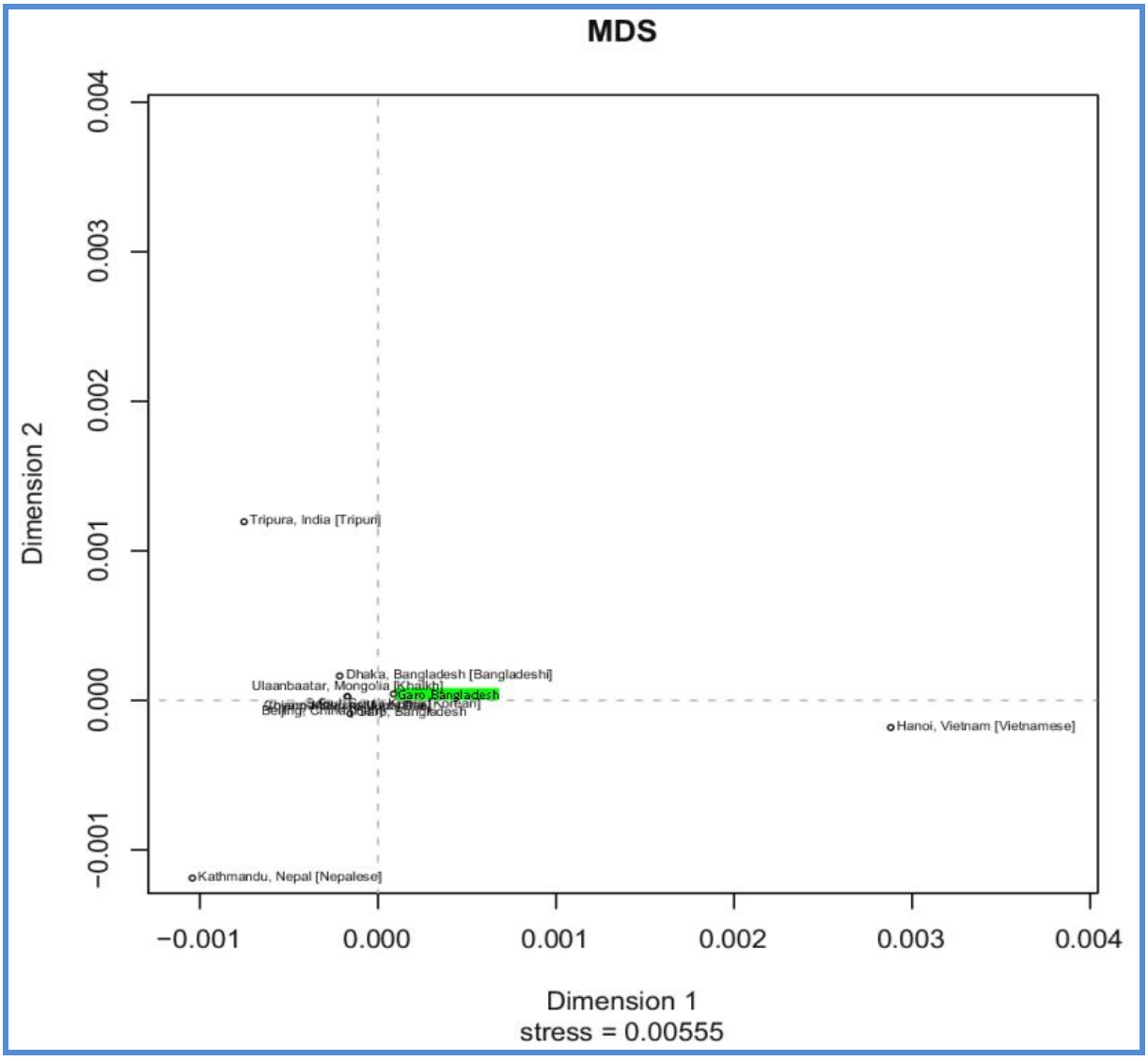


**Figure 1.6:** MDS plot based on pairwise Φst values calculated for the Garo (colored) and nine reference populations of Asia

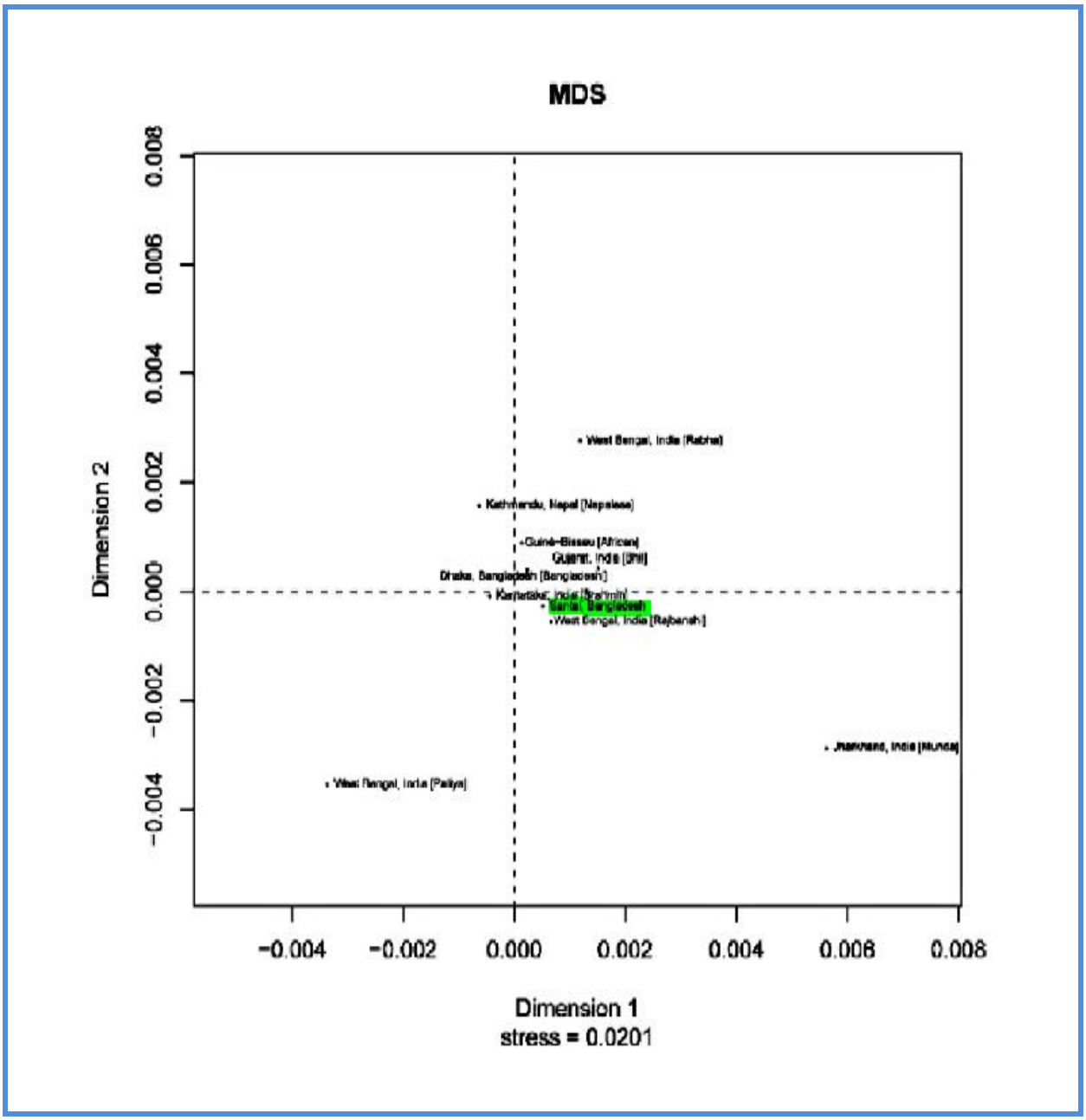


**Figure 1.7:** MDS plot based on pairwise Φst values calculated for the Santal (colored) and eight reference populations of Asia and one African reference population

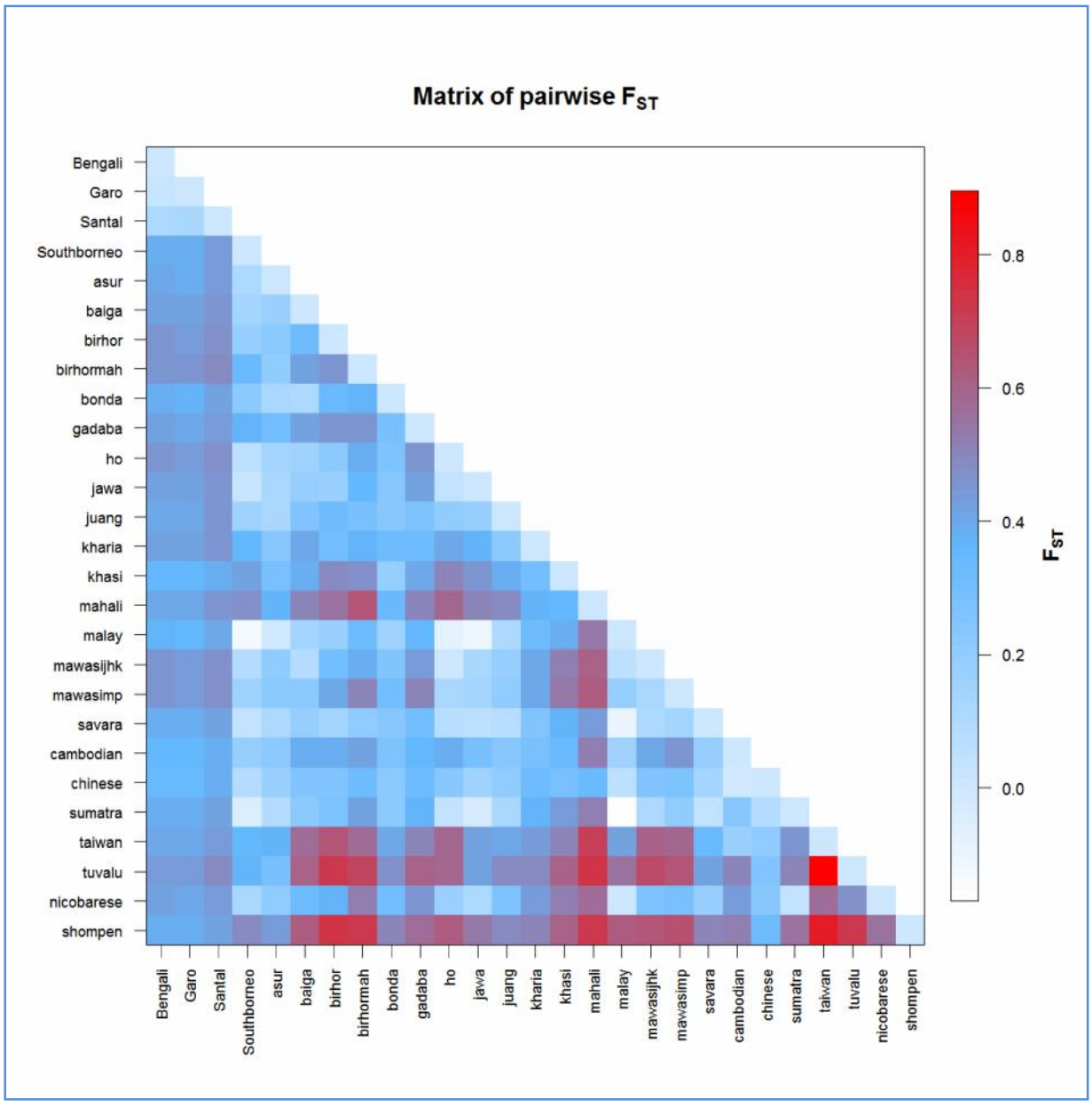


**Figure 1.8: Matrix of Pairwise $F_{ST}$ values indicates short-term genetic distances between populations:** Results indicate that significant genetic divergence with comparisons of $F_{ST}$ parameters among twenty populations. Low Fst value indicates little genetic differences. Here it is seen that pairwise $F_{ST}$ value of Garo, Chinese and Cambodian populations ranges from about 0.1 to 0.3, which means that Garo, Chinese and Cambodian populations are closely related to each other. On the other hand, Santal, Malay and Khasi population ranges from about 0.2 to 0.4, which means Santal, Malay and Khasi population are closely related to each other.

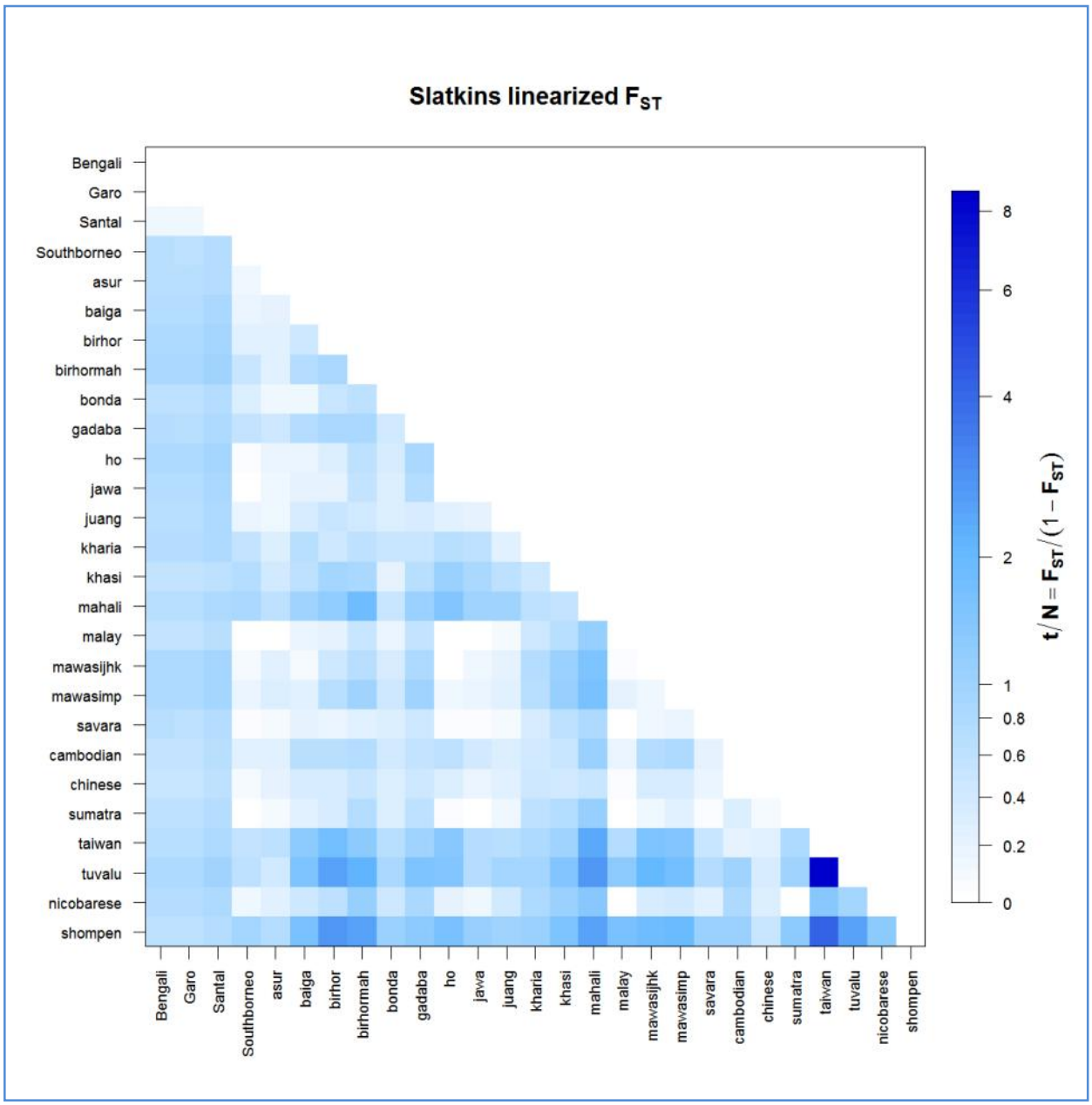


**Figure 1.9: Slatkins Linearized Fst value indicates short-term genetic distances between populations:** Lower the Slatkins Linearized Fst value, closer the two populations. Results indicates that the Slatkins Linearized Fst value among Garo, Chinese and Cambodian population is ranges from 0.4 to 0.6, which means that these three populations are closely related to each other. On the other hand, Santal, Malay and Khasi population ranges from about 0.5 to 0.7, which means these three populations are closely related to each other.

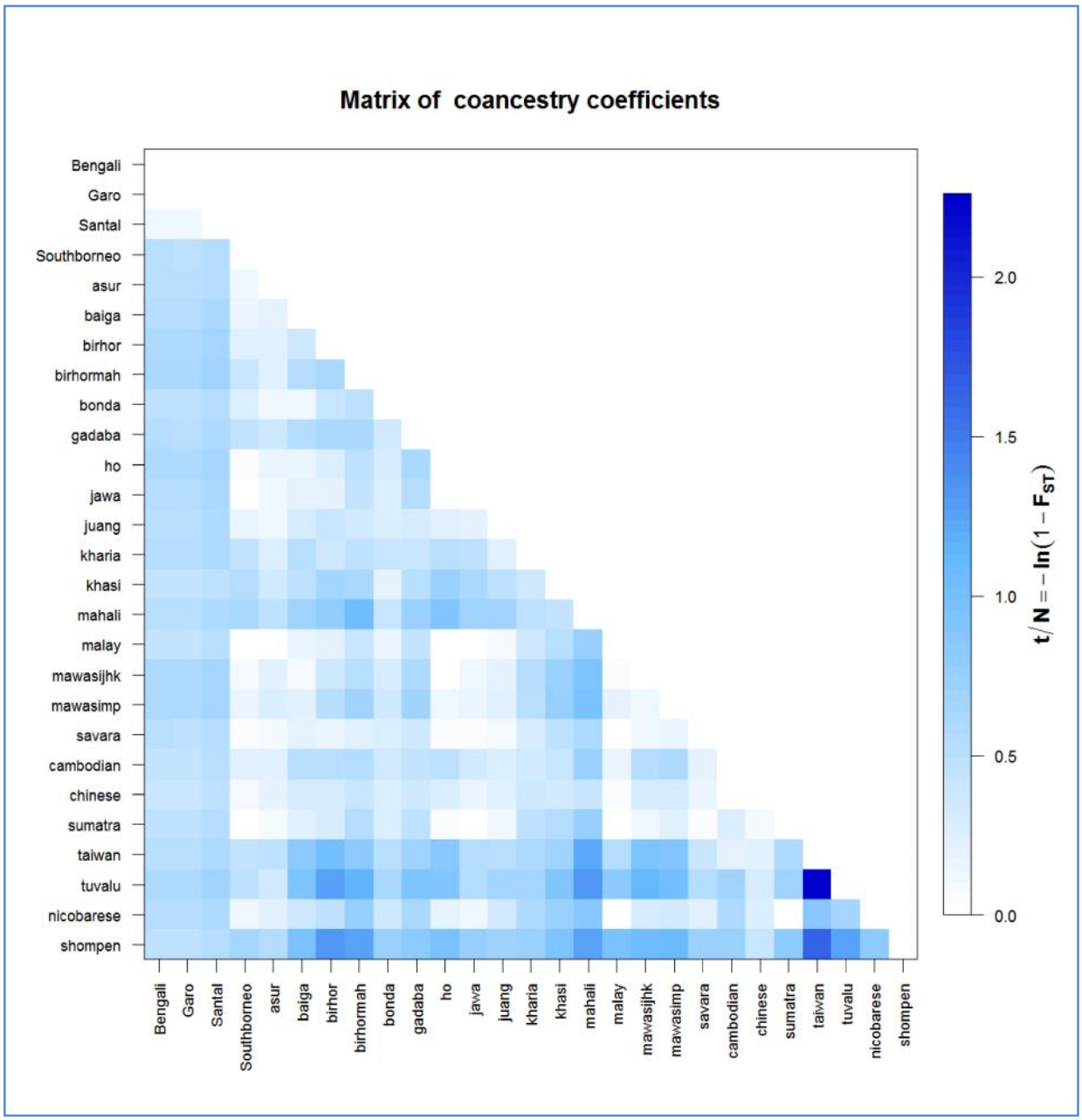


**Figure 1.10: Matrix of co-ancestry coefficients indicates the relationship between different populations:** Lower the co-ancestry coefficient means closer the two populations. Results showed that co-ancestry coefficients among Garo, Chinese and Cambodian populations ranges from about 0.2 to 0.5, which means that Garo, Chinese and Cambodian populations are closely related to each other. On the other hand, Santal, Malay and Khasi population ranges from about 0.5 to 0.6 means these three populations are closely related to each other.

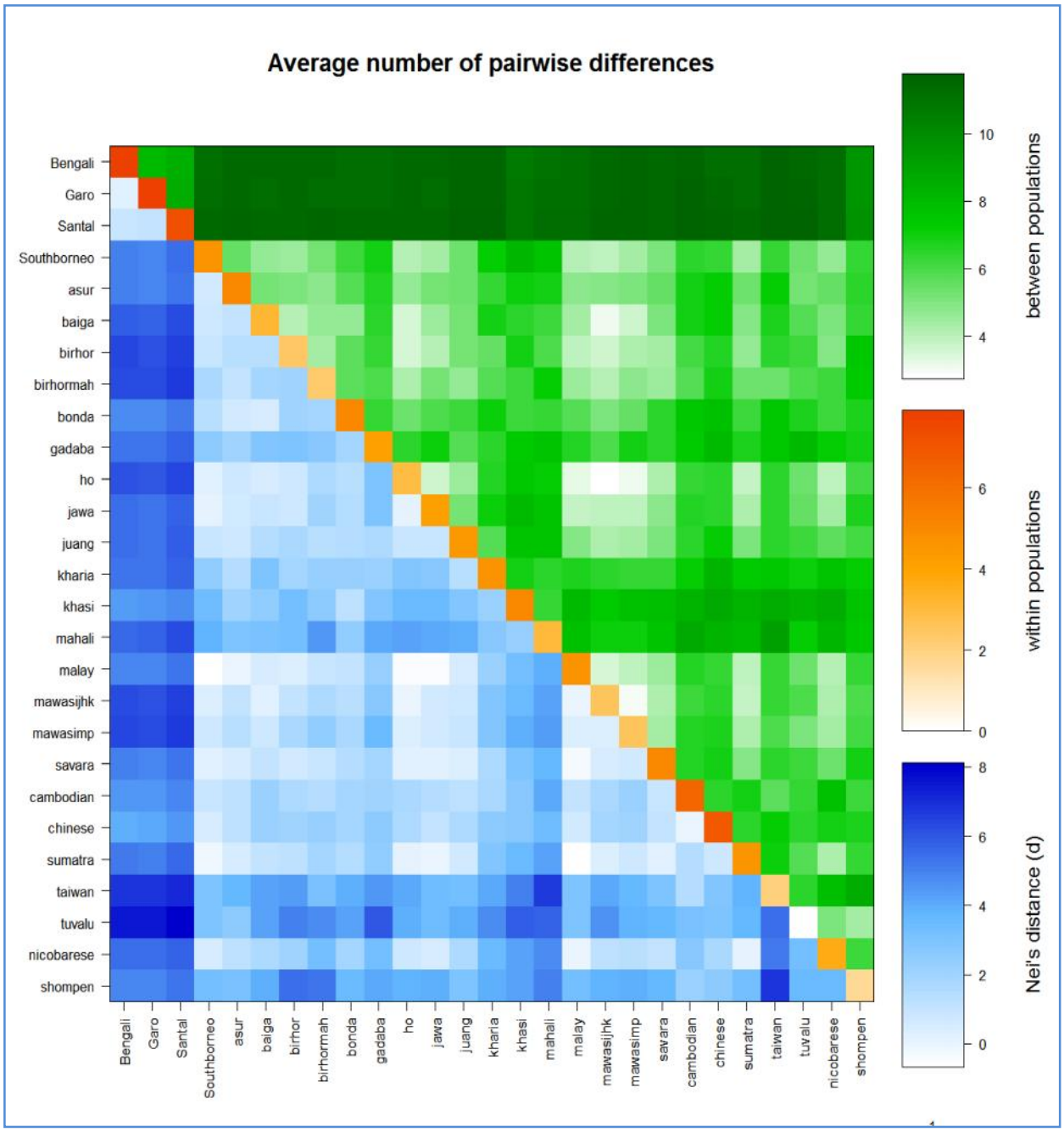


**Figure 1.11: Average number of pairwise differences indicates the relationship between different populations:** The Nei's distance indicates that Garo, Chinese and Cambodian populations are closely related. On the other hand, Santal, Malay and Khasi population are closely related to each other.

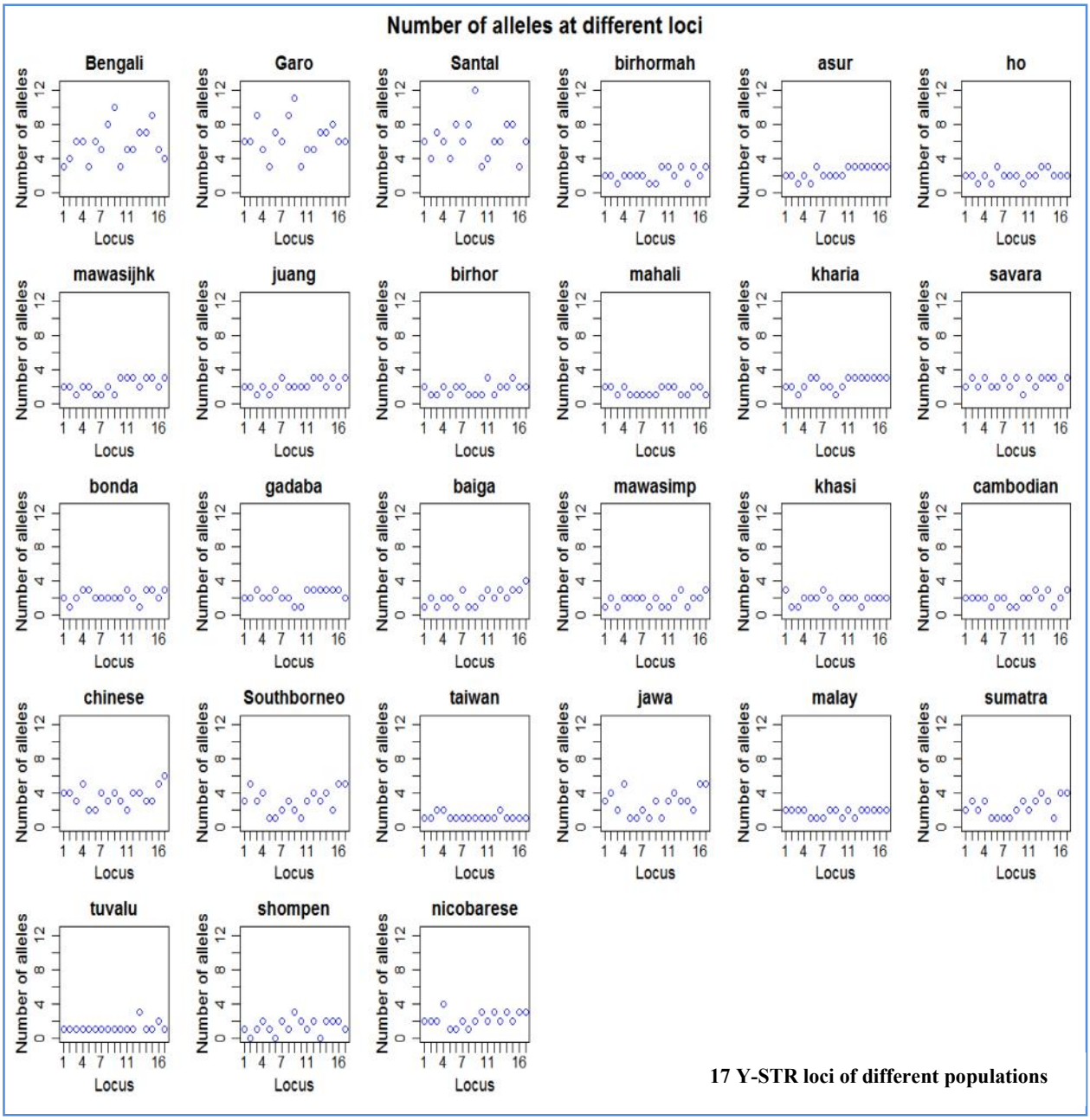


**Figure 1.12: Number of alleles at different loci shows the relationship between different populations:** Number of alleles per locus was compared between different populations. Allele distribution in different loci indicates that Garo and Chinese population populations and Santal and Khasi populations are closely related to each other, respectively.

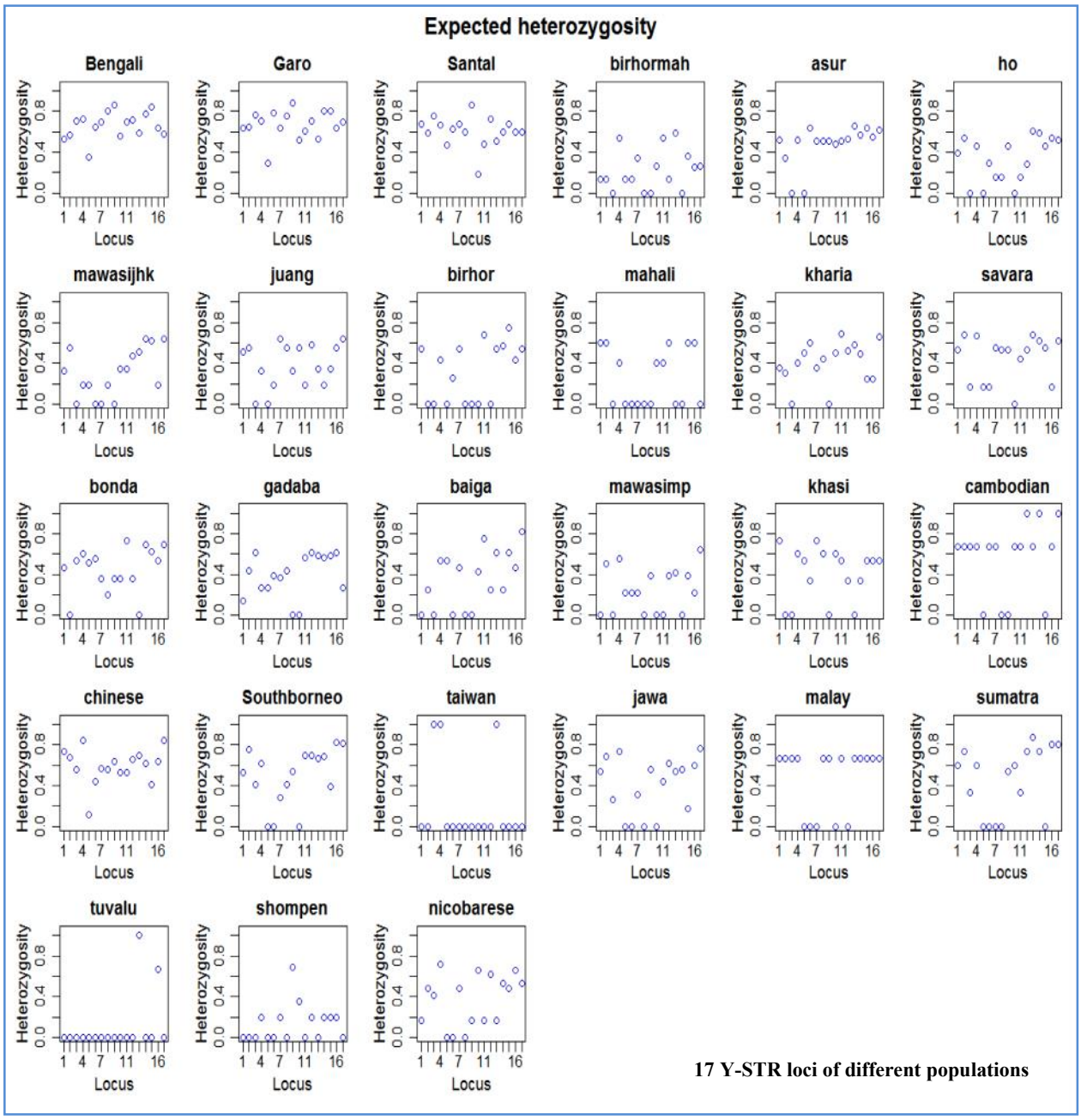


**Figure 1.13: Expected heterozygosity indicates the closely relationship between different populations:** Expected heterozygosity of most of the loci showed that Garo, Chinese and Cambodian populations have close relationship. Results also indicated that Garo is closely related to Bangladeshi mainstream population. On the other hand, results showed that Santal and Khasi populations are closely related.

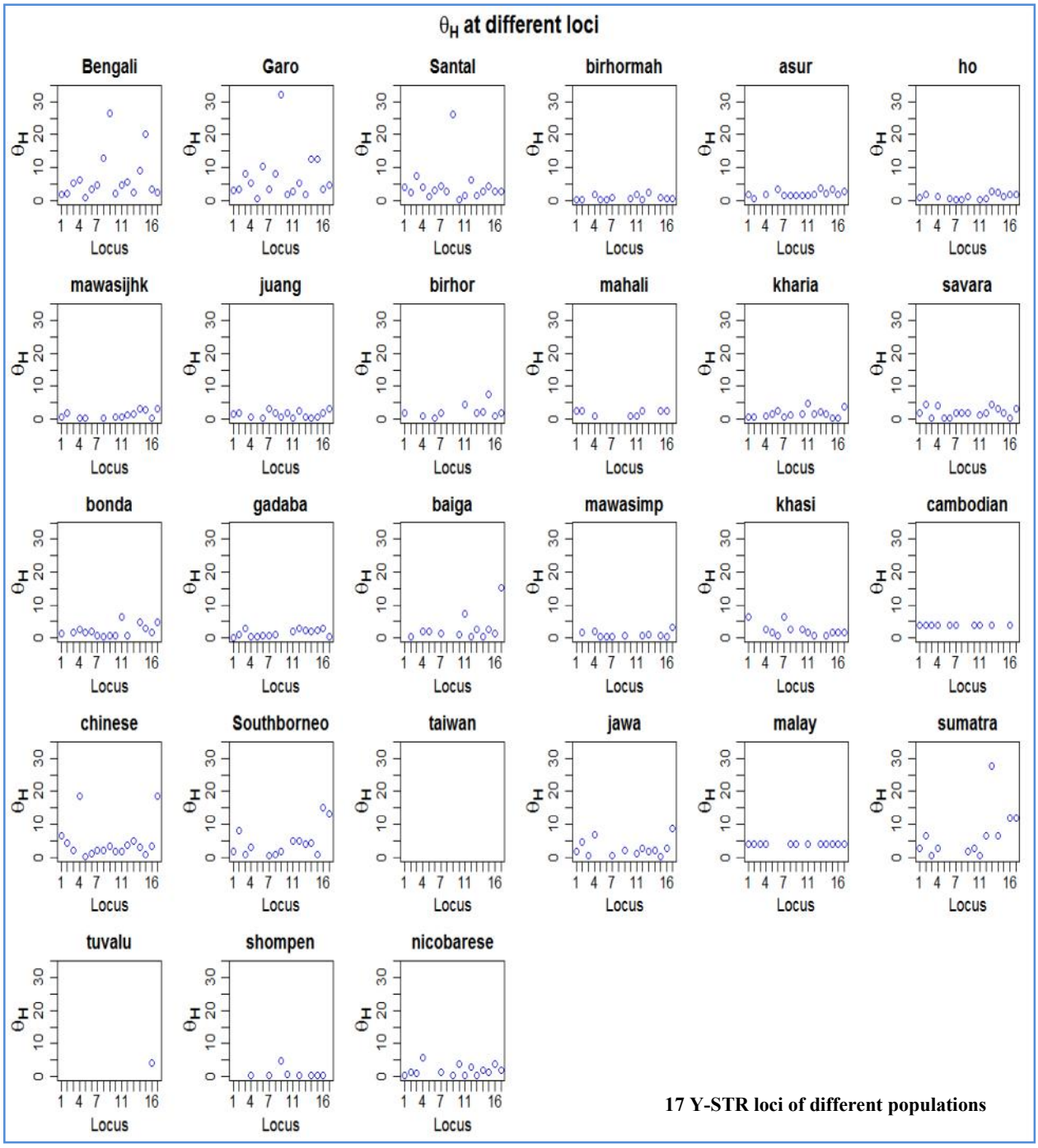


**Figure 1.14: Molecular distance at different loci ($\theta_H$) indicates the molecular distance among various populations based on expected heterozygosity:** $\theta_H$ value of most of the loci indicates Santal and Khasi populations are closely related. On the other hand, Garo, Chinese and Cambodian populations have close relationship.

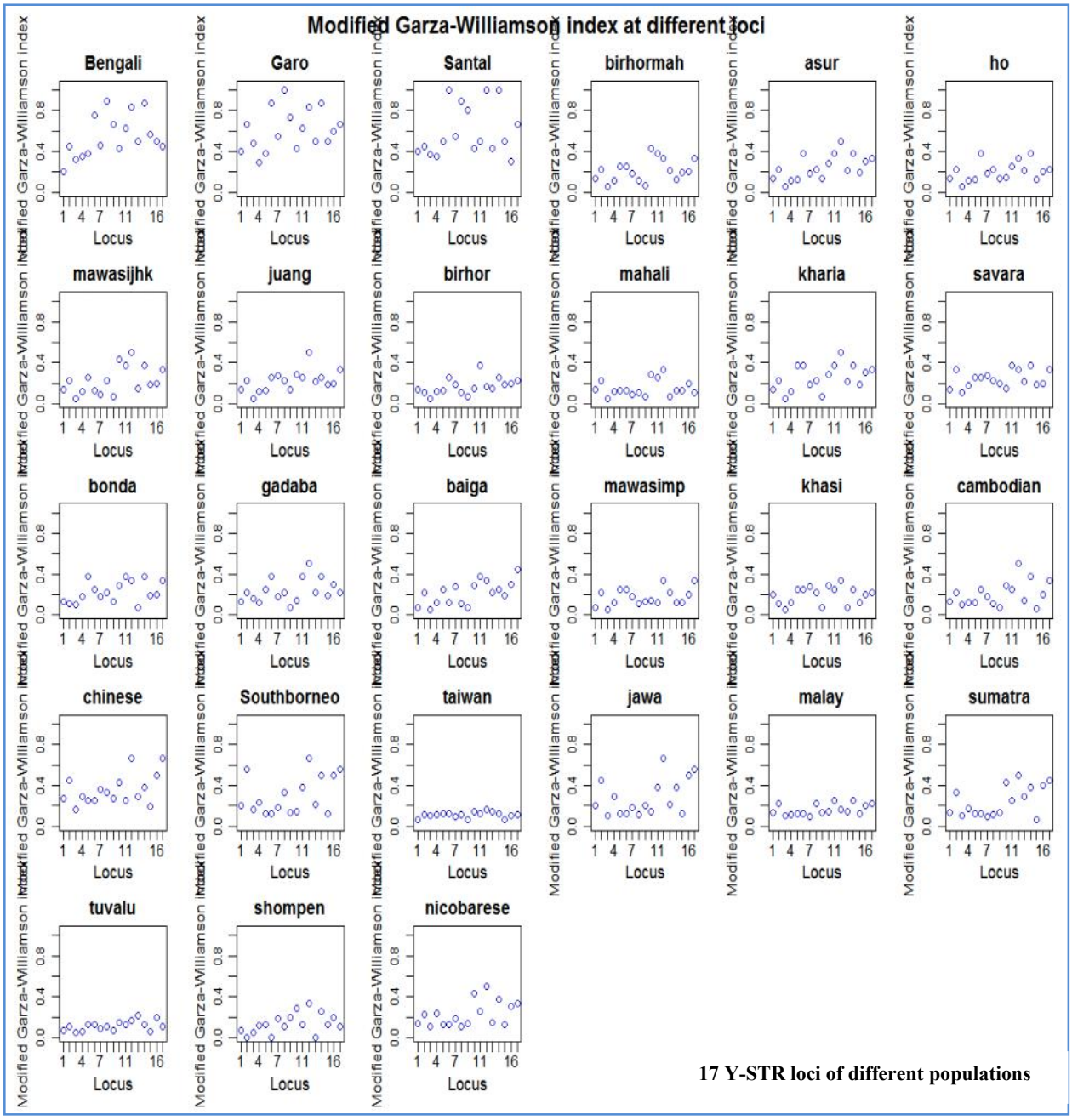


**Figure 1.15: Modified Garza-Williamson index at different loci indicates the molecular distance among different populations:** Modified Garza-Williamson value of Santal and Nicobarese population ranges from 0.4 to 0.9 means they are closely related. On the other hand, Garo and Chinese population ranges from 0.3 to 0.9 means close relationship between Garo and Chinese population.

## Chapter Four

# DISCUSSION

Any two unrelated people around the world are 99.9 percent identical at the genetic level. It means 99.9% (2 billion and 997 million) of our 3 billion base pairs are the same in all people, with 0.1% (3 million) being different because of genetic drift, mutations, natural selection, etc. The amount of variation (0.1%) is not constant throughout the human genome. DNA regions containing repeated nucleotide sequences such as autosomal STRs, X-STRs and Y-STR are of interest to the forensic scientist because its show a lot of variation, which makes them effective for human identification, purposes [Butler J.M., 2001].

The Y chromosome is inherited only from father to son and remains generally unchanged over generations except for gradual accumulation of mutations. This unique mode of inheritance and the absence of recombination with the non-recombining portion of the X chromosome during meiosis lead to the maintenance of polymorphisms inherited through men of the same paternal lineage. This property can be exploited for forensic purposes and paternity testing. The polymorphisms of Y-chromosomal short tandem repeat (Y-STR) loci are a powerful tool for identification and confirmation of shared paternity and the relatedness among individuals belonging to the same paternal lineage.

Based on the allelic frequencies, several statistical parameters of genetic and forensic efficiency were estimated. These include the gene or haplotype diversity, genetic distance with pairwise Fst value, Neighbour-Joining (NJ) Tree, multidimensional scaling (MDS) population plot designing and matrix of pairwise Fst value.

Seventeen Y-STRs were analyzed in 120 unrelated Garo males. A total of 99 different haplotypes were identified, of which 82 were unique, 14 were shared in two individuals, 2 were shared in three individuals and 1 was shared in four individuals. In total, 99 different haplotypes among 120 individuals which correspond to a discrimination capacity (DC) was

0.825 (99/120). The most frequent haplotype occurs 21 times (17.5%). The most frequent haplotype was Ht2, which was shared by 4 individuals.

The most common alleles of the 17 Y-chromosomal STR loci were 15 for DYS19, 12 for DYS389I, 28 for DYS389II, 23 for DYS390, 10 for DYS391, 14 for DYS392, 12 for DYS393, 11 for DYS438, 12 for DYS439, 14 for DYS437, 20 for DYS448, 15 for DYS456, 17 for DYS458, 21 for DYS635, 12 for Y_GATA_H4 and 13/18 for DYS385a/b. The genetic structure of the Garo has therefore been largely conserved, and over half of 17 Y-STR loci typed contained high frequency alleles. At the DYS391 and DYS437 loci, only 3 alleles were recorded with little polymorphic variation. Allele frequencies at DYS loci varied from 0.0083 to 0.8250. The highest allele frequency (0.8250) observed in locus DYS391.

Another 17 Y-STR haplotype were typed from 139 unrelated Santal males. A total of 129 different haplotypes were identified, of which 120 were specific, 8 were shared in two individuals and 1 was shared in three individuals. In total of 129 different haplotypes among 139 individuals which correspond to discrimination capacity (DC) was 0.928 (129/139). The discrimination capacity of the Garo population was lower than that of 0.928 for Santal population (129/139) and 0.976 for Bangladeshi Bengali population (211/216), shown by other studies on individuals that had some degree of relatedness. Such findings might result from the fact that the Garo tribe rarely interact with other peoples in Bangladesh. The most frequent haplotype occurs 10 times (7.19%). The most frequent haplotype was Ht6, which was shared by 3 individuals. The most common alleles of the 17 Y-chromosomal STR loci were 15 for DYS19, 13 for DYS389I, 29 for DYS389II, 25 for DYS390, 10 for DYS391, 13 for DYS392, 14 for DYS393, 10 for DYS438, 12 for DYS439, 14 for DYS437, 18 for DYS448, 15 for DYS456, 16 for DYS458, 21 for DYS635, 11 for Y_GATA_H4 and 15/20 for DYS385a/b. The genetic structure of the Santal has therefore been largely conserved, and over half of 17 Y-STR loci typed contained high frequency alleles. At the DYS437 and Y_GATA_H4 loci, only 3 alleles were recorded with little polymorphic variation. Allele frequencies at DYS loci varied from 0.0072 to 0.8993. The highest allele frequency (0.8993) observed in locus DYS437.

In Garo samples, two microvariant alleles, typed as 17.1 and 15.3 based on fragment size was observed at DYS385a/b locus of haplotype-39 (Ht39) and haplotype-76 (Ht76), respectively.

For the Santal samples, one microvariant allele, typed as 15.3 based on fragment size was observed at DYS385a/b locus of haplotype-102 (Ht102). This variant was found to be reproducible and follows the guidelines of and recommendations by Gusamao, *et. Al.* 2006.

The gene diversity values ranged from 0.2939 (DYS391) to 0.9505 (DYS385a/b). DYS635 (0.8035), DYS458 (0.8022), DYS392 (0.7858) and DYS389II (0.7579), in order of those, are highly diverse in a Garo population. The highest gene diversity (0.9505) has been found in multi-copy locus DYS385a/b (the most polymorphic marker), wherein the most frequent allele has been allele 25 with a frequency of 0.1333. Haplotype diversity value highest (0.9505) indicating a high potential for differentiating between male individuals. The lowest gene diversity (0.2939) has been found in DYS391 locus (the least polymorphic marker), wherein the most frequent allele has been allele 10 with a frequency of 0.8250 (Appendix 2). DYS19, DYS389I, DYS393, DYS438, DYS437, DYS456 and Y_GATA_H4 did not significantly increase the number of different haplotypes and the haplotype diversity. We observed that the haplotype diversity was 0.9505, indicating a high potential for differentiating between male individuals.

On the other hand, in Santal males, gene diversity values ranged from 0.1849 (DYS437) to 0.9212 (DYS385a/b). DYS389II (0.7595), DYS439 (0.7259), DYS635 (0.6799) and DYS393 (0.6740), in order of those, are highly diverse in a Santal population. The highest gene diversity (0.9212) has been found in DYS385a/b locus (the most polymorphic marker), wherein the most frequent allele has been allele 32 with a frequency of 0.1871. The lowest gene diversity (0.1849) has been found in DYS437 locus (the least polymorphic marker), wherein the most frequent allele has been allele 14 with a frequency of 0.8993 (Appendix 4). DYS389I, DYS391, DYS392, DYS438, DYS448, DYS456, DYS458 and Y_GATA_H4 did not significantly increase the number of different haplotypes and the haplotype diversity.

This study showed that, the gene or haplotype diversity values for Garo (0.9505) are higher than the values for Santal (0.9212) and Bangladeshi Bengali population (0.9315) of Bangladesh. These results indicate a high potential for differentiating between male individuals. The gene diversity of DYS385a/b is usually highest among Y-chromosomal STR loci. In this study, from Garo and Santal population, we also found that gene diversity is

highest at DYS385a/b locus. The lowest gene diversity has been found in DYS437 locus in Santal males compared with other Bangladeshi Bengali and Garo males (Figure 1.3).

Our present haplotype data of Garo ethnic group were compared with the previously published data available by means of analysis of molecular variance (AMOVA) based on pairwise Fst comparisons in samples from Beijing, China; Thai individuals from Chiang Mai, Thailand; Bangladeshi individuals from Dhaka, Bangladesh; Vietnamese individuals from Hanoi, Vietnam; Nepalese individuals from Kathmandu, Nepal; Korean individuals from Seoul, South Korea; Japanese individuals from Tokyo, Japan; Tripuri individuals from Tripura, India and Khalkh individuals from Ulaanbaatar, Mongolia and the results are presented in Table 1.6.

Distance and phylogenetic statistics indicated a much closer relationship between Garo population and Khalkh population of Ulaanbaatar, Mongolia and near close to Tripura population of India (Figure 1.4). This study also indicated that Garo and Mongolian population have been evolved from the same Most Recent Common Ancestor (MRCA), while among the ten (10) population the initial oldest MRCA was split into three population: Chinese (Han), Thailand (Thai) and another population which is the MRCA for rest eight populations [i.e. Bangladeshi Bengali (Bangladesh); Vietnamese (Vietnam); Nepalese (Nepal); Korean (South Korea); Japanese (Japan); Tripuri (India); Khalkh (Mongolia) and Garo (Bangladesh)] .

In this study, another haplotype data of Santal ethnic group were compared with the previously published data available by means of AMOVA based on pairwise Fst comparisons in samples of Bangladeshi individuals from Dhaka, Bangladesh; African individuals from Guiné-Bissau; Bhil individuals from Gujarat, India; Munda individuals from Jharkhand, India; Brahmin individuals from Karnataka, India; Nepalese individuals from Kathmandu, Nepal; Paliya individuals from West Bengal, India; Rabha individuals from West Bengal, India and Rajbanshi individuals from West Bengal, India and the results are presented in Table 1.7. Pairwise analysis confirmed that Santal population is more close to Rajbanshi population of West Bengal, India and Rabha population of West Bengal, India. This study also indicated that Santal (Bangladesh) and Rajbanshi (West Bengal, India) population have been evolved from the same MRCA while among this ten population, the initial oldest MRCA was

split into three population: Guine-Bissau (African), Bangladeshi population and another population which is the MRCA for rest eight populations populations [Bangladeshi Bengali (Bangladesh); Guiné-Bissau (African); Bhil (Gujarat, India); Munda (Jharkhand, India); Brahmin (Karnataka, India); Nepalese (Nepal); Paliya (West Bengal, India); Rabha (West Bengal, India) and Rajbanshi (West Bengal, India)] shown in Figure 1.5.

Methods of multidimensional scaling (MDS) or ordination seek a parsimonious representation of individuals in a space of low dimensionality. Parsimony in the context implies that the distances between individuals in ordination space optimally represent their dissimilarities in variable space. MDS is specifically designed to graphically represent relationships between objects in multidimensional space. Based on AMOVA analysis as illustrated with the MDS plot in Figure 1.6 has shown Garo population is closely related to Khalkh individuals from Ulaanbaatar, Mongolia. On the contrary, MDS plot in Figure 1.7 has shown that Santal population is closely related to Rajbanshi population from West Bengal, India.

Matrix of Pairwise Fst can be used as short-term genetic distances between populations, with the application of a slight transformation to linearize the distance with population divergence time. Although Fst has a theoretical range of 0 to 1.0, the observed maximum is usually much less than 1.0. Lower the Fst value, smaller the genetic differences. A measure of the extent of genetic differentiation among populations can range from 0.0 (no differentiation) to 1.0 (complete differentiation). Figure 1.8 indicates that pairwise $F_{ST}$ value of Garo, Chinese and Cambodian populations ranges from about 0.1 to 0.3, which means that Garo, Chinese and Cambodian populations are closely related to each other. After that, Garo has close relation with Bangladeshi, Sumatra and Shompen populations; the Fst value ranges from about 0.2 to 0.4. On the other hand, the pairwise Fst value of Garo and Birhormah population ranges from about 0.5 to 0.6, which means that Garo and Birhormah population are distantly related to each other. On the contrary, Santal, Malay and Khasi population ranges from about 0.2 to 0.4, which means these three populations are closely related to each other. After that, the pairwise Fst value of Santal, Birhormah, Tuvalu and Mawasimp populations ranges from about 0.5 to 0.6, which means that these populations are distantly related to each other.

Slatkins Linearized Fst value indicates short-term genetic distances between populations shown in Figure 1.9. Results indicates that the Slatkins Linearized Fst value among Garo, Chinese and Cambodian population is ranges from 0.4 to 0.5, which means that these three population are closely related to each other. On the other hand, Santal, Malay and khasi population ranges from about 0.5 to 0.7, which means these populations are closely related to each other.

Matrix of co-ancestry coefficients indicates the relationship between different populations. Lower the co-ancestry coefficients, closer the two populations. Figure 1.10 indicates that, co-ancestry coefficients among Garo, Chinese and Cambodian populations ranges from about 0.2 to 0.5, which means that Garo, Chinese and Cambodian populations are closely related to each other. On the other hand, Santal, Khasi, Malay and Nicobarese populations ranges from about 0.5 to 0.6 means that these populations are closely related to each other. Results showed that Garo and Birhormah population are distantly related. On the contrary, Santal, Birhormah, Tuvalu and Mawasimp are distantly related to each other.

Average number of pairwise differences indicates the relationship between different populations with represents three different colors. In Figure 1.11, the Nei's distance indicates that Garo, Chinese and Cambodian populations are closely related. There are significant genetic distances found within Garo population. But there are close genetic distance between Garo and Bangladeshi population. On the other hand, Santal, Malay and Khasi population are closely related to each other. There are close genetic distance among Santal, Birhormah, Tuvalu and Mawasimp populations.

Number of alleles per locus was compared between different populations. Allele distribution in different loci indicates that Garo, Chinese, Cambodian and Bangladeshi mainstream populations have close relationship to each other. On the other hand, Santal and Khasi populations are closely related to each other.

Expected heterozygosity indicates the closely relationship between different populations. Figure 1.13 stated that expected heterozygosity of most of the loci showed that Garo, Chinese and Cambodian populations are closely related to each other. Results also indicated that garo is

closely related to Bangladeshi mainstream population. On the other hand, Santal and Khasi populations have close relationship.

Molecular distance at different loci ($\square_H$) indicates the molecular distance among various populations based on expected heterozygosity. Figure 1.14 illustrated that $\square_H$ value of most of the loci on Garo, Chinese, Cambodian and Bangladeshi population ranges from 0 to 9. This result indicates that these populations are closely related. After that it is closely related to Sumatra population. On the other hand, on Santal and Khasi population it was most of 0 to 8, means these two populations are closely related.

Modified Garza-Williamson index at different loci indicates the molecular distance among different populations. Figure 1.15 showed that Modified Garza-Williamson value of Santal and Nicobarese population ranges from 0.4 to 0.9 means they are closely related. On the other hand, Garo and Chinese population ranges from 0.3 to 0.9 means these two populations are closely related.

## Conclusion

The main aim of this research was to study 17 Y-STR loci in the Garo and Santal population in order to investigate the genetic relationship between the tribes and the neighboring populations. The results also obtained from this study could provide molecular genetic evidence for human settlement of the Pacific and to evaluate the usefulness of these loci for forensic genetic purposes.

The conclusions that have been drawn from the present study for **Garo population** include:

- The number of different alleles observed across the population was found to be 60 alleles. Two microvariants alleles, typed as 15.3 and 17.1 based on fragment size was observed at DYS385a/b locus.
- The highest allele frequencies occurred in the allele 10 for DYS391 locus (0.8250) and the lowest allele frequencies mostly occurred in DYS385a/b (0.0083).

- The DYS385a/b locus showed the largest number of different allelic class (35 allelic class) and DYS391 & DYS437 loci represented the smallest number of different alleles (3 alleles).
- The highest gene or haplotype diversity (0.9505) observed in DYS385a/b locus and the lowest gene diversity (0.2939) observed in DYS391locus. High gene diversity indicating a high potential for differentiating between male individuals.
- The gene or haplotype diversity values for Garo (0.9505) are higher than the values for Santal (0.9212) and Bangladeshi Bengali population (0.9315) of Bangladesh.
- The arrangement of the polymorphic markers in a decreasing order was DYS635, DYS458, DYS392, DYS389II, DYS439, DYS390, DYS448, DYS389I, Y_GATA_H4, DYS393, DYS19, DYS438, DYS437, and DYS456.
- Haplotype data of Garo ethnic group were compared with the YHRD published data by means of AMOVA based on pairwise F*st* comparisons indicated a much closer relationship between Garo population (Bangladesh) and Khalkh population (Mongolia) and near close to Tripura population of India.
- Based on AMOVA analysis as illustrated with the MDS plot has shown Garo population is closely related to Khalkh individuals from Ulaanbaatar, Mongolia.
- Matrix of Pairwise Fst showed that Garo, Chinese and Cambodian populations are closely related to each other.
- According to the statistical parameters, the combined analysis of these 17 Y-STR systems are powerful forensic tools especially for the area of sexual assault evidence and are superior to autosomal systems as well as paternity testing in the Garo ethnic population. This Y-STR system is also investigating the genetic relationship between the tribes and the neighboring populations.

The conclusions that have been drawn from the present study for **Santal population** include:

- The number of different alleles observed across the population was found to be 56 alleles. One microvariant allele, typed as 15.3 based on fragment size was observed at DYS385a/b locus.
- The highest allele frequencies occurred in the allele 14 for DYS437 locus (0.8993) and the lowest allele frequencies mostly occurred in DYS385a/b (0.0072).

- The DYS385a/b locus showed the largest number of different allelic class (31 allelic class) and DYS437 & Y_GATA_H4 loci represented the smallest number of different alleles (3 alleles).
- The highest gene or haplotype diversity (0.9212) observed in DYS385a/b locus and the lowest gene diversity (0.1849) observed in DYS437locus. High gene diversity indicating a high potential for differentiating between male individuals.
- The gene or haplotype diversity values for Santal (0.9212) are lower than the values for Garo (0.9505) and Bangladeshi Bengali population (0.9315) of Bangladesh.
- The arrangement of the polymorphic markers in a decreasing order was DYS389II, DYS439, DYS635, DYS393, DYS19, DYS390, DYS392, Y_GATA_H4, DYS458, DYS448, DYS389I, DYS456, DYS438 and DYS391.
- Haplotype data of Santal ethnic group were compared with the YHRD published data by means of AMOVA based on pairwise F*st* comparisons indicated a much closer relationship between Santal population (Bangladesh) and Rajbanshi population (West Bengal, India) and near close to Rabha population (west Bengal, India).
- Based on AMOVA analysis as illustrated with the MDS plot has shown Santal population from Bangladesh is closely related to Rajbanshi population from West Bengal, India.
- Matrix of pairwise Fst indicates that Santal, Malay and Khasi populations are closely related to each other.
- According to the statistical parameters, the combined analysis of these 17 Y-STR systems are powerful forensic tools especially for the area of sexual assault evidence and are superior to autosomal systems as well as paternity testing in the Santal inhabitants. This Y-STR system is also investigating the genetic relationship between the tribes and the neighboring populations.

## Population Comparisons

Subsequent analysis of every data from the two ethnic groups, it was found that there is great dissimilarity between the data. Moreover, Pair wise values of Φst were calculated to measure the genetic distance corresponding to 17 Y-STR loci of Garo and Santal population as well as from other 17 submitted Asian populations published in YHRD whose most of them are

neighbor to Bangladesh geographically, using online AMOVA tool including significance tests for all pairwise comparisons indicated that the genetic makeup among these groups are almost different. In the study is consistent with previous analyses of the Bangladeshi male population demonstrating the strong effect of population substructures in this region. In this study also shows the inhomogeneity of the Garo and Santal ethnic group which is widely dispersed across Asian population.

## Recommendations

- The results of the current study indicate that the examined 17 Y-STR loci are useful genetic markers for the area of sexual assault evidence to criminal identification and paternity testing of the Garo and Santal ethnic community in Bangladesh.
- Y-STR testing can provide extremely valuable genetic information when other testing has failed, is inconclusive, or not appropriate based on the case or sample type.
- When motherless conditions this Y-STR system has great importance for paternity testing. Because, paternal lineages possess the same Y-STR haplotype (barring mutation).
- This Y chromosome STR system can help us for historical and genealogical research as well as human migration and evolutionary studies.
- Consequently, these Y-STR loci can be used for the establishment of a DNA database that will be beneficial for the population in terms of resolving social and moral disputes and will contribute to improvements in the judiciary system.
- Further studies are needed to characterize the level of polymorphism in X-chromosomes STRs, mitochondrial DNA and SNPs in the ethnic populations in Bangladesh. The stratified population structure must be observed by construction of a national population database for Bangladeshi Bengali and other indigenous population whether planned for autosomal or linear markers.

# REFERENCES

Ahmed, F., Ali, M.E., Alam, S., Hossain, T., Hany, U., Dissing, J. and Skhteruzzaman, S. (2006). Genetic data on 10 autosomal STR loci in the Bangladeshi Population. Legal Medicine; 8:297-299

Aisha, M., Qasim, A., Raheel, Q., Tatiana, Z., Agnar, H., Qasim, M., Chris, T.S. Y-chromosomal STR Haplotypes in Pakistani Populations. Forensic Science International, Volume 118, Issues 2–3, 15 May 2001, Pages 141–146

Alam, S., Ali, M.E., Ahmed, F., Hossain, T., Hasan, M.M. and Akhteruzzaman, S. (2010). Haplotype diversity of 17 Y-chromosomal STR loci in the Bangladeshi population. Forensic Science International: Genetics 4 (2010) e59–e60

Balding, D.J., Greenhalgh, M. and Nichols, R.A. (1996). Population genetics of STR loci in Caucasians. International journal of Legal Medicine; 108:300-305.

Bartlett and Stirling (2003)—A Short History of the Polymerase Chain Reaction. In: Methods Mol Biol. 226:3-6

Betz, A., Bassler, G., Dietl, G., Steil, X., Weyermann, G. and Pflug, W. (2001): DYS STR analysis with epithelial cells in a rape case; Forensic Sci Int 118:126

Bhoopat, T., Hohoff, C., Steger, H.F. (2003) : Identification of DYS385 allele variants by using shorter amplicons and Northern Thai haplotype data., *J Forensic Sci* 48(5), 1108-12

Bosch, E., Calafell, F., Pérez-Lezaun, A., Comas, D., Izaabel, H., Akhayat, O., Sefiani, A., Hariti, G., Dugoujon, J.M. and Bertranpetit, J. (2000): Y chromosome STR haplotypes in four populations from northwest Africa. Int J Legal Med.; 114(1-2):36-40.

Botstein, D.; White, R.; Skolnick, M.; Davis, R. (1980). "Construction of a genetic linkage map in man using restriction fragment length polymorphisms". American Journal of Human Genetics 32 (3): 314–331.

Brookes, A.J. (1999) *Gene*, 234, 177-186.

Bruce, B., Mike, A., Xavier, G.A., Charles, B., Ranajit, C., Dan, C., Bradley, D., Arthur, E., Roger, F., Ann, M.G., Carll, L., Hee-Suk Lee, Scott, C., Carole, M., Mechthild, P., Melanie, L.,

Gabriela, S., Amy, A., Lori, V., Benjamin, E.(2005) : Twelve short tandem repeat loci Ychromosomehaplotypes: Genetic analysis on populations residing in North America. Forensic Science International Volume 150, Issue 1, Pages 1–15

Butler JM, Devaney JM, Marino MA, Vallone PM: Quality control of PCR primers used in multiplex STR amplification reactions; Forensic Sci Int 119:87; 2001.

Butler JM, Ruitberg CM, Vallone PM: Capillary electrophoresis as a tool for optimization of multiplex PCR reactions; Fresenius J Anal Chem 369:200; 2001.

Butler JM, Schoske R, Vallone PM, Kline MC, Redd AJ, Hammer MF: A novel multiplex for simultaneous amplification of 20 Y-chromosome STR markers; Forensic Sci Int 129:10; 2002.

Chang YM., Perumal R., Keat PY., Kuehn DLC. (2007), 'Haplotype diversity of 16 Y-chromosomal STRs in three main ethnic populations (Malays, Chinese and Indians) in Malaysia.', Forensic Sci Int 167(1), 70-6

Collins, J.R., Stepens, R.M., Gold, B., Long, B.m dean, M. and Burt, S.K. (2003) *Genomics*, 82, 10-19

Cox, T.F. and Cox, M.A.A.(1994): Multidimensional Scaling. London: Chapman and Hall

Dettlaff-Kakol A, Pawlowski R: First Polish DNA "manhunt" - an application of Y-chromosome STRs; Int JLegal Med 116:289; 2002.

Felsenstein, J. 2005. PHYLIP (Phylogeny Inference Package) version 3.6. Distributed by the Author. Department of Genome Sciences, University of Washington, Seattle.

Foster EA, Jobling MA, Taylor PG, Donnelly P, de Knijff P, Mieremet R, Zerjal T, Tyler-Smith C: Jefferson fathered slave's last child; Nature 396:27; 1998.

Frankham, Richard (Nov. 2005). Genetics and Extinction: Biological conservation 126(2):131-140.

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(4), 216-23

Ghosh T , Sharma AK, D. Kalpana, Mukerjee S, Mukherjee M , Nath S, Rathod VR, MK Thakar, Ganga Nath Jha.: Genetic diversity of 17 Y-short tandem repeats in Indian population.

Forensic Science International: Genetics; Volume 5, Issue 4, August 2011, P 363–367

Gitschier J. Inferential genotyping of Y chromosomes in Latter-Day Saints founders and comparison to Utah samples in the HapMap project. Am J Hum Genet. 2009 Feb; 84(2):251-8.

Groom, M.J., Meffe, G.K. and Carroll, C.R. (2006). Priciples of Conservation Biology (3rd edi.). Sunderland, MA: Sinauer Associates. Website with additional information: http://www.sinaur.com/groom/

Gusmão, L, J.M. Butler, A. Carracedo, P. Gill, M. Kayser, W.R. Mayr, N. Morling, M. Prinz, L. Roewer, C. Tyler-Smith, P.M. Schneider, DNA Commission of the International Society of Forensic Genetics (ISFG): an update of the recommendations on the use of Y-STRs in forensic analysis, Forensic Sci. Int.157 (2006) 187-197.

Harlt, Daniel (2007). Principles of population Genetics. Sinaur Associates. P.95.

Henke J., Henke L., Chatthopadhyay P., Kayser M., Dulmer M., Cleef S., Poche H., Felske Zech H. (2001), 'Application of Y-chromosomal STR haplotypes to forensic genetics.', Croat Med J 42(3), 292-7

Hou YP, Zhang J, Li YB, Wu J, Zhang SZ, Prinz M. (2001). Allele sequences of six new Y-STR loci and haplotypes in the Chinese Han population. Forensic Sci Int. 2001 May 15;118 (2-3):147-52.

Huang TY, Hsu YT, Li JM, Chung JH, Shun CT. Polymorphism of 17 Y-STR loci in Taiwan population. Forensic Sci Int. 2008 Jan 30;174(2-3):249-54. Epub 2007 Jul 13

Jeffreys, A.; Wilson, V.; Thein, S. (1985). "Hypervariable 'minisatellite' regions in human DNA". Nature 314 (6006): 67–73.

Jobling M.A., Pandya A., Tyler-Smith,C. The Y chromosome in forensic analysis and paternity testing, Int.J. Leg. Med. 110 (3) (1997) 118– 124.

Jobling MA, Samara V, Pandya A, Fretwell N, Bernasconi B, Mitchell RJ, Gerelsaikhan T, Dashnyam B, Sajantila A, Salo PJ, Nakahori Y, Disteche CM, Thangaraj K, Singh L, Crawford MH, Tyler-Smith C: Recurrent duplication and deletion polymorphisms on the long arm of the Ychromosome in normal males; Hum Mol Genet 5:1767; 1996.

Jobling MA, Tyler-Smith C: New uses for new haplotypes the human Y-chromosome, disease and selection; Trends Genet 16:356; 2000.

Jobling MA: In the name of the father — Surnames and genetics; Trends Genet 17:353; 2001

Jobling, M.A. and Tyler-Smith, C. "Fathers and Sons: the Y chromosome and Human Evolution," *Trends in Genetics*, 11, 449-456 (1995).

Joseph Sambrook and David W. Russel (2001). Molecular Cloning: A Laboratory Manual, 3rd ed., Cold Spring Harbor, N.Y.: Cold Spring Harbor Laboratory Press. ISBN 0-87969-576-5. Chapter 8: In vitro Amplification of DNA by the Polymerase Chain Reaction

Kang Long Li , Liu Kai, Ma Yuming. Y chromosome STR haplotypes of Tibetan Living Tibet Lassa. Forensic Science International 172 (2007) 79–83

Karry Mullis. Nobel Lecture, December 8, 1993

Kayser M, Brauer S, Schädlich H, Prinz M, Batzer MA, Zimmerman PA, Boatin BA, Stoneking M. Y Chromosome STR Haplotypes and the Genetic Structure of U.S. Populations of African, European, and Hispanic Ancestry. Genome Res. 2003 Apr;13(4):624-34.

Kim YJ., Shin DJ., Kim JM., Jin HJ., Kwak KD., Han MS., Choi SK., Kim W. (2001), 'Y-chromosome STR haplotype profiling in the Korean population.', *Forensic Sci Int* **115**(3), 231-7

Kwak K.W, Suren G, Tundewrentsen, Jin Han J, Hong S.B., Han M.S, Kim W 2006: Y-chromosome STR haplotype profiling in the Mongolian population. J. Leg. Med. Vol.8; No.1; Page.58-61(2006)

Kwak KD., Jin HJ., Shin DJ., Kim JM., Roewer L., Krawczak M., Tyler Smith C., Kim W. (2005), 'Y-chromosomal STR haplotypes and their applications to forensic and population studies in east Asia.', *Int J Legal Med* **119**(4), 195-201

Lincoln, P.J. and Thomson, J. (1998). Methods in molecular Biology, Vol: 98: Forensic DNA Profiling Protocols

Manfred K, Michael K, Laurent E, Patrick D, Daniel C, Vincente P, Christian G, Luigi F. Bi, Jørgen J, Egbert B, Lutz R, and Peter de K, Am. J. Hum. An Extensive Analysis of Y-Chromosomal Microsatellite Haplotypes in Globally Dispersed Human Populations. Genet. 68:990–1018, 2001

Mizuno N., Nakahara H., Sekiguchi K., Yoshida K., Nakano M., Kasai K. (2008), '16 Y

chromosomal STR haplotypes in Japanese.', Forensic Sci Int 174(1), 71-6

Mulero J.J, C.W. Chang, L.M. Calandro, R.L. Green, Y. Li, C.L. Johnson, L.K. Hennessy, Development and validation of the AmpFlSTRYfiler PCR amplification kit: a male specific, single amplification 17 Y-STR multiplex system. J. Forensic Sci. 51 (2006) 64–75.

Nagy M., Henke L., Henke J., Chatthopadhyay PK., Volgyi A., Zalan A., Peterman O., Bernasovska J., Pamjav H. (2007), 'Searching for the origin of Romanies: Slovakian Romani, Jats of Haryana and Jat Sikhs Y-STR data in comparison with different Romani populations.', Forensic Sci Int 169(1), 19-26

Nei M, Molecular Evolutionary Genetics, Columbia University Press, New York, 1987, pp. 176–179.

Page,R.D.M. 1996. TREEVIEW: An application to display phylogenetic trees on personal computers. Computer Applications in the Biosciences 12: 357-358.

Park MJ., Lee HY., Yoo JE., Chung U., Lee SY., Shin KJ. (2005), 'Forensic evaluation and haplotypes of 19 Y-chromosomal STR loci in Koreans.', *Forensic Sci Int* **152**(2-3), 133-47

Parkin EJ., Kraayenbrink T., Opgenort JRML., Van Driem GL., Tuladhar NM., De Knijff P., Jobling MA. (2007), 'Diversity of 26-locus Y-STR haplotypes in a Nepalese population sample: isolation and drift in the Himalayas.', Forensic Sci Int 166(2-3), 176-81

Parkin EJ., Kraayenbrink T., Van Driem GL., Tshering Of Gaselo K., De Knijff P., Jobling MA. (2006), '26-Locus Y-STR typing in a Bhutanese population sample.', Forensic Sci Int 161(1), 1-7

Pontes ML, Cainé L, Abrantes D, Lima G, Pinheiro MF. (2007). Allele frequencies and population data for 17 Y-STR loci (AmpFlSTR Y-filer) in a Northern Portuguese population sample. Forensic Sci Int. 2007 Jul 20; 170 (1):62-7.

Postlethwalt, John (2009). Modern Biology. Holt, Rinehart and Winston. P.317

Rosa A, Ornelas C, Jobling MA, Brehm A., and Villems R . (2007).Y-chromosomal diversity in the population of Guinea-Bissau: a multiethnic perspective. BMC Evol Biol. 2007 Jul 27;7:124.

Roy S., Noor S., Haque I. (2010), 'Genetic analysis of 17 Y-chromosomal STRs haplotypes of

three ethnic groups residing in West Bengal, India.', Forensic Science International: Genetics Epub ahead of print, in press

Sánchez C, Barrot C, Xifró A, Ortega M, de Aranda IG, Huguet E, Corbella J, Gené M. (2007). Haplotype frequencies of 16 Y-chromosome STR loci in the Barcelona metropolitan area population using Y-Filer kit. Forensic Sci Int. 2007 Oct 25;172(2-3):211-7.

Sharma G, Tamang R, Chaudhary R, Singh VK, Shah AM, et al. (2012) Genetic Affinities of the Central Indian Tribal Populations. PLoS ONE 7(2): e32546. doi:10.1371/journal.pone.0032546

Singer-Sam, J., and Tanguay, R. 1989. Use of Chelex to improve the PCR signal from a small number of cells. Amplifications 3(1989)11.

Skoog D.A.; Holler, F.J.; Crouch, S.R "Principles of Instrumental Analysis" 6th ed. Thomson Brooks/Cole Publishing: Belmont, CA 2007

Tilford CA, Kuroda-Kawaguchi T, Skaletsky H, Rozen S, Brown LG, Rosenberg M, McPherson JD, Wylie K, Sekhon M, Kucaba TA, Waterston RH, Page DC: A physical map of the human Y-chromosome; Nature 409:943; 2001.

Tisdell, C. (2003). Socioeconomic causes of loss of animal genetic diversity: analysis and assessment. Ecological Economics: 45(3):365-376.

UNESCO, 2005, First Language First: Community-based Literacy Programmes for United Nations, 2007, “The situation of human rights and fundamental freedoms of indigenous people” general assembly.

Walsh, P.S., Metzger, D.A., and Higuchi, R. 1991. Chelex 100 as a medium for simple extraction of DNA for PCR-based typing from forensic material. Biotechniques 10:506-518.

William E. Frank, Heather C. Ralph,and Mohammad A. Tahir,: Y Chromosome STR Haplotypes and Allele Frequencies in a Southern Indian Male, J Forensic Sci, January 2008, Vol. 53, No. 1

Willuweit S., Roewer L. (2007), 'Y chromosome haplotype reference database (YHRD): Update', Forensic Science International: Genetics 1(2), 83-7

Yong, R.Y.Y, Lee, L.K.H, Yap, E.P.H. (2006). Y-chromosome STR haplotype diversity in

three ethnic populations in Singapore. Forensic Science International 159 (2006) 244–257

Zerjal T, A Pandya, K Thangaraj, EYS Ling and J Kearley: Y-chromosomal insights into the genetic impact of the caste system in India. Human genetics (2007), 121:137-144

Zhu B, Wu Y, Shen C, Yang T, Deng Y, Xun X, Tian Y, Yan J, Li T. Genetic analysis of 17 Y-chromosomal STRs haplotypes of Chinese Tibetan ethnic group residing in Qinghai province of China. Forensic Sci Int. 2008 Mar 5; 175(2-3):238-43. Epub 2007 Jul 30.

**Appendix 1: A list of Y Chromosome STR haplotypes detected in 120 unrelated males in Garo population**

| Haplo type | *N* | DYS 19 | DYS 389I | DYS 389II | DYS 390 | DYS 391 | DYS 392 | DYS 393 | DYS 385 a/b | DYS 438 | DYS 439 | DYS 437 | DYS 448 | DYS 456 | DYS 458 | DYS 635 | Y_GATA_H4 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Ht1 | 1 | 15 | 12 | 28 | 24 | 10 | 14 | 12 | 13-19 | 11 | 12 | 15 | 20 | 15 | 19 | 22 | 12 |
| Ht2 | 4 | 13 | 13 | 29 | 24 | 10 | 11 | 13 | 14-19 | 11 | 10 | 14 | 17 | 15 | 15 | 22 | 11 |
| Ht3 | 2 | 15 | 12 | 28 | 22 | 10 | 13 | 12 | 13-21 | 11 | 13 | 15 | 20 | 15 | 17 | 21 | 12 |
| Ht4 | 1 | 14 | 12 | 29 | 22 | 10 | 14 | 11 | 16-16 | 11 | 14 | 15 | 19 | 15 | 15 | 24 | 12 |
| Ht5 | 1 | 19 | 12 | 26 | 23 | 10 | 14 | 12 | 14-21 | 11 | 11 | 15 | 20 | 15 | 18 | 20 | 13 |
| Ht6 | 1 | 15 | 13 | 30 | 23 | 11 | 12 | 14 | 11-17 | 9 | 12 | 14 | 19 | 16 | 18 | 18 | 12 |
| Ht7 | 1 | 14 | 12 | 28 | 21 | 9 | 13 | 14 | 10-13 | 9 | 10 | 14 | 17 | 15 | 17 | 20 | 11 |
| Ht8 | 1 | 14 | 12 | 28 | 23 | 10 | 14 | 12 | 13-20 | 11 | 11 | 15 | 20 | 16 | 17 | 20 | 11 |
| Ht9 | 1 | 14 | 12 | 24 | 24 | 10 | 15 | 12 | 13-18 | 11 | 13 | 15 | 20 | 15 | 17 | 20 | 12 |
| Ht10 | 2 | 14 | 13 | 29 | 24 | 10 | 14 | 12 | 13-13 | 11 | 12 | 15 | 20 | 15 | 17 | 20 | 12 |
| Ht11 | 1 | 14 | 12 | 25 | 24 | 11 | 10 | 13 | 13-18 | 11 | 12 | 15 | 21 | 15 | 19 | 20 | 13 |
| Ht12 | 1 | 14 | 14 | 29 | 25 | 11 | 13 | 15 | 15-18 | 10 | 12 | 14 | 18 | 15 | 15 | 21 | 12 |
| Ht13 | 1 | 17 | 14 | 31 | 25 | 10 | 11 | 13 | 11-14 | 11 | 10 | 14 | 20 | 15 | 15 | 23 | 12 |
| Ht14 | 1 | 14 | 12 | 29 | 22 | 10 | 14 | 11 | 15-15 | 10 | 13 | 15 | 19 | 15 | 15 | 23 | 12 |
| Ht15 | 1 | 15 | 12 | 28 | 24 | 10 | 15 | 12 | 13-19 | 11 | 12 | 15 | 20 | 15 | 19 | 22 | 12 |
| Ht16 | 1 | 15 | 13 | 31 | 25 | 11 | 11 | 13 | 11-14 | 11 | 10 | 14 | 20 | 15 | 15 | 23 | 13 |
| Ht17 | 2 | 15 | 14 | 30 | 22 | 10 | 11 | 12 | 15-18 | 9 | 11 | 14 | 19 | 17 | 17 | 21 | 12 |
| Ht18 | 1 | 15 | 13 | 29 | 24 | 10 | 11 | 15 | 11-21 | 10 | 11 | 14 | 21 | 16 | 17 | 21 | 10 |
| Ht19 | 1 | 14 | 12 | 28 | 24 | 11 | 14 | 12 | 13-19 | 12 | 12 | 15 | 19 | 15 | 16 | 21 | 12 |
| Ht20 | 2 | 16 | 13 | 30 | 25 | 11 | 11 | 12 | 11-14 | 11 | 11 | 14 | 20 | 15 | 16 | 23 | 14 |
| Ht21 | 1 | 14 | 12 | 28 | 23 | 10 | 14 | 12 | 13-21 | 11 | 12 | 15 | 19 | 15 | 16 | 21 | 13 |
| Ht22 | 1 | 14 | 12 | 28 | 22 | 10 | 14 | 12 | 13-19 | 11 | 11 | 15 | 20 | 17 | 20 | 20 | 12 |
| Ht23 | 1 | 15 | 12 | 28 | 24 | 11 | 14 | 13 | 13-16 | 11 | 13 | 14 | 19 | 15 | 17 | 22 | 12 |
| Ht24 | 2 | 15 | 13 | 30 | 25 | 10 | 13 | 14 | 15-20 | 10 | 11 | 14 | 18 | 16 | 16 | 22 | 11 |
| Ht25 | 1 | 15 | 12 | 29 | 23 | 11 | 13 | 13 | 12-12 | 10 | 12 | 15 | 22 | 15 | 17 | 19 | 8 |
| Ht26 | 1 | 15 | 13 | 29 | 24 | 10 | 13 | 16 | 15-16 | 10 | 13 | 14 | 18 | 16 | 16 | 22 | 11 |
| Ht27 | 1 | 15 | 13 | 29 | 25 | 10 | 13 | 14 | 15-20 | 10 | 12 | 14 | 18 | 16 | 16 | 22 | 11 |
| Ht28 | 1 | 15 | 12 | 28 | 23 | 10 | 12 | 12 | 12-16 | 10 | 13 | 15 | 19 | 13 | 18 | 19 | 11 |
| Ht29 | 1 | 14 | 12 | 28 | 23 | 10 | 14 | 12 | 14-21 | 11 | 11 | 15 | 20 | 16 | 19 | 20 | 12 |
| Ht30 | 2 | 15 | 15 | 30 | 23 | 10 | 10 | 14 | 13-21 | 11 | 11 | 14 | 19 | 16 | 17 | 27 | 11 |
| Ht31 | 1 | 15 | 13 | 29 | 24 | 10 | 11 | 14 | 11-21 | 10 | 11 | 14 | 21 | 16 | 17 | 21 | 10 |

| | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Ht32 | 1 | 13 | 12 | 30 | 23 | 10 | 10 | 12 | 9-13 | 11 | 13 | 15 | 20 | 15 | 17 | 21 | 12 |
| Ht33 | 1 | 14 | 12 | 28 | 23 | 10 | 14 | 12 | 13-21 | 11 | 12 | 15 | 20 | 15 | 16 | 20 | 13 |
| Ht34 | 1 | 15 | 14 | 30 | 22 | 10 | 13 | 12 | 14-16 | 9 | 10 | 14 | 19 | 15 | 18 | 21 | 11 |
| Ht35 | 2 | 15 | 12 | 29 | 23 | 10 | 12 | 12 | 12-16 | 10 | 12 | 15 | 19 | 15 | 18 | 19 | 11 |
| Ht36 | 1 | 14 | 13 | 27 | 24 | 10 | 13 | 14 | 11-20 | 10 | 12 | 14 | 21 | 16 | 18 | 21 | 10 |
| Ht37 | 1 | 15 | 14 | 26 | 22 | 10 | 11 | 12 | 15-16 | 9 | 11 | 14 | 19 | 16 | 17 | 21 | 12 |
| Ht38 | 1 | 15 | 11 | 28 | 24 | 10 | 8 | 13 | 13-16 | 11 | 12 | 14 | 19 | 15 | 17 | 22 | 12 |
| Ht39 | 1 | 14 | 12 | 32 | 23 | 10 | 10 | 12 | 17.1-19 | 11 | 11 | 15 | 20 | 14 | 15 | 20 | 12 |
| Ht40 | 3 | 15 | 12 | 28 | 24 | 10 | 14 | 12 | 13-18 | 11 | 12 | 15 | 20 | 15 | 19 | 21 | 12 |
| Ht41 | 1 | 15 | 15 | 28 | 24 | 11 | 14 | 12 | 13-18 | 11 | 12 | 15 | 20 | 12 | 19 | 21 | 12 |
| Ht42 | 1 | 13 | 12 | 29 | 23 | 10 | 10 | 12 | 13-18 | 11 | 12 | 14 | 19 | 15 | 17 | 20 | 13 |
| Ht43 | 2 | 14 | 15 | 30 | 25 | 11 | 13 | 15 | 15-19 | 10 | 12 | 14 | 18 | 15 | 15 | 21 | 12 |
| Ht44 | 1 | 14 | 12 | 28 | 23 | 10 | 14 | 12 | 13-21 | 11 | 12 | 15 | 20 | 15 | 16 | 21 | 13 |
| Ht45 | 1 | 14 | 12 | 28 | 24 | 11 | 8 | 12 | 13-19 | 12 | 12 | 15 | 19 | 15 | 17 | 21 | 12 |
| Ht46 | 2 | 14 | 12 | 29 | 23 | 10 | 14 | 12 | 13-20 | 11 | 13 | 15 | 20 | 15 | 18 | 20 | 12 |
| Ht47 | 1 | 14 | 12 | 28 | 23 | 10 | 14 | 12 | 14-21 | 11 | 12 | 15 | 20 | 15 | 16 | 20 | 13 |
| Ht48 | 1 | 15 | 13 | 29 | 25 | 11 | 13 | 14 | 16-18 | 10 | 12 | 14 | 18 | 15 | 15 | 21 | 11 |
| Ht49 | 1 | 15 | 13 | 30 | 24 | 10 | 13 | 14 | 16-21 | 10 | 12 | 14 | 18 | 17 | 16 | 21 | 11 |
| Ht50 | 1 | 14 | 12 | 28 | 23 | 10 | 14 | 12 | 13-18 | 11 | 13 | 15 | 20 | 15 | 17 | 20 | 12 |
| Ht51 | 1 | 15 | 10 | 25 | 22 | 11 | 13 | 13 | 11-14 | 10 | 11 | 15 | 18 | 15 | 17 | 21 | 11 |
| Ht52 | 1 | 15 | 13 | 29 | 23 | 10 | 13 | 13 | 12-16 | 10 | 13 | 14 | 19 | 17 | 19 | 19 | 12 |
| Ht53 | 1 | 15 | 13 | 30 | 23 | 11 | 8 | 13 | 12-17 | 10 | 12 | 14 | 19 | 15 | 16 | 19 | 11 |
| Ht54 | 2 | 14 | 12 | 28 | 23 | 10 | 14 | 12 | 13-19 | 11 | 11 | 15 | 20 | 15 | 18 | 20 | 13 |
| Ht55 | 1 | 15 | 13 | 30 | 24 | 10 | 10 | 15 | 17-21 | 10 | 13 | 14 | 20 | 16 | 16 | 23 | 11 |
| Ht56 | 1 | 15 | 13 | 27 | 25 | 11 | 13 | 12 | 11-14 | 11 | 11 | 14 | 20 | 15 | 16 | 23 | 12 |
| Ht57 | 1 | 14 | 13 | 26 | 24 | 10 | 15 | 12 | 12-19 | 11 | 12 | 15 | 20 | 15 | 18 | 20 | 12 |
| Ht58 | 1 | 14 | 12 | 28 | 23 | 10 | 11 | 12 | 13-21 | 11 | 13 | 15 | 20 | 14 | 17 | 21 | 12 |
| Ht59 | 1 | 14 | 14 | 30 | 23 | 10 | 11 | 12 | 13-19 | 9 | 12 | 15 | 21 | 15 | 16 | 23 | 12 |
| Ht60 | 2 | 14 | 14 | 30 | 23 | 10 | 11 | 12 | 13-17 | 9 | 12 | 15 | 21 | 15 | 16 | 23 | 12 |
| Ht61 | 1 | 14 | 13 | 30 | 22 | 10 | 13 | 12 | 15-17 | 9 | 13 | 14 | 19 | 15 | 16 | 19 | 12 |
| Ht62 | 1 | 13 | 13 | 30 | 25 | 10 | 11 | 13 | 11-14 | 11 | 10 | 14 | 20 | 15 | 15 | 23 | 12 |
| Ht63 | 1 | 15 | 12 | 29 | 23 | 10 | 12 | 12 | 12-16 | 10 | 13 | 15 | 19 | 13 | 18 | 19 | 11 |
| Ht64 | 3 | 14 | 12 | 27 | 23 | 10 | 13 | 12 | 13-18 | 11 | 12 | 14 | 19 | 15 | 17 | 20 | 13 |
| Ht65 | 1 | 13 | 13 | 29 | 24 | 10 | 11 | 13 | 14-19 | 11 | 10 | 14 | 17 | 15 | 15 | 22 | 12 |
| Ht66 | 1 | 15 | 14 | 31 | 23 | 10 | 12 | 13 | 12-16 | 10 | 11 | 15 | 19 | 15 | 16 | 19 | 11 |

| | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Ht67 | 1 | 14 | 12 | 28 | 23 | 10 | 14 | 12 | 13-21 | 11 | 11 | 15 | 20 | 18 | 19 | 20 | 12 |
| Ht68 | 1 | 14 | 12 | 28 | 23 | 10 | 14 | 12 | 13-21 | 11 | 12 | 15 | 20 | 16 | 18 | 20 | 13 |
| Ht69 | 1 | 16 | 12 | 28 | 24 | 11 | 14 | 12 | 13-18 | 11 | 12 | 15 | 20 | 15 | 19 | 21 | 12 |
| Ht70 | 1 | 15 | 13 | 30 | 25 | 10 | 11 | 13 | 11-14 | 11 | 10 | 14 | 20 | 15 | 15 | 24 | 12 |
| Ht71 | 2 | 14 | 12 | 28 | 23 | 10 | 14 | 12 | 13-18 | 11 | 11 | 15 | 20 | 15 | 17 | 20 | 12 |
| Ht72 | 2 | 17 | 13 | 30 | 25 | 10 | 11 | 13 | 11-15 | 12 | 11 | 14 | 20 | 17 | 14 | 24 | 12 |
| Ht73 | 1 | 15 | 13 | 29 | 23 | 10 | 10 | 14 | 13-19 | 11 | 10 | 16 | 19 | 15 | 17 | 24 | 12 |
| Ht74 | 1 | 15 | 14 | 29 | 22 | 10 | 11 | 12 | 15-16 | 9 | 11 | 14 | 19 | 16 | 16 | 21 | 12 |
| Ht75 | 1 | 13 | 13 | 29 | 24 | 10 | 11 | 13 | 14-19 | 11 | 10 | 14 | 17 | 15 | 15 | 21 | 11 |
| Ht76 | 1 | 15 | 12 | 30 | 23 | 10 | 11 | 11 | 12-15.3 | 10 | 11 | 14 | 19 | 15 | 16 | 19 | 12 |
| Ht77 | 2 | 15 | 12 | 28 | 24 | 10 | 11 | 12 | 13-17 | 9 | 11 | 15 | 19 | 13 | 19 | 22 | 11 |
| Ht78 | 1 | 14 | 12 | 27 | 23 | 10 | 13 | 12 | 12-17 | 11 | 12 | 14 | 19 | 15 | 18 | 20 | 13 |
| Ht79 | 1 | 15 | 13 | 29 | 24 | 10 | 11 | 14 | 11-20 | 10 | 11 | 14 | 21 | 16 | 17 | 21 | 10 |
| Ht80 | 1 | 15 | 12 | 28 | 24 | 10 | 14 | 12 | 13-18 | 11 | 11 | 15 | 20 | 15 | 19 | 21 | 12 |
| Ht81 | 1 | 15 | 13 | 29 | 22 | 10 | 14 | 12 | 9-17 | 10 | 12 | 16 | 20 | 15 | 18 | 23 | 11 |
| Ht82 | 1 | 15 | 12 | 29 | 23 | 11 | 12 | 13 | 12-18 | 10 | 12 | 14 | 19 | 16 | 17 | 19 | 11 |
| Ht83 | 1 | 14 | 13 | 30 | 23 | 10 | 11 | 13 | 14-19 | 9 | 11 | 14 | 20 | 15 | 15 | 21 | 12 |
| Ht84 | 1 | 17 | 13 | 29 | 25 | 10 | 11 | 13 | 11-15 | 12 | 11 | 14 | 20 | 17 | 15 | 23 | 12 |
| Ht85 | 1 | 15 | 13 | 28 | 23 | 10 | 10 | 15 | 12-18 | 10 | 12 | 14 | 21 | 15 | 15 | 21 | 11 |
| Ht86 | 1 | 15 | 12 | 29 | 23 | 10 | 12 | 14 | 12-16 | 10 | 11 | 14 | 20 | 15 | 18 | 19 | 13 |
| Ht87 | 1 | 15 | 12 | 28 | 24 | 10 | 14 | 12 | 13-18 | 11 | 12 | 15 | 20 | 15 | 19 | 21 | 12 |
| Ht88 | 1 | 15 | 12 | 30 | 23 | 10 | 12 | 13 | 12-16 | 10 | 11 | 14 | 19 | 15 | 16 | 19 | 12 |
| Ht89 | 1 | 15 | 12 | 28 | 23 | 10 | 14 | 12 | 13-21 | 11 | 11 | 15 | 20 | 15 | 18 | 20 | 12 |
| Ht90 | 1 | 15 | 13 | 29 | 24 | 10 | 11 | 14 | 11-20 | 10 | 11 | 14 | 21 | 17 | 18 | 21 | 10 |
| Ht91 | 1 | 15 | 13 | 29 | 22 | 10 | 13 | 14 | 11-14 | 10 | 11 | 15 | 17 | 15 | 17 | 21 | 11 |
| Ht92 | 1 | 14 | 14 | 30 | 24 | 10 | 14 | 12 | 13-13 | 11 | 12 | 15 | 20 | 15 | 16 | 20 | 12 |
| Ht93 | 1 | 14 | 13 | 29 | 24 | 10 | 13 | 12 | 13-13 | 11 | 12 | 15 | 19 | 15 | 17 | 20 | 12 |
| Ht94 | 1 | 15 | 14 | 32 | 25 | 10 | 10 | 13 | 11-14 | 11 | 10 | 14 | 20 | 15 | 16 | 23 | 13 |
| Ht95 | 1 | 16 | 13 | 31 | 24 | 10 | 11 | 13 | 11-11 | 11 | 10 | 14 | 20 | 15 | 19 | 23 | 12 |
| Ht96 | 1 | 14 | 12 | 28 | 23 | 10 | 14 | 13 | 13-20 | 11 | 11 | 15 | 20 | 16 | 21 | 20 | 11 |
| Ht97 | 1 | 14 | 14 | 30 | 23 | 10 | 10 | 14 | 13-18 | 11 | 10 | 14 | 19 | 15 | 17 | 25 | 12 |
| Ht98 | 1 | 15 | 12 | 29 | 23 | 10 | 12 | 14 | 12-16 | 10 | 11 | 14 | 20 | 15 | 19 | 19 | 13 |
| Ht99 | 1 | 15 | 13 | 29 | 22 | 11 | 11 | 12 | 15-16 | 8 | 12 | 14 | 19 | 16 | 17 | 20 | 11 |

N=Number of Individuals

**Appendix 2: Allele Frequencies and Gene Diversity for 17 Y-STR markers in 120 unrelated Garo males**

| Allele | DYS 19 | DYS 389I | DYS 389II | DYS 390 | DYS 391 | DYS 392 | DYS 393 | DYS 438 | DYS 439 | DYS 437 | DYS 448 | DYS 456 | DYS 458 | DYS 635 | Y GATA H4 | Allelic Class | DYS 385 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 8 | | | | | | 0.0250 | | 0.0083 | | | | | | | 0.0083 | 9-13 | 0.0083 |
| 9 | | | | | 0.0083 | | | 0.1167 | | | | | | | | 9-17 | 0.0083 |
| 10 | | 0.0083 | | | **0.8250** | 0.0917 | | 0.2750 | 0.1333 | | | | | | 0.0417 | 10-13 | 0.0083 |
| 11 | | 0.0083 | | | 0.1667 | 0.2750 | 0.0250 | 0.5583 | 0.3417 | | | | | | 0.2667 | 11-11 | 0.0083 |
| 12 | | 0.4667 | | | | 0.0833 | 0.5417 | 0.0416 | 0.3833 | | | 0.0083 | | | 0.5166 | 11-14 | 0.0833 |
| 13 | 0.0750 | 0.3583 | | | | 0.2083 | 0.2250 | | 0.1333 | | | 0.0333 | | | 0.1500 | 11-15 | 0.0250 |
| 14 | 0.3833 | 0.1167 | | | | 0.2917 | 0.1500 | | 0.0083 | 0.5083 | | 0.0167 | 0.0167 | | 0.0167 | 11-17 | 0.0083 |
| 15 | 0.4667 | 0.0417 | | | | 0.0250 | 0.0500 | | | 0.4750 | | 0.6917 | 0.1667 | | | 11-20 | 0.0250 |
| 16 | 0.0333 | | | | | | 0.0083 | | | 0.0167 | | 0.1667 | 0.2083 | | | 11-21 | 0.0167 |
| 17 | 0.0333 | | | | | | | | | | 0.0667 | 0.0750 | 0.3000 | | | 12-12 | 0.0083 |
| 18 | | | | | | | | | | | 0.0833 | 0.0083 | 0.1583 | 0.0083 | | 12-15.3 | 0.0083 |
| 19 | 0.0083 | | | | | | | | | | 0.3083 | | 0.1333 | 0.1167 | | 12-16 | 0.0750 |
| 20 | | | | | | | | | | | 0.4500 | | 0.0083 | 0.2667 | | 12-17 | 0.0167 |
| 21 | | | | 0.0083 | | | | | | | 0.0833 | | 0.0083 | 0.2916 | | 12-18 | 0.0167 |
| 22 | | | | 0.1250 | | | | | | | 0.0083 | | | 0.1250 | | 12-19 | 0.0083 |
| 23 | | | | 0.4167 | | | | | | | | | | 0.1250 | | 13-13 | 0.0417 |
| 24 | | | 0.0083 | 0.3000 | | | | | | | | | | 0.0417 | | 13-16 | 0.0167 |
| 25 | | | 0.0167 | 0.1500 | | | | | | | | | | 0.0083 | | 13-18 | **0.1333** |
| 26 | | | 0.0250 | | | | | | | | | | | | | 13-19 | 0.0750 |
| 27 | | | 0.0500 | | | | | | | | | | | 0.0167 | | 13-20 | 0.0333 |
| 28 | | | 0.3083 | | | | | | | | | | | | | 13-21 | 0.0917 |
| 29 | | | 0.3000 | | | | | | | | | | | | | 14-16 | 0.0083 |
| 30 | | | 0.2417 | | | | | | | | | | | | | 14-19 | 0.0583 |
| 31 | | | 0.0333 | | | | | | | | | | | | | 14-21 | 0.0250 |
| 32 | | | 0.0167 | | | | | | | | | | | | | 15-15 | 0.0083 |
| | | | | | | | | | | | | | | | | 15-16 | 0.0333 |
| | | | | | | | | | | | | | | | | 15-17 | 0.0083 |
| | | | | | | | | | | | | | | | | 15-18 | 0.0250 |
| | | | | | | | | | | | | | | | | 15-19 | 0.0167 |

| | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | | | | | | | | | 15-20 | 0.0250 |
| | | | | | | | | | | | | | | | | 16-16 | 0.0083 |
| | | | | | | | | | | | | | | | | 16-18 | 0.0083 |
| | | | | | | | | | | | | | | | | 16-21 | 0.0083 |
| | | | | | | | | | | | | | | | | 17.1-19 | 0.0083 |
| | | | | | | | | | | | | | | | | 17-21 | 0.0083 |
| GD | 0.6326 | 0.6436 | 0.7579 | 0.7040 | **0.2939** | 0.7858 | 0.6355 | 0.6023 | 0.7066 | 0.5200 | 0.6898 | 0.4907 | 0.8022 | 0.8035 | 0.6427 | | **0.9505** |

GD= Gene or haplotype diversity

**Appendix 3: A list of Y Chromosome STR haplotypes detected in 139 unrelated males in Santal population**

| Haplo type | N | DYS 19 | DYS 389I | DYS 389II | DYS 390 | DYS 391 | DYS 392 | DYS 393 | DYS 385 a/b | DYS 438 | DYS 439 | DYS 437 | DYS 448 | DYS 456 | DYS 458 | DYS 635 | Y_GATA_H4 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Ht1 | 2 | 15 | 14 | 30 | 25 | 10 | 13 | 14 | 15-19 | 11 | 13 | 15 | 18 | 17 | 16 | 21 | 11 |
| Ht2 | 1 | 14 | 13 | 29 | 24 | 9 | 13 | 14 | 15-19 | 10 | 13 | 14 | 18 | 15 | 16 | 21 | 12 |
| Ht3 | 1 | 14 | 12 | 29 | 23 | 10 | 11 | 14 | 15-16 | 10 | 12 | 14 | 19 | 15 | 16 | 18 | 13 |
| Ht4 | 1 | 16 | 12 | 27 | 23 | 10 | 11 | 13 | 12-16 | 11 | 12 | 14 | 21 | 15 | 15 | 20 | 12 |
| Ht5 | 2 | 15 | 13 | 29 | 25 | 10 | 13 | 15 | 15-20 | 10 | 12 | 14 | 18 | 15 | 16 | 21 | 11 |
| Ht6 | 3 | 15 | 14 | 31 | 22 | 10 | 11 | 12 | 15-16 | 9 | 11 | 14 | 19 | 15 | 17 | 19 | 13 |
| Ht7 | 1 | 15 | 13 | 30 | 25 | 10 | 11 | 13 | 10-15 | 11 | 10 | 14 | 20 | 18 | 16 | 23 | 12 |
| Ht8 | 1 | 17 | 13 | 30 | 25 | 10 | 13 | 15 | 14-18 | 10 | 11 | 14 | 18 | 16 | 17 | 21 | 11 |
| Ht9 | 1 | 12 | 13 | 29 | 25 | 10 | 10 | 14 | 15-20 | 10 | 13 | 14 | 18 | 15 | 16 | 21 | 12 |
| Ht10 | 1 | 16 | 13 | 31 | 25 | 11 | 13 | 13 | 11-15 | 11 | 10 | 14 | 20 | 15 | 16 | 23 | 13 |
| Ht11 | 1 | 16 | 12 | 27 | 23 | 10 | 12 | 13 | 12-16 | 11 | 12 | 14 | 21 | 15 | 15 | 20 | 12 |
| Ht12 | 1 | 15 | 12 | 28 | 25 | 10 | 10 | 14 | 15-20 | 10 | 14 | 14 | 18 | 16 | 17 | 21 | 11 |
| Ht13 | 2 | 16 | 14 | 31 | 25 | 11 | 13 | 14 | 15-20 | 10 | 12 | 14 | 18 | 15 | 16 | 23 | 11 |
| Ht14 | 1 | 15 | 13 | 29 | 25 | 11 | 13 | 14 | 16-21 | 10 | 12 | 14 | 18 | 16 | 17 | 21 | 13 |
| Ht15 | 1 | 14 | 13 | 30 | 23 | 10 | 11 | 12 | 17-17 | 9 | 11 | 14 | 19 | 16 | 15 | 21 | 12 |
| Ht16 | 2 | 15 | 14 | 30 | 25 | 10 | 12 | 14 | 15-19 | 10 | 12 | 14 | 18 | 15 | 17 | 21 | 11 |
| Ht17 | 1 | 15 | 13 | 29 | 24 | 10 | 13 | 15 | 15-20 | 10 | 13 | 14 | 18 | 15 | 16 | 21 | 11 |
| Ht18 | 1 | 15 | 12 | 29 | 25 | 10 | 11 | 14 | 15-20 | 10 | 13 | 14 | 18 | 16 | 17 | 21 | 11 |
| Ht19 | 1 | 15 | 13 | 29 | 26 | 10 | 11 | 13 | 10-15 | 11 | 10 | 14 | 20 | 17 | 16 | 23 | 13 |
| Ht20 | 1 | 15 | 13 | 28 | 23 | 10 | 13 | 15 | 15-20 | 10 | 14 | 14 | 18 | 15 | 16 | 21 | 11 |
| Ht21 | 1 | 16 | 13 | 30 | 24 | 11 | 13 | 14 | 15-20 | 10 | 12 | 14 | 18 | 15 | 16 | 23 | 11 |
| Ht22 | 1 | 15 | 14 | 30 | 25 | 10 | 12 | 14 | 15-20 | 10 | 11 | 14 | 18 | 15 | 16 | 21 | 11 |
| Ht23 | 1 | 16 | 14 | 29 | 21 | 11 | 14 | 14 | 13-18 | 10 | 12 | 15 | 18 | 17 | 17 | 21 | 11 |
| Ht24 | 1 | 15 | 14 | 31 | 21 | 11 | 11 | 14 | 13-13 | 10 | 12 | 15 | 18 | 17 | 17 | 21 | 11 |
| Ht25 | 1 | 15 | 14 | 32 | 21 | 11 | 12 | 14 | 13-13 | 10 | 12 | 15 | 19 | 17 | 17 | 22 | 11 |
| Ht26 | 1 | 14 | 12 | 28 | 23 | 10 | 12 | 12 | 13-21 | 11 | 11 | 15 | 20 | 16 | 20 | 20 | 12 |
| Ht27 | 1 | 15 | 13 | 29 | 25 | 10 | 13 | 14 | 15-19 | 10 | 13 | 14 | 18 | 15 | 16 | 21 | 12 |
| Ht28 | 1 | 16 | 13 | 31 | 25 | 10 | 11 | 13 | 11-11 | 11 | 10 | 14 | 21 | 15 | 17 | 23 | 13 |
| Ht29 | 1 | 15 | 13 | 28 | 25 | 10 | 13 | 15 | 14-20 | 10 | 12 | 14 | 18 | 15 | 16 | 21 | 11 |
| Ht30 | 1 | 14 | 12 | 29 | 23 | 11 | 14 | 14 | 15-16 | 10 | 13 | 14 | 19 | 15 | 16 | 18 | 12 |
| Ht31 | 1 | 16 | 13 | 28 | 25 | 10 | 15 | 15 | 14-14 | 11 | 12 | 14 | 18 | 16 | 16 | 21 | 11 |

| | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Ht32 | 1 | 15 | 13 | 29 | 25 | 10 | 13 | 14 | 15-19 | 10 | 13 | 14 | 18 | 15 | 16 | 21 | 11 |
| Ht33 | 1 | 15 | 13 | 29 | 25 | 10 | 13 | 14 | 15-19 | 10 | 13 | 14 | 18 | 15 | 17 | 21 | 11 |
| Ht34 | 1 | 15 | 13 | 29 | 25 | 11 | 12 | 14 | 15-19 | 10 | 13 | 14 | 18 | 15 | 18 | 21 | 11 |
| Ht35 | 1 | 15 | 13 | 29 | 25 | 10 | 12 | 15 | 14-19 | 11 | 12 | 14 | 18 | 16 | 16 | 21 | 11 |
| Ht36 | 2 | 15 | 14 | 30 | 25 | 10 | 13 | 15 | 14-19 | 10 | 12 | 14 | 18 | 16 | 16 | 21 | 11 |
| Ht37 | 1 | 13 | 13 | 28 | 25 | 10 | 10 | 14 | 15-19 | 10 | 12 | 14 | 20 | 15 | 16 | 21 | 11 |
| Ht38 | 1 | 14 | 13 | 28 | 25 | 11 | 13 | 14 | 15-20 | 10 | 9 | 14 | 18 | 16 | 17 | 21 | 11 |
| Ht39 | 1 | 14 | 14 | 29 | 26 | 10 | 13 | 15 | 14-18 | 10 | 12 | 14 | 18 | 16 | 18 | 21 | 11 |
| Ht40 | 1 | 15 | 13 | 28 | 25 | 10 | 12 | 14 | 15-21 | 10 | 12 | 14 | 18 | 15 | 16 | 21 | 11 |
| Ht41 | 1 | 16 | 13 | 29 | 24 | 10 | 11 | 14 | 15-18 | 11 | 12 | 14 | 18 | 16 | 17 | 21 | 11 |
| Ht42 | 1 | 14 | 14 | 28 | 21 | 11 | 13 | 14 | 13-19 | 10 | 12 | 15 | 19 | 17 | 17 | 22 | 13 |
| Ht43 | 1 | 15 | 13 | 29 | 24 | 10 | 13 | 15 | 15-19 | 10 | 13 | 14 | 17 | 15 | 16 | 22 | 12 |
| Ht44 | 1 | 15 | 14 | 30 | 25 | 10 | 13 | 15 | 15-19 | 11 | 13 | 14 | 18 | 16 | 16 | 21 | 11 |
| Ht45 | 1 | 15 | 14 | 29 | 21 | 12 | 11 | 14 | 13-13 | 10 | 12 | 15 | 19 | 15 | 18 | 21 | 12 |
| Ht46 | 1 | 15 | 14 | 30 | 25 | 10 | 13 | 14 | 15-19 | 11 | 13 | 14 | 18 | 16 | 16 | 21 | 11 |
| Ht47 | 1 | 15 | 13 | 30 | 23 | 10 | 11 | 13 | 13-16 | 9 | 11 | 14 | 19 | 14 | 19 | 24 | 12 |
| Ht48 | 1 | 15 | 14 | 30 | 22 | 10 | 11 | 12 | 15-16 | 9 | 12 | 14 | 19 | 15 | 17 | 19 | 13 |
| Ht49 | 1 | 15 | 13 | 29 | 25 | 10 | 13 | 15 | 14-19 | 10 | 12 | 14 | 18 | 16 | 17 | 21 | 11 |
| Ht50 | 1 | 16 | 14 | 30 | 22 | 10 | 11 | 12 | 15-16 | 9 | 11 | 14 | 19 | 15 | 15 | 19 | 12 |
| Ht51 | 1 | 15 | 13 | 28 | 24 | 10 | 13 | 16 | 15-21 | 10 | 13 | 14 | 18 | 16 | 16 | 21 | 11 |
| Ht52 | 1 | 14 | 12 | 29 | 22 | 10 | 11 | 12 | 15-17 | 9 | 11 | 14 | 20 | 15 | 16 | 21 | 13 |
| Ht53 | 1 | 13 | 13 | 29 | 23 | 10 | 10 | 13 | 13-19 | 11 | 10 | 16 | 19 | 15 | 17 | 25 | 12 |
| Ht54 | 1 | 17 | 13 | 29 | 25 | 10 | 13 | 15 | 14-18 | 10 | 12 | 14 | 18 | 15 | 17 | 21 | 11 |
| Ht55 | 1 | 14 | 12 | 28 | 23 | 10 | 11 | 14 | 15-16 | 10 | 12 | 14 | 19 | 15 | 16 | 18 | 13 |
| Ht56 | 1 | 15 | 14 | 28 | 26 | 10 | 13 | 15 | 14-18 | 10 | 12 | 14 | 18 | 16 | 18 | 21 | 11 |
| Ht57 | 1 | 13 | 13 | 29 | 24 | 10 | 14 | 13 | 11-14 | 11 | 10 | 14 | 20 | 16 | 16 | 23 | 12 |
| Ht58 | 1 | 15 | 13 | 29 | 25 | 11 | 13 | 14 | 15-20 | 10 | 13 | 14 | 18 | 15 | 16 | 23 | 11 |
| Ht59 | 1 | 13 | 13 | 28 | 25 | 11 | 13 | 14 | 15-20 | 10 | 11 | 14 | 18 | 15 | 16 | 23 | 12 |
| Ht60 | 1 | 14 | 13 | 29 | 25 | 11 | 12 | 14 | 15-20 | 10 | 12 | 14 | 18 | 15 | 16 | 23 | 11 |
| Ht61 | 1 | 15 | 12 | 28 | 25 | 10 | 13 | 14 | 15-20 | 10 | 14 | 14 | 18 | 16 | 17 | 21 | 11 |
| Ht62 | 1 | 14 | 14 | 30 | 22 | 10 | 13 | 12 | 14-19 | 9 | 12 | 14 | 18 | 15 | 18 | 20 | 11 |
| Ht63 | 1 | 15 | 12 | 30 | 23 | 11 | 14 | 14 | 14-16 | 10 | 12 | 14 | 19 | 15 | 16 | 18 | 12 |
| Ht64 | 1 | 17 | 13 | 32 | 24 | 10 | 11 | 13 | 11-14 | 11 | 10 | 14 | 20 | 16 | 16 | 23 | 12 |
| Ht65 | 1 | 16 | 13 | 31 | 25 | 11 | 13 | 14 | 15-18 | 10 | 12 | 14 | 18 | 15 | 16 | 23 | 11 |
| Ht66 | 1 | 14 | 12 | 29 | 23 | 11 | 9 | 14 | 14-16 | 10 | 12 | 14 | 19 | 15 | 16 | 18 | 12 |

| | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Ht67 | 1 | 17 | 13 | 32 | 24 | 10 | 11 | 13 | 11-14 | 11 | 10 | 14 | 20 | 16 | 16 | 23 | 13 |
| Ht68 | 1 | 16 | 13 | 29 | 25 | 10 | 13 | 15 | 14-18 | 10 | 12 | 14 | 18 | 15 | 17 | 21 | 13 |
| Ht69 | 1 | 16 | 13 | 30 | 25 | 10 | 13 | 13 | 15-18 | 10 | 12 | 14 | 18 | 15 | 16 | 21 | 12 |
| Ht70 | 1 | 15 | 14 | 30 | 25 | 10 | 13 | 14 | 15-19 | 10 | 13 | 14 | 18 | 15 | 18 | 21 | 11 |
| Ht71 | 2 | 15 | 13 | 30 | 25 | 11 | 13 | 14 | 15-20 | 10 | 11 | 14 | 18 | 15 | 16 | 23 | 11 |
| Ht72 | 1 | 15 | 13 | 28 | 24 | 10 | 13 | 16 | 15-23 | 10 | 13 | 14 | 16 | 15 | 16 | 21 | 11 |
| Ht73 | 1 | 15 | 13 | 29 | 25 | 10 | 13 | 14 | 15-21 | 10 | 12 | 14 | 18 | 15 | 16 | 21 | 11 |
| Ht74 | 1 | 14 | 12 | 29 | 23 | 11 | 11 | 14 | 15-16 | 10 | 12 | 14 | 19 | 15 | 16 | 18 | 13 |
| Ht75 | 1 | 14 | 12 | 29 | 23 | 11 | 11 | 14 | 14-17 | 10 | 13 | 14 | 19 | 15 | 16 | 18 | 12 |
| Ht76 | 1 | 15 | 14 | 30 | 22 | 10 | 11 | 12 | 15-17 | 9 | 11 | 14 | 19 | 15 | 16 | 20 | 12 |
| Ht77 | 1 | 16 | 14 | 30 | 25 | 10 | 13 | 14 | 15-15 | 10 | 12 | 14 | 18 | 16 | 17 | 21 | 13 |
| Ht78 | 1 | 16 | 13 | 29 | 25 | 11 | 13 | 14 | 15-20 | 10 | 11 | 14 | 18 | 16 | 16 | 23 | 11 |
| Ht79 | 2 | 14 | 12 | 29 | 23 | 11 | 11 | 14 | 15-16 | 10 | 12 | 14 | 19 | 15 | 17 | 18 | 12 |
| Ht80 | 2 | 16 | 13 | 30 | 25 | 11 | 13 | 14 | 15-20 | 10 | 12 | 14 | 18 | 16 | 16 | 23 | 11 |
| Ht81 | 1 | 16 | 13 | 30 | 25 | 10 | 13 | 14 | 15-21 | 10 | 12 | 14 | 18 | 15 | 16 | 23 | 12 |
| Ht82 | 1 | 14 | 12 | 30 | 23 | 10 | 11 | 14 | 13-13 | 10 | 10 | 14 | 19 | 15 | 15 | 18 | 11 |
| Ht83 | 1 | 16 | 13 | 31 | 25 | 11 | 11 | 13 | 11-13 | 11 | 10 | 15 | 20 | 15 | 17 | 23 | 12 |
| Ht84 | 1 | 15 | 13 | 29 | 25 | 10 | 13 | 14 | 15-20 | 10 | 11 | 14 | 18 | 15 | 16 | 21 | 12 |
| Ht85 | 1 | 16 | 14 | 30 | 26 | 11 | 13 | 14 | 16-19 | 10 | 13 | 14 | 17 | 17 | 14 | 22 | 11 |
| Ht86 | 1 | 15 | 13 | 29 | 24 | 10 | 13 | 14 | 15-19 | 12 | 12 | 14 | 18 | 17 | 16 | 21 | 11 |
| Ht87 | 1 | 16 | 13 | 31 | 25 | 11 | 11 | 13 | 11-15 | 11 | 10 | 14 | 20 | 15 | 16 | 23 | 12 |
| Ht88 | 1 | 15 | 13 | 31 | 22 | 11 | 11 | 12 | 17-17 | 10 | 12 | 14 | 19 | 17 | 18 | 20 | 11 |
| Ht89 | 1 | 15 | 13 | 30 | 24 | 10 | 13 | 16 | 14-20 | 10 | 12 | 14 | 18 | 15 | 16 | 21 | 11 |
| Ht90 | 1 | 16 | 13 | 30 | 25 | 11 | 13 | 14 | 15-20 | 10 | 12 | 14 | 18 | 15 | 16 | 23 | 11 |
| Ht91 | 1 | 16 | 13 | 30 | 25 | 11 | 13 | 14 | 15-20 | 10 | 12 | 14 | 18 | 15 | 16 | 22 | 11 |
| Ht92 | 1 | 17 | 13 | 29 | 25 | 10 | 13 | 15 | 14-18 | 10 | 12 | 14 | 18 | 16 | 17 | 21 | 11 |
| Ht93 | 1 | 15 | 13 | 29 | 25 | 10 | 13 | 14 | 15-18 | 10 | 14 | 14 | 18 | 15 | 16 | 21 | 11 |
| Ht94 | 1 | 14 | 12 | 29 | 23 | 11 | 11 | 12 | 15-16 | 10 | 12 | 14 | 19 | 15 | 16 | 18 | 12 |
| Ht95 | 1 | 15 | 12 | 27 | 24 | 10 | 13 | 16 | 15-21 | 10 | 13 | 14 | 17 | 15 | 16 | 21 | 12 |
| Ht96 | 1 | 15 | 13 | 29 | 22 | 10 | 11 | 12 | 14-17 | 9 | 12 | 14 | 17 | 15 | 17 | 21 | 13 |
| Ht97 | 1 | 15 | 13 | 29 | 25 | 9 | 13 | 15 | 15-19 | 10 | 12 | 14 | 18 | 15 | 16 | 21 | 11 |
| Ht98 | 1 | 15 | 13 | 29 | 25 | 10 | 13 | 14 | 15-19 | 10 | 13 | 14 | 18 | 15 | 19 | 21 | 11 |
| Ht99 | 1 | 15 | 13 | 29 | 25 | 10 | 13 | 14 | 15-21 | 10 | 13 | 14 | 18 | 15 | 17 | 21 | 11 |
| Ht100 | 1 | 16 | 10 | 28 | 23 | 10 | 11 | 14 | 15-16 | 10 | 11 | 14 | 19 | 15 | 13 | 18 | 11 |
| Ht101 | 1 | 14 | 12 | 29 | 22 | 10 | 11 | 12 | 14-17 | 9 | 11 | 14 | 20 | 15 | 16 | 21 | 12 |

| | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Ht102 | 1 | 15 | 13 | 28 | 24 | 10 | 13 | 15 | 15.3-21 | 10 | 13 | 14 | 18 | 15 | 16 | 21 | 12 |
| Ht103 | 1 | 13 | 14 | 31 | 22 | 10 | 11 | 12 | 15-16 | 9 | 11 | 14 | 19 | 15 | 17 | 19 | 12 |
| Ht104 | 1 | 15 | 13 | 28 | 25 | 10 | 13 | 14 | 16-21 | 10 | 12 | 14 | 18 | 16 | 20 | 21 | 11 |
| Ht105 | 1 | 17 | 13 | 30 | 25 | 10 | 13 | 15 | 14-18 | 10 | 12 | 14 | 18 | 16 | 17 | 21 | 11 |
| Ht106 | 1 | 13 | 12 | 28 | 24 | 11 | 11 | 12 | 13-17 | 9 | 11 | 15 | 19 | 13 | 19 | 21 | 11 |
| Ht107 | 1 | 13 | 13 | 32 | 24 | 10 | 13 | 14 | 15-19 | 10 | 13 | 14 | 18 | 15 | 17 | 21 | 13 |
| Ht108 | 1 | 15 | 13 | 28 | 24 | 10 | 13 | 15 | 16-21 | 10 | 13 | 14 | 18 | 15 | 16 | 22 | 11 |
| Ht109 | 1 | 15 | 13 | 29 | 25 | 10 | 13 | 15 | 15-21 | 10 | 12 | 14 | 18 | 15 | 16 | 21 | 12 |
| Ht110 | 1 | 16 | 14 | 27 | 25 | 11 | 11 | 12 | 11-12 | 11 | 10 | 14 | 20 | 15 | 16 | 23 | 12 |
| Ht111 | 1 | 15 | 14 | 30 | 25 | 11 | 13 | 14 | 15-21 | 10 | 13 | 14 | 18 | 15 | 16 | 22 | 12 |
| Ht112 | 1 | 15 | 13 | 29 | 25 | 10 | 13 | 15 | 14-19 | 10 | 9 | 14 | 18 | 16 | 16 | 21 | 12 |
| Ht113 | 1 | 15 | 13 | 30 | 23 | 11 | 10 | 14 | 13-19 | 11 | 11 | 16 | 19 | 15 | 16 | 24 | 12 |
| Ht114 | 1 | 15 | 13 | 28 | 24 | 10 | 13 | 16 | 15-20 | 10 | 13 | 14 | 21 | 15 | 16 | 21 | 11 |
| Ht115 | 1 | 13 | 12 | 29 | 23 | 11 | 11 | 14 | 15-16 | 10 | 13 | 14 | 19 | 15 | 17 | 18 | 13 |
| Ht116 | 1 | 14 | 12 | 29 | 21 | 11 | 12 | 11 | 13-16 | 10 | 12 | 15 | 19 | 16 | 15 | 24 | 11 |
| Ht117 | 1 | 14 | 14 | 31 | 22 | 10 | 8 | 12 | 15-16 | 9 | 11 | 14 | 19 | 15 | 17 | 19 | 13 |
| Ht118 | 1 | 17 | 13 | 30 | 25 | 10 | 12 | 15 | 14-18 | 10 | 12 | 14 | 18 | 16 | 17 | 21 | 13 |
| Ht119 | 1 | 16 | 13 | 31 | 25 | 11 | 11 | 13 | 11-13 | 11 | 10 | 15 | 20 | 15 | 17 | 23 | 13 |
| Ht120 | 1 | 15 | 13 | 26 | 25 | 11 | 11 | 13 | 11-14 | 11 | 10 | 14 | 20 | 15 | 16 | 23 | 12 |
| Ht121 | 1 | 14 | 14 | 31 | 25 | 10 | 13 | 14 | 15-15 | 10 | 9 | 14 | 18 | 16 | 17 | 21 | 13 |
| Ht122 | 1 | 14 | 12 | 29 | 22 | 10 | 11 | 12 | 15-17 | 9 | 9 | 14 | 20 | 15 | 15 | 21 | 11 |
| Ht123 | 1 | 15 | 14 | 30 | 25 | 10 | 13 | 14 | 15-18 | 10 | 9 | 14 | 18 | 15 | 17 | 21 | 13 |
| Ht124 | 1 | 15 | 13 | 32 | 25 | 11 | 13 | 14 | 15-21 | 10 | 10 | 14 | 18 | 15 | 16 | 22 | 12 |
| Ht125 | 1 | 13 | 13 | 31 | 25 | 11 | 11 | 13 | 11-14 | 11 | 10 | 14 | 20 | 15 | 16 | 23 | 12 |
| Ht126 | 1 | 15 | 13 | 28 | 25 | 10 | 13 | 15 | 14-19 | 11 | 12 | 14 | 18 | 17 | 16 | 20 | 12 |
| Ht127 | 1 | 15 | 13 | 31 | 23 | 10 | 11 | 14 | 15-16 | 10 | 11 | 14 | 19 | 15 | 13 | 18 | 11 |
| Ht128 | 1 | 15 | 13 | 29 | 24 | 11 | 13 | 14 | 15-20 | 10 | 12 | 14 | 18 | 16 | 17 | 21 | 13 |
| Ht129 | 1 | 15 | 13 | 29 | 25 | 10 | 13 | 14 | 15-21 | 10 | 14 | 14 | 18 | 15 | 17 | 21 | 13 |

N= Number of Individuals

**Appendix 4: Allele Frequencies and Gene Diversity for 17 Y-STR markers in 139 unrelated Santal males**

| Allele | DYS 19 | DYS 389I | DYS 389II | DYS 390 | DYS 391 | DYS 392 | DYS 393 | DYS 438 | DYS 439 | DYS 437 | DYS 448 | DYS 456 | DYS 458 | DYS 635 | Y_ GATA H4 | Allelic Class | DYS 385 a/b |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 8 | | | | | | 0.0072 | | | | | | | | | | 10-15 | 0.0143 |
| 9 | | | | | 0.0144 | 0.0072 | | 0.1151 | 0.0360 | | | | | | | 11-11 | 0.0072 |
| 10 | | 0.0072 | | | 0.6547 | 0.0360 | | 0.6906 | 0.1151 | | | | | | | 11-12 | 0.0072 |
| 11 | | | | | 0.3237 | 0.3022 | 0.0072 | 0.1871 | 0.1655 | | | | | | 0.5108 | 11-13 | 0.0143 |
| 12 | 0.0072 | 0.1727 | | | 0.0072 | 0.0863 | 0.1367 | 0.0072 | 0.4387 | | | | | | 0.3022 | 11-14 | 0.0359 |
| 13 | 0.0647 | 0.5755 | | | | 0.5252 | 0.1223 | | 0.2087 | | | 0.0072 | 0.0144 | | 0.1870 | 11-15 | 0.0143 |
| 14 | 0.1727 | 0.2446 | | | | 0.0287 | 0.5107 | | 0.0360 | **0.8993** | | 0.0072 | 0.0072 | | | 12-16 | 0.0143 |
| 15 | 0.5108 | | | | | 0.0072 | 0.1871 | | | 0.0863 | | 0.6547 | 0.0504 | | | 13-13 | 0.0288 |
| 16 | 0.1942 | | | | | | 0.0360 | | | 0.0144 | 0.0072 | 0.2446 | 0.5467 | | | 13-16 | 0.0143 |
| 17 | 0.0504 | | | | | | | | | | 0.0288 | 0.0791 | 0.2950 | | | 13-17 | 0.0072 |
| 18 | | | | | | | | | | | 0.5827 | 0.0072 | 0.0504 | 0.1007 | | 13-18 | 0.0072 |
| 19 | | | | | | | | | | | 0.2302 | | 0.0215 | 0.0504 | | 13-19 | 0.0216 |
| 20 | | | | | | | | | | | 0.1223 | | 0.0144 | 0.0504 | | 13-21 | 0.0072 |
| 21 | | | | 0.0432 | | | | | | | 0.0288 | | | 0.5179 | | 14-14 | 0.0072 |
| 22 | | | | 0.1007 | | | | | | | | | | 0.0576 | | 14-16 | 0.0143 |
| 23 | | | | 0.1583 | | | | | | | | | | 0.1942 | | 14-17 | 0.0216 |
| 24 | | | | 0.1367 | | | | | | | | | | 0.0216 | | 14-18 | 0.0576 |
| 25 | | | | 0.5324 | | | | | | | | | | 0.0072 | | 14-19 | 0.0504 |
| 26 | | | 0.0072 | 0.0287 | | | | | | | | | | | | 14-20 | 0.0143 |
| 27 | | | 0.0287 | | | | | | | | | | | | | 15-15 | 0.0143 |
| 28 | | | 0.1583 | | | | | | | | | | | | | 15-16 | 0.1223 |
| 29 | | | 0.3741 | | | | | | | | | | | | | 15-17 | 0.0216 |
| 30 | | | 0.2662 | | | | | | | | | | | | | 15-18 | 0.0359 |
| 31 | | | 0.1295 | | | | | | | | | | | | | 15-19 | 0.1295 |
| 32 | | | 0.0360 | | | | | | | | | | | | | 15-20 | **0.1871** |
| | | | | | | | | | | | | | | | | 15-21 | 0.0719 |
| | | | | | | | | | | | | | | | | 15.3-21 | 0.0072 |
| | | | | | | | | | | | | | | | | 15-23 | 0.0072 |

| | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | | | | | | | | | 16-19 | 0.0072 |
| | | | | | | | | | | | | | | | | 16-21 | 0.0216 |
| | | | | | | | | | | | | | | | | 17-17 | 0.0143 |
| GD | 0.6695 | 0.5833 | 0.7505 | 0.6647 | 0.4697 | 0.6276 | 0.6740 | 0.4782 | 0.7259 | **0.1849** | 0.5950 | 0.5087 | 0.6124 | 0.6799 | 0.6172 | | **0.9212** |

GD= Gene or haplotype diversity