LOYOLA COLLEGE
AUTONOMOUS
Chennai-600 034

Tel : 2817 8200
Fax : 2817 5566
E-mail: welcome@loyolacollege.edu
Home Page : www.loyolacollege.edu

CERTIFICATE

This is to certify that this dissertation entitled ," ***Genetic study of Bhumij Tribe of Jharkhand using mitochondrial and Y chromosome DNA markers***", has been carried out by **Miss. Smita Bernadet Kujur, Reg. no: 08-PBT-26**, at **Centre for Cellular and Molecular Biology**, Hyderabad, under the supervision and guidance of **Dr. K. Thangaraj** for the partial fulfillment of degree of Master of Science in Biotechnology to Madras University, Loyola College, Chennai. This work is original and has not been submitted in part or full to any other university or college for any other degree or diploma.

Date:

Internal Examiner

Dr. P. Agastian,M.Sc., Ph.D.
Examiner

External

Head of the Department

Dept. of Plant Biology and Biotechnology

Loyola College

Chennai: 600034

Thesis Report On

Genetic Study of Bhumij Tribe of Jharkhand using mitochondrial and Y chromosomal DNA markers

A thesis submitted in partial fulfillment of the requirements of the degree of

Masters of Science in Biotechnology

By: Smita Bernadet Kujur of Loyola College.

Work done at CCMB

ACKNOWLEDGEMENT

I am heartily thankful to **Dr. Ch Mohan Rao**, Director of CCMB, Hyderabad, India for granting me his kind permission to work in CCMB. I am highly obliged to him for providing me with all the excellent facilities, rich source of knowledge and a healthy competitive environment.

With a deep sense of veneration and obligation to **Dr. K.Thangaraj**, Scientist , CCMB, Hyderabad, India, whose encouragement, guidance and support from the initial to the final level enabled me to develop an understanding of the subject.

I thank my co-guide **Miss. Sakshi Singh** [JRF], for her continuous support in my dissertation program. She was always there to listen and to give advice.She taught me how to express my ideas. She showed me different ways to approach a research problem and the need to be persistent to accomplish any goal. She guided me professionally as well as personally. She helped me day and night with my thesis preparation.May God bless her n may she come out with flying colours. I wish her all the success in her life ahead.Thank you maam.

I owe my sincere thanks to my Head of Department of Plant Biology and Biotechnology, **Dr. Agastian. S. Theoder** M.Sc., M.Phil., Ph.D. (Head of the Department), my department co-ordinator **Mrs. Mary Dorothy Anitha Sebastian**(SET) M.Sc. , my guide **Ms. D. Jacintha Jasmine,** staff of my department **Ms. P. Margaret Sangeetha** M.Sc., M.Phil, **Dr. (Ms). Shirly George Panicker** M.Sc. (Agri) Ph.D, **Ms. Sally Gloriana** M.Sc.,M.Phil, **Ms. D. Jacintha Jasmine,** M.Sc., M.Phil, **Mr. Preetam**, **Mr. Victor**.I am greatly indebted to them for their encouragement.

I thank **Mr. A.Govardhan Reddy** [Technical officer], **Mr. Surya Narayan** ,**Mr. Rakesh Tamang** [Proj. Asstt**], Aditya Nath Jha** [JRF], **Sharath Anugula** [Proj. Asstt], **Mr. Nizam** , **Mr. Haneef**. They were always available listen and talk about my ideas, to provide understanding, provide reagents and mark up my papers and chapters, and to ask me good questions to help me think through my problems (whether philosophical, analytical or computational).

Last but not the least, I owe special thanks to my respected father **Mr. Srimanim Belkhas Kujur**, my mother **Mrs. Susheela Kujur**, **Mrs. Goretti**, **Mrs. Milaani Kullu**, , **Mr. Ananta Lal Tud**u, , **Mr. N.P.C Sardar**, , **Dr. Daisy** , **Mr. Surya Narayan** ,**Mr. Kuldip Khalkho, Mrs. Bina**, **Mr. Bhujang, they all helped me volunteerilly.**

That's not all. People with great personalities took out time from their busy schedule and helped me in blood sample collection. I am immense pleased to introduce to you a very honored person Shree. **Dr Balram Singh Sardar (BISM),** (village – Tirildih , Dist West Singbhum, Jharkhand) is a renowned

and poplar doctor from the Bhumij Community. He is holding the post of Secretary in the **OYON AKHRA** (in their language) which is the Central Executive Body of **AADIM BHUMIJ MUNDA SAMAJ KAYYAN SAMITI.**

Shri. Subodh Singh Sardar, (village – Bhatin, Dist West Singbhum, Jharkhand) is popular **Congress party leader**. He contested the 2009 Assembly election of Jharkhand from Congress ticket. He is a Graduate.

Gunadhar Singh Sardar, (village – Gitilata, Dist West Singbhum, Jharkhand) is renowned social worker and community leader. He is one of the Trustee member of **AADIM BHUMIJ MUNDA SAMAJ KALYAN SAMITI and Ex-Secretary.** Presently he is one of the Advisor to the Samiti. He is a Graduate.

Niranjan Singh Sardar (village – Tirildih, Dist West Singbhum, Jharkhand) is the **NYA** (in their language) that means Community Priest. He is also **PRADHAN** (village Head Man) of Tirildih village.

Amulya Singh Sardar (village – Bunudih, Dist West Singbhum, Jharkhand) is a renowned and veteran politician of **Jharkhand Mukti Morcha (JMM).** He is Ex-MLA of Jharkhand Assembly. He is also the **Secretary** of Bunudih Branch of **AADIM BHUMIJ MUNDA SAMAJ KALYAN SAMITI.**

Shatrudhan Singh Sardar (villge – Tentla, Dist West Singbhum, Jharkhand) is prominent distinguished member of Bhumij community. He holds the post of President of **OYON AKHRA** (in their language) which is the Central Executive Body of **AADIM BHUMIJ MUNDA SAMAJ KALYAN SAMITI.** He is just a matric but very active in social activities.

Miss. Mona Bhumij (village – Ghaghidih, Dist West Singbhum, Jharkhand) is the daughter of a retired TISCO employee Mr. Ghasiram Bhumij. She has all round qualitative skills. As a **brilliant student** she is doing her PG in Economics from Women's Collge Jamshedpur. She is a **talented sports woman and athlete** participated at **national level events of Hand Ball, Kabbaddi and Javelin Throw** and won Medals. Her social and community life is also full of self service activities as she organizes classes for the children as well as grown ups under **SARVSHIKSHA ABHIYAN** (a educational scheme of the govt.) from the Govt. Primary School, Ghaghidih as the centre. Miss Mona is presently the **treasurer** of the local Committee of this educational scheme. Besides this, she imparts tuitions to the local children free of cost. Last but not least, she is very much fond of gardening flowers, singing and listening music as extra curricular activities. She took lot of pain and cooperated to help me in collecting the blood samples by arranging meetings and convincing people to come forward for giving blood samples.

Mrs. Daisy is a local Senior Nurse working in the Govt. Health Centers located in the village habitats of majority Bhumij tribe. She is well known and gracefully respected in the Bhumij community. Another compounder, **Mr. Surya Narayan** also cooperated along with Mrs. Daisy. They also arranged the permission of the respected authorities an allowed to use the room of health Centre for collection of Blood samples.

Collection of blood was a very tedious job . People of this community are very superstitious and orthodox. The people named above helped me both directly and in directly in convincing the villagers to co-operate and gather at a arranged venue. All of them contributed in helping CCMB in there population genetic research. Thank you a lot for supporting me.

I offer my regards and blessings to all my friends of those who supported me in any respect during the completion of the project ," **Joe Zacharias, Manju Kashyap, Vinee Khanna, Sapna Narvariya, Dheepa Kaliyaperumal, Devi.K ,Manimaran .M., Vidya, Anushah, Hema, Shobna, Pavitra, Richa, Satrupa,Upasna Saranji, Gayitri.** They all are a blessing in my life.

Their presence made the CCMB Lab a wonderful workplace and home for the 6 months by
indulging my ever expanding bookshelf space and computer equipment needs. Also thanks to the
folks at the CCMB Lab for interesting discussions and being fun to be with. Thanks,
friends !

May God be there to support and guide you always in any form you as you were there for me. I express my heartiest gratitude to all of them for being a source of inexhaustible encouragement, unconditional love and inspiration to build up my educational career. Their influence is all over these pages and will stay all through my days to come.

Smita Bernadet Kujur

CONTENTS

LIST OF FIGURES

LIST OF ABBREVIATIONS

%	-	Percentage
□C	-	Degree Celsius
ATP	-	Adenosine 5'-triphosphate
bp	-	base pair(s)
cm	-	centimeter
cpm	-	counts per minute
dATP	-	2'-deoxyadenosine 5'-triphosphate
dCTP	-	2'-deoxycytidine 5'-triphosphate
DDW	-	Double distilled water
dGTP	-	2'-deoxyguanosine 5'-triphosphate
D-loop	-	The displacement loop
DNA	-	Deoxyribonucleic acid
dNTP	-	2'-deoxynucleotide 5'-triphosphate
ddNTP	-	2',3'-dideoxynucleotide 5'-triphosphate
dTTP	-	2'-deoxythymidine 5'-triphosphate
EDTA	-	Ethylene diamine tetra acetic acid
Et.Br	-	Ethidium bromide
Extn	-	Extension
Figure	-	Figure
g	-	gram
kb	-	kilo base
M	-	molarity
mA	-	milli ampere
mg	-	milligram
min	-	minutes
ml	-	millilitre
mm	-	millimeter
mM	-	millimolar
mtDNA	-	mitochondrial DNA
mtRNA	-	mitochontrial RNA
rRNA	-	ribosomal RNA
tRNA	-	transfer RNA
N	-	Normality
nm	-	nanometer
NaOH	-	sodium hydroxide
ng	-	nanogram
OD	-	Optical density

O_H	-	Origin of heavy chain replication
O_L	-	The L-strand origin
PCR	-	Polymerase chain reaction
pM	-	picomole
RNA	-	ribonucleic acid
rpm.	-	Revolutions per minute
SDDW	-	Sterile Double Distilled water
SDS	-	Sodium dodecyl sulphate
Sec.	-	Seconds
SNPs	-	Single Nucleotide Polymorphisms
SRY	-	Sex-Determining Region On Y Chromosome
SSC	-	Sodium saline citrate
STR	-	Short Tandem Repeat
TAE	-	Tris-Acetate-EDTA
TE	-	Tris-EDTA
Tris	-	Tris (ecogniz methyl) amino methane
TMRCA	-	Time to the most recent common ancestor
U	-	unit
UEP	-	unique event polymorphism
UV	-	Ultraviolet
V	-	Volts
v/v	-	Volume/Volume
w/v	-	Weight/Volume
g	-	Microgram
l	-	Microlitre
MW	-	Micro molar Watts
YAP	-	Y-*Alu* polymorphism
YCC	-	Y Chromosome Consortium

ABSTRACT

India is a conglomeration of various ethnicities with 4693 communities, 325 languages, 25 scripts and numerous endogamous groups. It is a home of several tribal pockets, which represents different genetic isolates and thus provides unique wealth to understand human evolution. These autochthonous tribal populations reveal striking diversities in terms of language, marriage practices as well as in their genetic architecture. The origin and settlement of the Indian people still remain intrigues for the scientist studying the impact of the past and modern migration of the genetic diversity and structure of contemporary populations. Indian populations are stratified as tribe, caste and religious community. Endogamy has probably been a major reason for genetic diversification among the people of this region. Taking geographical and ethnic diversity into account and to answer the question of origin and evolution of maternal and paternal lineages of Indian population .Above 400 base pairs of the HVR-1 region and selected coding regions of the mitochondrial DNA (mtDNA) and Y chromosome markers in 102 individuals of Bhumij, an Austro-Asiatic tribe of Jharkhand, was analyzed and compared with the data available from the Indian subcontinent. Based on the mutations observed in the HVR -1 and selected coding region of mitochondrial DNA, haplogroups were assigned to each of the individual. It was observed that most of the individuals of Bhumij tribal population were falling in Indian specific macro haplogroup M, displaying the array of South Asian specific lineages. On the other hand, Y chromosomal analysis is showing 70% percentage of individuals falling into O2a-M95 haplogroup, found frequently among Austro-Asiatic. Moreover, it is evident that our investigation of the small population is a snapshot with respect to the peopling of the Indian subcontinent. In future, detailed phylogeographic and phylogenetic analysis of more tribal population can reveal the detailed account of maternal and paternal lineages as well as genetic affinity of the Indian population.

Chapter 1

Introduction To The Study

INTRODUCTION:

Tracing about the origin and ancestral links of *homo sapiens* have been the subject of curiosity for various scientists. And a number of scholars have devoted themselves to disclose these hidden mysteries of Human origin and dispersal on earth.

Where did we come from, and how did we get here? This is the question which genetic anthropology field is seeking an answer for. DNA studies indicate that all modern humans share a common female ancestor who lived in Africa about 140,000 years ago, and all men share a common male ancestor who lived in Africa about 60,000 years ago. These were not the only humans who lived in these eras, and the human genome still contains many genetic traits of their contemporaries. Humanity's most recent common ancestors are identifiable because their lineages have survived by chance in the special pieces of DNA that are passed down the gender lines nearly unaltered from one generation to the next. These ancestors are part of a growing body of fossil and DNA evidence indicating that modern humans arose in sub-Saharan Africa and began migrating, starting about 65,000 years ago, to populate first southern Asia, China, Java, and later Europe. Each of us living today has DNA that contains the story of our ancient ancestors' journeys.

When DNA is passed to our next generation, the processes that make each person unique from their parents is the combination of both their genomes. Some special pieces of DNA, however, remain virtually unaltered as they pass from parent to offsprings. One of these pieces are carried by Y chromosome. It is passed only from father to son. Secondly, mitochondrial DNA (mtDNA), is passed (with few

exceptions) only from mother to child. Since the DNA in the Y chromosome does not undergo crossing over, it is like a genetic surname that allows scientists to trace back their paternal lineages. Similarly, mtDNA allows both men and women to trace their maternal lineages. Both the Y chromosome DNA and mtDNA are subject to occasional harmless mutations that become inheritable genetic markers. After several generations, almost all male and female inhabitants of the region in which it arose carry a particular genetic marker. When people leave that region, they carry the marker with them. By studying the genes of many different indigenous populations, scientists can trace when and where a particular marker arose. Each marker contained in a person's DNA represents a location and migration pattern of that person's ancient ancestors. For example, roughly 70% of English men, 95% of Spanish men, and 95% of Irish men have a distinctive Y-chromosome mutation known as M173. The distribution of people with this mutation, in conjunction with other DNA analysis, indicates that they moved north out of Spain into England and Ireland at the end of the last ice age (genomics.energy.gov).

Information about the history of our species comes from two main sources: the paleo-anthropological record and historical inferences based on current genetic differences observed in humans. Although both sources of information are fragmentary, they have been converging in recent years on the same general story (Underhill *et; al.)*.

Since the 1990s, it has become common to use multilocus genotypes to distinguish different human groups and to allocate individuals to groups (Bamshad et al. 2004). These data have led to an examination of the biological validity of races as evolutionary lineages and the description of races in cladistic terms. The technique of multilocus genotyping has been used to determine patterns of human demographic history. Thus, the concept of "race" afforded by these

techniques is synonymous with ancestry broadly understood (Berg *et al.,)*.

Y chromosome and mitochondrial DNA are transmitted uni-parentally through father and mother, respectively and don't under go any recombination. Hence, markers present on both are useful to trace the paternal and maternal lineages. Haplotypes can be constructed by combining the allelic status of multiple markers, which would provide adequate information for establishing paternal lineages. The non-coding region (D-loop) of mtDNA, which harbors two hyper variable regions (HVR I and HVRII), shows variation between different populations. A large number of studies have been conducted on various populations using Y chromosome markers and mtDNA D-loop region to understand their origin, evolution and migration.

Indian populations reveal striking diversities in terms of language, marriage practices as well as in their genetic architecture. The social structure of the Indian population is governed by the hierarchical caste system. In India, there are nearly 5,000 well-defined endogamous populations. In addition to the native populations, there are a few migrant populations inhabiting various parts of India. Several important historical migrations into India caused amalgamation of migrant populations with the local population groups. Major demographic event like migrations, population bottlenecks and population expansion leave genetic imprints and alter gene frequencies. These imprints are passed onto successive generations, thus preserving the population's history within the population. Therefore, we have undertaken to disclose the genetic information about how different caste and tribal populations of India help to construct ecognize and help to construct the evolutionary tree (Cavalli-Sforza et al.,).

Two major routes have been proposed for the initial peopling of East Asia; one via Central Asia to Northeast Asia, which subsequently expanded towards Southeast Asia and beyond, and the other through

India to Southeast Asia and further to different regions of East Asia.[1] It is pertinent in this context that the Indian subcontinent has been considered as a major corridor for the migration of human populations to East Asia.[2-4] Given its unique geographic position, Northeast India is the only region which currently forms a land bridge between the Indian subcontinent and Southeast Asia, hence hypothesized as an important passage for the initial peopling of East Asia. This region is inhabited by populations belonging to Indo-European, Tibeto-Burman and Austro-Asiatic linguistic families.

‘‘BHUMIJ TRIBE’’ come under austro-asiatic linguistic population. Austro-Asiatic speakers, hypothesized as probably the earliest settlers in the Indian subcontinent ([5] and references their in), are also found in other parts of India as well as in East/Southeast Asia. Therefore, if Northeast India had served as an initial corridor, it is likely that the Austro-Asiatic tribes of this region should provide hitherto missing genetic link, which may reflect genetic continuity between Indian and East/Southeast Asian populations. Based on mitochondrial DNA (mtDNA) and Y-chromosome markers, Cordaux et al. [6] observed genetic discontinuity between the Indian and southeast Asian populations and inferred that Northeast India might have acted as a barrier rather than the facilitator of the movement of populations both into and out of India.

However, this study include only ‘‘BHUMIJ’’ Tribe of Jharkhand region from Jamshedpur district. Further evidence is needed by way of determining the mtDNA and Y-chromosome haplogroups/lineages of the Austro-Asiatic tribes of the northeastern region and their comparison with appropriate set of South and Southeast Asian populations. Jharkhand is basically an agricultural land. Geographically it is covered by jungles, mountains, rivers and Chotanagpur plateau etc.

1.2 BACKGROUND :

HUMAN GENOME DIVERSITY PROJECT (HGDP) :-

The HGD Project was started internationally on mid-September of 1993 and it has 13 countries participating in it. The Human Genome Diversity Project is an international project that seeks to understand the diversity and unity of the entire human species.

The Human Genome Diversity Project (HGDP) aims to collect biological samples from different population groups throughout the world, with the aim of building up a representative database of human genetic diversity. This seems a laudable aim, but the Project has been enmeshed in massive controversy since it was first proposed in 1991, with violent reactions from many of the indigenous people's groups it proposes to study.

The eminent geneticist ***Luigi Luca Cavalli-Sforza of Stanford University*** first conceived by the HGDP. For many years, he and other geneticists and anthropologists have been visiting different ethnic groups around the world, collecting samples, and trying to build up a picture of how different human populations are related to each other. The samples are seen as immensely valuable, but they are in laboratories spread around the world. In 1991, Cavalli-Sforza and a number of colleagues wrote a letter to the scientific journal, Genomics, pointing out the need for a systematic study of the whole range of human genetic diversity, within the context of the Human Genome Project. They pointed to a problem: 'The populations that can tell us most about our evolutionary past are those that have been isolated for some time, are likely to be linguistically and culturally distinct and are often surrounded by geographic barriers. Such isolated populations are being rapidly merged with their neighbours, however, destroying irrevocably the information needed to reconstruct our evolutionary history. It would be tragically ironic if, during the same decade that

biological tools for understanding our species were created, major opportunities for applying them were squandered.

Major demographic events like migration, population bottlenecks and population expansion leave genetic imprints where gene frequency of the genome is altered (Thangaraj et, al., 1998). These imprints are passed onto successive generations thus preserving the population history within the population. ***In general, human beings group themselves into units in such a way that members between units rarely exchange genes due to cultural and geographical barriers resulting in genetic divergence of population***. The Human Genome Diversity Project proposed in early nineties is a combined effort preceded by anthropologists, geneticists, doctors, linguists and other scholars from around the world aims at collecting the blood samples from different ethnic populations throughout the world aiming at building up a representative database of human genetic diversity.

The reason lying behind selecting only tribes for sampling is that they are believed to have been isolated during an evolutionary time, linguistically and culturally distinct and are often isolated by geographic barriers and thus prove to be best tools for study.

IN THIS PROJECT, THE SUBJECT OF GENETIC STUDY IS ''BHUMIJ TRIBE'' FROM JHARKHAND (CHOTANAGPUR PLATEAU), INDIA .

1.3 STATEMENT OF PURPOSE :

How does DNA helps us to trace back?

Y chromosome and ***mitochondrial DNA*** are transmitted uni-parentally through father and mother respectively and do not undergo any recombination. Hence, markers present on both are useful to trace the paternal and maternal lineages. Haplotypes can be constructed by combining the allelic status of multiple markers, which would provide adequate information for establishing paternal lineages. The non-coding