GENETIC AFFINITY OF NEWA POPULATION OF KATHMANDU VALLEY AND CHANGPA POPULATION OF LADHAK

M.Sc. Thesis
2016

Submitted to
Central Department of Biotechnology
Tribhuvan University
Kirtipur, Kathmandu, Nepal

BY

Rajdip Basnet

Registration No: 5-2-37-998-2007

Supervisors

Prof. Dr. Tilak R. Shrestha
Central Department of Biotechnology
Tribhuvan University,
Kirtipur, Nepal

Prof. Kumarasamy Thangaraj
Scientist & Group Leader
Evolutionary and Medical Genetics
Laboratory.
Centre for Cellular and Molecular
Biology, Hyderabad, India

Acknowledgement

This study was accomplished by the help of so many people and institutions, who provided their constant support, guidance and encouragement. First, I would like to express my sincere gratitude to my thesis supervisor Prof. Dr. Tilak R. Shrestha for providing me the opportunity to complete a M.Sc. dissertation under his supervision, for his constant support, patience, understanding, guidance from the time of development of the proposal through the experimental protocols for the lab work and write up of the thesis. I am indebted to him for facilitating a collaborative research with Prof. K Thangaraj of CCMB (Hyderabad, India), which provided the opportunity for me to conduct the present research at CCMB Hyderabad. I feel highly honored to work under the guidance of eminent scientist Prof. K Thangaraj, who inspired me in the field of molecular biology. I am grateful to CCMB and Prof. K Thangaraj, for funding the laboratory costs of this research and ensuring that chemicals were available on time.

I am indebted to Dr. Niraj Rai for his relentless support, motivation, enthusiasm and inspiration throughout my Dissertation. I am grateful to him for the comments made on the experimental protocols and the review of this thesis during most of the stages when it was being written. My sincere thanks to Dr. Maanasa Raghavan, A. Govardhana Reddy, Mr. Sheikh Nizzamuddin, Mr. Satya Prakash for helping me in learning techniques and also for developing my knowledge. I would also like to thank Sri, A. Govardhana Reddy, Mr. Suryanarayan and Mr. Hanif for providing me the supplies for smooth conduct of my experiments. I want to take this opportunity to also thank all the members of Dr. K. Thangaraj laboratory and Center Department of Biotechnology for their cooperation and support.

I would like to express my gratitude to Prof. Dr. Rajani Malla, former HOD of CDBT-TU (HOD at our time) and Prof. Dr. Krishna Das Manandhar (HOD Of CDBF-TU) for their official support during my whole M.Sc. period. Worries in life are halved and happiness is doubled when shared among friends. I thank Ram Reddy, Deepak Kashyap, Shubadra jena, Venkatapati challa, P. Meghna, Usha Kabilan, Nagendra Awasthi, Nagendra Verma, Sandesh Maharjan, Ankit Jaiswal, Ram Dhami, Ganga Phuyal, Isha Pradhan, for their helping nature and friendly companionship.

The only rock I know that stays steady, the institution I know that works, is the family. My immense gratitude to my family, my parents for their patience, understanding, love and unwavering support throughout my dissertation.

I am grateful to the many volunteers who helped during the blood collection. I would like to Thank Niraj Rai for providing DNA samples belonging to Ladhak populations. Last but not the least it is a worthwhile to mention words of special thanks to all those who had given blood samples for this study, without their help this project would have not been possible. It is with a sense of pride and pleasure when I look to acknowledge those who have been a source of encouragement in my entire endeavor. To all others who helped me along the way — Thank You

Table of Contents

Chapters

List of Abbreviations

AMH: Anatomically modern Human

AA: Austro Asiatic

aDNA: Ancient DNA

Alt: Altitude

BP: Before Present

CMS: Composite of multiple signals.

CmtDNA: Complete mitochondrial DNA

CRS: The Cambridge Reference Sequence

D-loop: displacement loop

DNAsp: DNA Sequence Polymorphism.

EHH: extended haplotype homozygosity

EPAS1: Endothelial PAS domain-containing protein 1

EDAR: Ectodysplasin A Receptor

ExoSAP: exonuclease shrimp alkaline phosphatase

Fst: Wright's fixation index.

gDNA: genomic DNA

GenBank: NIH genetic sequence database, an annotated collection of all publicly available DNA sequences

GW: Genome Wide

Hg(s): Haplogroup(s)

HGNC: HUGO Gene Nomenclature Committee

His: Integrated haplotype score.

HIF2A: Hypoxia-Inducible Factor 2-Alpha (EPAS1/HIF2A)

HVR: Hyper variable region.

HVS-1/HVS-II: first or second hypervariable segment of mtDNA.

IE: Indo European

Kya: Thousand years ago

Kyr: Thousand years

LRH: long-range haplotype.

LSBL: locus-specific branch length metric.

MAF: Minor allele frequency.

MRCA: Most recent common ancestor

MPD: Mean Pairwise Difference

MSEA: Mainland Southeast Asia

mtDNA: mitochondria DNA

ND: Nucleotide Diversity.

OOA: Out of Africa.

PCR: Polymerase chain reaction

Pmol/ul : Picomol/microliter

rCRS: Revised Cambridge reference sequence.

SDS: Sodium Dodecyl Sulfate

SNP: Single Nucleotide polymorphism

TB: Tibeto-Burman

TAE: Tris bae-Acetic Acid-EDTA

TE: Tris EDTA

TMRCA: Time to the most recent common ancestor

XP-EHH: cross-population extended haplotype homozygosity.

XP-CLR: cross-population composite likelihood ratio

YPB: Years before present.

List of Figures

List of Tables

Abstract

Nepal harbors a variety of ancient and indigenous haplogroups that have been largely ignored in several recent studies. To evaluate the matrilineal gene pool of the Himalayan populations and their genetic adaptation to the high-altitude hypoxia environment, 695 (508 newly generated) samples belonging to three different populations of Nepal (Newar, Magar and Brahmin) and 3 different populations from India (North India, Changpa and general individuals of Ladhak) were analyzed. Investigation was performed based on the high resolution mtDNA sequencing followed by the most precise identification of East Asian, Southeast Asian, South Asian and West Eurasian Haplogroups. Further, Phylogeographic analysis of the genealogical lineages where obtained by comparing the whole mtDNA sequences of the studied populations with the other Asian populations. Desired variants of Epas1 and EDAR gene were obtained through direct sequencing. Newar are considered to be the original and ancient inhabitants of the Kathmandu valley but when and where the Newar people originated from remains contentious. Hypothesis are proposed to have shared ancestry with the East Eurasian, South Asian and West Eurasian. Present study revealed the presence of albeit limited, ancient and deep time depth South Asian lineages in Newar, Magar and Brahmin. Further, Newar experiences several waves of migration adding several East Asian, South Asian as well as West Eurasian lineages into Newar via multiple dispersal, from several distinct sources in different time, rather than just one or two major admixture events in the Neolithic/Bronze age. Most of the East Eurasian haplotype, observed in Newar, Magar and Brahmin branched off directly from the nodes occupied by the Tibetan lineage, albeit a few haplotypes were shared in between Nepali and Northeast Indian. Taking into account the previous studies on East Eurasian maternal components, now it is convincing that the major Genetic Influx from East Asia across the Himalayas into Nepal occurred in between 3-6 Kya and these lineages has made a hefty contribution to the modern gene pools of the Newar and Magar. Hence, our analysis supports a close genetic relationship between Newar and East Asian population with substantial genetic contribution from South Asia and West Eurasia. This was further supported by the analysis of the EPAS1 gene and 1540C variants of the EDAR gene. In a good agreement with the previous studies on Tibetans and Sherpa, high-altitude adaptation in Ladhak also identified positive signal of selection in genetic variants of EPAS1 gene. Further, a strong linear correlation between genotyped alleles of EPAS1 gene and different level of altitude were detected in the Himalayan populations, suggesting that extremely high-altitude hypoxia environment exert a selective effect on Epas1 variants. Changpa shares most of their matrilineal lineages with the indigenous Tibetans suggesting that the Tibetans are the ancestral populations of Changpa and the adaptive traits for the high-altitude adaptation has been recently inherited form their ancestors in Tibet. Whereas the general individuals from Ladhak shows higher proportion of West Eurasian maternal components with substantial genetic contribution from Tibet and India.

1 Introduction

1.1 Background

Over 7 Billion people lives in our planet. The human species with individual of many shapes, sizes, and colors appear to be very diverse. How did we get to where we are today? People have always been curious about their origins. For decades the only clues were the scattered bones and artifacts/artefact our ancestors left behind on their journeys. Genetic studies and fossil record/evidence indicate that archaic humans evolved to anatomically modern humans (AMH) solely in Africa. A tiny group of AMH left Africa ≈between 125- 60 KYA. Over time, these AMH spread throughout the world and replaced the other existing populations of the genus Homo such as Homo erectus and Neanderthals. This assumption was solely based on the several scientific evidences. The major evidence came from the DNA analysis of the modern human as well as the analysis of the ancient DNA (aDNA), in which most of us (modern day humans) can trace our origins to one branch of AMH who left Africa. As the AMH spread throughout the world, new mutations get accumulated on the way. Hence, DNA present on contemporary humans have accumulated different pattern of mutations and these differences/ polymorphism/ genetic variants/ markers, provide a record of our relatedness and genetic history. Analysis of the genomic data also provides the insights on the spread of culture, whether the spread is due to the movement of people or ideas. Due to the advancement in DNA sequencing and methods for the enrichment and extraction of ancient DNA, we are able to answer the several answers such as, gene flow between AMH and the Hominins, the origins of the native Americans (still debated), the relative importance of the movement of people and ideas in cultural transitions (such as where did the agriculture emerged and spread on the other regions of the world). (Fu et al., 2016; Maanasa Raghavan et al., 2015; Stewart & Chinnery, 2015).

Y-chromosome only detected in males are inherited along the paternal side (father to sons). Whereas, mtDNA is only inherited along the maternal side (mother to daughter and son). Hence, these markers also called as uniparental markers are widely used in population genetics study. Further, Y and mtDNA remains constant from generation to generation – Except for the accumulated mutations. The mutations in the mtDNA occurs so rarely and always originate as unique events in a single individual at some point in the past. Hence, they define unique lines of descent—like a clan. If a person shares mtDNA mutation with someone, then that person must have shared an ancestor at some point in the past. These mtDNA genome analysis will assign individuals to different clans/groups. In genetic terms, these clans are known as haplogroups (Hg)—a group of people who share a set of genetic markers and therefore share an ancestor (Ingman et al. 2000). By examining these regions and constructing trees on the basis of their relationships, we can infer where we came from and hopefully how we got to where we live today. Similarly, we can also make inference about the various demographic events in the past, such as; bottleneck (dramatic reduction in number of population), Founder effect (population expansion in the past) and Genetic drift (losses of genetic variation).

As the archaic human start to spread out of Africa, they slowly increased in numbers and start to inhabit different places in Earth, where they get exposed to myriad (countless) new environments, pathogens, food/diets and they were forced to adapt, leading to the great

diversity observed today (Nielsen et al., 2017; Nielsen, Hellmann, Hubisz, Bustamante, & Clark, 2007; Stewart & Chinnery, 2015; Vitti, Grossman, & Sabeti, 2013). Natural selection was first described by Darwin and Wallace in 1858, which explain about the idea of beneficial traits. In a population, over the time the frequency of traits which improve an individual's chances to survive and reproduce will be more frequent.

1.2 Nepal | Facts and History

1.2.1 Geography

Himalayan Kingdom Nepal, is situated in the middle of Asia making top of the Indian sub-continent separated from Eurasia by great inaccessible towering mountains called Himalayas in the north. Nepal is Landlocked by India on three sides and china in north; sandwiched between the two-great ancient civilization of China and India (Whelpton, 2005). The area of the country is 147,181 sq. km with around 900km east west stretch and 250 km North south stretch having population size of nearly 30 million ("Central Bureau of Statistics of Nepal: Major Highlights.," 2011). According to the National census 2011, Nepal a land of diversity, includes a patchwork of different religions and harbors a total of 125 different ethnic groups with 123 linguistic communities

The two largest groups include the Chettri and Brahmin. Nepali is the most commonly spoken language and is related to Indo-European (IE) language family. Whereas, Nepal Bhasa (Newari) is a Tibeto-Burman (TB) language, related to the Sino-Tibetan language family. Modern day Nepal is primarily dominated by the Hinduism, while Buddhism is second. Nepal is the steepest country with tremendous geographic diversity in the world with the elevation ranging from 60m above the sea level to the highest point on Earth. Hence, Nepal harbor wider climatic zones (diverse environments) ranging from tropical to artic climate. The physiographic region of Nepal ranges from tropical forest in the south plain to the Himalayas in the north of which the later includes the world tallest (8848 meters/29,028 feet) mountain, Mount Everest called as Sagarmatha or Chomolungma in Nepali and Tibetan respectively (Pranjal, 2015). However, the southern Nepal (Terai plain) is a tropical monsoonal lowland extends nearly 800 Km from East to West to the Indo-Gangetic plain. The upper Himalayan region is snowy covered with ice. Terai, the southernmost region of Nepal, is predominantly Hindu and consists of low-altitude plains, which comprise the northern edge of the Gangetic Plain that extends into North India. Between these two extremes lie the intermediate hills and valleys that are home to the majority of the Nepalese population, whose cultural practices incorporate both Buddhist and Hindu traditions (Whelpton, 2005).

1.2.2 Caste system of Nepal

Nepalese caste system follows the classical Chaturvarnashram model in which the society is divided into four major social classes or varna: A) Brahmin, B) Kshatriya, C) Vaishya, D) Sudra, and E) untouchables (Levine, 1987). Though, the caste system is still intact today, the rules are not as rigid as they were in the past.

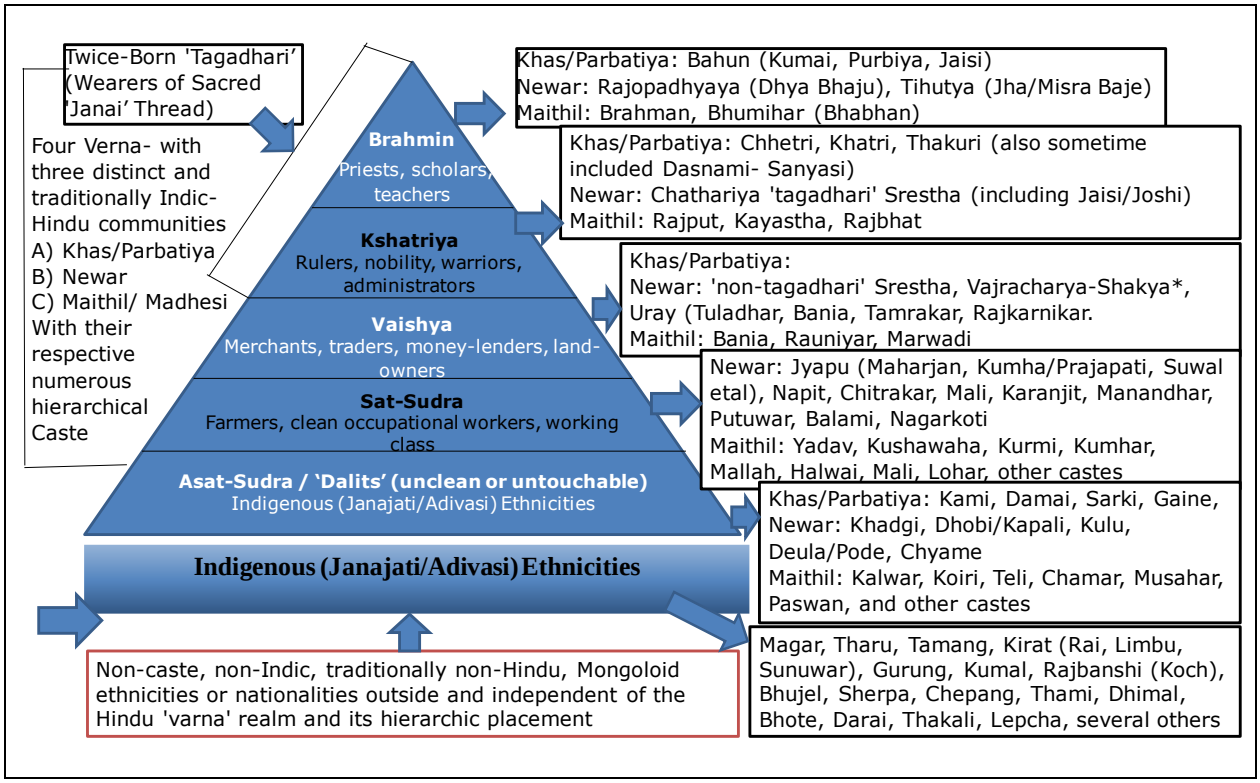


Figure 1.1 | Caste system of Nepal. Social stratification was traditionally only prevalent in Hindu societies of the Khas, Newar and Madhesi. However, several indigenous "Adivasi/Janajati" tribes of Nepal were also incorporated into the caste system/hierarchy after the unification of Nepal (18th century) (Levine, 1987).

1.3 Population under study

1.3.1 Newar (Newa or Newah)

1.3.1.1 History

From the prehistoric time the term Nepal referred the valley Kathmandu of area nearly 600 sq. km. Although Newars are found in every part of the country, historically Newars are the main habitat of the Kathmandu valley as well as its surrounding places. Kathmandu is famous for fine woodcarving, sculpture, bronze and statues. The extremely beautiful Temples, stupas, monasteries, palaces, Shrines and monuments are glorious treasure of Nepal as well as heritage of the whole mankind (Pranjal, 2015). There are no any archaeological studies to clarify the history of Newar, inferences about early periods are based on Nepalese chronicles and legends. In the prehistoric time, the present-day Kathmandu valley was submerged into water making a vast lake. Though some tribal populations such as Nag Bansi (serpentine clan) were inhabited the surrounding hill around the lake. The most common legend of settlement states as; Manjushri came from Tibet, China to the valley and drained the water and made the valley habitable (Dixit, 2012) . According to the Hogduon 1847, the earliest inhabitant of Nepal was the group of people belonging to proto Australoid origin known as Kusunda (Sarkar & Ghosh, 2003). Proto Australoid refers to the ancient hunter gatherer who are descended from the first major wave of AMH to leave Africa. Historians suggest the Newars consists several mixed groups including the Austroids, Dravidian, and then to Mongols and to Aryans. However, according to Dr. Regmi (Regmi, 1965) valley of Kathmandu was first inhabited by the Kol and Bhil tribes of Southern plain immigrated

through the route of Baghmati River. To this group a significant number of Dravidian speakers were also added (Regmi, 1965). Kol and Bhil are the ancient populations know from the Ramayana, one of the great epics of India. Bhil are the largest tribal populations of the Rajasthan-Gujarat region and second largest in the Maharashtra-Madhya Pradesh region(Singh, 1994). Then Mongoloid people from North of the Himalayas immigrated. They were Kiratas and so powerful with vast in number. Rai and Limbu of the contemporary Eastern Nepal claim to be the Kiratas. They started ruling of the valley. Tibeto-Burman Kiratas, who reigned until the 4th century C.E. were subsequently replaced by the Licchavi Dynasty (400–800 C.E.), a group that migrated from northern India (present-day Bihar state) and introduced Hindu culture and traditions to the people of Nepal (Regmi, 1965). The Valley witnessed several different waves of migrations and increase in the settlements grew and prospered because of the fertility of the soil, favorable climate, malaria-free zone and ancient trade location/ route between the China and India (Dixit, 2012; Lewis, 1988). Newars were must probably the descendent of the groups who initially inhabited the Kathmandu valley. According to the historians the term Newars referred to anyone who live in the valley irrespective of their racial origin and features. It is believed that Newars existed as early as sixth century BC (Regmi, 1965). The Urbanization of the Kathmandu valley was started from 4th Century AD (Regmi, 1965). In the 6th century BC, Prince Siddhartha Gautama was born into the Shakya royal family of Kapilvastu, Lumbini, later famously known as Buddha.

After the IE language, TB language family is considered as second largest language family in the world based on number of native speakers. It has been speculated that the TB language family originated in China and spread from the Yellow river of China into Myanmar and the greater Himalayan region (Degener, 2005). TB language family with no doubt is considered to have arrived from East Asia into the South Asia and is mainly present in Nepal, Bhutan, Bangladesh, Northeast India and in some parts of Pakistan. The Newar and Magar group analyzed in this study speaks the Branch of the Tibeto-Burman family. In Nepal, the earliest dated epigraph was found at Bajrayogini (Vajrayogini) Sankhu and dated A.D. 1172. Later during the Malla period (1200-1768), the Nepal Bhasa flourished, both in inscriptions and in manuscripts. Vast collections of classical Nepal Bhasa literature are preserved in Nepal in the National Archives, the Kesher library and the Asa archives, whereas thousands of manuscripts from Nepal have been carried away by western scholars and deposited in different libraries of the world. As architects, sculptors and painters Newar created innovative and powerful forms. Throughout Southeast and Central Asia they were acknowledged masters at casting bronze and copper, arts which they still practice (Regmi, 1965; Whelpton, 2005).

Geographically, Kathmandu valley lies in the central Himalayas at the altitude of 1200-1400m. The land is so fertile with fairly sufficient amount of rainfall that creates an ideal place for agriculture. Fairly cold climate, available of black soil and alluvium which are ideal for brick making, are the ideal factor for the immigration since prehistoric time into the valley from beyond the Himalayas as well as from southern plain (Regmi, 1965). Most of the Newar people live in Kathmandu valley and rest of them live mostly in the urban area of the country. Newars are socioeconomically and politically most advance community with 1,321,933 (5.0% of the total populations) populations throughout the country making 6th largest community in the country. After the Unification of Nepal in 18th Century, Newar started to emigrate towards various trade routes within Nepal as well as in North East India and settled there (Gurung, 1980). The language

of Newar community is commonly known as Nepal Bhasa or Newari, belongs to Tibeto-Burman family scripted in 12 different script prominently Ranjana Script. However, unlike other Tibeto-Burman family this language is also heavily influenced by the Sanskrit and other Indo-Aryan Languages (Tuladhar, 2000).

1.3.1.2 Newar Caste System

Newar caste system can also be grouped on the basis of occupation as well as religious affiliation as shown in the **Table 1.1** & **Table 1.2**. Rajopadhyaya Brahmins occupied the highest social position in the Hindu side whereas in the Buddhists is represented by the Vajracharya.

Table 1.1 | A simplified model of Caste system of Newar based on the occupations. Newar caste system is based on the Vedic Varna model which involves the grouping of populations based on their hereditary occupations. (Furer-Haimendorf, 1956; David N. Gellner, 2009; David N Gellner & Quigley, 1995; Regmi 1960; Whelpton, 2005)

Varna/Caste	Sub-Caste	Family Surname or Sur Groups
I. Brahmin	Upadhyaya Brahman-	Rajopadhyaya, Rimal, Sharma, Subedi - Royal Priests and Family Purohits
	Tirhute/Maithil Brahman-	Misra, Bhatta, Jha — Temple Priests
II. Kshatriya	Non-Brahmin Priests-	Joshi, Daiwagya — Assistant Priests and Astrologers
		Karmacharya, Guruwacharya, Acharya — Tantric temple priests and assistants
	Chatharia Shrestha-	Thakur: Malla, Simha, Dev, Pradhan, Vardhan, Varman, Pradhananga — Malla royals and nobles
		Chatharia: Amatya, Bhaju, Dhaubhadel, Hada, Kayastha, Mahapatra, Maske, Mool, Rajbaidya, Rajbanshi, Rajbhandari, Piya, Sainju, etc. — Malla nobles, courtiers and administrators
III. Vaisya	Panchtharia Shrestha-	Achaju, Bhadra, Kacchipati, Madhikarmi, Mulmi, Naeju, Nyacchon, Sahu, Shrestha —Administrators or Merchants and Traders
	Udas/Uraye -	Udas/Uraye: Tuladhar, Kansakar, Bania, Sthapit, Rajkarnikar, etc. - Merchants and Traders, Craftsmen
IV. Clean (Sat) Occupational Castes	Jyapu –	Maharjan, Dangol, Suwal, Shilpakar, Awale, Prajapati, etc.: Farmers and related fields.
	Sayami –	Manandhar, Sayami - Oil pressers
	Kau –	Nakarmi - Blacksmiths
	Nau –	Napit - Barbers and Nail cutters
	Gathu –	Malakar, Mali - Gardeners and flowers specialists
	Pu –	Chitrakara - Painters and Artists
	Bha –	Karanjit - Funeral Specialists
	Cipa –	Ranjitkar - Dyers of clothes Various other small castes
V. Unclean (Asat) Occupational Castes (Sudra)	Kusah –	Kapali, Kusule - Musicians and tailors, temple cleaners
	Naye –	Khadgi, Shahi - Butchers and milksellers
	Kulu –	Kulu - Cobblers and Drummakers
	Chama: Khala	Pode, Dhyola, Pujari - Sweepers, fishermen.

Newar community is divided into Buddhist and Hindu section however most of the rituals, customs and most importantly languages are same for both sector. Intermarries between two sectors of similar Hierarchical group is common. Both of these group of Newar society consist a number of hierarchically arranged caste, which structure resembles more closely to North India and Madhesis than that of Khas (Parbatiyas), in which all four Varna (Brahmin, Kshatriya, Vaishya and Shudra) and untouchables are represented (Levine, 1987; Whelpton, 2005).

Table 1.2 | Caste system of Newar based on Religious affiliation. The Newar caste system is unique than those prevalent in the Hindu caste systems (Khas and Madhesi societies), in which the Buddhist "ex-monks" forms a "double-headed" element to the Newar caste system.

<table>
<tr><td colspan="4">Caste system of Newar based on religious affiliation (only included the major caste)</td></tr>
<tr><td colspan="2"></td><td>Shivamargi (Hindu)</td><td>Baudhamargi (Buddhist)</td></tr>
<tr><td rowspan="4">Pure Castes entitled to secret Tantric initiation</td><td>I. Priestly Castes</td><td>A).Brahman— Rajcpadhyaya, Dyabhaju (Hindu priests and portents)</td><td>A). Gubhaju— Bajracharya (Buddhist priests and purohits)</td></tr>
<tr><td rowspan="3">II. High Castes</td><td>B). Kshatriya - Chathariya Shrestha (Nobility, administrators) Malla, Thakura, Pradhan, Pradhananga, Joshi, Kayastha, Hada, Amatya, Rajvamshi, Patravamshi, Rajvaidhya, Rajbhandari, Rajlavat, Karmacharya, etc.</td><td>B). Bare —Shakya (Buddhist temple priests, goldsmiths)</td></tr>
<tr><td>C). Parnchthariya Shrestha (Traders, merchants) — 'Shrestha', Acharju, Sahukhala; Dhuhlthel, Mimi, Dolakha Shresthas, etc.</td><td>C). Uraya (Traders, merchants, craftsmen) —Tuladhar, Tamrakar, Kansakar, Baniya, Rajkarnikar, Halwai, Sthapit, etc</td></tr>
<tr><td></td><td></td></tr>
<tr><td rowspan="6">Pure service castes</td><td>III. Middle ranking farming caste</td><td colspan="2">D). Jyapu (farmers, potters, craftsmen)
Maharjan, Dangol, Suwal, Singh, Prajapati, Awaley, Kumhaa. Gopali, Vyanjankara, Khusah (Tandukar),Shilpakar, etc.</td></tr>
<tr><td rowspan="5">IV. Lower service castes</td><td colspan="2">E) “Ekthar” castes</td></tr>
<tr><td>Pun-Chitrakara (painters)</td><td>Bha - Karanjit (funeral ritual</td></tr>
<tr><td>Nau - Napit (barbers)</td><td>Putuwar-Dali (farmers)</td></tr>
<tr><td>Gathu -Mali (gardeners)</td><td>Kau – Nakarmi (blacksmiths)</td></tr>
<tr><td>Balami, Pahari (farmers of Valley outskirt)</td><td>Cipa - Ranjitkar (dyers)</td></tr>
<tr><td></td><td>Sayami-Manandhar (oilpressers)</td></tr>
<tr><td rowspan="6">Impure castes from which upper castes cannot accept water; served by their own priest</td><td rowspan="3">V. Impure but touchable</td><td colspan="2">F. Jogi - Kapali (death ritual specialists)</td></tr>
<tr><td colspan="2">g. Dhobi -Kuale (washermen, tailors)</td></tr>
<tr><td colspan="2">h. Naya -Khadgi -Kasain (butchers, musicians)</td></tr>
<tr><td rowspan="3">VI. Impure and untouchable</td><td colspan="2">I. Kulu-Suvarnakar (cobblers, leather workers)</td></tr>
<tr><td colspan="2">j. Pode- Deula (sweepers, cleaners)</td></tr>
<tr><td colspan="2">K. Chyame- Chamarkhala (sweepers, cleaners)</td></tr>
</table>

A short information on the major selected sub-caste of Newa is given below.

Bajracharya/ Vajracharya: Bajracharya (also called Gubhaju/ guru bhaju) are the highest-ranking Buddhist priest among the Newar communities. Priesthood is considered too be the only exclusive occupation of Bajracharya (Regmi 1960).

Shakya: Similar to the Bajracharya Shakya are also the higher ranked Buddhist priest. The best known Shakya was Siddhartha Gautam Shakya born in Lumbini, Nepal, who was the founder of Buddhism and came to be known as Gautama Buddha. Shakya word is derived from Sanskrit which means "the one who is capable. Unlike Bajracharya who only follows the occupation of priesthood, Shakya also follows the hereditary occupation of goldsmiths(Shakya, 2000).

Udaya / Urays: Uraya are the Traders/ merchants/craftsmen who follow the Newar Buddhism. Udaya are considered to be The early primary carriers of trade between Nepal and Tibet (Bista, 1972).

Shrestha or Syasyah: In Newar society the Shrestha are ranked below the priestly caste Rajopadhyay Brahmin. Shrestha is a Sanskrit word adopted by the Newar high caste Syasyah, which simply means the best or excellent. It is believed that the word Shrestha is derived from the Newar word Syasyah, which itself is derivation of a Sanskrit word syasta. The first use of the word syasta is found in the oldest chronicle of Nepal, the Gopalarajavanmsavai, which dates from fourteenth century. Shrestha was a title given to those who served as administrators at the Malla courts. Although many Syasyah began to adopt Shrestha as their caste name as early as the eighteenth century, it has become more common from the 1950s (David N. Gellner, 2009). Although there are no exact data, the Shrestha are believed to be the second largest in number after the **Jyapu**. Although the Shrestha are renowned as traders and administrators, they are found engaged in all sorts of occupations. A large section of them are farmers, especially in rural areas. For instance, in Sankhu, the Shrestha occupy the largest area of land (67.3%). Among the Newars, the Shrestha are considered to be the most educated caste(Whelpton, 2005). Unlike other Newar castes, the Shrestha are found in every district of Nepal. One of the reasons behind it is the adoption of “Shrestha” as one’s surname once a family belonging to any of the Newar caste moves to settle far off places from the Kathmandu Valley. It is believed that Shrestha are the most accommodating castes in Newar society. Sankhu (Sankharapur) is one of the ancient historical Newar town in Kathmandu where the Shrestha form the largest group (David N. Gellner, 2009; Whelpton, 2005).

Jyapu: In Nepal Bhasa, the word Jyapu literally refers to the competent worker. Numerically, they are considered to be the largest group of the Newar community. Jyapu group consist several sub-castes, such as Maharjan, Dangol, Suwal, Prajapati etc. as listed briefly in the table no 1 and 2. In the caste system, Jyapu belong to middle working, farming occupational caste. It is believed that most of the existing indigenous people were incorporated under the Shudra varna of farmers and working-class population(Bista, 1972). Hence, Maharjan are considered as a true descendent of the various original settlers of the Kathmandu valley including Licchavis, Ahirs, Karata, Gopaalas. Kirtipur an ancient city of Nepal, is the center for Newar culture where the Maharjan forms the largest group. History of Kirtipur dates from 1099 A.D (Kirkpatrick, 1811). It is believed that Jyapus were turned into a lower caste group during the Malla period.

Manandhar/ Sayami: On the category of the pure service castes Manandhar are categorized to lower service castes below to Jyapu. Manandhar are oil pressers by occupation (Bista, 1972).

1.3.2 Brahmin

Brahmins are distributed in scattered patterns all over the country, which are considered the highest castes in Nepal. Brahmin speak Nepali (Gorkhali) an Indo-Aryan language derived from Sanskrit. Origin of Brahmins is uncertain however many believes their connections to the south west Asia (the so called “Indo-Aryan invasions”). According to 2011 Nepal census, with 12.2% of Nepal's population (or 32,26,903 peoples) Brahmin forms the second largest groups in Nepal ("Central Bureau of Statistics of Nepal: Major Highlights.," 2011). Social practices of Brahmin are generally based on the Hindu religious epics. Brahmins are further divided into different sub clans (caste and gotras. The term gotras generally denotes all individual who trace descent in an unbroken male line from a common male ancestor. Brahmins, are mostly monogamous and marry within the same sub caste but in different gotras. The Brahmins observe prohibitive rule in case of marriage with cousins and relatives (Whelpton, 2005).

1.3.3 Magar

Magar is one of the indigenous ethnic nationalities of Nepal. As per the 2011 Nepal census, 7.13% of Nepal’s population (or 1,622,421 peoples) belong to the Magar ethnolinguistic group making it the largest indigenous ethnic group in the country. Further, majority of Magar are Buddhists whereas only a few them are non-Buddhist Hindu. The Kham Magar residing on the Rapti Zone are believed to have migrated from Siberia. Rapti Zone is considered to be the original homeland of Magar from were Magar are believed to have spread to the south and east part of Nepal. Magar have their own languages which belongs to the Sino-Tibetan family (Tibeto-Burman branch followed by Magaric). Tibetan influences in the language and culture of Magar indicates their relatedness with the Tibetans from the North. The shamanistic practices possibly brought from Siberia are practiced by Magar(Watters, 1975)

Table 1.3 | Classification of Magar into the several clans on the basis of their Linguistic affiliation.

Language	Septs
Magarkura speakers	Ale, Thapa, Rana
Khamkura/Kham language speakers	Budha, Gharti, Roka, Pun, Jhankri
Kaike speakers	Tarali Magar of Dolpa/ Budha, Gharti, Rokaya, Jhankri

1.3.4 Changpa and general individuals of Ladhak, India

The Changpa are a tribes found mainly in the Changtang (4000m), Ladakh (Jammu and Kashmir of India). They are high altitude pastoralists, raising mainly yaks and goats. Ladhak range is the most dangerous mountain range in the world when it comes to High Altitude Sickness (Hypoxia).

1.4 Hypothesis

Newar are considered to be the original inhabitant of the Kathmandu valley. All the people who had inhabited the valley at any point of time were either Newar or progenitors of Newar. Some believe they have migrated from north of the Himalayas (Tibet) whereas others believe they came from South Asia. The Newa population might have settled in the Kathmandu valley since the prehistoric times. Newa might have gathered several gene flow from several distinct geographic regions (East Eurasia, West Eurasia and South Asia) at different times. No doubt, recent immigrants from both region (Tibet and India) at different times from distinct sources might have exerted a considerable influence on Newar culture and have ultimately absorbed into Newar society. South Asia is the earliest place inhabited by the AMH after leaving Africa. Regarding the geographic proximity of India and Nepal there are every reason to believe that the bulk of contemporary Newar might contain Indigenous and deep time depth Hg. Changpa, the semi nomadic tribes living in the high altitude hypoxia region of Ladhak, India might share their matrilineal lineages with the indigenous Tibetans.

1.5 Objectives

1.5.1 General objective

1. To study the genetic affinity of Newar along with the samples of Brahmin and Magar.
2. To understand the genetic structure of the populations from Ladhak (India) and their genetic adaptation to the high-altitude hypoxia.

1.5.2 Specific Objective

1. Basic Molecular Diversity indices and Neutrality test
2. Haplogrouping and complete mtDNA sequencing of the selected samples
3. Principle component analysis
4. Phylogeography of Dominant haplogroups (Median joining Network, PhyloTree, Contour map and TMRCA)
5. Frequency of the adaptive variant 1540T>C of EDAR gene.
6. To explore the signal of selection and to identify the candidate gene that might have undergone selection in Ladhak region of India.
7. Correlation between the genotyped SNPs and altitude of the studied populations

2 Literature review

2.1 Mitochondrial DNA (mtDNA)

Mitochondria are unusual organelles that contain their own genome. Mitochondria consists a small amount of DNA of 16,569 base pairs (16.5kb). Mitochondria are normally inherited exclusively from the mother; the mitochondria in mammalian sperm are usually destroyed by the egg cell after fertilization (Stewart & Chinnery, 2015).

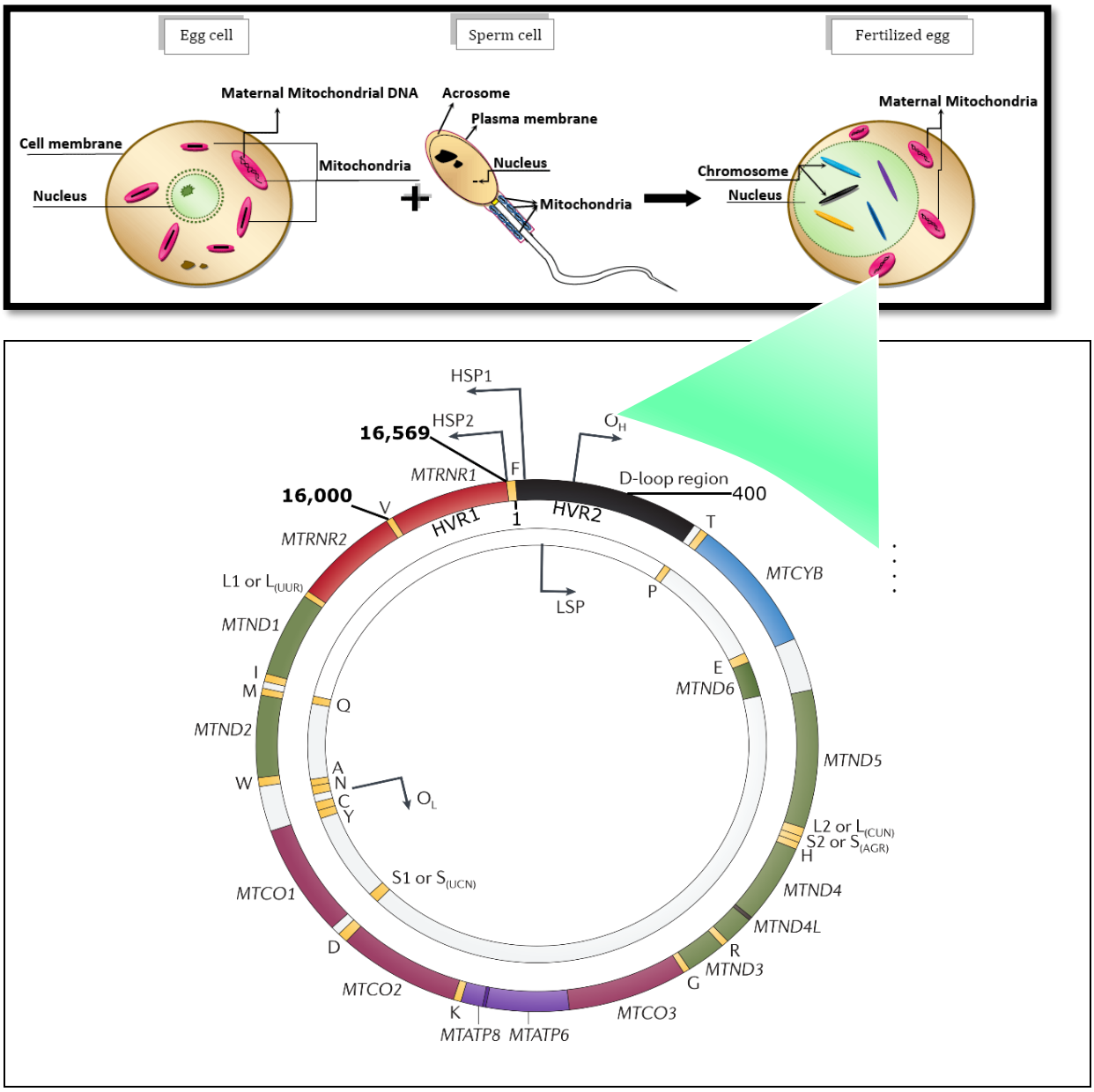


Figure 2.1 | Human Mitochondrial DNA and maternal inheritance of mtDNA. The mitochondria in mammalian sperm are usually destroyed by the egg cell after fertilization**.** In total, 37 genes are present in mtDNA of which, thirteen are involved in manufacturing the enzymes essential in oxidative phosphorylation. Whereas, the other genes play important role in making transfer RNAs (tRNAs) and ribosomal RNAs. Figure used here has been modified from the source (Stewart & Chinnery, 2015).

Some of the important features of mtDNA are (Stewart & Chinnery, 2015):

1. High copy number
2. Maternal inheritance
3. Lack of recombination, and
4. Generally higher mutations rate; almost 10 times higher than in Nuclear DNA

2.1.1 Human mtDNA and its origin in Africa

New mutations arose (and continue to arose) simultaneously with the expansion and migration of the AMH. In this process, initially, mutations must have arisen within a germ cell followed by heteroplasmic and might have survived the mtDNA genetic bottleneck over several generations and eventually have become fixed (homoplasmic) within a maternal line (Stewart & Chinnery, 2015). The mtDNA unique characteristics being uniparental inheritance and associated lack of intermolecular recombination mean that these unique variants have remained confined (restricted) to specific ethnic groups in some geographical locations. Hence mtDNA present in our cells are almost similar to our mother's mtDNA and can be traced back to the mother mother's mtDNA from several hundred/thousands of generations ago. This proposed common female ancestor of the present-day population is termed as "Mitochondrial Eve" and is believed to have lived in Africa over ≈100,000 years ago. Uniparental Inheritance and lack of recombination has lead into the accumulation of mutation in mtDNA over several thousand years. The accumulation of mutations in mtDNA over time have subdivided the human population into several discrete clans (Haplogroups). A haplogroup (hg) labeled from A to Z is a group of people who all share the same pattern of SNPs in their mtDNA (van Oven & Kayser, 2009). These data have been used by population geneticists to infer about the past migration and colonization of the Earth, supporting the 'out of Africa' hypothesis, which proposes that the human mtDNA had its origins in Africa.

Revised Cambridge Reference Sequence (rCRS): Mutations are determined based on comparison with rCRS. When a variation in mtDNA is detected, it actually shows the regions of DNA that differ from the rCRS. Hgs are the main "trunks" of the mtDNA phylogenetic tree and represent extremely ancient family groups which arose tens of thousands of years ago. Over time, the descendants of each Hg formed further subgroups, called Subclades; named using numbers and letters. For example, Subclades of Hg H include H1, H2, H3, H4… and so on. The Subclade H2 can be further classified as H2a, H2b, etc. Further it can be classified into H2a1, H2b1, etc.(van Oven & Kayser, 2009).

2.2 Origin of Anatomically Modern Human (AMH) in Africa

Archaeological sites found in Africa are linked with AMH remains in Africa, which provide the evidence for African origin of AMH. Some of the important archaeological sites are:

A. Jebel Irhoud (Morocco)

   Skeletal remains of five Archaic homo sapiens along with middle stone age tools found in a cave called Jebel Irhoud, in Morocco dated to 315 ±34 thousand years old is the oldest homo sapiens fossils yet discovered (Hublin et al., 2017).

B. Ethiopia

1. **Omo Kibish:** Omo Kibish is a name of an archaeological site in Ethiopia which Contains the partial skeleton of Homo sapiens dated to ≈195,000 years ago(Aubert et al., 2012).
2. **Bouri:** Skeletal remains of four archaic hominins date 2.50,000 and 160,000 years ago. Among these three Hominin crania were identified as Homo sapiens dated 160,000 years before present (BP), including middle stone age tool such as hand axes, cleavers, scrapers, Levallois flake tools, cores and blades(White et al., 2003)

2.3 Potential routes Out Of Africa (OOA)

Comparison of DNA differences among the present-day people, the analyzed DNA samples retrieved from the bones of ancient hominids, archaeological evidence, fossil finds and studies of ancient climates all suggest the Out of Africa (abbreviated as OOA) or African replacement Hypothesis, which argues that every human being is descended from a small group of AMH that originated in Africa, from where they expanded outwards into the different parts of world where they replace other species of Homo such as Neanderthals and Denisovans. Alternative to OOA, multiregional model argues that modern humans evolved simultaneously in multiple times (most probably from Homo erectus), in multiple regions, facilitated by gene flow that results from ongoing migration between locations. This suggested hypothesis of Multiregional model was widely rejected. Phylogenetic tree based on human mtDNA sequences have a root in Africa(van Oven & Kayser, 2009), which is in consistent to the OOA hypothesis. This is further supported by the presence of highest level of diversity among the Africans as compared to any contemporary populations (Rosenberg et al., 2002; Tishkoff et al., 2009). Since AMH after their origins in Africa has existed continuously in Africa longer than any other geographic region and have maintained relatively large effective population size, thus resulting in the high levels of diversity within-populations (Tishkoff et al., 2009). Moreover, all the non-African maternal (mtDNA) lineages (M and N) which populate the rest of the world are derived from a single branch (L3 linage) which is exclusively present in Africa (van Oven & Kayser, 2009). OOA is further strengthened by the several recent studies (Fu et al., 2016; Hublin et al., 2017; Malaspinas et al., 2016; Mallick et al., 2016; Pagani et al., 2016; Pankratov et al., 2016; Sahakyan et al., 2017) which suggests that most of the non-Africans have inherited genes from people who left Africa in a single or multiple pulse around 50,000-120,000 years ago, although those dates are changing rapidly as new studies are unfolded.

Potential routes out of Africa are summarized as:

A) Earliest route (Out of Africa 2): Earliest wave of migration discovered very recently is considered as a failed dispersal because only a handful of Homo sapiens sites has been discovered as being this old outside of Africa. This migration was believed to have occurred in between 130,000-115,000 years ago, along the Nile corridor and into Levant (present day Israel). The presence of anatomically modern human based on fossil evidence has been reported from Middle East (Three burial site: Skhul, Qafzeh and Tabun of Israel) falling within the time range of 100-130 Kyr (Grun et al., 2005) supporting the idea of Early wave of dispersion. However, these early humans did not survive and the skhul and Qafzeh (Israel) hominins (Grun et al., 2005) are probably the remnants of the failed exodus. Surprisingly, a recent study based on the analysis of

teeth believed to be of AMH, claimed the presence of anatomically modern human in southern china about 80 Kyr ago (Y. C. Li et al., 2015).

B) Northern route: Migration from northern Africa might have occurred ≈60,000 years ago through Arabia (Western Asia situated northeast of Africa on the Arabian plate).

C) Southern route: Southern dispersal hypothesis proposed that AMH left Africa at least 70,000 years ago following the costal route of Arabia, across the Indian subcontinent and then into South-East Asia and finally into Australia. Through these place modern humans are thought to have spread rapidly into Southeast Asia and Oceania. This route is supported by various studies based on mtDNA (Behar et al., 2008; Kivisild et al., 2004; Macaulay et al., 2005; Mishmar et al., 2003; Thangaraj et al., 2005), Archaeological and Fossils studies. The stone tool kit assemblages unearthed from above and below the Toba ash deposit in India's Jawalapuram Valley (Andra Pradesh), shows the affinities to the middle stone age African assemblages (Patnaik & Chauhan, 2009). The eruption of Toba, in what is now Indonesia is the largest volcano eruption of the last two million years. Eruption of Toba lead into the voluminous ash deposits across much of the Indian Ocean, Indian Peninsula, and South China Sea (Lane, Chorn, & Johnson, 2013). The presence of modern humans in India at the time of the Toba super-eruption would be consistent with humans having used the southern route, but would remain speculative till further excavations because currently there are no any archaeological evidence of such ancient human migrations along India's west coast and into southern Tamil Nadu. Similarly, another archaeological evidence found in the Jebel Faya site in the Arabian Peninsula allowed authors to consider that the manufacturers of these tools could have dispersed into India as early as 1.25 Kyr ago (Armitage et al., 2011; Petraglia, 2011).

A recent study which include whole genome sequencing data from more than 270 locations worldwide suggested a single dispersion OOA and on leaving Africa, possibly may have immediately separated into two waves of dispersal, one wave is the founder of Australasia and New Guinea and another wave is the founder of the present-day mainland Europeans (Pagani et al., 2016). Another recent study proposed that the Aboriginal Australians and New Guinea's native Papuans diverged from Eurasians around 51-72 Kya, following a single OOA dispersal and subsequent admixture with the other Archaic humans. This study also concluded that all the Aboriginal Australians descend from a single founding population that differentiated around ≈10-32Kya (Malaspinas et al., 2016).

2.4 Intermixing of Modern Humans with Neanderthals and Denisovans

Geneticists proposed that the common ancestor of Homo sapiens, Neanderthals and Denisovans, lived around 6, 00,000 years ago. Whereas, humans and our closest cousins, the Neanderthals (named after the Neander valley in Germany were the species was discovered) split at least 500,000 years ago (Hublin, 2009). It is now well accepted fact that all the AMH who left Africa admixed with Neanderthals (Green et al., 2010), as suggested by the presence of substantial amount (2%) of Neanderthal ancestry found in all contemporary non-African populations studied so far (Prufer et al., 2014; Sankararaman et al., 2014; Sankararaman, Patterson, Li, Paabo, & Reich, 2012; Vernot & Akey, 2015; Wall et al., 2013). Further, comparatively Neanderthals ancestry

contributed more to the East Asians (20% more Neanderthal sequence) than the Europeans, which suggests a further admixture between the Neanderthals and the ancestors of the modern day East Asians who separated from the Europeans (Vernot & Akey, 2015). Recent study whose dataset include 142 distinct human populations, proposed/concluded that ancient human who left Africa interbred with Neanderthals in the Middle East before splitting into distinct groups that headed into Asia and Europe (Mallick et al., 2016).

2.5 Timeline of the major events in human evolutionary genomics

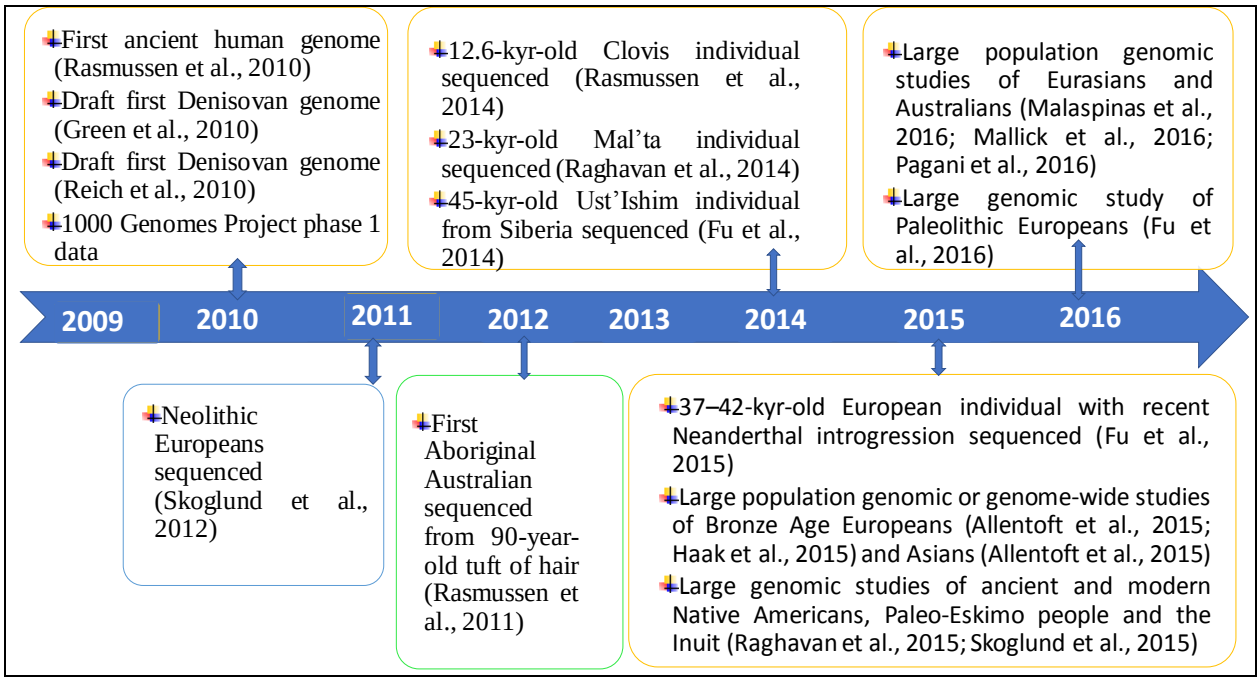


Figure 2.2 | Timeline of the important events in human evolutionary genomics. Large number of studies based on the genomic data has provided important insights into the past human origin/movement. The above timeline provides a short information about the important studies carried out in the past.

As revealed by the analysis of the genes of the modern people including the aDNA, three major movements out of Africa has been proposed by the several researchers as shown in the **Figure 2.3**: first a group that wound up in what is now Australia and the Pacific Islands, then a population that settled in East Asia, and finally western Eurasians or Europeans (Nielsen et al., 2017). The African Hg L3 give rise to the two-major super Hgs, M and N, around 55,000–65,000 YBP. This Hg branches are present in all of the non-African contemporary populations. After the OOA super Hg N is believed to be directed to Eurasia and another super Hg M moved to Asia, giving rise to the several Hgs A, B, C, D, G and F. Super Hg N in Europe led to Hg R most probably in the Caucasus, which is the root of the European Hgs H, J, T, U and V, which emerged 39,000–51,000 YBP. Hg U2I (U2a, U2b and U2c) present in South-Asia might have diverged to its European version U2e ≈45-50 Kyr BP. Hgs S, P, and Q are found in Australasia and were formed ≈48,000 YBP, and Hgs A, B, C and D arose <20,000 YBP and populated East Asia and the New world (Americas) (Allentoft et al., 2015; M. Raghavan et al., 2014; Stewart & Chinnery, 2015).

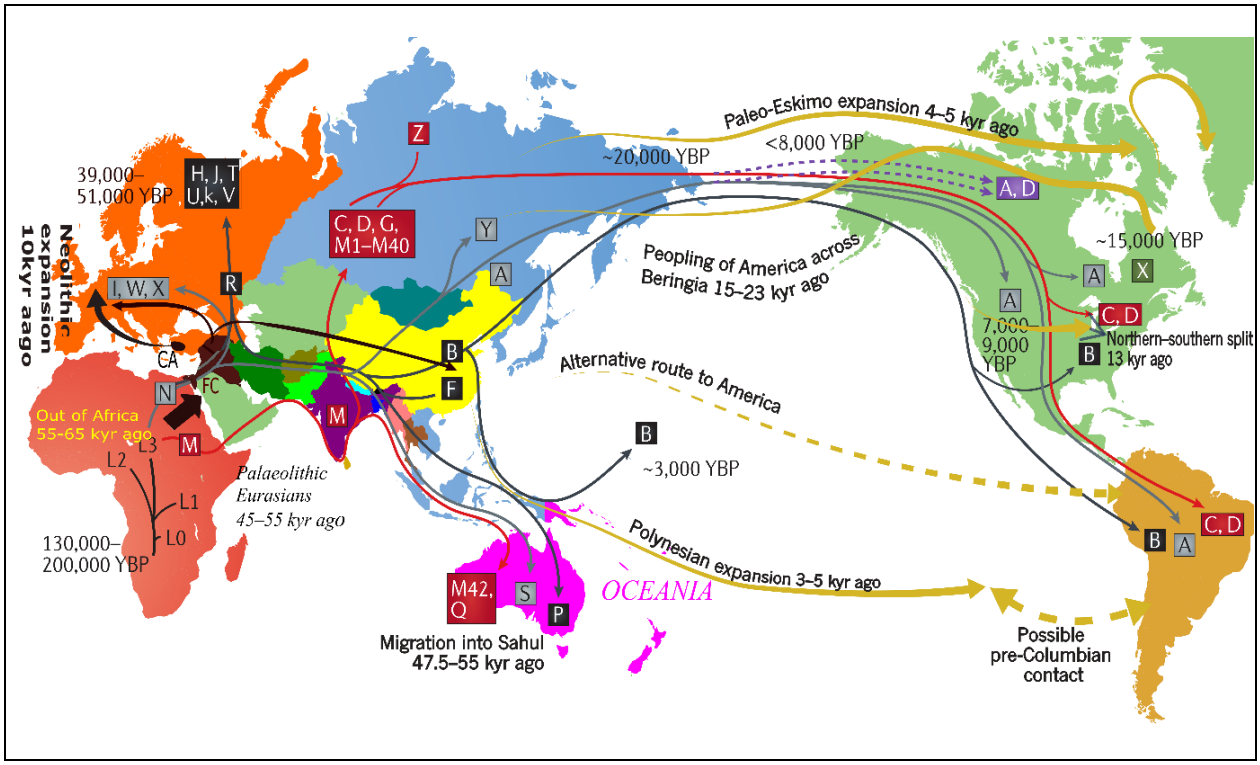


Figure 2.3 | Major human migration across the world showing the distribution of major mtDNA Hgs. Proposed routes of migration that remain controversial are indicated by dashed lines. Several terms used in the map, such as CA: Central Anatolia; FC: Fertile Crescent; IP: Iberian Peninsula; PCS: Pontic–Caspian steppe will be discussed in another section (Indo-European language). Map modified from source (Stewart & Chinnery, 2015)

2.6 The peopling of Asia

Although few study suggests the single dispersal event (Mallick et al., 2016), most evidence supports at least two early waves of migration might have colonized Asia. The first wave includes the ancestors of Australian and the Papuan people, whereas the second waves includes the East Asians. Further A subsequent gene flow occurred into the East Asians but not the Europeans from the remnants of an earlier migration into Asia of Aboriginal Australians ancestors at some point before 20kyr BP (Rasmussen et al., 2011). Ancient genome study which included the genetic analyses of a 37,000-year-old (named as Kotenski 14) human bone (leg bone) form European Russia shows a close relation with the contemporary West Eurasians but not the East Asians (Seguin-Orlando et al., 2014), implies that Western Eurasians and East Eurasians diverged outside Africa between 45-36.2 Kyr BP (Nielsen et al., 2017).

2.6.1 Spread of the Indo-European Languages

It was debated for decades, whether the spread of Indo European language occur due to the migration of people during the Bronze age or during the Earlier spread of agriculture (late Neolithic). The most widely accepted hypothesis for the origin of Indo-European language is the steppe hypothesis which proposed PCS (Ponte-Caspian steppe) as a homeland for the proto Indo-European speakers, who are responsible in spreading of Indo-European language In Europe and parts of Asia. . PCS extends from Northern Romania to Ukraine and to Russia. This hypothesis has been further supported by a recent study based on the ancient DNA (aDNA), mitochondrial as

well as Y-DNA evidence. Inference based on the sequencing of 101 ancient genomes, suggests the large-scale movement and replacement of populations have occured in the Bronze Age (period in between 3000-1000 BC). Further this process was responsible in shaping the contemporary (present day) demographic structure, languages and certain phenotypic traits in both Europe and Asia (Allentoft et al., 2015).

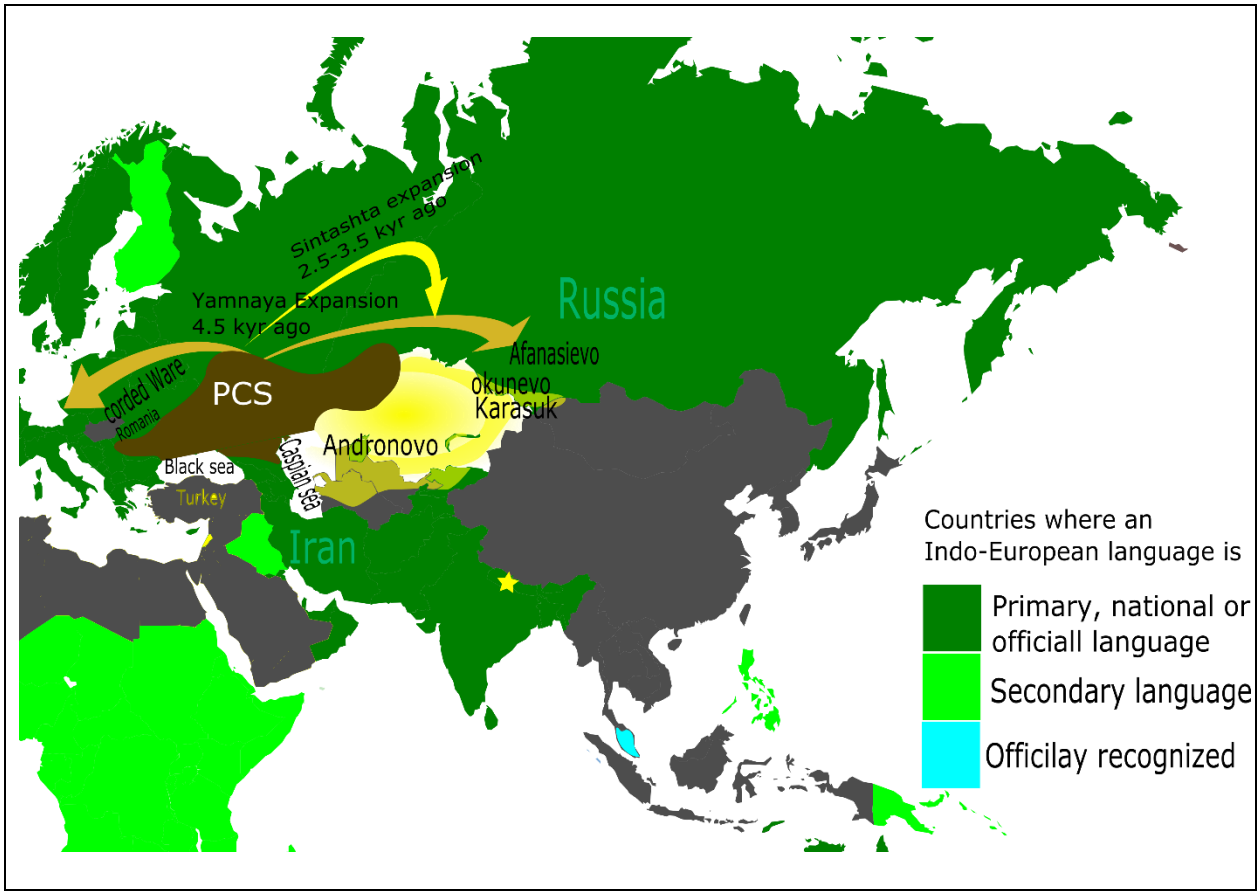


Figure 2.4 | Map showing approximate distribution of Early Bronze Age cultures Yamnaya, Corded Ware, and Afanasievo cultures (Allentoft et al., 2015). The star like shape indicates the geographic position of Nepal in the map. All colored boundaries in Map are approximate. The two arrows (dark yellow) from the PCS directed towards Europe and central Asia shows the Yamnaya expansions occurred at the same time during the Early Bronze Age. Sintashta, Andronovo, Okunevo and Krasuy the middle and Late Bronze age cultures with the east ward migration are shown (indicated by light Yellow) in the map. Contemporary speakers of IE speakers include a broader area of Eurasia as represented by several green colors in the map. Map was created in the present study using Corel Draw Suite X8 (version 18).

Genomic evidence suggests the spread of Yamnaya/Afanasievo culture during the Bronze Age occurred from PCS directed towards Europe and central Asia. These studies showed two later population expansions occurred into the central Asia from Europe and Western Asia. The first expansion occurred around 5 Kyr ago from the Yamnaya herders to the Central Asia and Europe at similar time. Further these Yamnaya people in central Asia formed the Afanasievo culture and were replaced by the second wave of migration from the individual belonging to Sintashta culture (Allentoft et al., 2015), who move in from the Europe and Urals. As the Sintashta culture spread towards Altai it evolved to Andronovo culture. This culture than gradually admixed and replaced by the East Asians people belonging to Mezhovskaya and Karasuk culture. Hence this culture first expanded towards East and later into South towards central Asia in the Bronze Age ≈3.8 Kyr ago.

The Andronovo (possibly the Sintashta groups before Andronovo) might have infiltrated and dominated the soma-using Bactrian Margiana Archaeological Complex (BMAC) in Turkmenistan/northern Afghanistan by 3.5-4 Kyr BP. BMAC gradually came in contact and admixed with the Indus valley civilization in Baluchistan ≈4kyr BP, especially the pastoralism dominated group dispersed further ≈3.5Kyr ago in two directions; one across South Asia and another to the northern Iran including Syria. Previously it was debated for decades, whether the spread of Indo European language occur due to the migration of people during the Bronze age or at the time of the spread of agriculture. Large-scale migrations and population replacement resulted into cultural transitions during the Early bronze age was most probably the likely scenario for the most hypothesized Indo-European migrations which is most probably responsible in shaping the contemporary (present day) demographic structure, languages and certain phenotypic traits in both Europe and Asia (Allentoft et al., 2015).

However, Genome wide studies have strictly denied these hypothesis (Metspalu et al., 2011; Moorjani et al., 2013). In the last few years, recent large-scale study based on Genome-wide (GW) has been carried out. These study shows India harbors two major ancestral components: one component ASI (Ancestral South India) is restricted to South Indians and the other component ANI (Ancestral North Indian) shows more closer affinity to the population from central Asia, middle East, the Caucasus and Europe (Metspalu et al., 2011; Moorjani et al., 2013). These haplotype diversities observed in both of the Indian ancestry were older than the hypothesized Indo-Aryan Invasions ≈3,500 years ago (Metspalu et al., 2011). East Asian lineages are very common among the extant Central Asian. If the proposed immigration from Central Asia to South Asia have taken place, East Asian component which is common in extant Central Asians, should be evident in the Subcontinent if it had experienced large-scale Bronze Age immigration from Central Asia (Metspalu et al., 2011). However, a recent study based on the aDNA denies the presence of East Eurasian component in the relevant source region on the Central Asia during the early Bronze Age (Allentoft et al., 2015). Contrary to the previous study (Underhill et al., 2010; Wilson et al., 2001) a recent study (Batini et al., 2015; Karmin et al., 2015; Poznik et al., 2016) based on the Y-DNA analysis supports the Bronze age genetic influx into the South Asia from the Central Asia.

2.6.2 Peopling of South Asia

South Asia refers to the southern geographic region of the Asian continent comprising modern-day nation's territories of Sri Lanka, Bangladesh, Bhutan, Nepal, India, Maldives and Pakistan. A map showing the geographic territory of the South Asia, Central Asia, Caucasus, East Asia and Mainland South East Asia (MSEA) are shown in the Appendix H (**Figure 8.1**). South Asia is one of the earliest regions considered to be occupied by the early AMHs (Macaulay et al., 2005; Majumder, 2010; Mellars, Gori, Carr, Soares, & Richards, 2013). The early out-of-Africa migrants colonized South Asia, dominated by India, and including Pakistan, Nepal, Bhutan, Bangladesh, Sri Lanka and parts of South-East Asia, especially Myanmar. Southern dispersal hypothesis proposed that anatomically modern humans (AMH) left Africa at least 70,000 years ago following the costal route of Arabia, across the Indian subcontinent and then into South-East Asia and finally into Australia. This route is supported by various studies based on mtDNA (Behar et al., 2008; Kivisild et al., 2004; Macaulay et al., 2005; Mishmar et al., 2003; Thangaraj et al., 2005). The presence of

high genetic diversity among the south Asian, probably the second highest after the sub-Saharan populations also supports the above point. South Asian populations shows relatively higher affinities with the Caucasus and Central Asian populations rather than other western Eurasian populations.

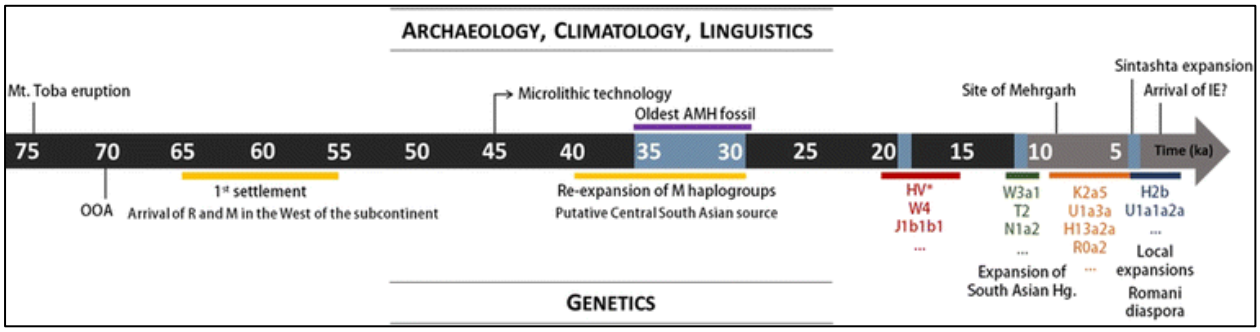


Figure 2.5 | Timeline showing the evolution of AMH in South Asia based on the genomic data including Archaeology, climatology and linguistic evidence. Pleistocene and Holocene are represented by the Black and grey colored portions in the arrow. Blue color indicates the climate change in which periods between 35 and 30 Kyr represents drier period, ≈18 Kyr corresponds to Last glacial maximum, and the `12kyr and 4.2 Kyr point the younger Dryas. Lineages in red are shown as an indicator for the putative Glacial/postglacial genetic flux from West Eurasia; green color-coded lineages indicate gene flow from West Eurasia around the Pleistocene/Holocene transition, orange directs Neolithic period and blue for the genetic events in the last 4ka (Silva et al., 2017).

2.6.2.1 Indigenous South Asian specific mtDNA Hgs (Pleistocene Modern Human Settlement in South Asia)

According to the recent study older lineages in South Asia belonging to super hg M are most like to be originated either in South/West, East and Central India. Whereas the Hg lineages belonging to R (R30, R7, R31, U2) are most likely to be originated in South/ west India (Silva et al., 2017). The younger dating of south Asian specific hg M and hg R suggests a later arrival of these hg in South Asia than the Mount Toba Eruption (Mellars et al., 2013; Silva et al., 2017). It is estimated that the 30% of the Earth surface was covered by ice during the Pleistocene resulting in the decrease in the sea level of the entire Earth surface ≈by 100 meters or even more. History of Using Micro lithic technology (stone tool) in India can be traced back to ≈45Kyr ago (Mishra, Chauhan, & Singhvi, 2013).

- M2, M6, M32'56, M36, M39 originated in **south/west** India.
- M4'67, M35, M52 originated in **Central India**.
- M13b, M31, M42b, M61, M49, M50 and M60 originated in **East India**
- M38, M65, M45, M5b, M5c, M34, M57, M33a displays a signal of dispersion from the East and central India. The age of these lineage is ≈45-35 Kyr.
- M4'67: This hg shows an extraordinary multi branching are also observed in other south Asian population including Nepal (Fornarino et al., 2009; Gayden et al., 2013; Silva et al., 2017; Wang et al., 2012).
- Other Hgs that are likely to be originated within India are listed below (Silva et al., 2017):
    - R = R5, R6, R7, R8, R30, R31 and U2i (U2a, U2b and U2c)
    - N (N1'5).

2.6.2.2 West Eurasian specific mtDNA Hg in South Asia

The presence of west Eurasian Hg in south Asia is the indicator of the re-peopling after the Last Glacial Maximum (LGM). Proto Dravidian language might have originated ≈15Kyr BP in Western Asia. This was followed by Invent of agriculture which occurred ≈11Kyr ago. It has been proposed that a multiple dispersal from the North-West, from several distinct sources, rather than just one or two major admixture events in the Neolithic /Bronze Age has contributed to the South Asian gene pool (west Eurasian hg makes 20% of the south Asian gene pool). Here is the list of non-autochthonous N lineages present in South Asia most likely the subclades belonging to West Eurasian origin (Silva et al., 2017). Subclades of West Eurasian Hgs (found in India): H2b, H7b, H13, H15a, H29, HV, I1, J1b, J1d, K1a, K2a, N1a, R0a, R1a, R2, T1a, T2, U1, U7, V2a, W and X2

2.6.2.2.1 Last Glacial Maximum (LGM)

After the first initial settlement ≈45-65 Kyr ago the earliest movement into the South Asia occurred from West Eurasia (most probably from Near East). Whereas the genetic influx from the East Asia to South Asia occurred very lately. South Asian populations shows relatively higher affinities with the Caucasus and Central Asian populations rather than other western Eurasian populations. The evidence for the movement from west Eurasia to the South Asia is provided by the lineage N1a1b1. Other similar Hgs in this category include pre-HV2, HV+146! HV+9716, HV+73! Pre-U1c, J1d and a basal clade within T2). These LGM lineages corresponds to the approximate 2% of the south Asian lineage. Some other lineages arrived in the late glacial period: 16–13 Kyr (HV12b, U7a, HV + 16311!, W4, I1, and J1b1b1)- all belongs to Near Eastern clades. Previous study associated the presence of west Eurasian Hg among higher caste Brahmin and religious Muslims and their rarity among the lower caste Indians, with the arrival of Indo-Aryan speakers. These findings were based on the presence of high frequencies of U7 lineage among the Brahmin and religious Muslims (Palanichamy et al., 2015)

2.6.2.2.2 Early postglacial arrivals

At the end of the younger Dryas, the movement of the population increased/intensified as suggested by the arrival of many more west Eurasian lineages in South Asia; U7a3a + 6150, W3a1 + 143, W3a1b, T2 + 195 + 4225, U1a3 + 10253, N1a2, U7a + 12373 and T2e2. Further other lineages included in this category are: (U7b + 16309!, T2d1a, W6, T2b, , and K1a1b2a (Silva et al., 2017). This time (≈12 Kyr BP) also corresponds with the expansion of indigenous (autochthonous) lineage across the south Asia.

2.6.2.2.3 Neolithic Arrival

Neolithic (last stage of the stone tool) is associated with the spread of agriculture along with significant migration with genetic consequences and language replacement. Southwest Asian crops from the Fertile Crescent are the founder of the earlier crops present in South Asia (Fuller, 2007; Kingwell-Banham, Petrie, & Fuller, 2015). These putative Neolithic lineages entered to the South Asia at ≈9.5Kyr from Anatolia, the Caucasus and Iran (Silva et al., 2017). Over the time distinctive nested south Asian specific branch are present in these lineage (3.4% frequency): H13a2a + 8952, HV14 + 150, U1a3a, K2a5 + 2831, X2 + 153! + 7109 and K2a5 + 2831+ 189. The

earliest Neolithic sites, on the Indus Valley around Mehrgarh in Baluchistan, date to before 9 Kyr BP (Jarrige J-F, 2006; Petrie, 2015).

2.6.2.2.4 Bronze Age arrivals

As discussed in the above section, Bronze Age arrivals was most probably responsible for the spread of Indo-European language in South Asia. However, mtDNA study showed most of the genetic influx was restricted to Pakistan. Gene flow in Bronze Age are represented mainly by the lineages

1) From west to south Asia: H29 + 9156 + 4689, R2a + 7142 and U1a1a2a. (Silva et al., 2017).
2) Form South Asia to West: M5a2a4, U2c1b + 146 and M3a1b + 13105 (Silva et al., 2017).

Recent study based on GW concluded that two major founding populations; Ancestral South Indian (ASI) and Ancestral North Indian (ANI) involved in a major mixing in between 1,900-4,200 years BP resulting into the contemporary admixed Indian populations (Moorjani et al., 2013). These mixing events between the highly differentiated populations brought a major profound change characterized by the de-urbanization (people move from urban areas to rural areas) of the Indus civilization (Harappan civilization) causing the increase in population density, most likely introduction of Indo European language and Vedic religion in India and Nepal.

2.6.2.2.5 Peopling of Nepal

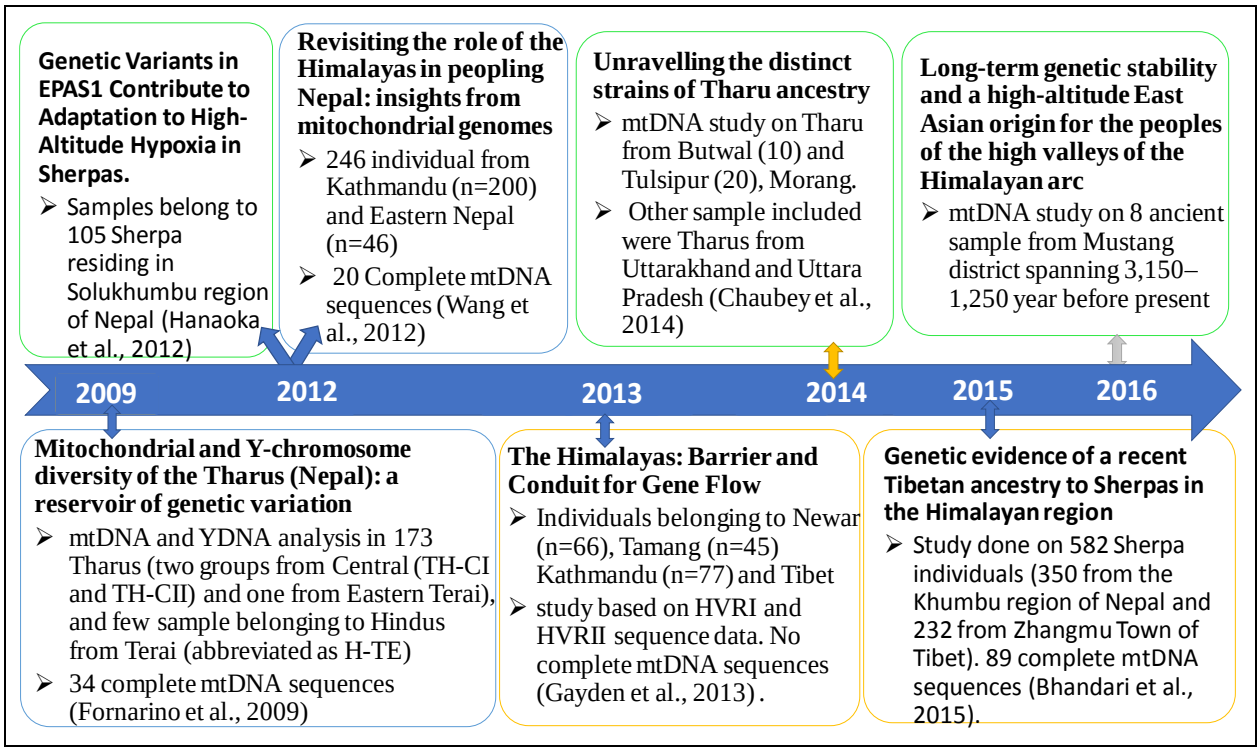


Figure 2.6 | Timeline of the major studies in human evolutionary genomics related to Nepali population. Only a few number of studies are conducted so far on the Nepali populations regarding the past human origin/movement.

The geographical location of Nepal is located at the Historical crossroads of the major cultural event in Asia, the Middle East, Central Asia and Europe. Majority of the East Asian lineages

observed in the Nepali populations were also found among the Tibetans (Wang et al., 2012), Similarly, another study (Gayden et al., 2013) proposed a higher proportion of East Asian component among the Tamang whereas Newar and Kathmandu contains both East and South Central Asian lineages (Gayden et al., 2013; Wang et al., 2012).

Although highly debated, many studies claimed the Paleolithic colonization and Neolithic expansion of modern humans on the Tibetan Plateau. The first wave of migration reached the Tibetan plateau in between 18-30 Kya (upper Paleolithic) followed by the second major (distinct) wave of migration around 7-10 Kyr ago. This was followed by the rapid population expansion along with the establishment of farming and Yak pastoralism on the plateau in the early Neolithic (Xuebin Qi et al., 2013). Previous studies indicated a recent gene flow from East Asia to Nepal across the Himalayan mountain range ≈6 Kyr ago (Gayden et al., 2013; Wang et al., 2012). Similarly, Nepal could be inhabited during the initial peopling of South Asia, because of presence of several deep-rooted south Indian lineages, including M2 (50Ky), R (50Kyr) and U2 (40Kyr) (Gayden et al., 2013). However, these studies claims were based on the analysis of HVR region of mtDNA, which needs to be analyzed further based on complete mtDNA phylogeny.

2.7 Adaptation to the Local environment

As the Early Modern Human migrate into the different parts of world, they are exposed to the different local environment. The Geographic distribution of genomic signature in several modern populations was established relatively late in Human history (Allentoft et al., 2015). Human skin pigmentation shows a strong positive correlation with the ultraviolet radiation intensity. Those population which settled near the equator have comparatively a darker skin to protect against the skin damage and photolysis. Whereas those population at the higher latitudes, receives lower reduction in the exposure of the sunlight (that means lower vitamin D production) has a lighter skin. These facts suggest that variation in skin color might be due to the adaptation via natural selection. As the population move into higher latitude from Africa, several genes including SLC24A5, SLC45A2 and MC1R might have under gone selection which might favor the diminished number, size and density of melanosomes resulting in a lighter skin pigmentation (Lamason et al., 2005; Norton et al., 2007).

Similarly, the population living in the higher altitude regions such as Tibet, Nepal, Ladhak (India), shows decrease in the red blood cell production, which might be due to the adaption via natural selection in response to the low oxygen environment, also termed as Hypoxia. Several genomic studies in the past has suggested these changes in population is driven by the alteration in the allele frequency mainly in the two genes involved in the hypoxia response pathway: **EPAS1 (Endothelial PAS Domain Protein 1)** and EGLN1. The EPAS1 gene, also called as HIF2A, provides instructions for making a protein called **hypoxia-inducible factor 2-alpha (HIF-2α).** HIF-2 α is a subunit of a protein complex called HIF, which plays a critical role in the body's ability to adapt to changing oxygen levels. HIF controls several important genes including a hormone called erythropoietin, which regulates red blood cell production. These proteins are constantly produced in our body. When the level of oxygen is normal HIF-2α are targeted by other proteins and are degraded, so it does not build up. Likewise, when the level of oxygen decreases below normal (hypoxia), HIF-2α is degraded at a slower rate. This results in the increase in the availability

of HIF and hence stimulates the formation of new blood vessels and the production of red blood cells. These activities help maximize the amount of oxygen that can be delivered to the body's organs and tissues. Studies suggest that the EPAS1 gene is involved in the body's adaptation to high altitude. At higher altitudes, such as in mountainous regions, air pressure is lower and less oxygen enters the body through the lungs. Over time, the body compensates for the lower oxygen levels by changing breathing patterns and producing more red blood cells and blood vessels. Tibetan populations living in the high altitude are believed to have adapted hypoxia environment. EPAS1 gene is among the several gene believed to have undergone selection in the Tibetan populations. Tibetan populations living in the high altitude do not show markedly elevated red blood cell production thereby avoiding high blood viscosity creating cardiovascular risk (Beall et al., 2010; Bigham et al., 2010; Simonson et al., 2010; Yi et al., 2010).

EDAR gene is involved in the development of hair follicles, teeth and sweat glands. This gene harbors a nonsynonymous single nucleotide polymorphism (SNP) (rs3827760) that results in a valine to alanine substitution at position 370 of the amino acid sequence (V370A). Positive selection on the EDAR gene resulted into the higher frequency of 370A among the East Asians and the Americans. Phenotypically, 370A is associated with the thickness of the hair. The positive selection on EDAR gene acted prior to 10,000 years ago. List of several genes under selection, genomic as well as functional evidence for selection are provided in the **Table 2.1** (Bersaglieri et al., 2004; Cook et al., 2009; Enard et al., 2009; Enard et al., 2002; Ge et al., 2012; Grossman et al., 2013; Grossman et al., 2010; Hamblin & Di Rienzo, 2000; Kamberov et al., 2013; Miller, Mason, Clyde, & McGinniss, 1976; Perez-Morga et al., 2005; Simonson et al., 2010; Udpa et al., 2014).

2.7.1 Detecting Natural selection

Genome wide scan of sequence variations is considered to be a powerful tool to detect the molecular signatures of selection in the population (Nielsen et al., 2017).

Selective sweep: As a result of strong positive natural selection, prevalence or rise in frequency of a new beneficial mutation will occur in a population, thus the linked nearby alleles (hitchhiker variants) of the beneficial allele will also rise in frequency. Such as a neutral allele, which is linked to adaption will increase in frequency (in some case, until they become fixed in population), at the same time other non-advantageous allele (ancestral) will fall in frequency (in some cases until the extinction). Due to selection acting on an allele in one population but not in another creates a marked difference in the frequency of that allele between the two populations, which ultimately cause the increase in wright's fixation index (F_{st}) between the populations. Wright's fixation index is the most commonly used metric for population differentiation, which compare the variance of allele frequencies within and between the populations. FST has a central role in population and evolutionary genetics and has wide applications in fields that range from disease association mapping to forensic science (Holsinger & Weir, 2009). High value of F_{st} indicates a stark differentiation between the populations, which is suggestive of directional selection.

Composite method simultaneously integrates information from multiple signals of selection. Such as the cross-population composite likelihood ratio (XP-CLR) simultaneously integrates/models the signal of selection from the multi-locus test of allele frequency differentiation between the two population as a function of selection intensity and the genetic distances between the

Beneficial/causal allele and neutral markers surrounding the target region. The XP-CLR extends Fst to many loci and identifies genetic regions in which change in allele frequency over many sites occur too quickly(Chen, Patterson, & Reich, 2010).

Table 2.1 | Using selection scans to study human evolution.

Gene under selection	Population(s)	Genomic evidence for selection	Functional evidence	Putative adaptive role	References
FOXP2	All (selection predates out-of-Africa migration)	Accelerated evolution in coding region, D, H	Mouse transgenic	Affects development of corticobasal ganglia circuits; thought to be involved in mechanics of speech	(Enard et al., 2009; Enard et al., 2002)
LCT	Northern Europeans, East Africans (pastoralist societies)	EHH, iHS; Fst analysis	Human association study; in vitro lactase expression assay	Confers lactase persistence; allows digestion of lactose into adulthood	(Bersaglieri et al., 2004), (Cook et al., 2009)
EDAR	East Asians and Native, Americans	CMS	Human association study; mouse transgenic	Affects morphology of hair, sweat glands, and mammary glands	(Grossman et al., 2010), (Kamberov et al., 2013)
TLR5	West Africans	CMS	In vitro assay of NF-κB pathway activation	Modulates immune response to bacterial flagellin	(Grossman et al., 2013) ,(Grossman et al., 2010)
DARC	African populations in malaria-endemic regions	Fst	Human association study	Heterozygosis reduces susceptibility to malaria	(Hamblin & Di Rienzo, 2000), (Miller et al., 1976)
APOL1	African populations in trypanosome-endemic regions	CMS	In vitro assay of response to trypanosome invasion	Modulates susceptibility to trypanosomiasis	(Grossman et al., 2010), (Perez-Morga et al., 2005)
HBB	African populations in malaria-endemic regions	LRH LRH similarity	Human association study	Heterozygosis reduces susceptibility to malaria	
EPAS1, EGLN, et al.	Tibetans	iHS, XP-EHH	Human association study	Selected variants decrease hemoglobin concentration and modulate hypoxia response	(Ge et al., 2012), (Simonson et al., 2010)
SLC24A5, SLC45A2	Europeans	Fst analysis,XP-EHH, CMS	Human association study; in vitro assay of melanocyte cultures; zebrafish transgenic	Decreases melanin pigmentation in skin	(Cook et al., 2009)
CBARA1, VAV3, et al.	Ethiopian-highland populations	LSBL, iHS, XP-EHH	Human association study	Selected variants decrease hemoglobin concentration and modulate hypoxia response	(Udpa et al., 2014)

3 Materials and methods

3.1 Ethics Statement

The current study was approved by the Nepal Health Research Council (Kathmandu, Nepal) **(Appendix L)**. The protocol was explained to each individual and an informed consent written in Nepali was obtained by either signature or fingerprint if the subject could not write.

3.2 Sample collection

3.2.1 Nepal

5ml of blood samples were collected from 186 healthy male individuals from Nepal belonging to Maharjan (n=100), Shrestha (n=36) and Newa-mix (50). Sample from Various sub caste of Newar were collected from Kathmandu valley, Bhaktapur and Banepa. Genomic DNA of healthy male individuals belonging to Brahmin (n=48) and Magar (n=36) were obtained from CCMB lab (The Centre for Cellular & Molecular Biology, Hyderabad, India). Magar and Brahmin sample were previously collected from Tanahun district of Nepal (**Figure 3.1**). Genealogical information of sampled individuals was recorded for at least 2 previous generations to carefully choose blood unrelated individuals. The individuals were healthy and unrelated as recognized by the personal interrogation. All ethical guidelines were followed as stipulated by Nepal Health and Research council (NHRC). All the study protocol was approved by review board of NHRC, Nepal. The information about the ethnic population, sample size as well as the sample location are summarized in the **Table 3.7**. For more information about the Newa caste included on the group "Newa mix" see **Table 8.5** (Appendix G).

Along with the Maharjan (MHJ), Shrestha (STH), and Newar mix (NMIX), unpublished raw ABI sequenced file (CCMB and CDBT) belonging to Manandhar (93), Udaya (58), Bajracharya (20) and Shakya (19) populations were also included and reanalyzed in this study. Previously study (M.Sc. thesis) on Newar sub caste was carried out by (Awasthi N., 2014) (Newa sub caste included: Udaya, Shakya and Bajracharya) and (Pradhan I., 2014) (Newa sub caste: Manandhar). These unpublished raw sequences were included to reanalyze the mtDNA sequences according to the updated mtDNA phylogenetic tree Build 17 (Phylotree.org – mtDNA tree Build 17, 18 Feb 2016) (van Oven & Kayser, 2009). PhyloTree is the reference for worldwide phylogeny for mtDNA studies, and its latest update has recently been released (Build 17, 18 February 2016). In total 460 sample belonging to 9 Nepali population where analyzed in this study.

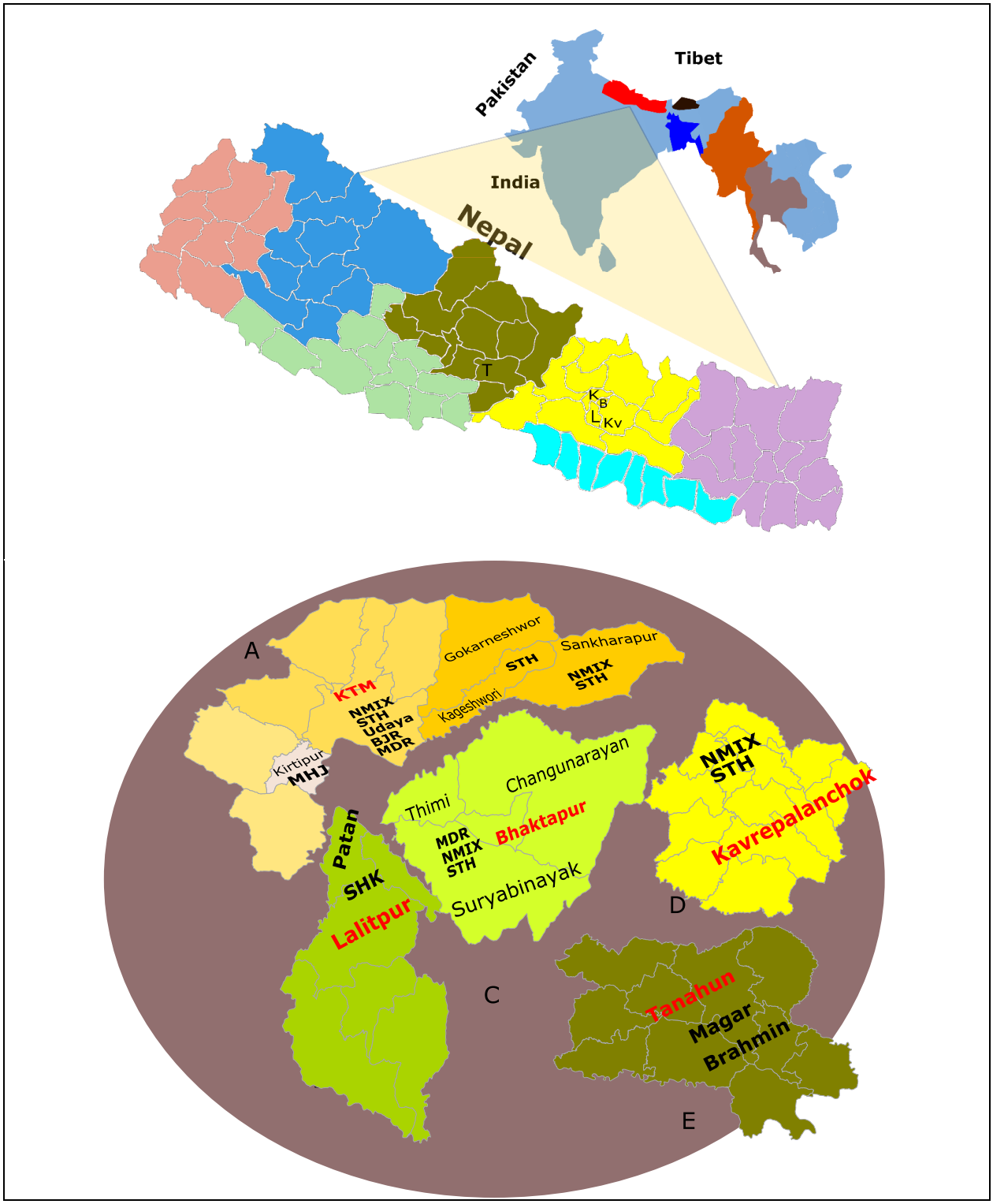


Figure 3.1 | Map of Nepal indicating approximate locations of populations used in this study. In the map T (Tanahun), K (Kathmandu), Kv (Kavrepalanchok) and L (Lalitpur) indicates the name of the district from where the sample was collected. Similarly, Name of the sample and their corresponding districts are also shown in the the map. Likewise, letter A (Kathmandu/KTM), B (Lalitpur), C (Bhaktapur) and D (Tanahun) represent the name of the district.

3.2.2 India

To dissect the matrilineal lineage and to understand the genetic adaption of people living in high altitude regions, 220 stock gDNA (genomic DNA) of healthy male individuals belonging to Ladhak and North India were obtained from CCMB lab. The samples from the Ladhak (**Figure 3.2**) and the Nepali (**Figure 3.1**) populations, was combined to perform the correlation study between altitude and the genotyped variant of the Epas1 gene.

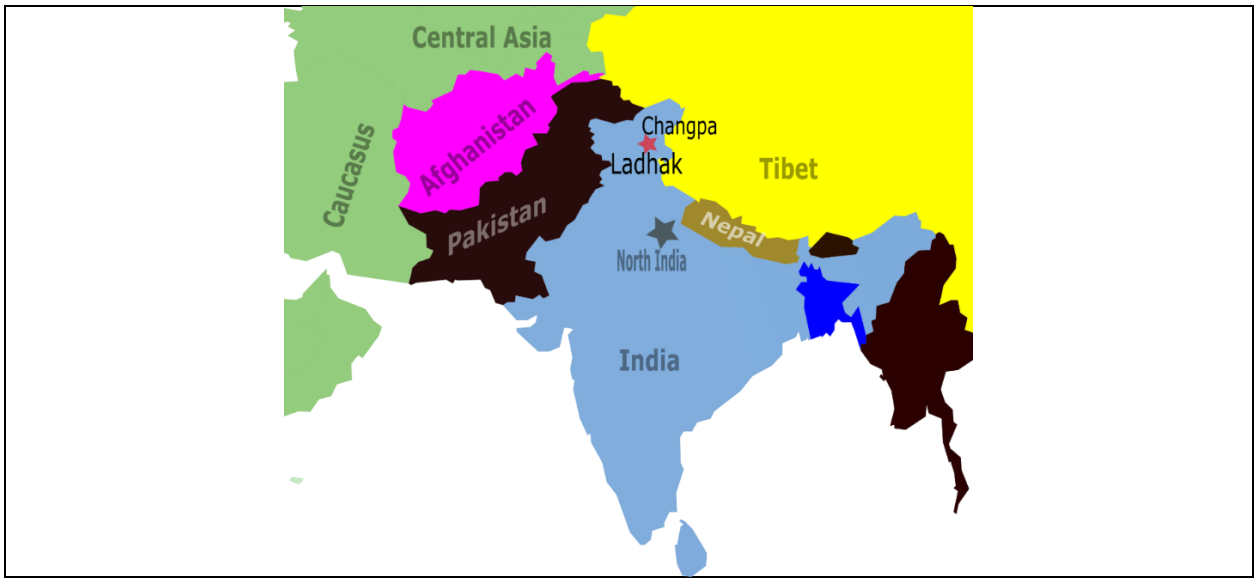


Figure 3.2 | Map of the Asia indicating approximate geographic location of the population used in this study. The star shape in the map shows the approximate geographic location from where the samples were collected. 96 samples (male individual) were collected from the general population of Ladhak, 96 belonging to Changpa Tribe and 46 sample from North India.

3.3 Reagents

3.3.1 Following reagents were prepared and sterilized.

1. Reagent A (Lysis buffer 1)
   Lysis buffer 1 was prepared by adding the following reagents:
   a. Sucrose 109.54 gms (320 mM)
   b. 1 M MgCl2 5.0 ml (5 mM)
   c. Triton X 10.0 ml (1% Triton X 100)
   d. 1 M Tris-HCl (pH-8) 10.0 ml (10mM)
   Total volume was made up to 1000 ml by adding autoclaved double distilled water (DDW). The solution was then autoclaved.

2. Reagent B (Lysis buffer 2)
   Lysis buffer 2 was prepared by adding the following reagents:
   a. 1 M Tris-HCl (pH-8) 40 ml (400mM)
   b. 0.5 M Na-EDTA 12 ml (60mM)
   c. 1 M NaCl 15 ml (150mM)
   Total volume was made up to 95 ml by adding autoclaved DDW (Double distill water). After autoclave 5 ml of 20% SDS (1%) was added to the above solution.

3. Reagent C
   5 M Na-per chlorates (E. Merck, Darmstadt, Germany) of 100 gms was added and then DDW was added to make a total volume of 142 ml. The solution was not autoclaved.

4. TE-Buffer
   a. 1 M Tris-HCl 1.0 ml (10.0mM)
   b. 0.5 M EDTA 0.2 ml (1mM)
   TE buffer was prepared by adding the above reagents and DDW was added to make a final total Volume of 100 ml.
5. Tris Saturated Phenol: Following reagents were added to prepare Tris saturated phenol: Phenol, 0.1% 8-Hydroxy Quinoline, 0.5 M Tris HCl (pH 8.0), 0.1 M Tris HCl (pH 8.0).
6. Chloroform: Isoamyl alcohol (24:1) 24ml of Chloroform was added to 1ml Isoamyl alcohol.
7. 20% SDS: 20g of SDS was dissolved in 80ml DDW at 65^{o}C. The final was volume made up to 100ml.
8. 80% Ethanol: was prepared by adding 80ml of absolute alcohol in 20 ml DDW.
9. 10X TAE Buffer: 48.4g Tris base, 20ml 0.5M EDTA (pH 8.0), 11.402ml glacial acetic acid was mixed and volume was made up to 1 liter.
10. 6x loading Dye: 0.125g of Bromophenol Blue, 0.125g of Xylene Cyanol FF, 15 ml of glycerol. Diluted with DDW to make up volume to 50ml.
11. 10X TBE Buffer: weighed 108g Tris base, 7.44g EDTA, 55g boric acid and added DDW to make volume to 1 liter, pH was adjusted to 8.
12. Ethidium Bromide: 10mg of Ethidium Bromide was added in 1 ml DDW. The solution was then stored in dark bottles.

3.3.2 Reagents used for PCR, plate processing and sequencing

a) 10X PCR Amplification Buffer: 500mM KCl, 100mM Tris (pH 8), 15mM MgCl2, 0.1% gelatin.
b) 25mM MgCl2, 2.5mM dNTPs, Primers 10ng/µl, Taq Polymerase (1 unit)
c) ExoSAP
d) Big Dye ™, 50% HiDi Formamide, 70% ethanol.
e) 3M Sodium Acetate: Dissolved 24.612g of Sodium Acetate in 80ml DDW, pH was adjusted to 5.2 with conc. HCl

3.4 DNA Isolation

Each blood sample was subjected to DNA isolation using standard protocol. Erythrocytes were lysed with 20 mL of erythrocyte lysis buffer (10 mM Tris pH 8.0, 320 mM sucrose, 5 mM MgCl2, 1% Triton X-100; Sigma Chemical Company, St Louis, Mo) for 5 minutes. After complete lysis of erythrocytes, leukocytes were pelleted by centrifugation at 3000rpm for 10 minutes. The leukocyte pellet was dissolved in 2.5mL of leukocyte lysis buffer (400 mM Tris, 60 mM EDTA, 150 mM NaCl, and 1% SDS; Sigma) and mixed thoroughly. To this, 5 M sodium perchlorate (¼th of the volume of reagent B) (E. Merck, Darmstadt, Germany) was added and mixed thoroughly for 2–3 minutes. DNA was precipitated after extracting once with phenol: chloroform and once with chloroform. DNA was washed with 70% ethanol and dissolved in TE buffer (10 mM Tris pH 8.0, 1 mM EDTA). Detail protocol used for the isolation of DNA from Blood is given as below.

a) To the blood samples (5ml), four times the volume (20ml) of reagent A was added in a polypropylene tube. The tube was then mixed gently.

b) Samples were then centrifuged at 3000 rpm for 10 minutes to obtain the pellet free from RBC. The supernatant containing lysed RBC was discarded carefully.
c) The pellet was disturbed thoroughly and half the volumes (2.5ml) of reagent B was added. After adding reagent B, falcon tube was mixed thoroughly and gently inverted for 3-4 minutes to get viscous solution.
d) Then, reagent C (¼th of the volume of reagent B) was added and mixed gently for 3-4 minutes. Equal volume (as that of regents B+C) of Tris-saturated Phenol: Chloroform was added and mixed well and centrifuged at 3000 rpm for 10 minutes. This separate the solution into three layers viz., aqueous layer, protein layer, and solvent layer.
e) Aqueous layer was carefully transferred into other centrifuge tube using a broad mouth tip (protein layer should not be disturbed). Equal volume of Chloroform was added to the supernatant. And mixed gently for a minute and centrifuged for 2500 rpm for 5 minutes.
f) Aqueous phase was transferred into a fresh tube and 2 volumes of ice-cold (chilled) Isopropyl alcohol was added to precipitate DNA.
g) DNA lump was spooled out in a new Eppendorf tube and the alcohol was discarded. DNA was washed twice with 70% ethanol and short spin was given to remove the alcohol. Pellet was dried completely and 50-100µl of TE buffer was added. The stock DNA was then stored at 4^0C.

3.4.1 DNA dilution

Concentration of stock DNA was measured by using Nanodrop Spectrophotometer (Thermo Scientific™ NanoDrop™ spectrophotometer). Final dilution was made to 10 ng/µl for all samples by adding required volume of Mili-Q water. Working dilution was then run on 1% Agarose gel for final confirmation. Genomic bands were visualized under Gel Documentation System (**Figure 8.2**) (Model GGS 12/D-E Gene. Genius Classic, Syngene, UK).

3.4.2 Agarose gel electrophoresis (1%)

1.2 g of Agarose was dissolved in 120ml of 0.5X TAE buffer in a conical flask and was boiled to dissolve Agarose completely. 2µl of ethidium bromide was added from stock solution. Gel was cooled down to 60^o C and then poured onto a gel tray and was allowed to solidify. A standard DNA sample of known concentration was also loaded along with the samples to quantify DNA and electrophoresis was carried out at a constant voltage of 80V. As the sample run halfway of the gel, the gel was observed under ultraviolet light of UV trans-illuminator of gel doc (**Figure 8.3**).

3.5 Primer Design

3.5.1 Mitochondrial DNA primers

24 pairs of PCR primers were designed to amplify the whole mtDNA genome. 24 different set of overlapping primers were ordered and obtained from Bio serve India.

Table 3.1 | 24 sets of primers used to amplify complete mtDNA. 24 pairs of PCR primers were used to amplify the complete/partial mtDNA genome. Each fragment of ≈600bp was amplified independently.

ID	Seq 5'-3'	Length	ID	Seq 5'-3'	Length	ID	Seq 5'-3'	Length
1F	ctcctcaaagcaatacactg	20	9F	gaggcctaacccctgtcttt	20	17F	Tcactctcactgcccaagaa	20
1R	tgctaaatccaccttcgacc	20	9R	attccgaagcctggtaggat	20	17R	ggagaatgggggataggtgt	20
2F	cgatcaacctcaccacctct	20	10F	ctcttcgtctgatccgtcct	20	18F	Tatcactctcctacttacag	20
2R	tggacaaccagctatcacca	20	10R	agcgaaggcttctcaaatca	20	18R	Agaaggatataattcctacg	20
3F	ggactaacccctataccttctgc	23	11F	acgccaaaatccatttcact	20	19F	aaacaacccagctctccctaa	21
3R	ggcaggtcaatttcactggt	20	11R	cgggaattgcatctgttttt	20	19R	Tcgatgatgtggtctttgga	20
4F	aaatcttaccccgcctgttt	20	12F	acgagtacaccgactacggc	20	20F	Acatctgtacccacgccttc	20
4R	aggaatgccattgcgattag	20	12R	tgggtggttggtgtaaatga	20	20R	Agaggggtcagggttgattc	20
5F	tacttcacaaagcgccttcc	20	13F	tttccccctctattgatccc	20	21F	Gcataattaaactttacttc	20
5R	atgaagaatagggcgaaggg	20	13R	gtggccttggtatgtgcttt	20	21R	Agaatattgaggcgccattg	20
6F	tggctcctttaacctctcca	20	14F	cccaccaatcacatgcctat	20	22F	Tgaaacttcggctcactcct	20
6R	aaggattatggatgcggttg	20	14R	tgtagccgttgagttgtggt	20	22R	agctttgggtgctaatggtg	20
7F	actaattaatcccctggccc	20	15F	tctccatctattgatgagggtct	23	23F	tcattggacaagtagcatcc	20
7R	aatggggtgggttttgtatg	20	15R	Aattaggctgtgggtggttg	20	23R	gagtggttaatagggtgatag	21
8F	ctaaccggctttttgccc	18	16F	Gccatactagtctttgccgc	20	24F	caccatcctccgtgaaatca	20
8R	acctagaaggttgcctggct	20	16R	Ttgagaatgagtgtgaggcg	20	24R	aggctaagcgttttgagctg	20

3.5.2 Primer for EPAS and EDAR gene

The sequence for the specific SNP of interest were obtained from the Ensemble website (http://asia.ensembl.org). The Primers to amplify the specific SNP of a particular gene was then designed by using Primer3plus software. Following primers were used to amplify the specific target region of the EPAS and EDAR gene.

Table 3.2 | List of Primers used for EPAS1, and EDAR gene. 3 sets of primers were used to amplify the targeted variant SNP of EPAS1 gene. Similarly, one set of primer was used to amplify the EDAR gene.

Gene	SNP position	Primer	Seq text (5'-3')	Length
EPAS1	rs13419896 (**G**>A)	F	TCCCTCCTCCTTTTTGTTCA	20
		R	TTCCCTTGACCATCTTCCTG	20
	rs1868092 (**G**>A)	F	TGCTGGTGGCTCTTATTTCC	20
		R	AGGTCTGCTCACAAGCAGGT	20
	rs150877473 (**C**>G)	F	GGCTGCCAAGAAAAACTGCA	20
		R	CCCACAGCATACACACCCAT	20
EDAR	rs3827760 (A>G) / (T>C)	F	TCTGCACACAAGGACTCCAC	20
		R	GCGTTCTAGGTGTCGTTTGC	20

Detail information about the location of the SNP in the chromosome, allele frequency and target sequence are given below. The highlighted and underlined portion in the sequence indicates the Forward and Reverse primer respectively. Similarly, the R/S indicate the location of the major variant SNP. MAF indicate Minor allele frequency.

A. Gene: EPAS1 (Endothelial PAS domain-containing protein 1)

1. SNP: rs13419896
Allele: C/G |Ancestral: C | MAF: < 0.01 | (G)|Highest population MAF: 0.18 (A)
Location: Chromosome 2:46329206 (forward strand)
Target Sequence:

ATTCCCTGTTCCCTCCTCCTTTTTGTTCATAAAAAAAATTATTTTTTTCCAAATAAGCAAAATGGAGTCAAACTTTG
CATCCTCAAAATGTATGAAGTAAACATTATTCCTAATGAGCCTCTGGGAAAGTGCTCACCTTTGAACTTGGCCAA
GGATTATGCAGCAAAGAAAAAGTCTTAAGAACTTGATAGAGTGTTAGAGCTTCCTGGGTTTTTAAAGTCAAGTTG
CATATTACATTTCTTTTCCTTAATAGGGGCATTTCCAGAAACCCTTCCTGGTTGAGTAGGCCAGTGTCTGAAAGTG
AAGCRCTAGGATTGGTTACTGACTCTGGTTCAGGGATTGTCATCTGGGTGCGAGGCAACCACAGGGTAGGAGGCA
CCATTGCTAGAATGCTCTTTCTTTCCTAGGACTCAGCACCTATGCCGACAGTCCTTGCAAGACAGGAGGGGAACG
TGTAGGCTATCCTATTTATGGTAAGGGGAGGCCTATGGAAAAACAGGAAGATGGTCAAGGGAAGCTGGC

2. SNP: rs1868092
Allele: G/A |Ancestral: G|MAF: 0.24 (A)|Highest population MAF: 0.50
Location: Chromosome 2:46387063 (forward strand)
Target Sequence:

GATCCTACCACTGCTGGTGGCTCTTATTTCCTGTGGATTTATCATGGGCTATGCCAATGTAAAGTGAAGTATTAA
GTCGTTCATATAAATTGAACATTAGCACAGACCTTTATGAGACTACTGAAGTGAGCTGATAAGACTGGTGAAGG
AAAGACTGTGTTAAATACTTGGTAGTTGCCTAAGTCTGAAAATTCATATTGTATCTTCCATATTTCTGCAAGTAT
TTTTCTATCTATCTATCTATGTGCCTCGGGGCCATTTTTTGTATCCTATGGGGAACCCAAGGAAAGAAGTCATGG
TCTCTGCTTTCAGAAAATTACATGAAGTCGTTTGAGAGCCACACTGTTCTGTATATTCACATAGTGCACTTTGRG
ATGGAGGTCCCGAAGGGAAAGGGGCAGCCTGGGCATTAGGACTTCAATTTGGAGGCGGCTGGGCTGGAGTAGG
GTCTTCCCAGGTAGACTGTGATTGGGAAGAGCAGAGATCCCCTCGGCCAGAGACAGAGCACAGTGAAAGCCCA
GAACTGGACATGAGCATGATATGTTCCCCAATTCAAATTTCACTTTTCCAGATCATGACCTGCTTGTGAGCAGAC
CTTTGTCTCCATGGAGAGATGGGGGCAAGGGGAGAGACCGTGCAGCAGCCTGCGCTC

3. SNP: rs150877473
Allele: C/G |Ancestral: C | MAF: < 0.01 | (G)|Highest population MAF: 0.03
Location: Chromosome 2:46360880 (forward strand)
Target Sequence:

AGAGAGCAGATATTTGGAAAATATAAACAATAGGCTGCCAAGAAAAACTGCAGCTGGGCCCCTCTCATGAATA
TCCATATAAAACTGACTTCAGCTGGTTCTTCCCATCCTTCCACATCCAGGCTCTGGTTTTGGGAAAAAAAGCAAA
GACATGTCCACAGAGCGGGACTTCTTCATGAGGATGAAGTGCACGGTCACCAACAGAGGCCGTACTGTCAACCT
CAAGTCAGCCACCTGGAAGGTAGGGCAACATCAGGCCTGGGTTGGAGTCCCAGGTGTAGGGTAACGGCGGTGC
AGGGGATGCCTAAGGCCCTACCCCCACCCCCAGCACTCTCGGCTCCATGTCTGACCCTTCCACGCCTGTSTCAGG
TCTTGCACTGCACGGGCCAGGTGAAAGTCTACAACAACTGCCCTCCTCACAATAGTCTGTGTGGCTACAAGGAG
CCCCTGCTGTCCTGCCTCATCATCATGTGTGAACCAATCCAGCACCCATCCCACATGGACATCCCCCTGGATAGC
AAGACCTTCCTGAGCCGCCACAGCATGGACATGAAGTTCACCTACTGTGATGACAGGTAGGGGGCCATGGGTGT
GTATGCTGTGGGCAGAGATGGGTCTTACCTGTGTGTG

B. Gene: EDAR (Ectodysplasin A Receptor)
SNP**:** rs3827760
Allele: A/G |Ancestral: A|MAF: 0.24 (G)|Highest population MAF: 0.49
Location: Chromosome 2:46360880 (forward strand)
Target Sequence:

GCCCGCCCACTCCAGTATGTCTGCACACAAGGACTCCACAGCATCCAGCCGCTCAATCTGCACCAGTTTTGTGA
GTAGCTCAGGGATGCTGTAGCCTGCCGTGCTGATGCGGTCAAAGAGTTGCATGCCGTCTGTCATGCCCCCAATCT
CATCCCTCTTCAGGCCGAAGCTCTCGGCGAGGTGGCGCCACGTTTTCACARCAGCCTTCTCAGAGTTGTACGTGG
AGCTGAGCATTCGGCTAGTCTTCTCGAGGCAATCAAATGGCAGCTCCGTGGGGCTAAGACCTACAGACACCAAT
GGCCACAGTTAGATGTTGCAAGTCACAGTCAATAGAAGGTCAACACTGAAAAATTATTTAAATGAAATAAAAAT
TATTTGAAAAAAATTTTTTTAAAATTATAAAACATGCAGAAAGCCACCTAAAACAAATGCATAACTTAAGGCAA
ACGACACCTAGAACGCTGCCAGCCACAAGAGAGGGCTGTGCATGTACCCTGACCCCGGTAATAGCCCCTTCCCC
ACTAGGAAAATGACTGCTATATTGACAATGATAATAATTCCCTCATGACCATTTTTAG

The adaptive variant 1540T>C of EDAR gene in exon12 was genotyped by PCR-direct sequencing using the above-mentioned primer.

3.6 Protocols for PCR

DreamTaq Green PCR Master Mix (2X) is a ready-to-use solution containing DreamTaq DNA Polymerase, optimized DreamTaq Green buffer, MgCl2, and dNTPs ("ThermoScientific DreamTaq Green PCR Master Mix (2X)," 2016). A master mix was prepared in a 1.5ml Eppendorf tube and then equally dispensed into the 96 well plate containing 1-1.2 µl of the DNA template of the individual samples. Concentration of DNA for each sample was **10 ng/µl**. PCR was carried out in Veriti Thermal Cycler (Applied Biosystems). Stock concentration of Primers was made to **100 pmol/ µl**. Later these stocks were subjected to 1:10 dilutions for working stocks **(10 pmol/µl)**.

Table 3.3 | PCR 'reaction mix for the, mtDNA, EGLN, EPAS1 and EDAR

S.N.	Template DNA used (10 ng/µl)	PCR master mix	Concentration of Forward Primer (10 pmol/µl)	Concentration of Reverse Primer (10 pmol/µl)	Milli-Q	Total
For single sample	1-1.2 µl	5 µl	0.15 µl	0.15 µl	3.5 µl	10 µl

Table 3.4 | PCR conditions used in this study.

Steps	Stage	Temperature	Time	No. of cycle
1	Initial Denaturation	95˚C	5 min	
2	Denaturation	95˚C	1 min	35 cycles
	Annealing	52-58˚C	30 sec	
	Extension	72˚C	2 min	
3	Final Extension	72˚C	7 min	
4	Hold	4˚C	∞	

3.6.1 Agarose gel electrophoresis

PCR products were electrophoresed at 120V in 2% agarose gel. The PCR products (≈650 bp in size) were then visualized under Gel Doc **(Figure 8.3).** On obtaining a single band devoid of any primer-dimer bands the PCR products (650 bp) were proceed for purification using ExoSAP enzyme followed by Sequencing.

3.6.2 ExoSAP (Exonuclease Shrimp Alkaline Phosphatase) Treatment

The end product of PCR contains unwanted amplicons, DNA template, extra dNTPs, non- specific molecules and primers that obstruct the sequencing result. So, in order to obtain pure amplicons, the PCR product were purified using ExoSAP. Exonucleases degrades single stranded DNA i.e. extra primers. Shrimp alkaline phosphatase then removes phosphates from dNTPs, so that they cannot be used to synthesize again. **Protocol**: 0.8-1µl of PCR product was transferred to a 96 well sequencing plate using multi-channel pipette. 1µl of ExoSAP was added in every sample with the help of repeater. Then ExoSAP reaction was carried out in PCR machine by using following PCR conditions.

Table 3.5 | ExoSAP conditions used in the present study. ExoSAP enzyme was kept in -20^{o} C before and after use.

Incubation	enzyme inactivation	Hold
37°C for 15-20 min	80°C for 15-20 min	4°C

3.6.3 Sequencing PCR

ExoSAP purified samples are then proceeded for DNA sequencing with either forward or reverse primer. A master mix was prepared in a 1.5ml tube and then dispensed equally into the sequencing plate containing ExoSAP purified PCR amplicons. Sequencing PCR was performed on a ABI Prism 3700 DNA Analyzer (Applied Biosystems).

Table 3.6 | Reactions mix and conditions for sequencing PCR.

Requirement	Volume(µl)	Steps	Conditions	Cycle
Big Dye™	1.8	1	96°C for 10 secs	35
Primer forward/re	0.075	2	55°C for 6 secs	
Mili Q	2.13	3	60°C for 4 min	
DNA (10ng/µl)	0.8	4	4°C hold & end	

3.6.4 Processing of the sequencing PCR product

In this final step, dye bound amplicons were precipitated by removing unbound dye and unused primers

a. 3ml of absolute alcohol was added to 120µl of 3N sodium acetate (pH 5.2) in a tube.
b. The tube was mixed thoroughly. 25ul of the above mixture was added in each well of the plate.
c. The plate was centrifuged at 4000rpm for 15/20 min in Eppendorf (5810R) centrifuge at 25ºC.
d. The plate was then inverted to remove the supernatant. 100µl of fresh 80% ethanol was added to each well and again centrifuged at 4000rpm for 10min.
e. The plate covered by filter paper was inverted and a pop spin was performed for few seconds at 750rpm to removed alcohol.
f. The plate was covered properly with fresh foil. At the time of sequencing, 10µl of 50% HiDye™ formamide was added to all the wells.
g. The sample plates were kept and run in the ABI Prism® 3700 DNA genetic3 Analyzer (for sequencing).

3.7 Screening of sample

PCR products were directly sequenced using BigDye™ Terminator V3.1 cycle sequencing kit (Applied Biosystems) in ABI Prism 3700 DNA Analyzer following the protocol (as described above). The mtDNA HVS-I (range: 16,038–16,462) and HVS-II (range: 65–417) regions were first sequenced using primer set 23 and 24. Several other coding region sites were sequenced to define each individual within their respective mtDNA Hgs with minimum effort. After the complete sub haplogrouping, samples belonging to interested Haplogroup were selected for whole mtDNA sequencing. After the sequencing raw ABI sequences file were obtained. For better understanding of the population under study, unpublished raw ABI sequenced files (CCMB lab) belonging to Manandhar (93), Udaya (58), Bajracharya (20) and Shakya (19) populations were also included in this study.

3.7.1 Sequence alignment

In total, mtDNA sequences belonging to 695 samples were included for mtDNA analysis (sequences for 508 individuals were newly generated, whereas sequences of 187 individuals were obtained from lab).

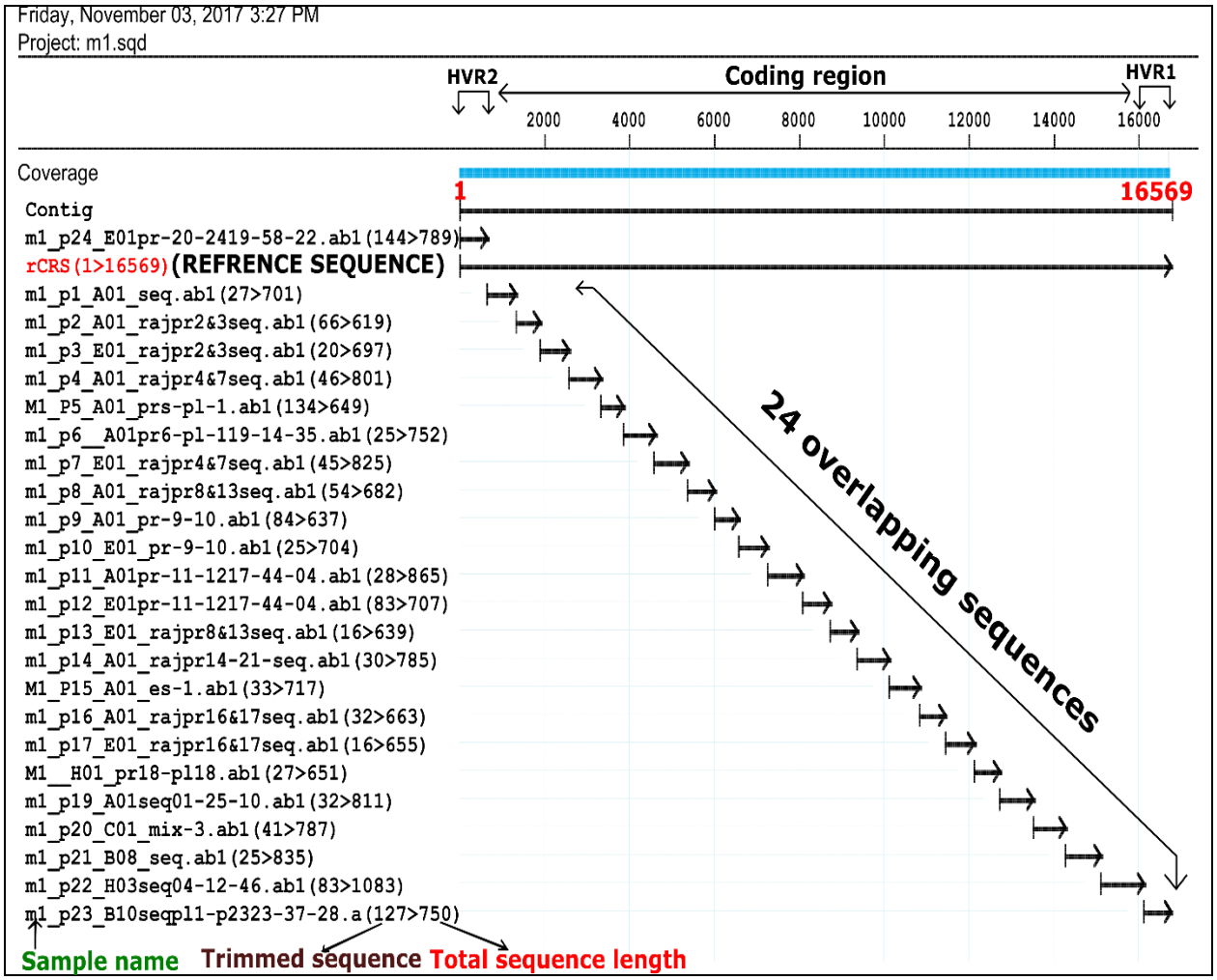


Figure 3.3 | Complete mtDNA representative sample (belonging to Sample-ID Maharjan 1) sequence as seen in DNASTAR. In the above figure, Contig and coverage area for the sample belonging to Maharjan 1 are shown along with the 24 sets of overlapping sequences generated separately by using 24 different sets of primers.

The mtDNA variants were identified by comparing the individual sequences with the rCRS. Higher degree of care has been taken to ensure the high-quality data. For this all the recorded mutations were re-confirmed by rechecking the electropherograms and/or resequencing the bad sequences.

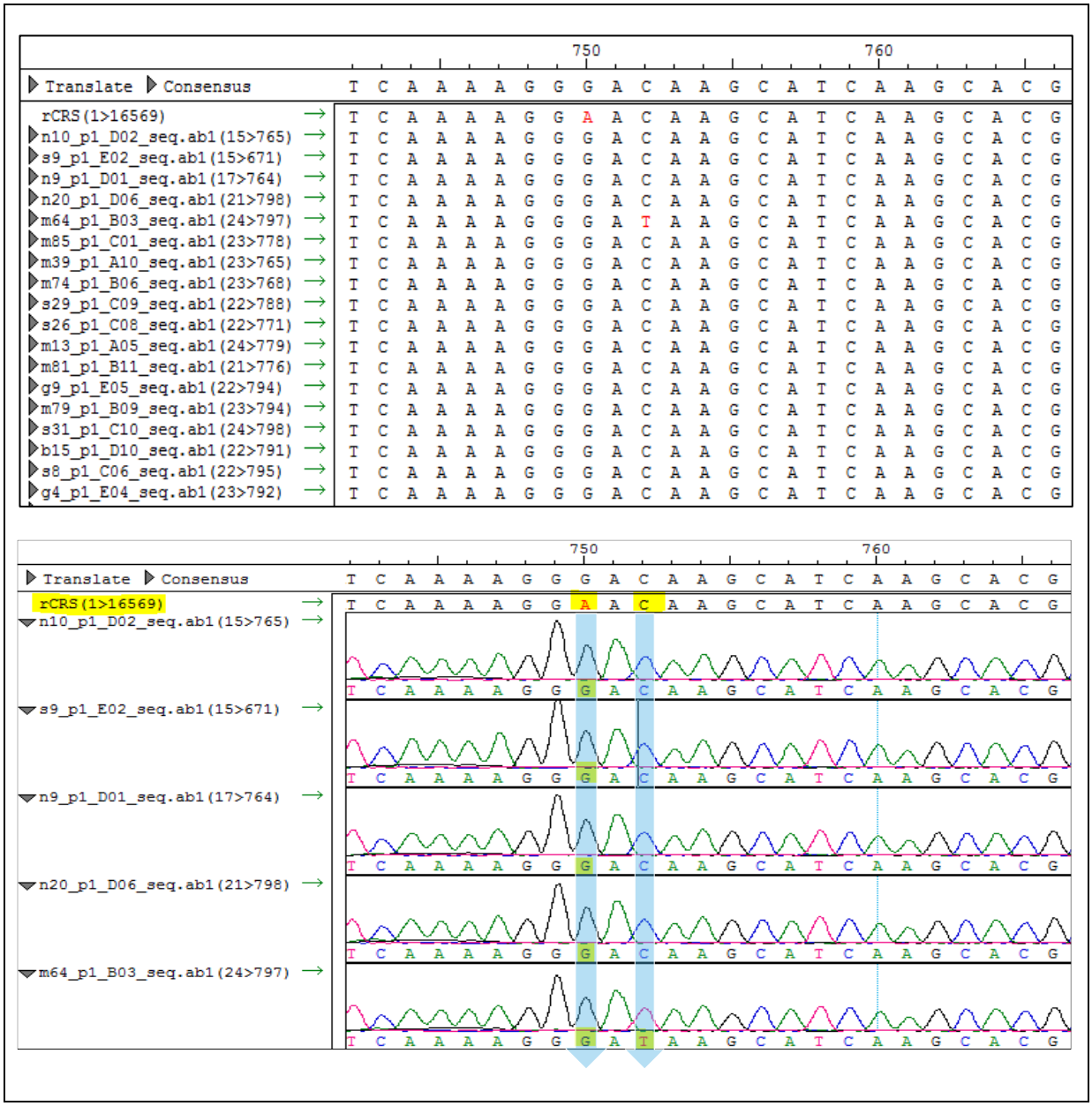


Figure 3.4 | mtDNA sequence aligned with the revised Cambridge reference sequences. The sequences were edited and aligned by DNASTAR (DNASTAR, Inc., USA) and mutations were recorded by comparing with the revised Cambridge reference sequence.

Deletions observed in the sequences are indicated by a letter "del or d" after the deleted nucleotide position (e.g., T15944d/del). Insertions are indicated by a dot (In some cases the dot was replaced by letter "Ins") followed by the position number and type of inserted nucleotide(s) (e.g., 5899.1C for a C insertion at the first inserted nucleotide position after position 5899 and 5899.2C for a subsequent C insertion, and these are abbreviated as 5899.1CC when occurring on the same branch). Polynucleotide stretches of unknown length were labelled as follows: example of multiple insertion of nucleotide "C" at position 573 is indicated as 573.XC. Sample from Nepal and Ladhak were used for two separate (independent) mtDNA studies.

3.7.2 Haplogroup Assignment

The mtDNA variants were used to assign the Hgs to each sample by searching the variants manually on the mtDNA phylogenetic tree Build 17 (Phylotree.org – mtDNA tree Build 17, 18 Feb 2016) (van Oven & Kayser, 2009). PhyloTree is the reference for worldwide phylogeny for mtDNA studies, and its latest update has recently been released (Build 17, 18 February 2016). Haplogrouping to each samples were also confirmed with the assistance of the software Haplogrep version 2 (Kloss-Brandstätter et al., 2011) which performs automated mtDNA classification using PhyloTree Build 17.

3.8 Data analysis

3.9 Part I (mtDNA)

3.9.1 Mitochondrial DNA data collection

To establish the genetic relationships between Newar, Brahmin, Magar and other Asian populations mitochondrial Hgs data of present study where compiled with the previously published mtDNA data from Nepal, Southeast Asia (Myanmar, Thailand, Vietnam, Malaysia), East Asia, South Asia, Caucasus and Central Asia. For this study, more than 18,350 mtDNAs belonging to 262 populations, were taken into account.

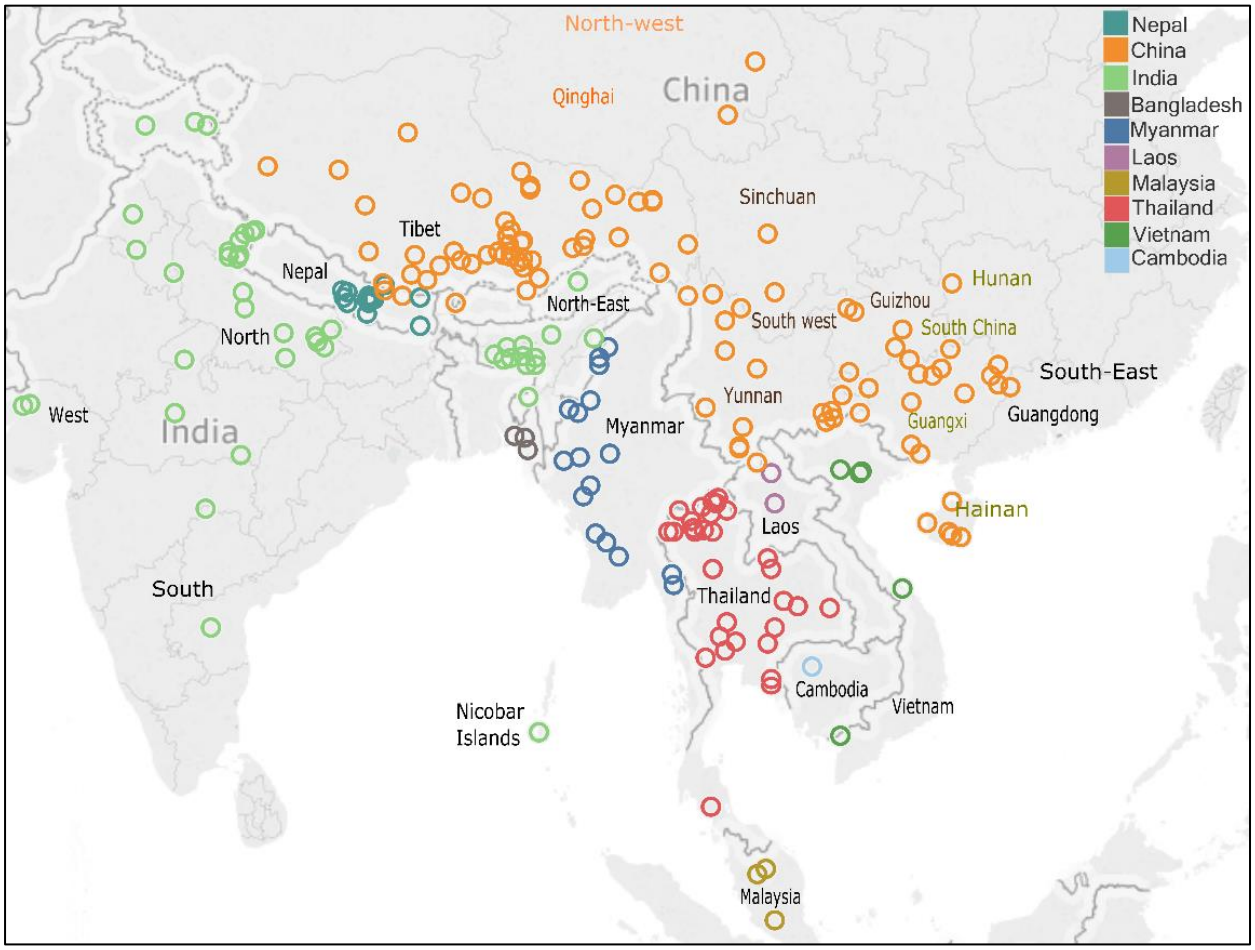


Figure 3.5 | Map showing the Geographic locations of 262 Asian populations analyzed in the present study. The Nepali populations (Newar, Brahmin and Magar) as well as the other Asian populations are indicated by different colors respective to their geographic location/country. The latitude and longitude in the map corresponds with the approximate geographical location of the sample. More details on the sample, geographic location, and linguistic affiliation is given in the **Figure 3.1** and **Table 3.7**

Several other populations from Siberia, Central Asia, Caucasus as given in **Table 3.7** but not shown in the Map (**Figure 3.5**) were also analyzed in the present study.

3.9.2 Molecular Diversity and Neutrality tests

Hg frequencies was calculated by simple gene counting method. Basic Molecular Diversity indices of populations such as the number of haplotypes, polymorphic sites, transition, substitution, transversions, haplotype diversity, nucleotide diversity (ND), mean number of pairwise difference (MPD) and their corresponding variance were obtained based on the mtDNA HVRI sequence using the program Arlequin version 3.5.2 (Excoffier & Lischer, 2010). Similarly, Fu's fs test (Y. X. Fu, 1997) and Tajima's test (Tajima, 1989) of selective Neutrality were calculated using Arlequin. The Nexus file format (.nex/.nxs) created based on the Hvr1 sequences was used to perform several analysis on Arlequin. Nei's d genetic distances and AMOVA (Analysis of Molecular Variance) were performed to evaluate the genetic structure of the populations, with the significance of variance components tested with 10,000 permutations.

3.9.3 mtDNA Phylogeny reconstruction

The mtDNA phylogeny was constructed by evaluating both all previously available published as well as unpublished complete mtDNA sequences and the complete mtDNA sequences generated in this study, aiming at the most parsimonious solution and aided by the software mtPhyl (Eltsov & Volodko, 2009) and Haplogrep (Kloss-Brandstätter et al., 2011). Mutations relative to the RCRS are indicated on the branches. Sample ID for each sequences are shown in the box with colour code. The colour are coded according to the geographic location of the samples. The accession number/ the publication source form which the sample were retrieved are shown just below the sample ID. Reversal mutations/Back mutations are suffixed by an exclamation mark (!). Recurrent mutation is underlined. Insertions are indicated by a dot followed by the position number and type of inserted nucleotide(s). Deletions are indicated by adding letter "d/del" after the deleted nucleotide position. Insertions are indicated by a dot (In some cases the dot was replaced by letter "Ins") followed by the position number and type of inserted nucleotide(s) (e.g., 5899.1C for a C insertion at the first inserted nucleotide position after position 5899 and 5899.2C for a subsequent C insertion, and these are abbreviated as 5899.1CC when occurring on the same branch). The highly variable site 16519, length variation in the poly-C stretches at nucleotide positions (nps) 303-315 and 16184-16194 were not used during the phylogeny reconstruction. A-C transversions at nps 16182 and 16183 were excluded because of their dependence on the presence of the C-T transition at np 16189. Heteroplasmies, Transitions and Transversions were not noted in the phylogeny. The number (1', 2'... n') corresponds to the reference for the analyzed sample. The corresponding reference for each number (1', 2'... n') are provided in Appendix E . While constructing the PhyloTree, Hg already predefined in the PhyloTree (Oven, 2015) are shown in Red color, whereas the Hg Newly defined are shown in Black color.

3.9.4 Principle component analysis (PCA)

PCA was performed, based on the frequencies of mtDNA Hgs using MVSP – A multivariate statistical package for Windows, ver. 3.22 software (Kovach, 1999) (www.kovcomp.com). The MVSP software package was used to identify the principal components (PCs) of mitochondrial variations that lead to form a Hg for every individual. This experiment was repeated to confirm the outcomes.

3.9.5 Median joining Network (MJN)

To analyze the detailed structure of the major mtDNA Hgs in Nepali population, median network for major Hg were constructed manually using complete/Hvr1 mtDNA sequences and checked by the program NETWORK 5.0 (www.fluxus-engineering.com/ sharenet.htm).

3.9.6 Contour Maps

Contour maps of spatial frequencies of major Hg observed in the studied populations were constructed using Golden Software Surfer 12.8 (Golden Software Inc., USA) with the Kriging algorithm.

3.9.7 Time to the Most Recent Common Ancestor (TMRCA)

The coalescence age or TMRCA of a basal mtDNA Hg was estimated using the Rho (ρ) statistics (Forster, Harding, Torroni, & Bandelt, 1996). Value of ρ was estimated with standard deviation (σ) as in Saillard et al (Saillard, Forster, Lynnerup, Bandelt, & Norby, 2000) using a mito- genome clock of one substitution every 3624 years further corrected for purifying selection (one mutation per 2,585 years for complete mtDNA genome) (Soares et al., 2009).

3.10 Part II (High altitude adaptation study)

3.10.1 Subjects and genotype data

Table 3.7 | Information of the populations included for high altitude adaptation study.

Population	Ethnic group	Altitude (m)	Country	Linguistic family	Sample Size
Maharjan	Newar	1400	Nepal	Tibeto-Burman	98
Shrestha	Newar	1400	Nepal	Tibeto-Burman	36
Udaya	Newar	1400	Nepal	Tibeto-Burman	59
Shakya	Newar	1400	Nepal	Tibeto-Burman	19
Bajracharya	Newar	1400	Nepal	Tibeto-Burman	20
Manandhar	Newar	1400	Nepal	Tibeto-Burman	89
Newarmix	Newar	1400	Nepal	Tibeto-Burman	50
Magar	Magar	1300	Nepal	Tibeto-Burman	40
Brahmin		1300	Nepal	Indo-European	48
Changpa	Tibetan	4000	India	Tibeto-Burman	96
Ladhak		3000	India	MIX	96
North India		2000	India	MIX	46

For the correlation analysis between genotyped SNPs of EPAS1 gene and altitude, 695 samples were included from various regions of India and Nepal. The population name/group, sample size, geographic location, Linguistic affiliation and elevation/altitude of each population are indicated in **Table 3.7**. To explore the signal of selection in Ladhak region, whole genome SNP data of 5 Changpa and 5 Tibetan refugee samples were utilized. Whole SNP data of these populations were obtained from CCMB Lab. F_{st} values between populations was calculated using Smart-PCA (Chen et al., 2010).

3.10.2 Signal of selection

Cross population composite likelihood ratio (XPCLR) was calculated to score and extract the genetic regions which are under natural selection (Chen et al., 2010). Since, XPCLR calculation requires phased haplotype data; initially phasing was performed with Beagle ver. 3.3.2 (Browning & Browning, 2007) and used haplotype of CDX (Chinese Dai in Xishuangbanna), CHB (Han Chinese in Beijing, China), JPT (Japanese in Tokyo, Japan) and KVH (Kinhin Ho Chi Minh City) as reference (http://hapmap.ncbi.nlm.nih.gov/index.html.en). To calculate XPCLR, 0.02 centi-Morgan window size, grid size 2000 base pairs was considered. Maximum number of SNPs in window is 50. The Pearson correlation coefficient was calculated using basic package of R (Gentleman, Ihaka, & Bates, 2009). For correlation study, samples from Nepal and India were stratified into the groups based on the altitude. P≤0.05 was considered to indicate a statistically significant. Allele and genotype frequencies were computed by the gene-counting method.

3.11 General Overview of Methodology

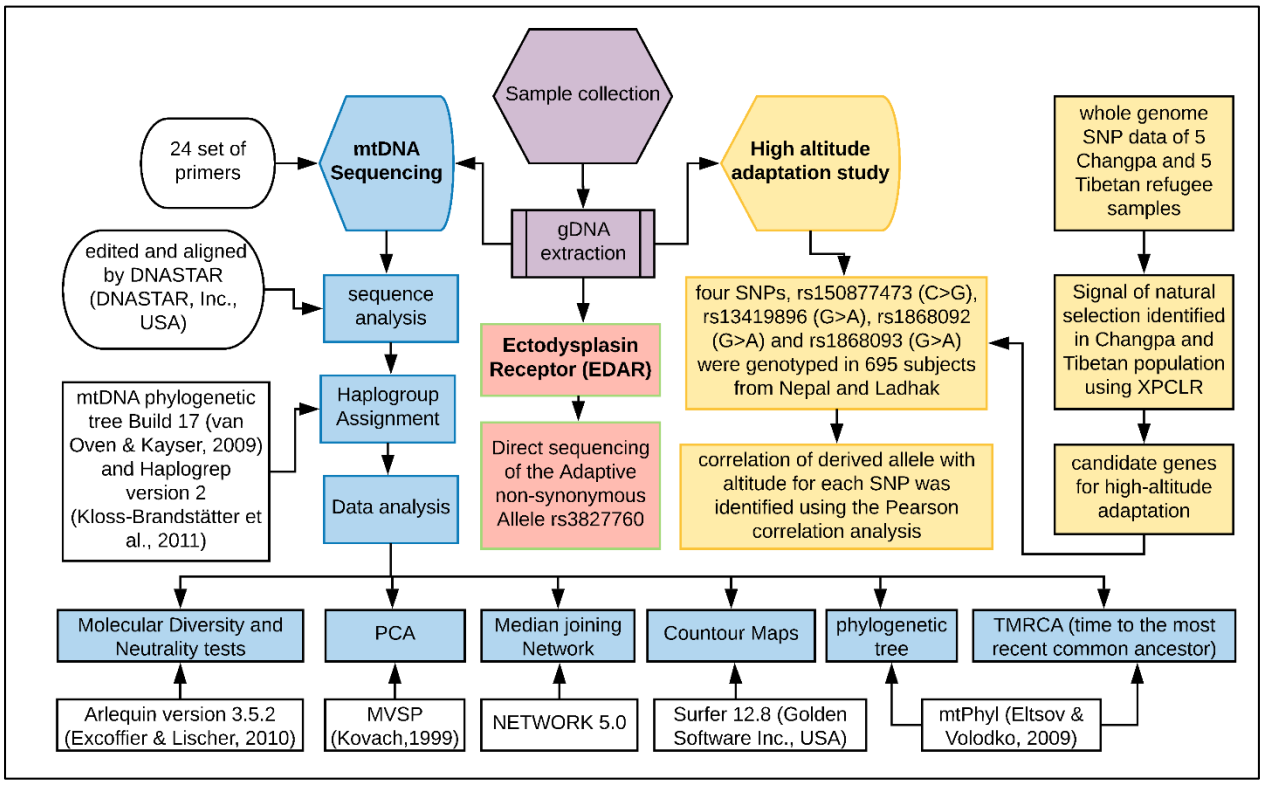


Figure 3.6 | General Overview of Methodology for mitochondrial and High altitude adaptation study.

4 Results

4.1 Basic molecular diversity indices and demographic parameter

Table 4.1 | Diversity Indices and tests of selective neutrality based on HVRI mtDNA sequence. Abbreviation used in the above table; n= sample size, s= number of substitution, h=number of haplotype, MPD= Mean Pairwise Difference, HD= Haplotype diversity, ND= Nucleotide diversity, SE= Standard Error. For more information about the populations used in the above table please see **Table 8.1**

Statistics	n	s	h	MPD ± SE	HD ± SE	ND ± SE	Tajima's D	Tajima's D p-value	Fu's fs	Fs p-value
Manandhar	91	44	37	5.878388 +/- 2.833826	0.9575 +/- 0.0083	0.017188 +/- 0.009179	-1.02	0.17	-16.81	0.00
Bajracharya	17	30	13	7.308824 +/- 3.599412	0.9632 +/- 0.0328	0.021371 +/- 0.011778	-0.72	0.26	-3.29	0.07
Shakya	16	23	12	5.816667 +/- 2.935650	0.9583 +/- 0.0363	0.017008 +/- 0.009615	-0.65	0.29	-3.51	0.04
Udaya	50	42	34	5.244082 +/- 2.578380	0.9829 +/- 0.0075	0.015334 +/- 0.008367	-1.35	0.07	-24.94	0.00
Newamix	50	169	29	10.981224 +/- 5.076108	0.9690 +/- 0.0111	0.032109 +/- 0.016472	-2.47	0.00	-6.29	0.04
Shrestha	36	42	25	5.579365 +/- 2.742980	0.9635 +/- 0.0192	0.016314 +/- 0.008915	-1.51	0.05	-14.16	0.00
Maharjan	100	54	43	5.701616 +/- 2.754982	0.9606 +/- 0.0090	0.016671 +/- 0.008922	-1.45	0.05	-24.37	0.00
newamix2	66	182	35	13.899301 +/- 6.317429	0.9725 +/- 0.0075	0.040641 +/- 0.020480	-2.09	0.00	-5.61	0.10
Newar_total	492	311	149	8.682621 +/ 4.016036	0.9807 +/- 0.0020	0.025388 +/- 0.012987	-2.13	0.00	-23.99	0.00
Magar	38	169	30	13.192034 +/- 6.069238	0.9801 +/- 0.0137	0.038573 +/- 0.019719	-2.41	0.00	-10.39	0.00
Brahmin	48	53	36	6.404255 +/- 3.086961	0.9832 +/- 0.0088	0.018726 +/- 0.010019	-1.61	0.03	-25.06	0.00
Tamang	44	49	25	6.261099 +/- 3.029369	0.9556 +/- 0.0152	0.018307 +/- 0.009835	-1.52	0.05	-9.57	0.00
Kathmandu	77	286	64	19.229323 +/- 8.605531	0.9942 +/- 0.0034	0.056226 +/- 0.027886	-1.76	0.01	-24.07	0.00
Tibet1	156	306	93	28.793879 +/- 12.660234	0.9868 +/- 0.0034	0.084193 +/- 0.040974	-0.65	0.28	-23.70	0.00
Other_Nepali	245	310	138	21.270191 +/- 9.415536	0.9912 +/- 0.0015	0.062194 +/- 0.030459	-1.19	0.08	-23.62	0.01

4.1.1 Molecular diversity

Higher Haplotype diversity was observed among the Nepalese ranging from 0.95 to 0.99. Similarly, nucleotide diversity among the Nepali population ranges from 0.015 to 0.062. Highest nucleotide diversity was observed in Nepali-other (sample from Kathmandu and Eastern Nepal) and Kathmandu group. Whereas, comparatively lower nucleotide diversity was observed among the several Newar groups except Newamix (present study) and Newamix2 (Gayden et al., 2013). These data are further strengthened by the analysis of Mean pairwise difference's (MPD). Similar to the Nucleotide diversity, value of MPD was also lower among the Newar groups, except Newamix and Newarmix2. Higher value of MPD and ND for Newamix and Newarmix2 samples might be as a result of combined sample included from multiple Newar sub caste.

4.1.2 Neutrality tests.

Fu's fs test (Y.-X. Fu, 1997) and Tajima's test (Tajima, 1989) of selective Neutrality were calculated using Arlequin. If the obtained values for Tajima's D and Fu's Fs are significantly negative, then the population might have expanded in the past. Fu's fs value was more negative in Maharjan (-24.36), Udaya (-24.9), Manandhar (-16.8) and Shrestha (-14.16). When all the Newar sub caste were treated as a single population and analyzed separately, Fu's Fs value was significantly negative -23.99. Fu's Fs values were statistically significant in all the Newar sub populations as well as in Magar and Brahmin for Hypervariable region I (HVRI) sequences, indicating historical population expansion. Tajima's D test also yielded significant negative values for each population. These results are further supported by the fact that the major Newar sub caste exhibit low Nucleotide diversity, which denotes the population expansion in the past.

4.1.3 Pairwise genetic distance

Genetic distance between the Nepali populations based on the **Fixation inde**x (F_{st}) were calculated as shown in the **Table 4.2**. F_{st} varies between 0 and 1. F_{st} value of 0 indicates that two populations are genetically identical whereas an increasing value from 0 indicates an increasing variation between the pair of populations. F_{st} value of 1 indicates maximum genetic diversity between the two populations.

Table 4.2 | A pairwise Fst distance matrix between each population. The F_{st} values in the Table are color coded in the background, with green (high Fst value), Yellow (intermediate F_{st} value) and Red (low Fst value) color gradient. Population name: Newar mix2 (N-mix2), KTM (Kathmandu), Tmg (Tamang) (Gayden et al., 2013); Nepali-other (N-OT) individual from Ktm and Eastern Nepal (Wang et al., 2012); Mdr=Manandhar, SHK= Shakya, BJR= Bajracharya, UDY= Udaya, MHJ= Maharjan, STH= Shrestha, Mgr= Magar, BRH= Brahmin; Newar total= includes all Newar sub caste. For more information about the populations used in the above table please see **Table 8.1**

	MDR	SHK	BJR	UdY	NMIX	MHJ	STH	NMIX2	Other_Nepali	KTM	MGR	BRH	TMG	Newar_total	Tibet1
MDR	0														
SHK	0.049	0													
BJR	0.077	-0.001	0												
UDAYA	0.048	0.009	0.035	0											
NMIX	0.061	0.012	0.015	0.021	0										
MHJ	0.071	0.043	0.069	0.047	0.008	0									
STH	0.078	0.053	0.062	0.054	0.003	0.001	0								
NMIX2	0.049	-0.003	0.011	0.017	0.006	0.021	0.006	0							
Other_nepali	0.043	0.01	0.005	0.026	0.013	0.03	0.016	0.004	0						
KTM	0.045	0.006	0.003	0.021	0.005	0.03	0.011	0	-0.004	0					
MGR	0.063	0.016	0.013	0.025	0.006	0.03	0.016	0.007	0.007	-0.002	0				
BRH	0.08	0.068	0.029	0.065	0.033	0.04	0.03	0.027	0.014	0.012	0.019	0			
TMG	0.114	0.065	0.055	0.063	0.028	0.061	0.048	0.02	0.018	0.014	0.013	0.045	0		
Newar TOTAL	0.026	0	0.018	0.008	0.005	0.007	0.003	0.005	0.029	0.021	0.018	0.025	0.031	0	
Tibet1	0.082	0.037	0.03	0.056	0.033	0.073	0.049	0.04	0.028	0.017	0.028	0.052	0.04	0.092	0

.

Distance Matrix of Fst based on the mtDNA HVR1 shows a higher genetic distance between Newar group (indicated as Newar_total) and Tamang (0.031) followed by Other_Nepali group (0.31), Manandhar (0.026), Brahmin (0.025) and Magar (0.018). Within the Newar group, Manandhar (0.026) and Bajracharya (0.018), sub caste shows highest genetic distance with the Newar group. Whereas the other sub caste; Udaya (0.08), Maharjan (0.07), Newar mix (0.005), Shrestha (0.003) and Shakya (0.000) shows comparatively less genetic distance, indicating a higher proportion of shared genetic component between these populations. The lower genetic distance for Shakya (0.000) and higher genetic distance for Bajracharya (0.018) was possibly due to the few sample number of these population. The Manandhar group has significant proportion of unshared genetic components with the other Newa sub caste as well as with the other Nepalese populations analyzed so far.

4.1.4 Mean pairwise differences (π) within and between populations

Nei's genetic distance (Nei, 1978) between the populations based on the net number of nucleotide differences was also computed as shown in the **Figure 4.1**.

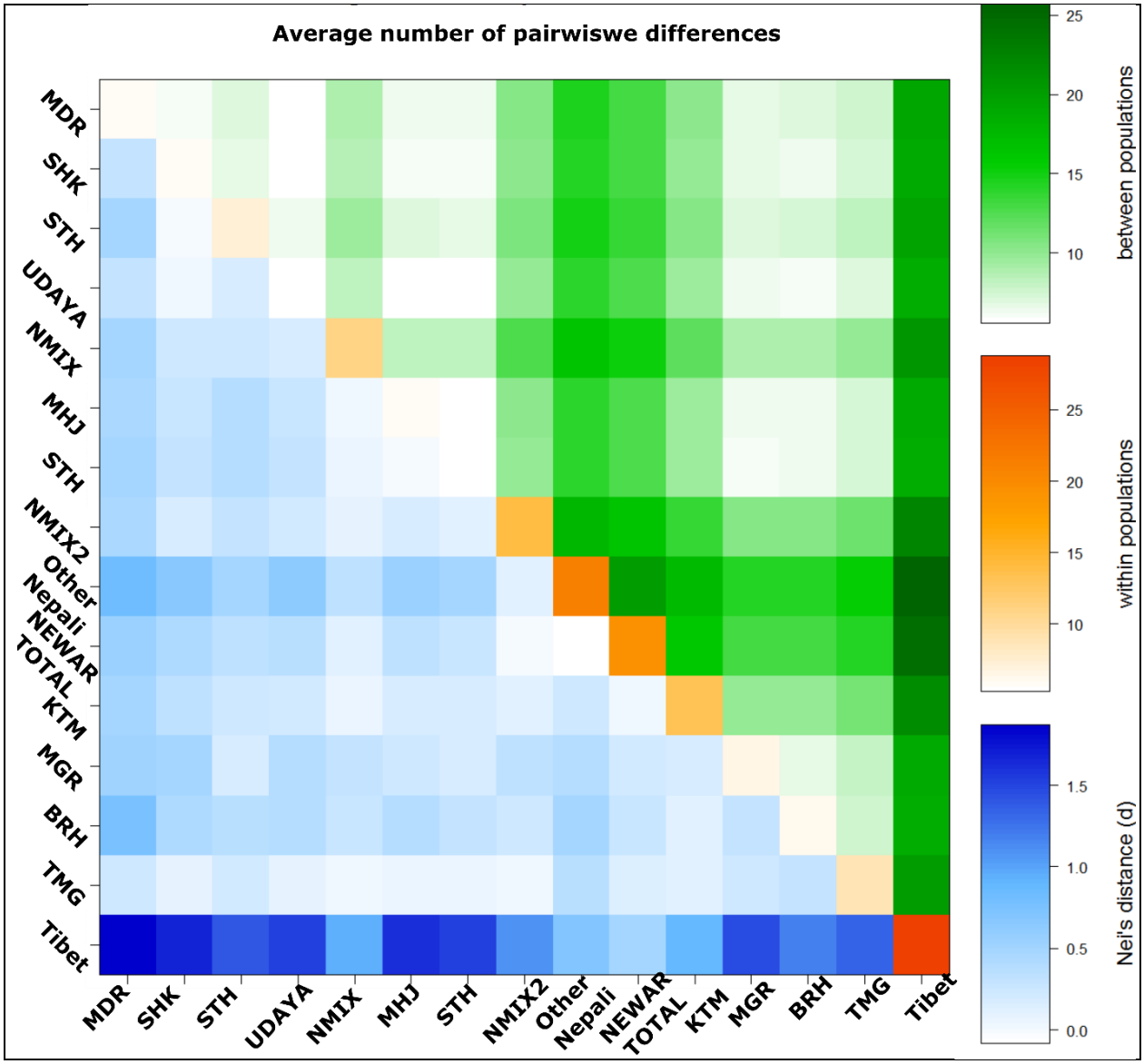


Figure 4.1 | Mean pairwise differences (π) within and between populations. Average number of pairwise differences (π) between the populations are shown by three different color scales. The Orange colour on diagonal of the graph shows the value of π_{xy} within the populations, Green colour just above the yellow color (diagonal) shows the value of π_{xy} between pairs of populations and the Blue colour on the graph just below the diagonal represents the net number of nucleotide differences between the populations**.** Populations are abbreviated as: Manandhar (MDR), SHK (shakya), STH (Shrestha), Nmix (Newar mix), MHJ (Maharjan), STH (Shrestha), KTM (Kathmandu), BRH (Brahmin), TMG (Tamang) and MGR (Magar). For more information about the populations used in the above table see **Table 8.1.**

4.2 Classification of mtDNA sequences.

In Newar, a total of 374 out of 376 individuals were classified into known mtDNA Hgs/sub-Hgs belonging to three main groups: **East Eurasian**, **South Asian** and **West Eurasian**. Surprisingly, two samples (complete mtDNA sequencing for these sample was done) were unable to allocate into already known Hg (van Oven & Kayser, 2009). The East Eurasian prevalent Hgs were grouped into East Eurasian ancestries (Xuebin Qi et al., 2013). Accordingly, south Asian and west Eurasian prevalent Hgs were also allocated into South Asian and West Eurasian ancestries (Silva et al., 2017).

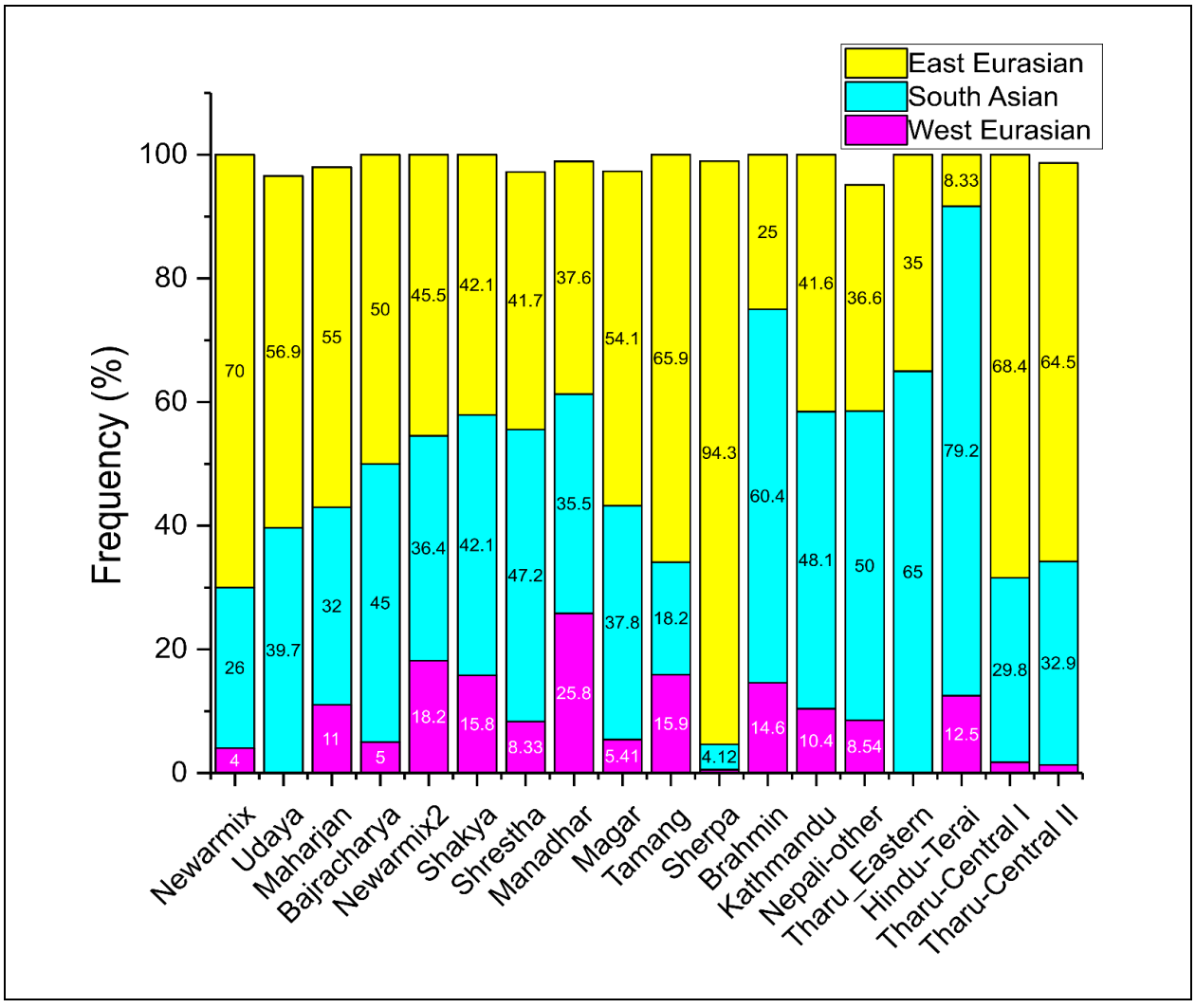


Figure 4.2 | mtDNA components observed in the studied populations and other Nepalese populations studied so far. mtDNA data for the other Nepali populations were computed based on the mtDNA Hg frequencies. For more information about the populations see **Table 8.1** & **Table 8.2**.

East Eurasian ancestries where higher among Sherpa (**95.3%**), Newar mix (**70%**), Tamang (**65.9**%), Tharu from Chitwan I (**68.4**%), Tharu from Chitwan II (**62.8**%), Udaya (**56.9**%), Maharjan (**55**%) and Magar (**54%**). The proportion for **South Asian ancestries** where higher among the Hindus form Morang (**79.2**%), Tharus from Morang (**62.5**%), Brahmin from Tanahun (**60.41**%) and Nepali-other (individuals from Kathmandu and Eastern Nepal) (**52.4**). Intriguingly, Manandhar shows

nearly equal proportion for all three ancestries. Whereas, Shakya, Shrestha, Bajracharya and Kathmandu shows nearly equal proportion for East Eurasian and South Asian ancestries.

4.2.1 Dominant Hgs in Newar and other Nepali population.

Major Haplogroup				Newar sub caste							Origin
Pop	Newar	BR-NP	MGR	MDR	STH	N-MIX	MHJ	BJR	SHK	Udaya	
size	376	48	37	93	36	50	100	20	19	58	
F	17.8	6.3	0.0	9.7	19.4	28.0	26.0	10.0	0.0	15.5	EE/SEA
Z	17.8	0.0	5.4	18.3	11.1	20.0	16.0	20.0	26.3	19.0	EE/CA
M5	9.6	2.1	2.7	11.8	2.8	4.0	10.0	5.0	15.8	13.8	SA
D	5.9	4.2	24.3	0.0	5.6	10.0	6.0	10.0	15.8	6.9	EE/EA
M3	5.3	8.3	2.7	3.2	5.6	2.0	7.0	10.0	10.5	5.2	SA
A	4.8	2.1	0.0	7.5	2.8	6.0	1.0	10.0	0.0	6.9	EE/EA
U2I	4.3	16.7	2.7	0.0	13.9	4.0	7.0	5.0	5.3	0.0	SA
U7	4.3	2.1	0.0	0.0	8.3	2.0	9.0	0.0	15.8	0.0	WE
M30	3.7	4.2	0.0	1.1	11.1	6.0	0.0	0.0	0.0	10.3	SA
I	3.2	0.0	0.0	12.9	0.0	0.0	0.0	0.0	0.0	0.0	WE
M38	3.2	0.0	0.0	10.8	2.8	0.0	0.0	5.0	0.0	0.0	SA
G	2.7	8.3	0.0	2.2	0.0	6.0	1.0	0.0	0.0	6.9	EE/EA
HV	2.7	0.0	0.0	10.8	0.0	0.0	0.0	0.0	0.0	0.0	WE
R6	1.6	10.4	2.7	0.0	0.0	2.0	1.0	15.0	5.3	0.0	SA
W	1.1	0.0	0.0	1.1	2.8	0.0	0.0	0.0	0.0	3.4	WE
M33	1.1	0.0	2.7	1.1	0.0	0.0	1.0	0.0	0.0	3.4	SA
M52	0.8	2.1	10.8	0.0	0.0	2.0	2.0	0.0	0.0	0.0	SA
H	0.8	2.1	5.4	1.1	0.0	2.0	1.0	0.0	0.0	0.0	WE
M9	0.8	0.0	13.5	0.0	0.0	0.0	2.0	0.0	0.0	1.7	EE/EA
R5	0.8	4.2	2.7	3.2	0.0	0.0	0.0	0.0	0.0	0.0	SA
U3/U4/U5	0.5	2.1	0.0	1.1	0.0	0.0	1.0	0.0	0.0	0.0	WE
C	0.3	2.1	8.1	0.0	0.0	0.0	1.0	0.0	0.0	0.0	EE/EA
T	0.3	4.2	0.0	0.0	0.0	0.0	0.0	5.0	0.0	0.0	WE
N1	0.0	4.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	WE

Table 4.3 | List of Dominant Hgs observed in the studied populations. Geographic origin of the population are abbreviated as; EE=East Eurasia, CA=Central Asia, EE=East Eurasia, SEA=South East Asia, SA= South Asia and WE=West Eurasia. Population are abbreviated as; BR-NP=Brahmin, MGR=Magar, MDR= Manandhar, STH= Shrestha, N-Mix= Newar-mix, BJR= Bajracharya and SHK= Shakya.

Gene flow in Newar is dominated by East Asian (50.8%) haplogroups followed by South Asian (35.9%) haplogroups. Whereas the contribution from West Eurasia (12.8%) is very low in Newar. Two East Asian haplogroups Z and F contribute 35.6% of the total mtDNA gene pool of Newar. Hg Z and F both are also detected in Tibetan population. The other East Asian hgs with appreciable frequency include only hg D (5.9%) and A (4.8%). Hg A and D are found in higher frequency diversity among the Tibetans.

Among the South Asian (35.9%) hg, only two hgs M5 (9.6%) and M3 (5.3%) shows significant frequency contribution on Newar. Whereas the other south Asian hgs are present in very low frequency. West Eurasian hg U7 has 4.3% contribution to the total mtDNA gene pool of Newar.

In overall, the dominant Hgs observed in Newar are **Z** (17.82%), **F** (17.82%), **M5** (9.57%), **U** (9.04%), **M3** (5.32%), **D** (5.85%) and **A** (4.79%). Likewise, the dominant Hgs observed in Brahmin are **U** (27.03%), **R6** (13.51%), **M3** (10.81%), **G** (10.81%), F (8.11%), R5 (5.41%) and M35b2 (5.41%)

which altogether account for 86.48% of the total Brahmin sample. Likewise, mtDNA pool of Magar were dominated by East Asian specific Hgs D, M9 and C.

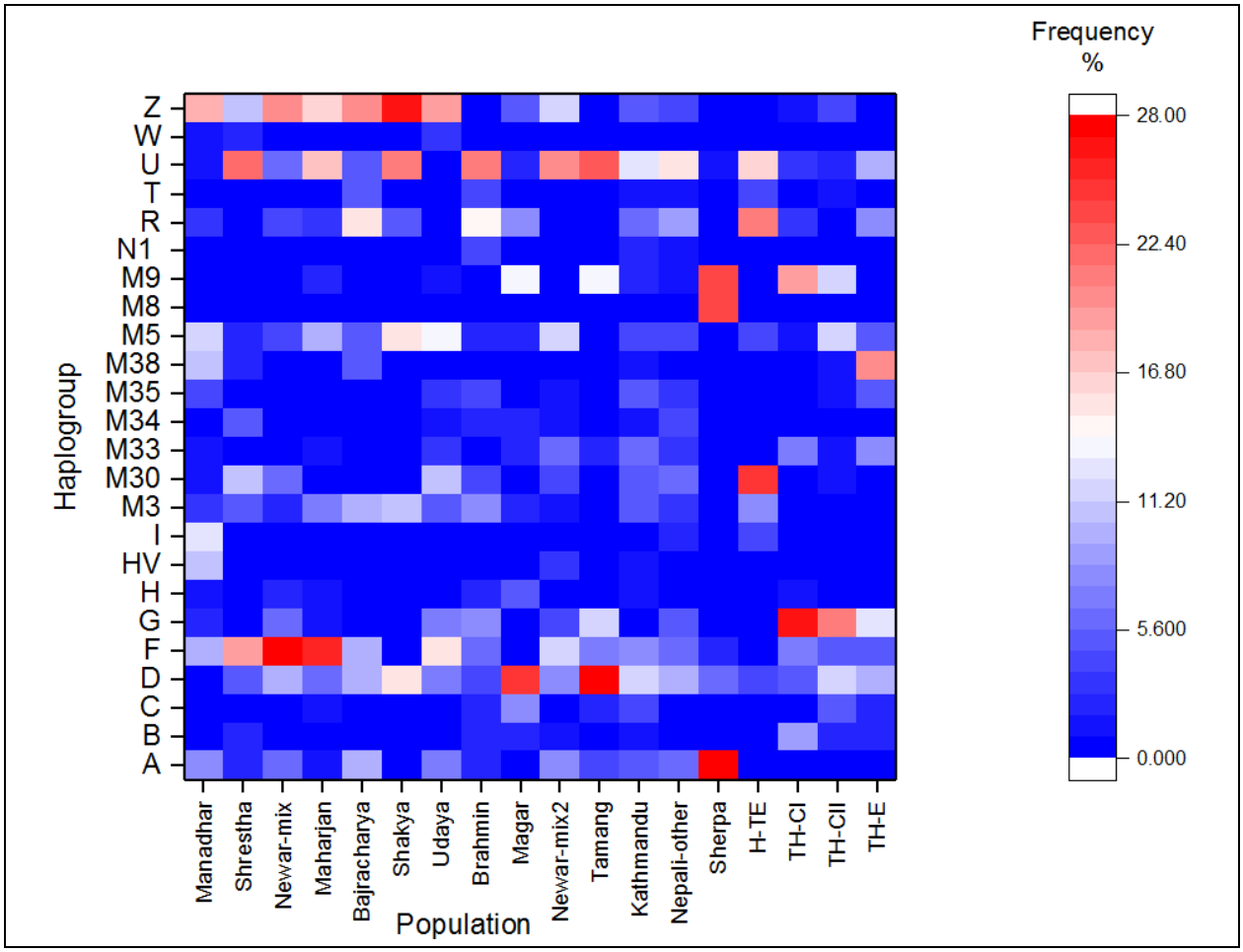


Figure 4.3 | Heat map based on the major mtDNA Hg frequencies observed in the Nepali populations. Population is abbreviated as: Newar-mix [Present study], Newar mix2 [Gayden et al., 2013], Nepali-other [KTM and Eastern Nepal], TH-CI &TH-CII [Tharu from Chitwan], TH-E [Eastern/ Morang Tharu] and H-TE [Hindu Terai]. For more information about the population see **Table 8.1** & **Table 8.2**.

Hg Z believed to be originated in **Central Asia** was observed in high to moderate frequency in all Newar sub caste, but completely absent or present in very low frequency among the other Nepali populations. Hg F present in high to moderate frequency among the majority of Newar caste were also present in moderate to low frequency among the other Nepali populations. Highest frequency of F1d1 was observed in Maharjan and Newamix. This shows, Hg F has a wider distribution than the Hg Z, in which the later has a relatively more restricted-distribution/confined within the Newa caste only. Hg U is one of the most prevalent Hg and has a wider distribution among the several Nepali populations. Similarly, South Asian hg M5 and M3 is more prevalent in Newar population compared to other Nepali population studied so far. **East Asian specific** hg D, prevalent in higher frequency in Tamang and Magar were also observed in moderate (Shakya) to low frequency in Newar and other populations. Several Hgs were found to be restricted/confined within a single population, indicating their restricted distribution. Especially, West Eurasian hg I1e and HV were present in significant frequency in Manandhar but was either completely absent or present in low frequency among the other Nepali populations. This finding is in accordance with

the high genetic distance observed between the Manandhar and other Newar sub caste (**Table 4.2**) as discussed previously. All the Newar sub caste were included as a single group (Newar total) and their geographic origin were analyzed as shown in the table below.

Table 4.4 | (A) Contribution to the total mtDNA gene pool of Newar form East Eurasia, South Asia and West Eurasia. (B) Major Haplogroup, count and Frequency (%) observed in Newar group.

Geographic origin	Count	Frequency (%)
East Eurasian	191	50.8
South Asian	135	35.9
West Eurasian	48	12.8
Unable to define hg	2	0.5
Total sample	376	

(A)

South Asian	Count	Freq (%)
M18c	1	0.3
M2	4	1.1
M3	20	5.3
M30	14	3.7
M31c	2	0.5
M33	4	1.1
M34	3	0.8
M35	6	1.6
M38	12	3.2
M43	2	0.5
M5	36	9.6
M52b	3	0.8
R30b2a	3	0.8
R5a2b	3	0.8
R6b	6	1.6
U2b	8	2.1
U2c	8	2.1
Total South A	**135**	**35.9**
Total sample	376	

East Eurasian	Count	Freq (%)
A	18	4.8
B5a1d	1	0.3
C	1	0.3
D	22	5.9
F	67	17.8
G	10	2.7
M10	2	0.5
M9	3	0.8
Z	67	17.8
Total East Eurasian	**191**	**50.8**
Total sample (N)	376	

West Eurasian	Count	Freq (%)
H13	3	0.8
HV12b	10	2.7
I1e	12	3.2
T2a	1	0.3
U3	1	0.3
U5	1	0.3
U7	16	4.3
W3	4	1.1
Total West Eurasian	**48**	**12.8**
Total sample (N)	376	

(B)

Gene flow in Newar is dominated by East Asian (50.8%) haplogroups followed by South Asian (35.9%) haplogroups. Whereas the contribution from West Eurasia (12.8%) is very low in Newar. Two East Asian haplogroups Z and F contribute 35.6% of the total mtDNA gene pool of Newar. Hg Z and F both are also detected in Tibetan population. The other East Asian hgs with appreciable frequency include only hg D (5.9%) and A (4.8%). Hg A and D are found in higher frequency diversity among the Tibetans.

Among the South Asian (35.9%) hg, only two hgs M5 (9.6%) and M3 (5.3%) shows significant frequency contribution on Newar. Whereas the other south Asian hgs are present in very low frequency. West Eurasian hg U7 has 4.3% contribution to the total mtDNA gene pool of Newar.

Table 4.5 | Classification of mtDNA sequences in nine Nepali populations. Among these, 7 caste belongs to Newar (376), remaining two belongs to Magar (36) and Brahmin (48). Number of Haplogroup present in each population are provided as a count data. For more information see **Table 8.1**, **Table 8.2** & **Table 8.3**.

Haplogroup	Newar							Brahmin	Magar
	Manandhar	Shrestha	Newamix	Maharjan	Bajracharya	Shakya	Udaya		
A+152+16362	5	1	2	1	2	-	4	-	-
A14	-	-	-	-	-	-	-	1	-
A17	2	-	1	-	-	-	-	-	-
B4a4	-	-	-	-	-	-	-	1	-
B5a1d	-	1	-	-	-	-	-	-	-
B5a1b1	-	-	-	-	-	-	-	-	1
C4a1a1a	-	-	-	-	-	-	-	-	1
C4a2b2	-	-	-	1	-	-	-	-	-
C4a2c2	-	-	-	-	-	-	-	1	-
C7b	-	-	-	-	-	-	-	-	2
D4	-	-	3	-	1	2	-	-	-
D4a5	-	-	1	-	-	-	-	-	-
D4b2b	-	-	1	1	1	-	1	-	-
D4e1a2	-	-	-	-	-	-	-	-	1
D4h	-	1	-	1	-	-	-	-	-
D4i	-	-	-	-	-	-	3	-	2
D4j1a1	-	-	-	1	-	-	-	1	-
D4j1b	-	-	-	-	-	-	-	-	2
D4o	-	1	-	-	-	-	-	-	-
D4q	-	-	-	1	-	-	-	-	-
D5a2a1	-	-	-	-	-	-	-	-	2
D5a2a1+ @1617	-	-	-	-	-	-	-	-	2
D5a2b	-	-	-	2	-	1	-	1	-
F1b1c	-	-	-	-	-	-	-	1	-
F1c1a2	8	-	-	-	-	-	1	-	-
F1d1	-	7	10	21	2	-	5	2	-
F1g	1	-	2	5	-	-	1	-	-
F2b1	-	-	2	-	-	-	2	-	-
G2	1	-	-	-	-	-	1	-	-
G2a	-	-	-	-	-	-	-	1	-
G2a1d2	1	-	3	1	-	-	1	-	-
G2b2a	-	-	-	-	-	-	2	-	-
G3a1'2	-	-	-	-	-	-	-	3	-
H*	-	-	-	-	-	-	-	-	2
H13a2a	1	-	-	-	-	-	-	-	-
H13a1d	-	-	1	1	-	-	-	-	-
H2a	-	-	-	-	-	-	-	1	-
HV12b1	10	-	-	-	-	-	-	-	-
I1e	12	-	-	-	-	-	-	-	-

Haplogroup	Newar							Brahmin	Magar
	Manandhar	Shrestha	Newamix	Maharjan	Bajracharya	Shakya	Udaya		
M*	-	-	-	1	-	-	-	-	-
M10a1b	-	-	-	2	-	-	-	-	-
M18	-	-	-	-	-	-	-	-	1
M18c	-	1	-	-	-	-	-	-	-
M2a1a+207	-	-	-	1	-	-	-	-	-
M2a3a	-	-	-	-	1	1	1	-	-
M3	-	-	-	-	1	-	-	-	-
M3a1+204	-	2	-	2	-	-	1	-	-
M3a2	-	-	-	-	-	-	1	-	-
M3c1a	-	-	-	1	1	2	1	1	-
M3c2	1	-	-	-	-	-	-	-	-
M3d	-	-	-	2	-	-	-	-	1
M3d1	-	-	-	-	-	-	-	1	-
M3d1a	2	-	1	2	-	-	-	2	-
M30+16234	-	-	-	-	-	-	3	-	-
M30b	1	3	3	-	-	-	1	-	-
M30d1	-	1	-	-	-	-	2	1	-
M30g	-	-	-	-	-	-	-	1	-
M31b2	-	-	-	-	-	-	-	1	-
M31c	-	-	1	1	-	-	-	-	-
M33a	-	-	-	-	-	-	-	-	1
M33a1a	-	-	-	1	-	-	2	-	-
M33a2a	1	-	-	-	-	-	-	-	-
M34a1a	-	-	-	-	-	-	1	-	-
M34a2	-	2	-	-	-	-	-	-	-
M34'57	-	-	-	-	-	-	-	-	1
M34b	-	-	-	-	-	-	-	1	-
M35b2	-	-	-	-	-	-	-	2	-
M35b4	4	-	-	-	-	-	2	-	-
M38+199	10	-	-	-	-	-	-	-	-
M38a	-	1	-	-	1	-	-	-	-
M4"67	-	-	-	-	-	-	-	1	-
m43a1	-	-	1	-	-	-	-	-	-
M43b	-	1	-	-	-	-	-	-	1
M5a	9	-	1	3	-	-	3	-	-
M5a2	-	-	-	3	-	-	-	-	-
M5a2a1a1	-	-	-	-	-	-	-	1	-
M5a2a1a2	-	-	-	-	-	1	-	-	-
M5a'd	-	-	-	1	-	-	-	-	-
M5b2b	-	-	1	3	1	1	1	-	-
M5b2b1	-	-	-	-	-	-	1	-	-
M5b2b1a	-	-	-	-	-	-	-	-	1
M5c1	-	1	-	-	-	-	-	-	-
M5c2	1	-	-	-	-	1	3	-	-
M5d	1	-	-	-	-	-	-	-	-

Haplogroup	Newar							Brahmin	Magar
	Manandhar	Shrestha	Newamix	Maharjan	Bajracharya	Shakya	Udaya		
M52b	-	-	1	2	-	-	-	1	4
M58	-	-	-	-	-	-	-	-	1
M65b	-	-	-	-	-	-	-	1	-
M9a1a2	-	-	-	2	-	-	-	-	5
M9a1b1	-	-	-	-	-	-	1	-	-
N*	-	-	-	1	-	-	-	-	-
N1a1b1	-	-	-	-	-	-	-	2	-
R2'JT	-	-	-	-	-	-	-	-	1
R30b2a	-	-	1	2	-	-	-	-	-
R5a2b	3	-	-	-	-	-	-	2	1
R6a2	-	-	-	-	-	-	-	1	1
R6b	-	-	1	1	3	1	-	4	-
T2a1a	-	-	-	-	1	-	-	-	-
T2b4+152*	-	-	-	-	-	-	-	1	-
T2d1	-	-	-	-	-	-	-	-	-
T2g1*	-	-	-	-	-	-	-	1	-
U2a1a	-	-	-	-	-	-	-	1	-
U2b1	-	-	2	-	-	-	-	-	-
U2b1a	-	-	-	3	-	-	-	-	-
U2b2	-	3	-	-	-	-	-	2	-
U2c	-	-	-	-	-	-	-	5	-
U2c1	-	2	-	4	1	1	-	-	1
U3a	-	-	-	-	-	-	-	1	-
U3b	1	-	-	-	-	-	-	-	-
U5a1b1f	-	-	-	1	-	-	-	-	-
U7	-	-	-	1	-	1	-	-	-
U7a1a	-	1	-	-	-	-	-	-	-
U7a2	-	-	-	-	-	-	-	1	-
U7a3	-	-	1	-	-	-	-	-	-
U7a3a	-	1	-	7	-	2	-	-	-
U7b	-	1	-	1	-	-	-	-	-
W3a1b	1	1	-	-	-	-	2	-	-
Z3a1a	17	4	10	15	3	3	11	-	2
Z7	-	-	-	1	1	2	-	-	-
Total (N)	93	36	50	100	20	19	58	48	37

4.3 Principle Component Analysis (PCA)

4.3.1 Between the Nepali populations

The first component (PC1) which explains 26.5% of the genetic variance, basically segregates Newar with the other Nepalese populations. Whereas the PC2 which explains the 16.76% of the genetic variation separates the Newar populations away from the Sherpa and IE (Indo-European) speaking populations (Eastern Tharu, Brahmin and Hindu-Terai).

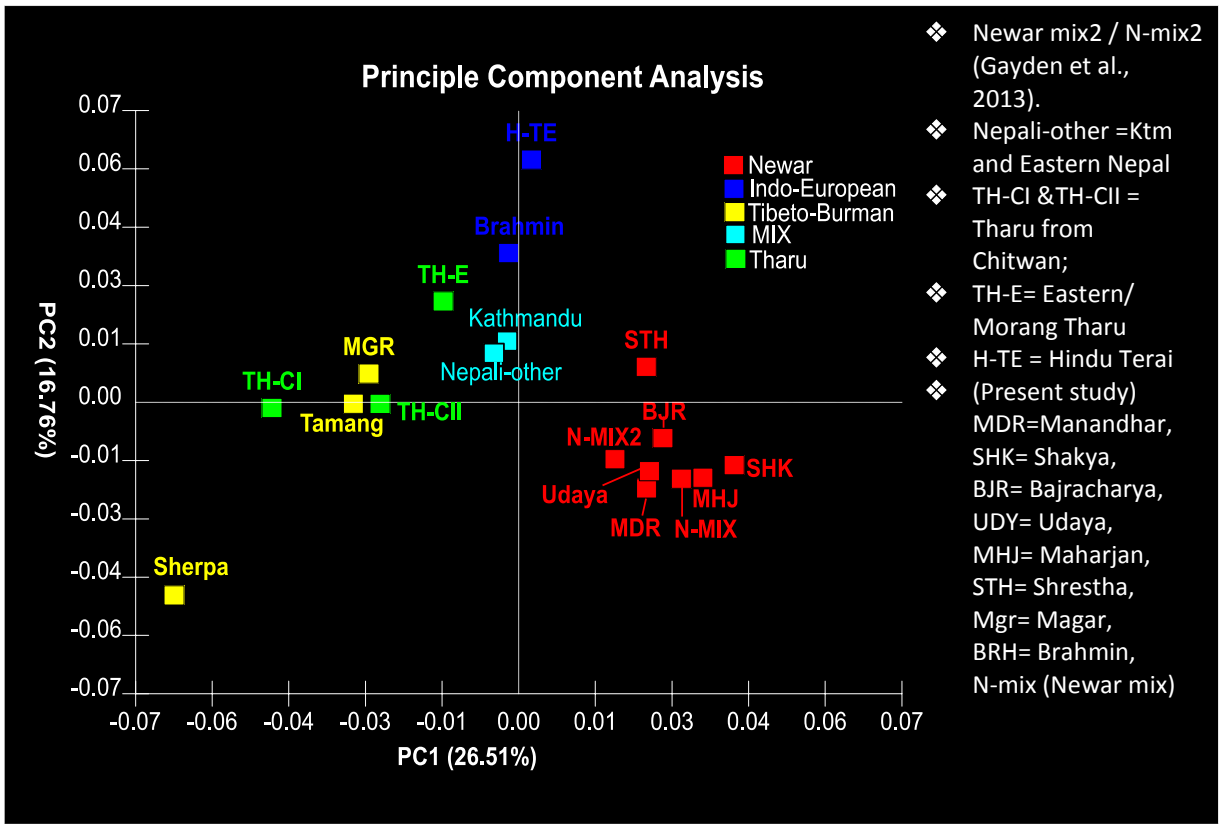


Figure 4.4 | Map of the principle component analysis among the Nepali populations. PCA was computed utilizing the mtDNA frequencies of the 19 Nepali populations. The first (PC1) and the second (PC2) Principle components explain 26.5% and 16.76% of the genetic variance respectively. Except Newar and Tharu all other remaining caste are grouped based of their linguistic affiliation. Newar speaks TB language, whereas Tharu speaks IE language. For more information see **Table 8.1** & **Table 8.2**.

In overall, all the Newa populations cluster tightly with each other indicating a shared genetic component between the Newa sub castes. Brahmin is closely placed with Kathmandu and Eastern Nepal including the Tharu (TH-E). Likewise, Magar, Tamang and Tharu from Chitwan cluster closely in the PCA map. Comparatively, Sherpa is placed away from other populations possibly indicating their distinct genetic components and less admixture from the other Nepali populations. In the PCA map, Vector Z, F1, M5 and U7 are mostly responsible for the relative positions of the Newar caste. These observations are concurrent with the data of the major Hgs observed in the Nepali populations (**Figure 4.3** and **Table 8.2**).

4.3.2 Between Nepali and other Asian populations

Further, map of Principle component analysis based on the mtDNA Hg frequencies of 18 Nepalese and 244 Different Asian populations (Appendix A: **Table 8.2**) was computed. The first component (PC1) which explains 21.6% of the genetic variance, basically segregates South Asian, East Asian and South East Asian populations. The second component PC2 which explains 11.62% of the genetic variance shows the variation/diversity within the Nepali populations. Intriguingly, the Nepali populations are clustered with the populations from MSEA (Myanmar and Thailand), South Asian and East Asian populations as shown in the PCA map.

In particular, majority of Newar sub caste are clustered with the populations from Myanmar, Thailand, China, N/NE Indian populations but relatively diverged away from the other Indian populations, except Shakya group which is placed nearer to the sample from Morang (Hindu-Terai) and Indian populations. The presence of comparatively high proportion of East Eurasian maternal components in Newa-mix (NMIX) and Maharjan group are supported by their positioning in the PCA map. Udaya, Bajracharya and Kshatriya (Uttarakhand, India) are closely clustered with the populations from Myanmar, Thailand and south west china. Among the North Indian populations, **Kshatriya** and **Shah** get partition away from the core Indian cluster, reflecting their distinct genetic composition as compared to the other North Indian group.

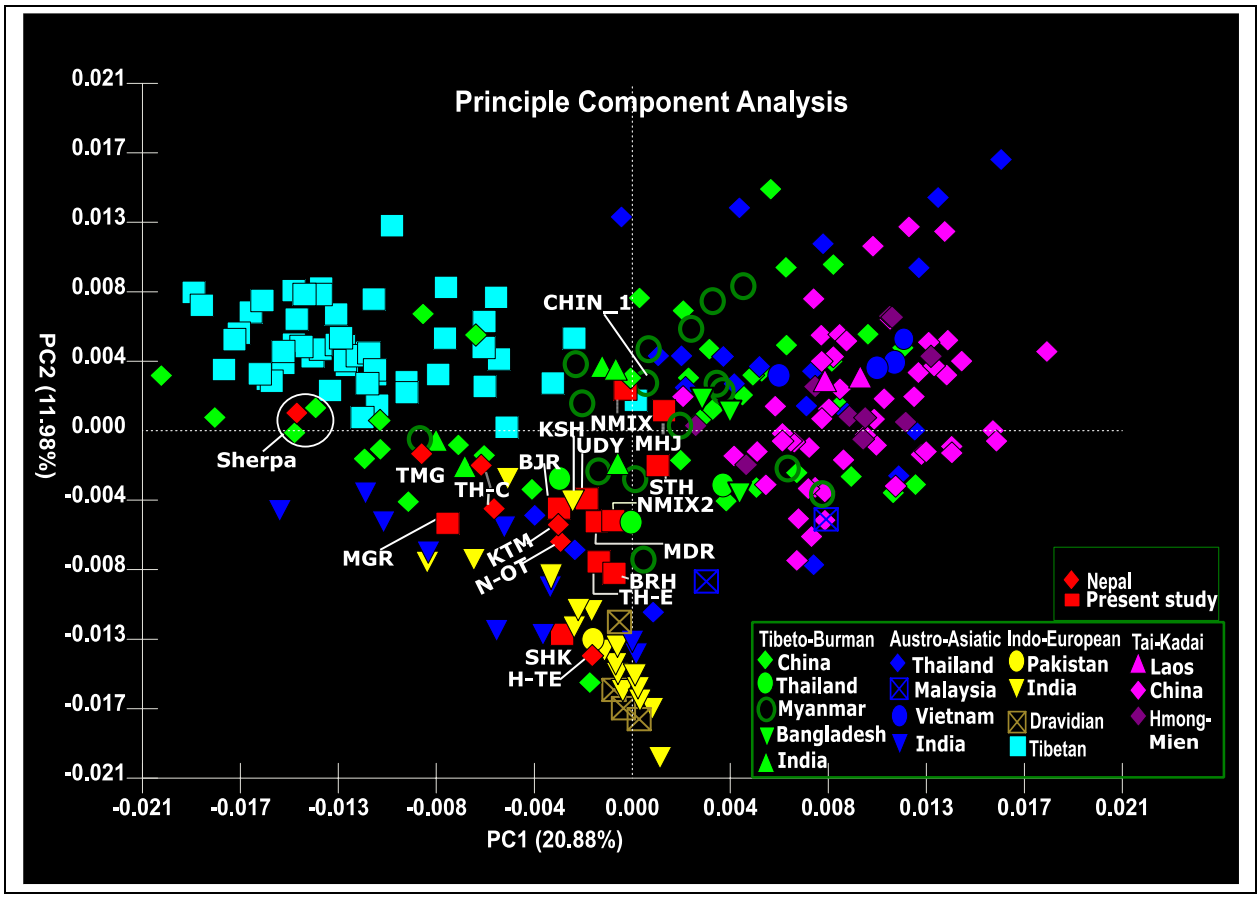


Figure 4.5| PCA of the 262 populations analyzed in the present study. Map of the Principal Component Analysis among the Nepalese and 244 other Asian populations. The first (PC1) and second (PC2) component explains 20.88% and 11.62% variation respectively. Except Nepali populations all are grouped linguistically and labelled with same color and their geographic location are indicated by different shape.

Terms used to denote the populations are given as: TMG, Tamang; TH-C, Tharu from Chitwan; SHK, Shakya; MDR, Manandhar; BRH, Brahmin; KSH, Kshatriya; MHJ, Maharjan; H-TE, Hindu Terai; Nmix, Newar mix (This study); Nmix2, Newar mix 2 (Gayden et al., 2013)

Previous study showed presence of higher proportions of East Eurasian maternal lineages among the Uttarakhand populations (Negi et al., 2016), indicating the gene flow from Nepal and/or other neighboring populations. Sherpa, Tamang, two central Tharu collection and Magar exhibit stronger genetic affinities with the Tibetan populations compared to the Newar group. One Austroasiatic group from Thailand is closely clustered with the Brahmin (Nepal) and several Indian populations (Dravidian, Indo-European and Austro-Asiatic). Magar cluster loosely with the Tamang and Tharu population from Chitwan. Likewise, Tharu from Chitwan (TH-CI) are tightly clustered with the Tibeto-Burman groups from Yunnan (South west china) and Shah from Uttarakhand.

Tibetan and Sherpa shows the highest frequency of hg M9, D4 and A. South china and MSEA populations shows highest frequency of Hg M7, B4, B5 and F1. Similarly, high frequency of hg U2I (U2a, U2b, U2c), M33 & M5 is observed in the majority of South Asian populations. Hg Z observed in highest frequency in Nepali populations were not observed as a major variable either in Central Asia/East Asia/MSEA. Hg Z believed to be originated somewhere in Central Asia is very rare and has been detected in very few populations till date. However, it is the most dominant Hg in Newar. Another dominant hg in Newar is Hg F1 which has a wider and high frequency distribution throughout East/SE Asia. The scarcity/unavailability of the mtDNA data of Hg Z have resulted in lower contribution from the vector Z. This might have possibly distorted the result of the PCA map. In the previous PCA map (Figure 4.4) Shakya was placed closely to the Newar cluster away from the Hindu Terai (H-TE) group. But in the second PCA map it is placed very closer to the Hindu-Terai group. Hg Z is present in high frequency in Newar group including the Shakya. Likewise, Hg F is present in high frequency in Newar except Shaky group. This has lead into the positioning of Shakya near to the Hindu Terai (H-TE) in the second PCA map. The PCA map would have been indeed very different for the Newar, with closer clustering of the Newar with the Central/East Asian/MSEA populations. However, the clustering pattern tends to supports the proposed antiquity of Newar (Regmi, 1965) in which the gene flow occurred from Central Asia/East Asia and South Asia.

4.3.3 Unweighted Pair Group Method with Arithmetic Mean (UPGMA)

Clustering was performed using the UPGMA based on the mtDNA Hg frequencies observed in the Nepali populations **Table 8.2**. In the Phenogram, Maharjan and Newar mix are similar to each other, in which Shrestha shows more inclination towards this group. Udaya are placed closer to the above cluster whereas Bajracharya followed by Shakya are distantly apart with Maharjan. Likewise, Kathmandu and Nepal-other are similar to each other and Brahmin are closer to this group. Tamang and Magar are closer to each other. Similarly, Tharu from Chitwan I (TH-CI) and Chitwan II (TH-CII) are closer to each other in which Tharu from Morang are also closer to the Tharu from Chitwan. As compared to the other population, Tharu shows more similarity with Tamang and Magar. The clustering analysis performed on Nepali populations yielded very similar results as those obtained in PCA map of Nepali populations (**Figure 4.4**).

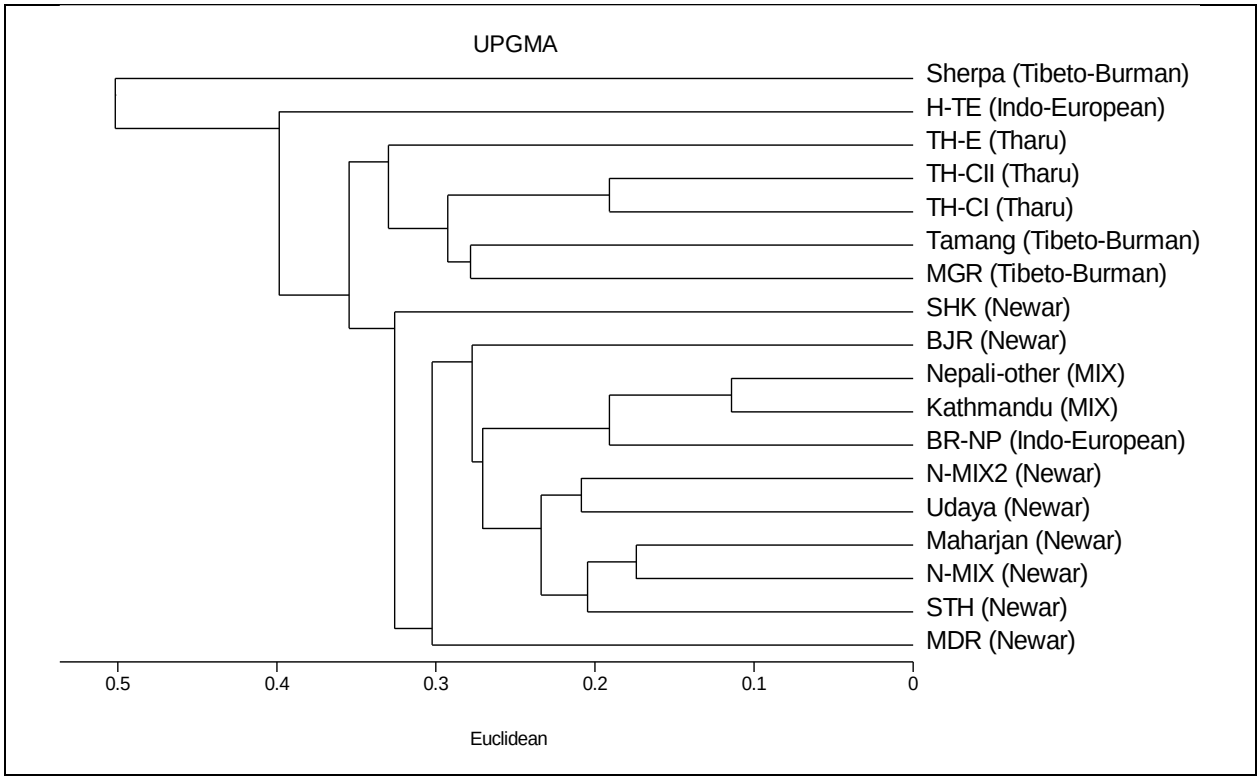


Figure 4.6 | Phylogenetic Tree showing the Genetic distance between the populations based on the distance matrix. Distance matrix was based on the UPGMA (Unweighted Pair Group Method with Arithmetic mean). Linguistic affiliation of each populations is shown in the bracket.

4.4 Phylogeographic pattern of the dominant Hgs within the Macro Hg N.

Super Hg N which populates half of the world includes 12 major groups.

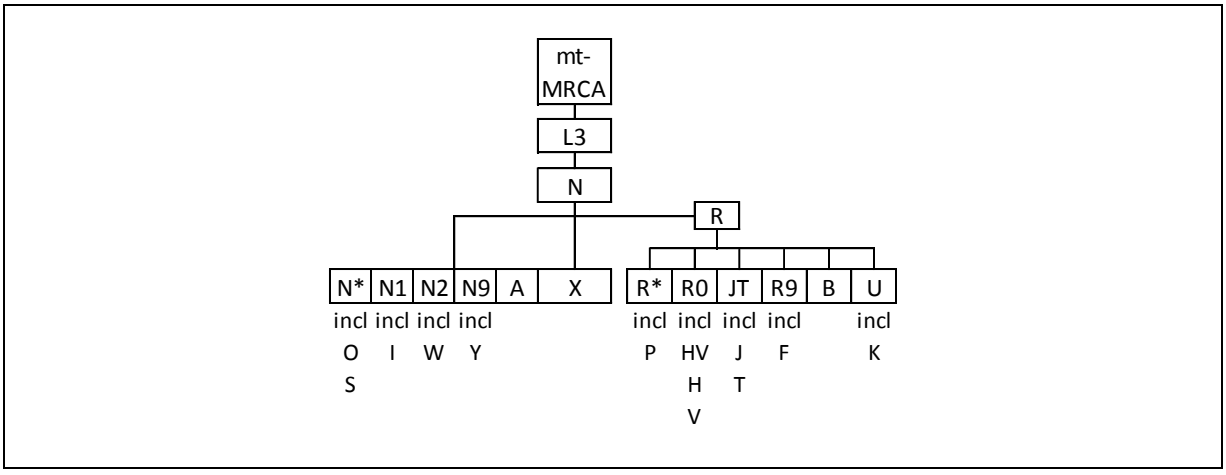


Figure 4.7 | Phylogeny of Macro Hg N. For convenient, Hg N is divided into 12 subtrees (Phylotree.org). In the figure, including is abbreviated as Incl.

4.4.1 Haplogroup R

Macro Hg R and its descendants (**Figure 4.7**) are distributed all over Europe, the Near East, south Asia, Oceania and the Americas.

Macro Hg R: Hg R contains the following sub hg

R2	
JT	J
	T
R5	R5a2b

R6	R6b
R9	F1
	F2
R30	R30b2a

In present study, single sample from Magar belong to hg R2" JT. Complete mtDNA sequencing of this sample was performed. But the sample didn't fall in any of the predefined sub Hg within R2'JT. Hg R2 and J were not observed in present study. However, traces frequency of hg J (1.29%) was observed in Kathmandu (Gayden et al., 2013).

4.4.1.1 Haplogroup R5

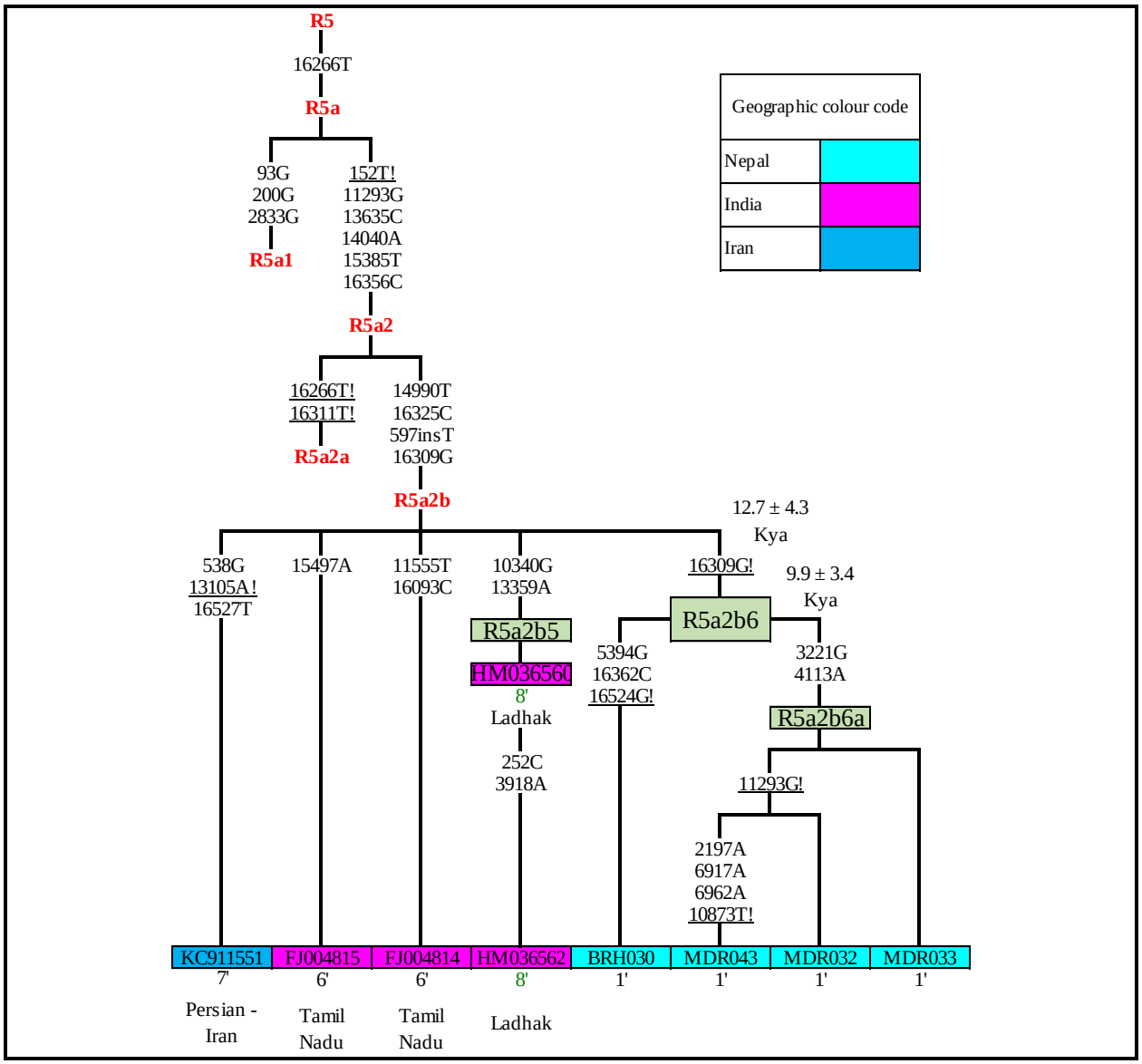


Figure 4.8 | Phylogenetic tree of sub hg R5a2b based on 8 complete mtDNA sequences. Mutations relative to the RCRS are indicated on the branches. Sample ID/GenBank accession number for each sequence are given in the box with geographic color code. The number (1', 2'... n') corresponds to the reference for the analyzed sample. The hg written in black color represents newly defined hg in this study, whereas red color indicates the predefined Hg in PhyloTree (Oven, 2015). Lineage R5a2b6 has been

differentiated within Nepal in which majority of basal variants are shared within the Nepali population, suggesting the ancient origin of R5a2b rather than recent entry via migration.

Hg R5 is considered to be one of the oldest lineage in South Asia. In India, Hg R5 has a wide distribution, especially in North India, Madhya Pradesh and West India. Sub Hg R5a1 are reported to be present in high frequencies among the Indo-European speaking population, whereas **R5a2** is present among the Dravidic groups of India and Sri Lanka (Chaubey et al., 2008). In present study, Sub Hg **R5a2b** was observed in **Manandhar**, Magar and **Brahmin** populations. Complete mtDNA sequencing of R5a2b sub clade shows Brahmin and Manandhar share a basal back mutation (16309G!). Further, HVR1 sequence of Magar also shares the basal variants (16189C! 16524G! 16362C) with the Brahmin populations.

4.4.1.2 Haplogroup T

Hg T, believed to be originated somewhere **Near East** contains two main branch T1 and T2. Both of these sub Hg have very different distributions. In present study, Hg T2 is present in **Brahmin** (4.08%), and **Bajracharya** (5%). Previous study reported traces of Hg T2 were present in Kathmandu and Eastern Nepal. In North India, Sub Hg T1 were observed in higher frequencies (Brahmins of Uttarakhand) (**Table 8.2**.)

4.4.1.3 Haplogroup R6

Hg R6 has been observed in small frequencies in India and Pakistan. In India, especially reported in North India, Madhya Pradesh and Kashmir **(Table 8.2)**.

In present study, sub Hg R6a2 was observed exclusively in Brahmin and Magar, whereas R6b was present in high to low frequencies among the Newar sub caste: **Bajracharya (15%)**, **Shakya (5.2%)**, Newa mix (2%) and Maharjan (1%). This Hg was also observed in **Brahmin (8.33%)**. Complete sequence analysis of samples belonging to subclade R6b shows, Nepalese sample cluster with a sample from Myanmar. Three basal variants (4751G, 2984A, 13812C) was shared between Myanmar and Newar sample (one sample each from Shakya and Newar mix, 3 sample belonging to Bajracharya). Sub Hg R6b1, R6b1a and R6b1a1 were newly defined in this study. The homogeneity in the sequence between Newar mix, Shakya and Bajracharya indicates a close matrilineal relationship between these Newar sub castes (**Figure 4.9**). Previous study reported Hg R5a2b2 and R6b were also present in Kathmandu (2.7%) (Gayden et al., 2013). Traces of sub Hg R8 were also reported to be present among the population from Kathmandu and East Nepal (Wang et al., 2012).

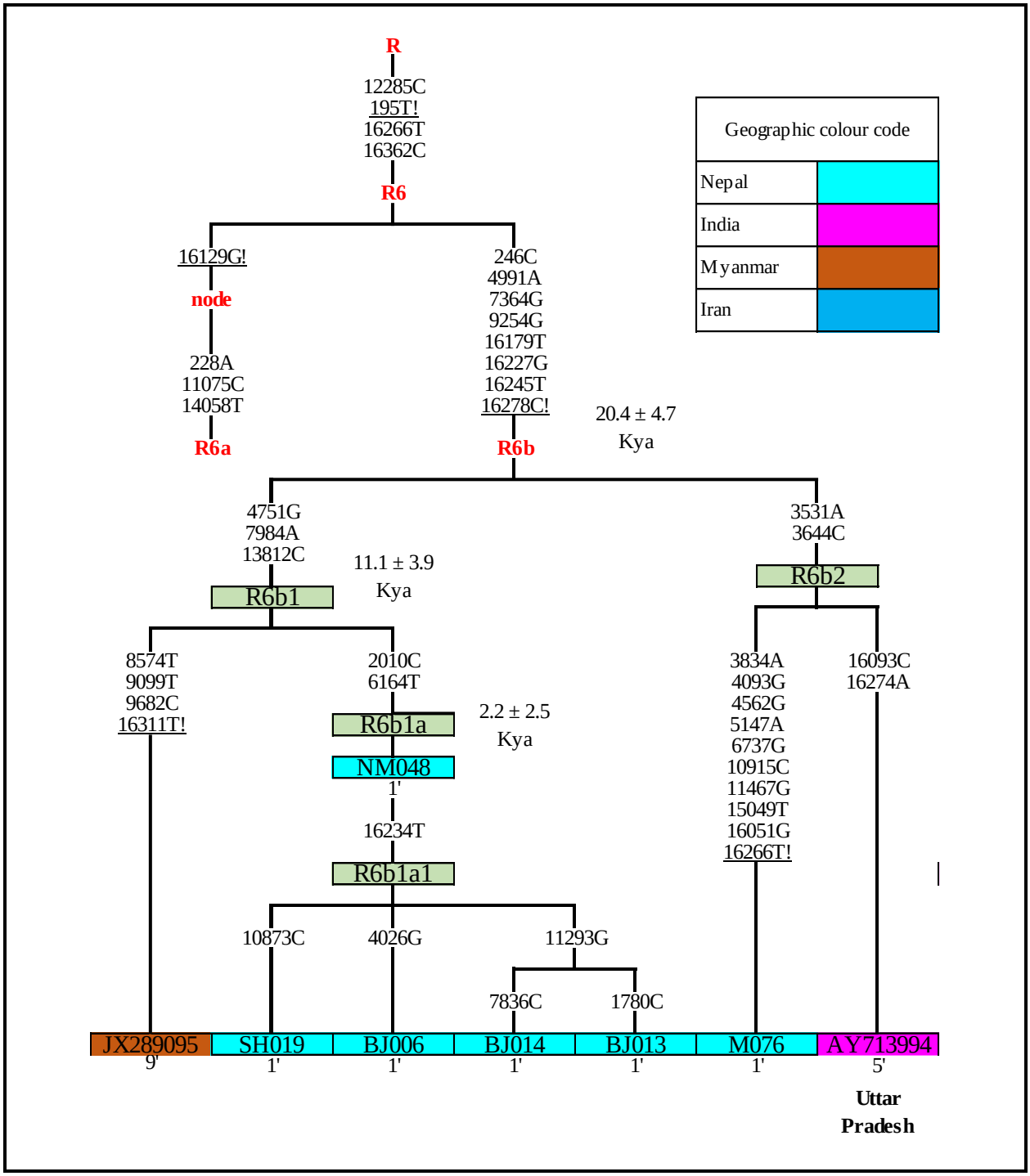


Figure 4.9 | Phylogenetic tree of sub Hg R6b based on 8 complete mtDNA sequences. Mutations relative to the RCRS are indicated on the branches. Sample ID/GenBank accession number for each sequence is shown in the box with geographic color code. The number (1', 2'... n') corresponds to the reference for the analyzed sample. The hg written in black color represents newly defined hg in this study, whereas red color indicates the predefined Hg in PhyloTree (Oven, 2015).

4.4.1.4 Haplogroup R30

Two sample, each belonging to Maharjan and Newar-mix belong to sub hg R30b2a. Sub hg R30b2a observed in Newar sample share an ancestry with the sample from India. R30 is considered to be the indigenous south Asian specific Hg (Silva et al., 2017)

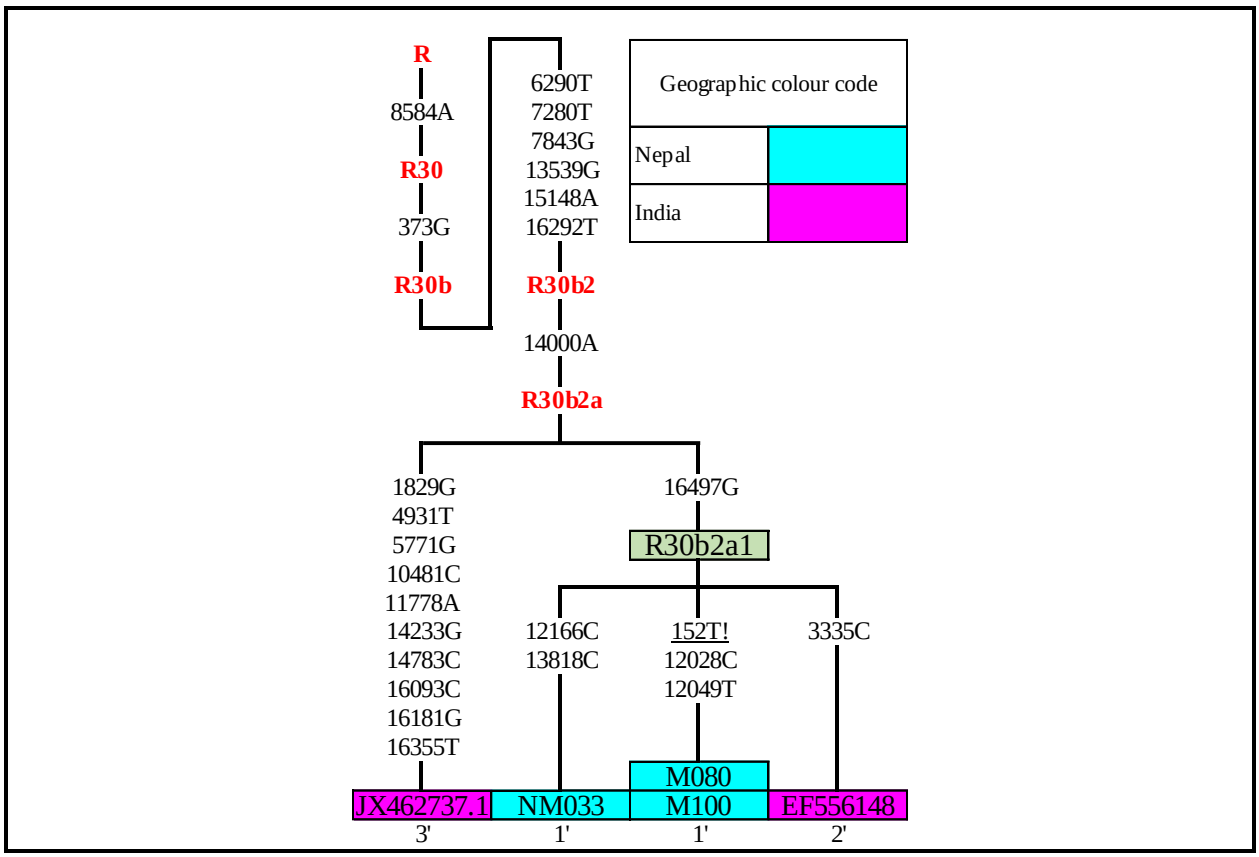


Figure 4.10 | Complete mtDNA phylogenetic tree based on 5 complete mtDNA sequences from South Asia. In present study, basal mutation (16497G) shared between Newar (Maharjan and Newamix) and India sample lead to define new sub hg (R30b2a1). Mutations relative to the RCRS are indicated on the branches. Sample ID/GenBank accession number for each sequence is shown in the box with geographic color code. The number (1', 2'... n') corresponds to the reference for the analyzed sample. The hg written in black color represents newly defined hg in this study, whereas red color indicates the predefined Hg in PhyloTree (Oven, 2015).

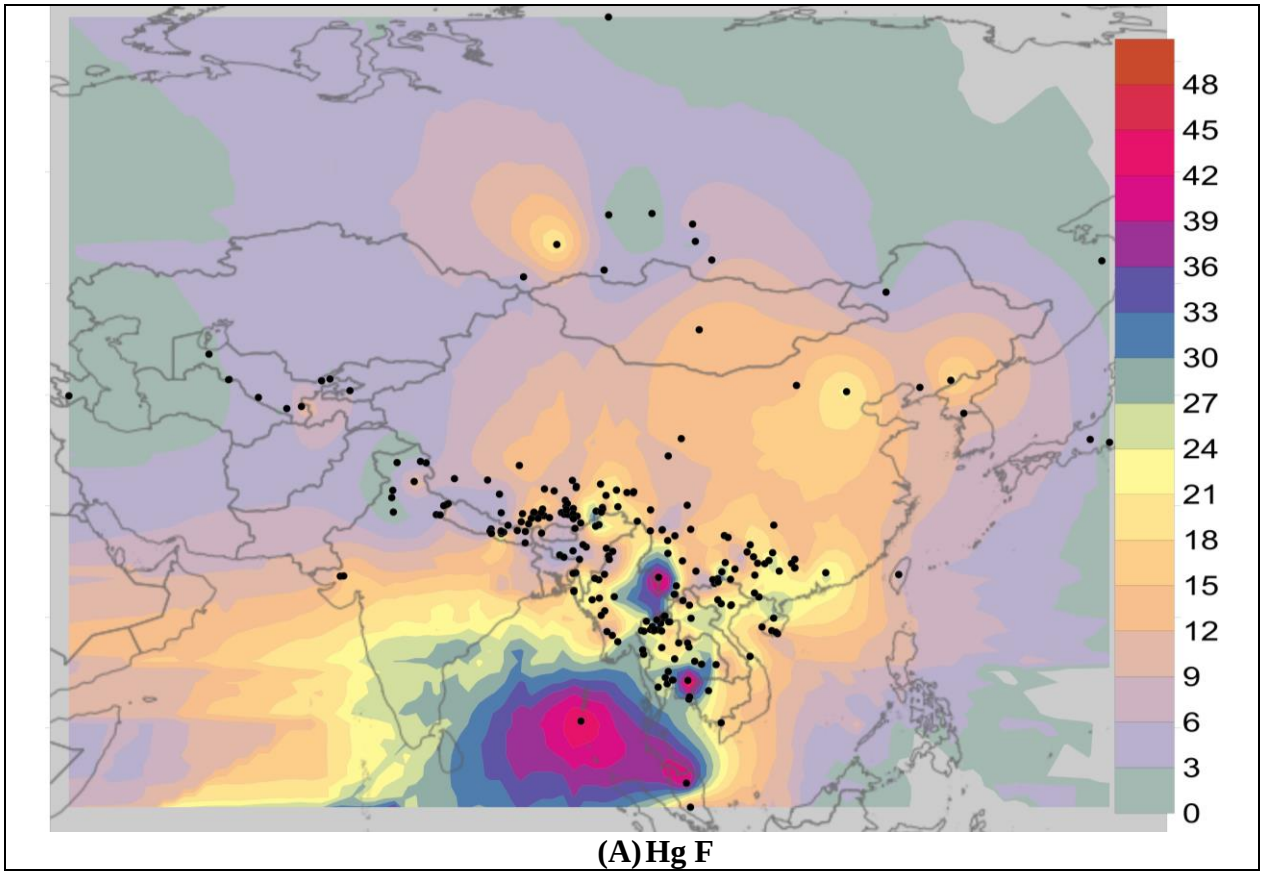
4.4.1.5 Haplogroup F

Hg F is derived from lineage R9. This lineage R9 has been observed mainly in Southeast Asia, found all over Indonesian, Malaysia and aboriginal Malays. Hg F has spread throughout East Asia, the Russian Far East and South-East Asia including the Philippines, Indonesia, Malaysia and some of coastal New Guinea. Highest frequency of hg F is reported to be present in places like Nicobar Island (50%), Arunanchal Pradesh (31%) of India and Shor people of Siberia (44%) (Delfin et al., 2014; Fedorova et al., 2013; Irwin et al., 2010; Kutanan et al., 2017; Xuebin Qi et al., 2013; Zhang et al., 2013). **This lineage represents one of the main maternal founding lineages in the Newar**.

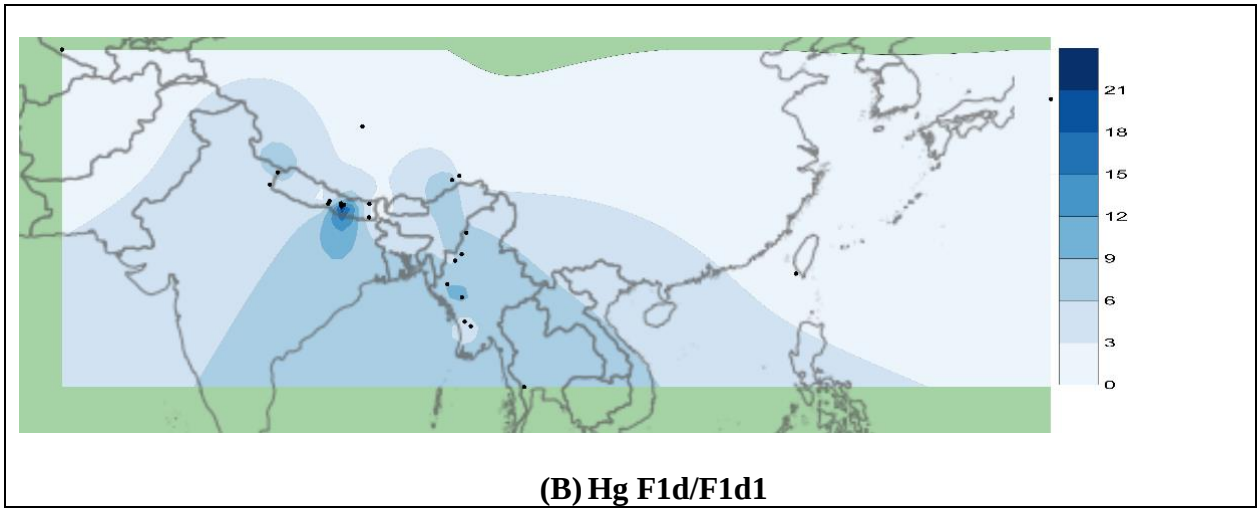
Hg F is characterized by sub hg F1, F2, F3 and F4 (Oven, 2015). Among these, F1a is the most dominant sub hg in South East Asia and Siberia. Overall frequency of hg F among the Tibetan population was 11.44% (Xuebin Qi et al., 2013). In present study, Nepali population where characterized by sub hg F1 (F1b, F1c, F1d1 and F1g) and F2. Although, hg F1 and F2 are shared between Nepali and East Eurasian, there are major differences in the distribution and frequencies of these sub hg.

4.4.1.5.1 Haplogroup F1d

F1d1 was the most predominant clade in Nepali population. Hg F1d/F1d1 is characterized by generally low population frequencies and limited sequence diversity/availability, despite a wide geographic distribution as shown in contour map (**Figure 4.11**). To get more insight into the spatial-frequency distributions of Hg F and F1d/F1d1 counter maps was created**.**



(A) Hg F



(B) Hg F1d/F1d1

Figure 4.11 | Contour map of A) hg F & B) hg F1d/F1d1 of Nepali and other Asian populations based on the mtDNA Hg frequencies (Table 8.1 & Table 8.2). The black dots indicate the geographic locations of the analyzed populations. The bar on the right side of the map indicate the haplotype frequency spectrum.

Traces of F1d has been reported to be present among the populations from South/South west China, however their frequency among the population from china is unavailable **(Table 8.1 & Table 8.2)**. Compared to other sub clades of hg F, both the phylogenetic structure and ancestral origin of hg F1d1 is obscure. Haplogroup F1d1 were observed in very low frequency with relict distribution among the Tibetans (1.2%), Dengs (1.08%) and Lhobas (2.7%) of Tibet (Kang et al., 2016). Hg F1d (4.2%) and F1d1 (4.3%) were also present among the Mon (MO2) people of North East Thailand and West Thailand (border between Thailand and Myanmar). The overall frequency of F1d/F1d1 in Burma (Myanmar) is ≈4.85%. Similarly, frequency of F1d in Taiwan was 1% (1/50). F1d was also reported to be present in traces frequency in Uzbekistan belonging to Turkmenistan origin (1%, 1/212) **(Table 8.2)**.

Hg F1d1 is distributed unevenly in Newar; observed in **high frequencies in Maharjan (21%)**, Newa-mix (20%), Shrestha (19.44%) and Udaya (8.62%). Whereas, completely absent in Manandhar and Shakya. This frequency pattern of F1 Hg might be responsible for the variation observed in the PCA map. Lineage F1d1 were also present in Brahmin (4.16%) but completely absent in Magar. Hg F1d were also observed among the Kshatriya (8.86%), Shah (2.85%) and Brahmin population from Uttarakhand, India (Negi et al., 2016). Median joining network for mtDNA hg F1d1 based on a complete mtDNA sequence are shown in the **Figure 4.12**.

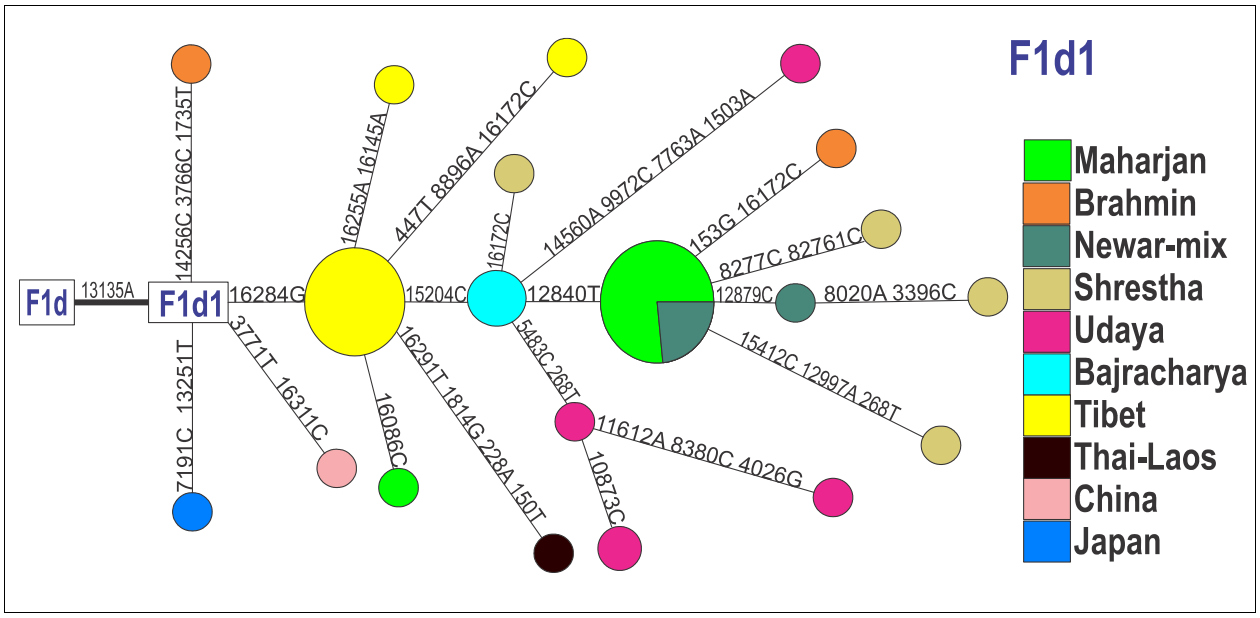


Figure 4.12 | Median Joining Network of Hg F1d1 based on complete mtDNA sequence. All the Nepali samples are grouped according to their caste and shown by different color. The size of the circles is proportional to the number of individual cmtDNA sequences. Nucleotide position number shown in the network are consistent with the Revised Cambridge Reference Sequence (rCRS). The geographic origin of sample is shown by different colors. The C stretch length polymorphism in regions 303–315, AC indels at 515–522, 16182C, 16183C, 16193.1C(C) and 16519 were disregarded for the network reconstruction.

As shown by the Median Joining Network, Thai-Laos and Nepali samples belonging to lineage F1d1 shares an ancestry with the Tibetans. On the basis of constructed networks, we can see that Nepali sample share some basal as well as internal haplotypes with the Tibetan samples. Nepali harbor's a number of unique (Nepal specific) haplotypes at the terminal level, most of which branched off directly came from the nodes occupied by Tibetan lineage.

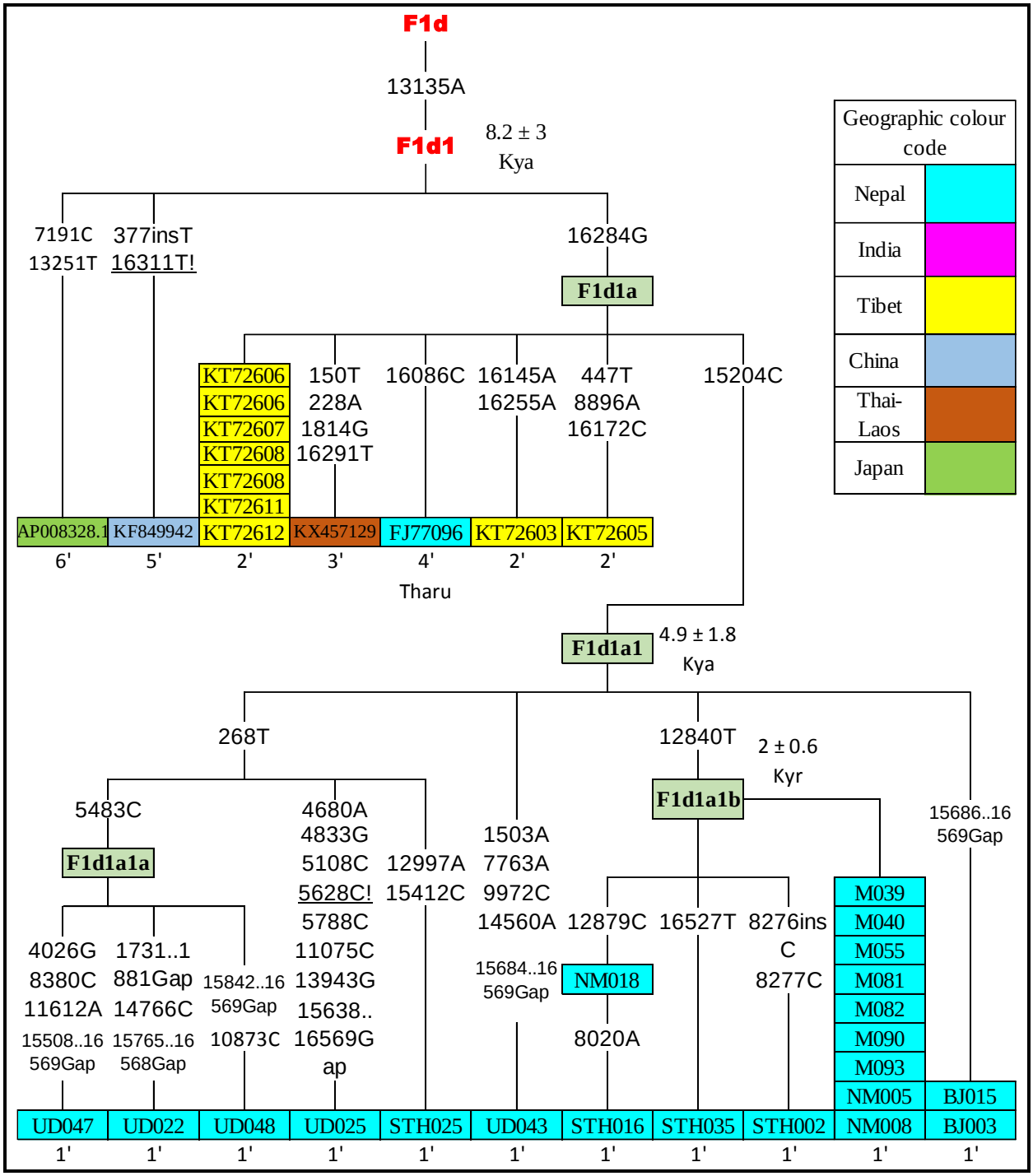


Figure 4.13 | Phylogenetic tree of sub hg F1d1 based on 36 complete mtDNA sequences of which 23 complete mtDNA sequences were generated in this study. Mutations relative to the RCRS are indicated on the branches. Sample ID/GenBank accession number for each sequence is shown in the box with geographic color code. The number (1', 2'... n') corresponds to the reference for the analyzed sample. The hg written in black color represents newly defined hg in this study, whereas red color indicates the predefined Hg in PhyloTree (Oven, 2015). Majority of the contemporary populations belonging to the lineage F1d1 from Nepal, Tibet, Thailand and Myanmar (only HVR data available) has been derived from the basal variant 16284G, suggesting that the most recent common ancestor for this sub lineage was originated in East/southeast Asia.

Further, complete sequence analysis (**Figure 4.13**) revealed the clustering of Nepali sample with the Tibet and Thailand sample, in which one basal polymorphism (16284G) is shared between

these populations to form a new lineage named herein as F1d1a. Basal Variant 15204C (only observed in Nepali populations) is the major founding (ancestral) variant of the contemporary Nepali populations belonging to lineage F1d1 (**Figure 4.13**) suggesting a reduced genetic diversity among the contemporary Nepali populations. In total Nepal specific ancestral variant 15204C has been observed in 43 individuals from Newar.

4.4.1.5.2 Haplogroup F1c1a2

Except Manandhar (8.62%) F1c1a2 is present in very low frequency among the other Nepali populations. Sub hg F1c1a has been previously reported to be present in Khon Muen (4%), Khamu (4%) and Paluang (8%) groups of Thailand. F1c1a2 was also observed in Ladhak, Myanmar 5.297% (44/845) and Tibet **(Table 8.2**). Highest frequency of F1c1a2 has been observed in Apatani people form Arunachal India. As indicted in contour map (**Figure 4.14**), Hg F1c is most likely to be originated in East Asia.

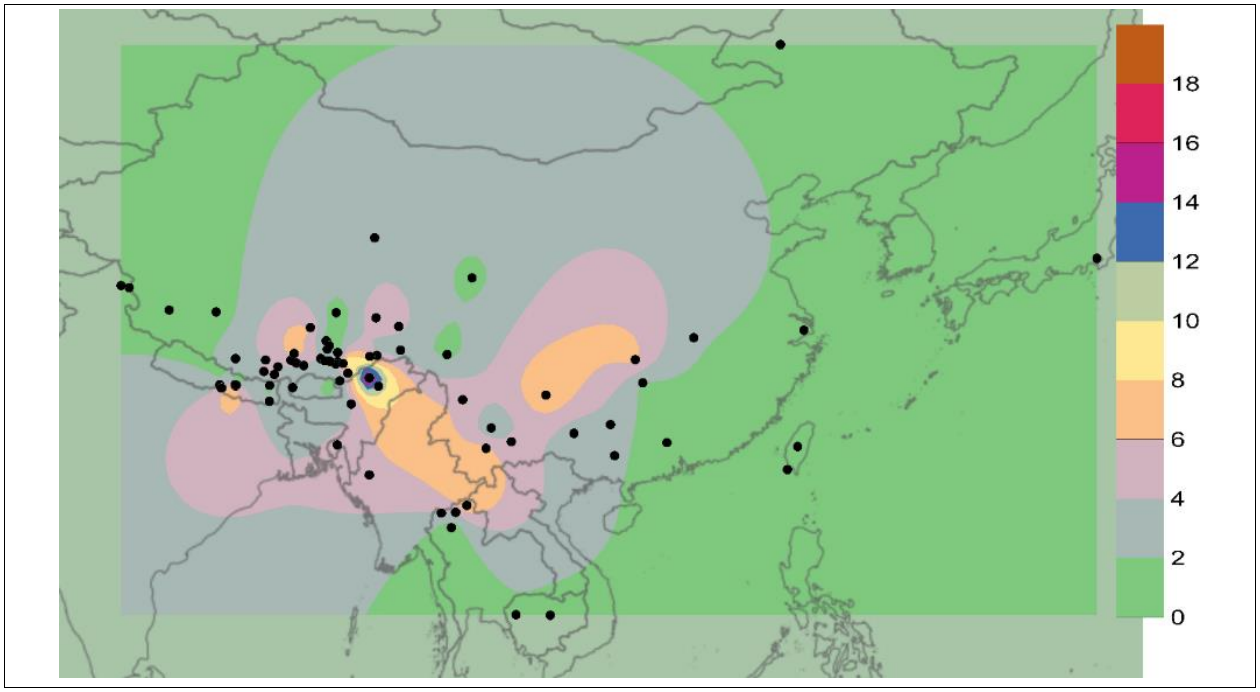


Figure 4.14 | Contour map of the Hg F1c of Nepali and other Asian populations. The black dots indicate the geographic locations of the analyzed populations. These spatial-frequency distributions were created using the Kriging linear model of the Surfer 12.8 package, based on the frequency of each Hg in different populations (Supplementary **Table 8.2**).

Complete mtDNA analysis of the sample belonging to lineage F1c1a2 shows the sharing of Basal mutation 16058C between the Tibetan and the Apatani people from Arunachal Pradesh, India. This basal mutation observed in Apatani people was absent in Nepali populations, indicating the independent migration from Tibet into both Nepal and North-East India. The founding variant (234G) of the contemporary populations of Nepal, Tibet and Thailand was also present in Myanmar (based on the HVR data) (Y. C. Li et al., 2015). Basal Variant 15388C (only observed in Newar populations) is the major founding (ancestral) variant of the contemporary Manandhar F1c1a2 lineage suggesting a recent populations expansion and hence a reduced genetic diversity among the contemporary Nepali populations. Nepal specific Variant 15388C was observed in

eight individuals from Manandhar and one individual from Udaya. Hence, the present-day Manandhar populations belonging to lineage F1c1a2 has been descended from a small number of colonizing ancestors (Founder effect) that arrived from East Asia through Tibet.

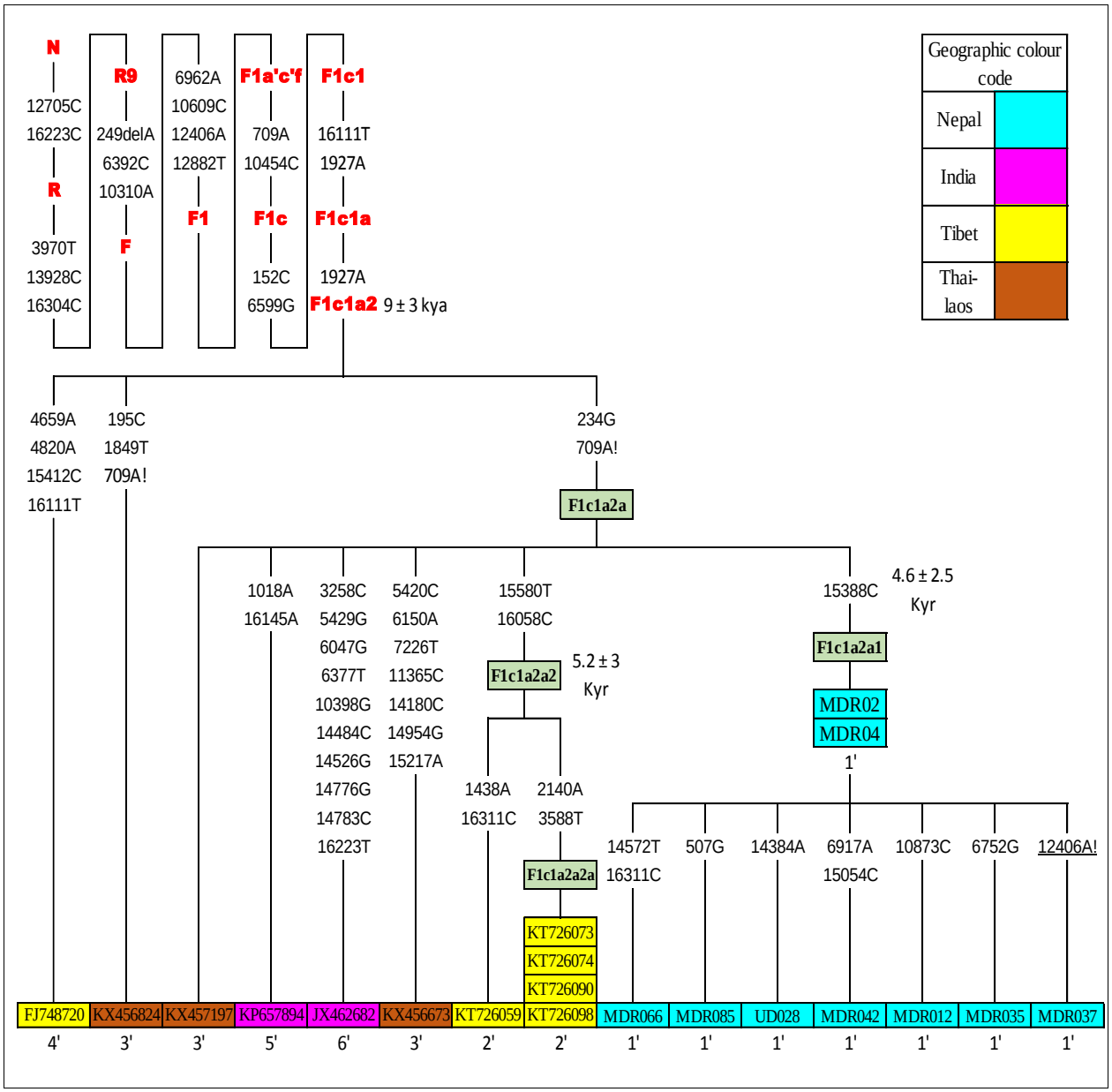


Figure 4.15 | Phylogenetic tree of sub hg F1c1a2 based on 20 complete mtDNA sequences of which 9 complete mtDNA sequences were generated in this study. Majority of the present-day populations from Tibet, Thailand, Myanmar (only HVR sequence available), India and Nepal belonging to lineage F1c1a2 has been derived from the ancestral variant 234G.

4.4.1.5.3 Haplogroup F1g

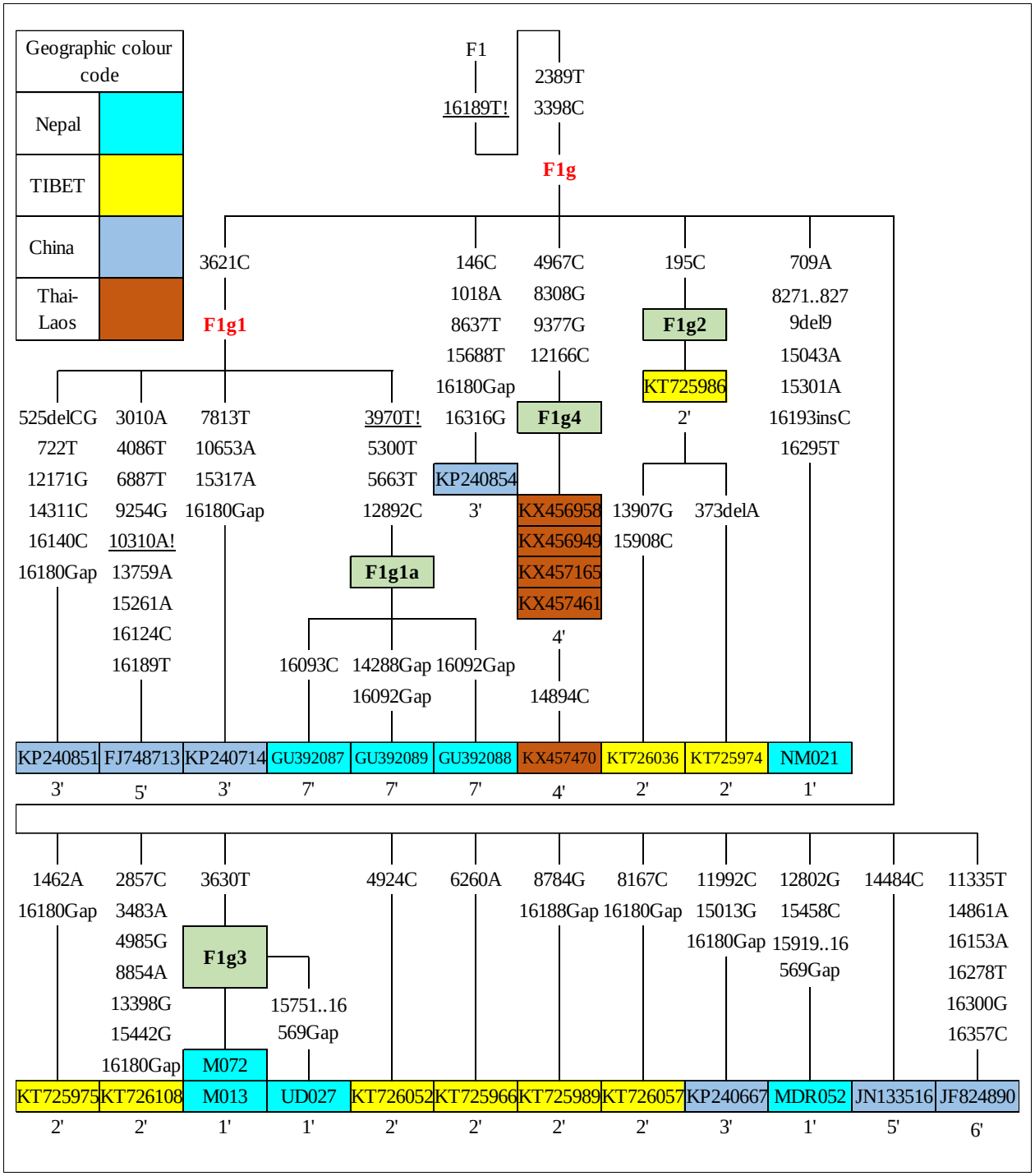


Figure 4.16 | Phylogenetic tree of sub hg F1g. PhyloTree was constructed using 29 complete mtDNA sequences of which 5 complete mtDNA sequences were generated in this study. All the Newar sample belongs to lineage F1g and the Individuals from the previous study (Wang et al., 2012) belong to lineage F1g1 were included in the analysis. Sample ID/GenBank accession number for each sequence is shown in the box with geographic color code. The number (1', 2'... n') corresponds to the reference for the analyzed sample. The hg written in black color represents newly defined hg in this study, whereas red color indicates the predefined Hg in PhyloTree (Oven, 2015).

Sub Hg F1g were present in low frequency among the Maharjan (5%), Shrestha (4%), Udaya (1.72%) and Manandhar (1%). F1g were also observed in low frequency among the population of

Tibet: Tibetans (4.21%) and Lhobas (2.7%) (**Table 8.2**). This hg was also detected among the Phuan people from Northern Thailand (Kutanan et al., 2017)

4.4.1.5.4 Haplogroup F2b1

In present study, hg F2b1 was observed in low frequency among Newa-mix (4%) and Udaya (3.44%). Complete mtDNA analysis indicate that the F2b1 lineage observed in Udaya share close maternal relationship with F2b1 of Thailand.

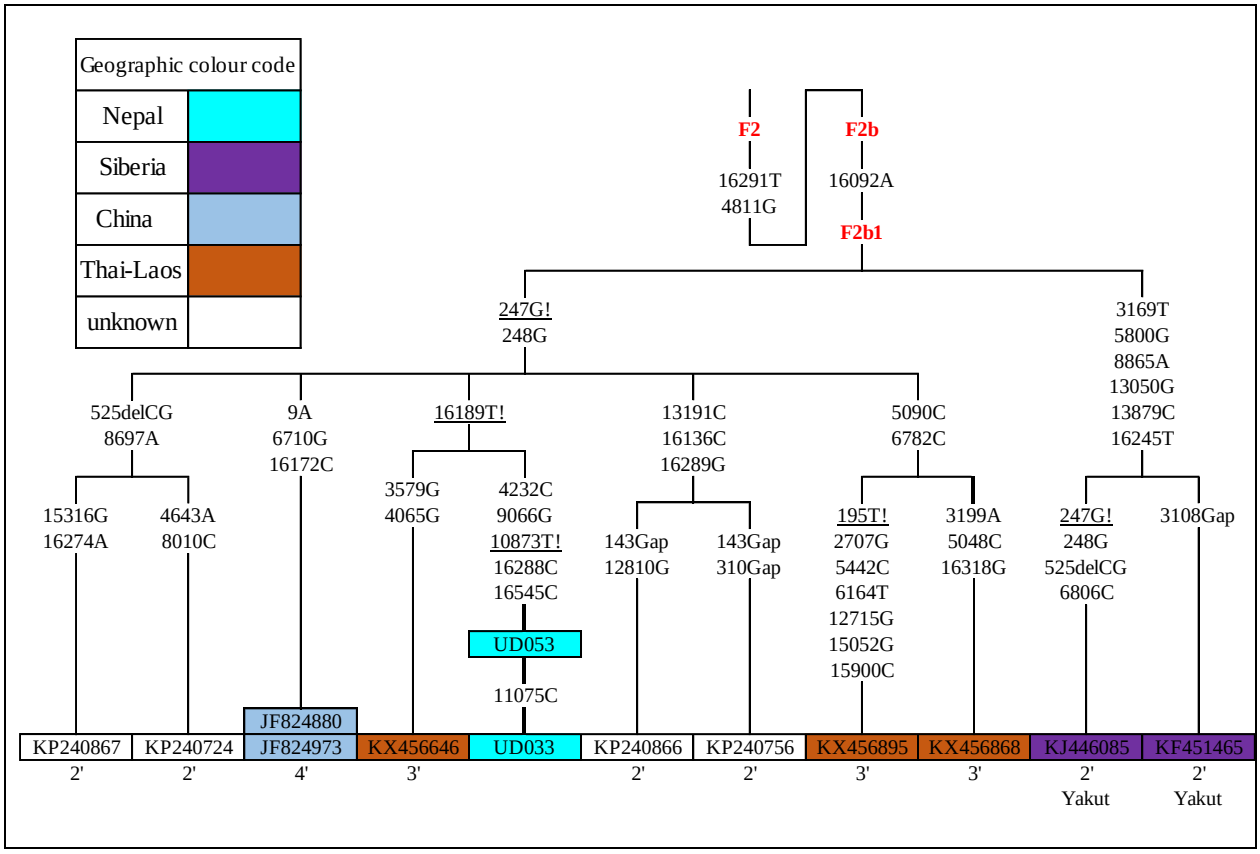


Figure 4.17 | Phylogenetic tree of sub hg F2b1 based on 13 complete mtDNA sequences. Mutations relative to the RCRS are indicated on the branches. Sample ID/GenBank accession number for each sequence is shown in the box with geographic color code. The number (1', 2'... n') corresponds to the reference for the analyzed sample. The hg written in black color represents newly defined hg in this study, whereas red color indicates the predefined Hg in PhyloTree (Oven, 2015).

4.4.1.6 Haplogroup HV

This Hg is present mainly in Western Asia, Southern Europe, Eastern Europe and North Africa. 24,000 year old AMH found in southern Italy either belong to Hg HV or R0, according to the study published on 2003 (Caramelli et al., 2003). Similarly, according to the study published on 2017, Hg HV was also been found among ancient **Egyptian mummies** excavated at the Abusr el-Meleq archeological site in the middle Egypt, which date from the Pre-Ptolemaic/late New Kingdom, and Roman periods (Schuenemann et al., 2017). **Half of the European population and 25-40% of the Near East populations belong to the lineage HV (Table 8.2)**.

Hg HV12 is mainly found in Iran. These Hg has been characterized by complete genome sequencing of the population from Iran, Armenia and Turkey (M. Derenko et al., 2013). HV12a and HV12b, subclades of Hg HV12 have been reported in the Near Eastern population **(Table 8.2)**. Whereas in India and Nepal only the lineage of HV12b (HV12b1 and HV12b1a) were observed. In the present study, Hg HV12b1 were observed in High frequencies exclusively in Manandhar (10.85%). Surprisingly, these Hg were completely absent among the other Nepali populations studied till date (Bhandari et al., 2015; Chaubey et al., 2014; Gayden et al., 2013; Wang et al., 2012).

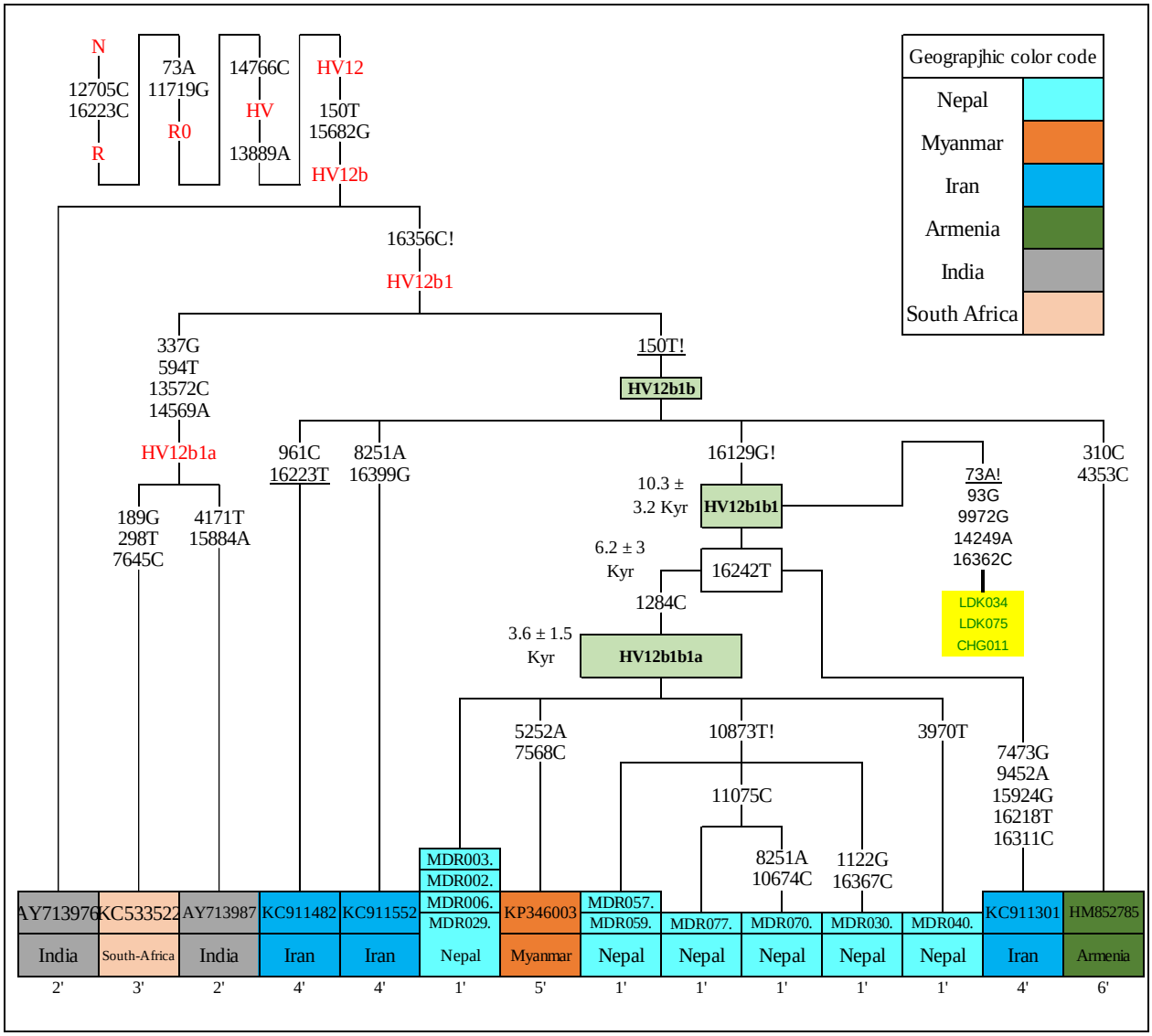


Figure 4.18 | Phylogenetic tree of Hg Hv12b based on 18 complete mitochondrial sequences. Manandhar sample belonging to lineage HV12b1 traces its ancestry to the populations from Iran and Armenia (Near East). Sample ID/GenBank accession number for each sequence is shown in the box with geographic color code. The number (1', 2'... n') corresponds to the reference for the analyzed sample. The hg written in black color represents newly defined hg in this study, whereas red color indicates the predefined Hg in PhyloTree (Oven, 2015).

Sub Hg HV12b1a share a deep ancestry with the Near Eastern (Iran and Armenia) populations (Schonberg et al. 2011; Derenko et al. 2013) as shown in the **Figure 4.18**. Similarly, sub Hg HV12b1 of Manandhar traces its ancestry back to the populations from Armenia and Iran. Back mutation

at position 150T has been observed in all sequences belonging to Iran, Myanmar, Armenian, Ladhak and Nepalese sample. Myanmar sample share a series of basal variant with the Manandhar sample suggesting a recent gene flow from Nepal to Myanmar.

4.4.1.7 Haplogroup H13

Sub Hg **H13a2a** believed to be originated on Iranian plateau is frequent in the Iranian populations (M. Derenko et al., 2013). In India, subclade H13a1a and H13a2a are present in Uttar Pradesh, Tamil Nadu and Andhra Pradesh (Palanichamy et al., 2015)

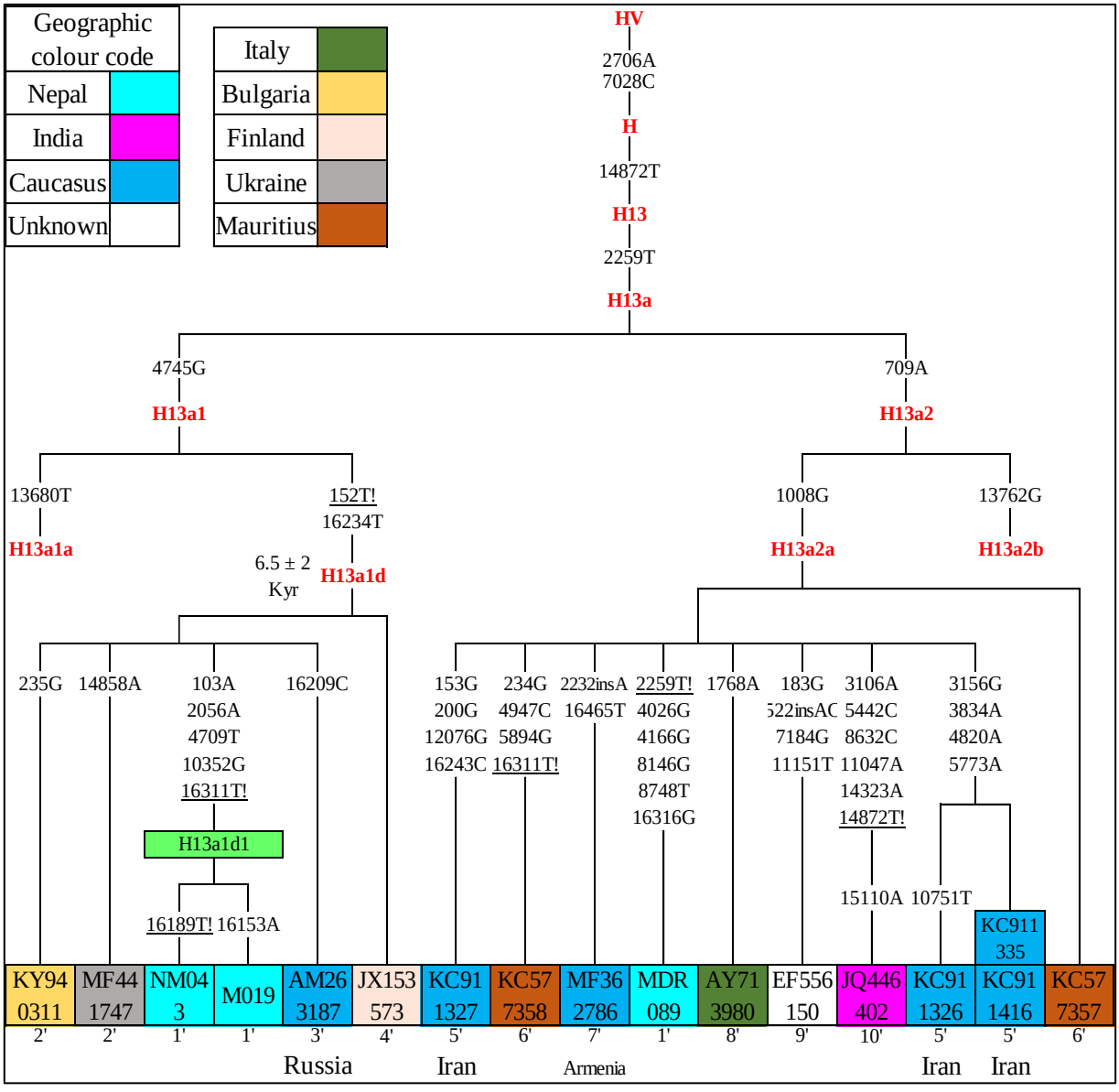


Figure 4.19 | Phylogenetic tree of Hg H13a1d based on 17 complete mitochondrial sequences. Mutations relative to the RCRS are indicated on the branches. Sample ID/GenBank accession number for each sequence is shown in the box with geographic color code. The number (1', 2'... n') corresponds to the reference for the analyzed sample. The hg written in black color represents newly defined hg in this study, whereas red color indicates the predefined Hg in PhyloTree (Oven, 2015). Maharjan and Newamix sample

belonging to lineage H13a1d are genetically closer with the populations from Russia, Ukraine, Bulgaria and Finland.

The Indian H13a2a lineage share an ancestry with Pakistani and the Near Eastern (Iran and Iraq) populations (Derenko et al. 2013.). In the present study, only a single individual from Manandhar caste belong to **H13a2a** lineage. Similarly, two **H13a1d** lineage belonging to **Maharjan** and Newa-mix group were present in this study. Complete mtDNA sequences for this lineage has been reported from Finland, Bulgaria, Russia and Ukraine as shown in the **Figure 4.19**. H13a2a and H13a1d lineage might have arrived in south Asia around ≈6.6-7.2 Kyr BP.

4.4.1.8 Haplogroup H2a

Hg H2a are found throughout Europe and the Caucasus. In our study only one sample belonging to Brahmin population belong to H2a linage. Whereas it was completely absent in Newar and Magar population. Complete mtDNA sequence of Brahmin shows sharing of three variants (13968A, 16271C, 16274A) with the individual from Kathmandu/East Nepal (Wang et al., 2012).

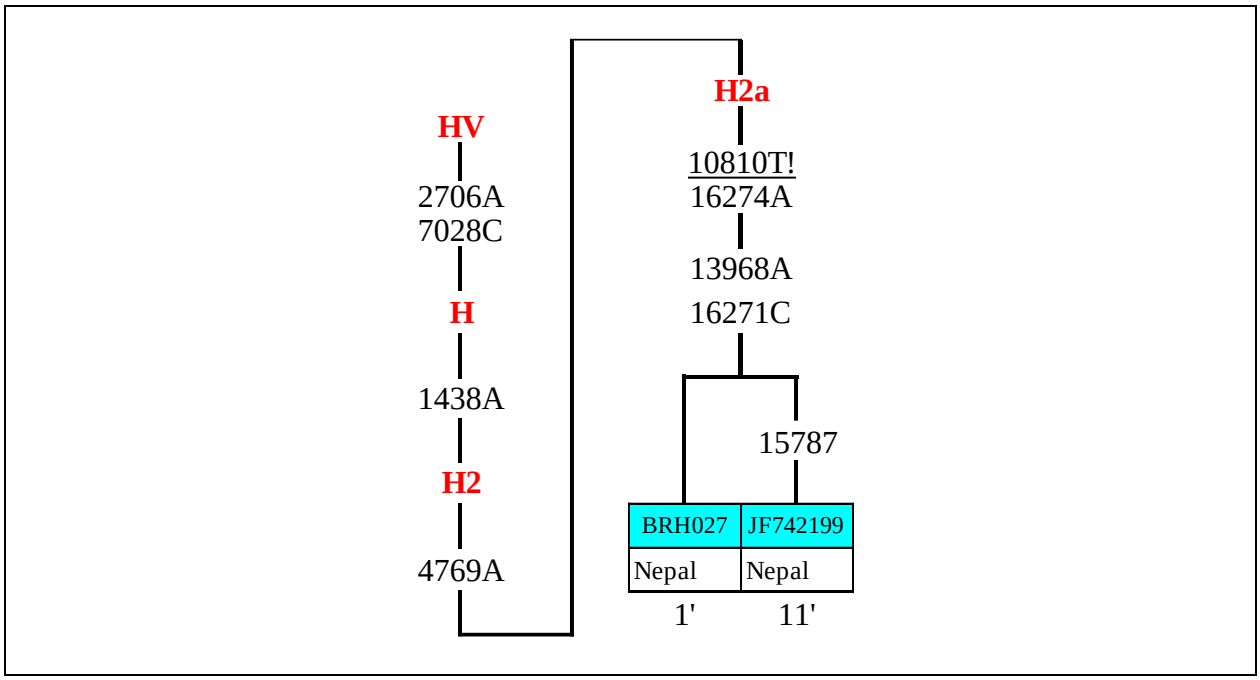


Figure 4.20 | PhyloTree of Hg H2a based on the complete mtDNA sequence. Mutations relative to the RCRS are indicated on the branches. Sample ID/GenBank accession number for each sequence is shown in the box with geographic color code. The number (1', 2'... n') corresponds to the reference for the analyzed sample. The hg written in black color represents newly defined hg in this study, whereas red color indicates the predefined Hg in PhyloTree (Oven, 2015).

4.4.1.9 Haplogroup U

Human mtDNA Hg U descendent from the Hg R is one of the initial maternal founders in **South West Asia** (it is a region at the crossroads between Asia, Africa and Europe. The terms have to some extent a similar notion as 'Middle East') **and Europe**. Around 40% of the mtDNAs type found in India fall into Western Eurasian Hg U. Hg U is subdivided into U1, U5, U6 and a fourth subclade U2'3'4'7'8'9 which is further divided into U2, U3, U4’9, U7 and U8 (including hg k). U2'3'4'7'8'9 is

a common ancestor defined by a common mutation A1811G, which arose between 42,000 and 48,000 years ago (Palanichamy et al., 2015).

Hg U has a wide distribution and many of these subclades shows region-specific distribution/frequency pattern in modern day population. Among these, hg U1 and U3 are largely restricted to the Near East (Sahakyan et al., 2017). Hg U2 is mostly common in South Asia where it shows the highest frequency and diversity. The overall frequency of U2 in India is largely contributed by the group U2i (U2i: U2a, U2b, U2c). Subclade U2d and U2e are confined to the Near East and Europe (**Table 8.2**). The Eurasian hg U2d and Indian Hg U2c are the sister clade of the common ancestor U2c'd defined by a common mutation C16234T. Whereas, the European variety of U2 is named as U2e.

In present study hg U was observed in Newar, Brahmin and Magar. It was found that two major branches U2 and U7 captures most of the U mitogenomes. Within clade U2, sub clade U2b and U2c were predominant in Newar and Brahmin. Whereas, Magar only contain U2c lineage. Hg U2a which is completely absent in Newar caste, is present in low frequency among the Brahmin and previously studied populations from Kathmandu and Morang, Nepal (Gayden et al., 2013; Wang et al., 2012). Hg U4 and U5 are present in Europe, Hg U6 is prevalent in the Circum-Mediterranean region, with frequency peaks in North Africa, U8 to the Near East and Europe and U9 is rare with only sporadic occurrence in Arabia, Ethiopia and India. Hg U7 is a west Eurasian specific Hg (Sahakyan et al., 2017).

PhyloTree of Hg U

<table>
<tr><td>U1</td><td></td><td></td><td></td></tr>
<tr><td>U5</td><td></td><td></td><td></td></tr>
<tr><td>U6</td><td></td><td></td><td></td></tr>
<tr><td rowspan="7">U2'3'4'7'8'9</td><td rowspan="3">U2</td><td>→ U2a, U2b, U2c</td><td>[U2I]</td></tr>
<tr><td>→ U2d</td><td></td></tr>
<tr><td>→ U2e</td><td></td></tr>
<tr><td>U3</td><td></td><td></td></tr>
<tr><td rowspan="2">U4"9</td><td>○ U4</td><td></td></tr>
<tr><td>○ U9</td><td></td></tr>
<tr><td>U7</td><td></td><td></td></tr>
<tr><td>U8</td><td></td><td></td></tr>
</table>

4.4.1.9.1 . Distribution and frequencies of Hg U in the studied populations.

1. **U2a:** Hg U2a is mainly found in Central Asia and South Asia (Pakistan, India). Hg U2a was also reported to be present in small frequency in Nepali population from Kathmandu and Eastern Nepal (Fornarino et al., 2009; Wang et al., 2012). In present study hg U2a were not observed in Newar and Magar, but present in very low frequency in Brahmin of Tanahun District, Nepal.
2. **U2b:** U2b is found in South Asia. Whereas sub hg U2b1 is also present in Myanmar, Thailand and Tibet (low frequency). In present study U2b1a was found in Newar mix and Maharjan in low frequency (**Figure 4.21**).
3. **U2b2**: In present study sub hg U2b2 was observed in Shrestha (8.3%) and Brahmin (4.16%). This hg was also found in Tamang (3.03%) and Kathmandu collection (2.59%) (Gayden et al., 2013).

4. **U2c:** In present study, Hg U2c were observed In Brahmin (10.4%). Whereas sub hg U2c1 were present in Shrestha (5.55%), Shakya (5.26%) and Magar (2.7%). HG U2c were also present in small frequency in Kathmandu and Hindu from Morang, Nepal (Fornarino et al., 2009; Gayden et al., 2013).

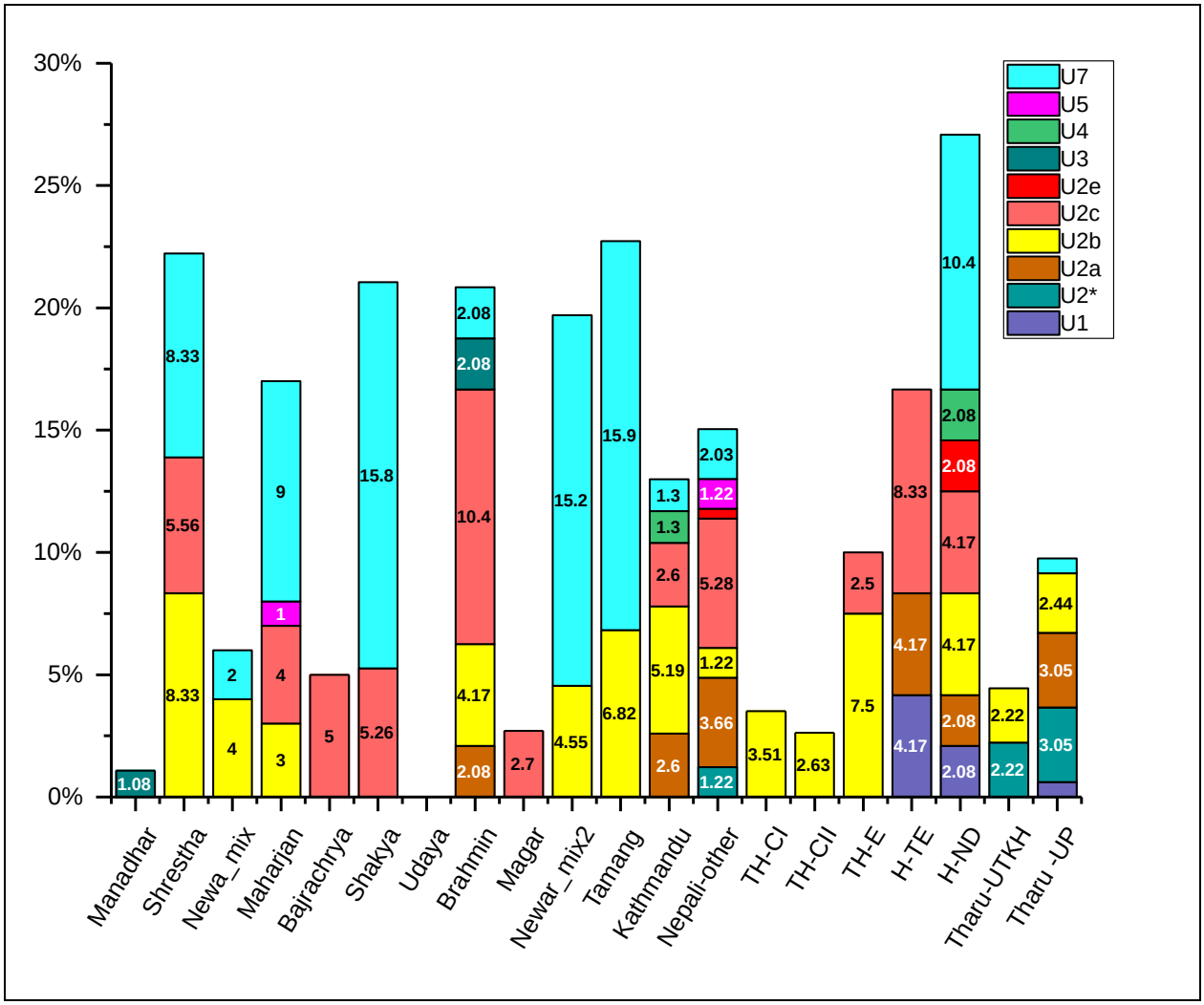


Figure 4.21 | mtDNA frequencies of Hg U in the Nepali populations. Sub lineage U2b, U2c and U7 were observed in majority of the Nepali populations.

Complete sequence analysis of Hg U2b1 shows Maharjan and Newamix share a basal variant (6620C, 10699A) with the sample from India. All the U2b1 lineages observed in Nepal (Maharjan and Newar mix), Tibet, Thailand and Myanmar share an ancestry with U2b1 observed in India.

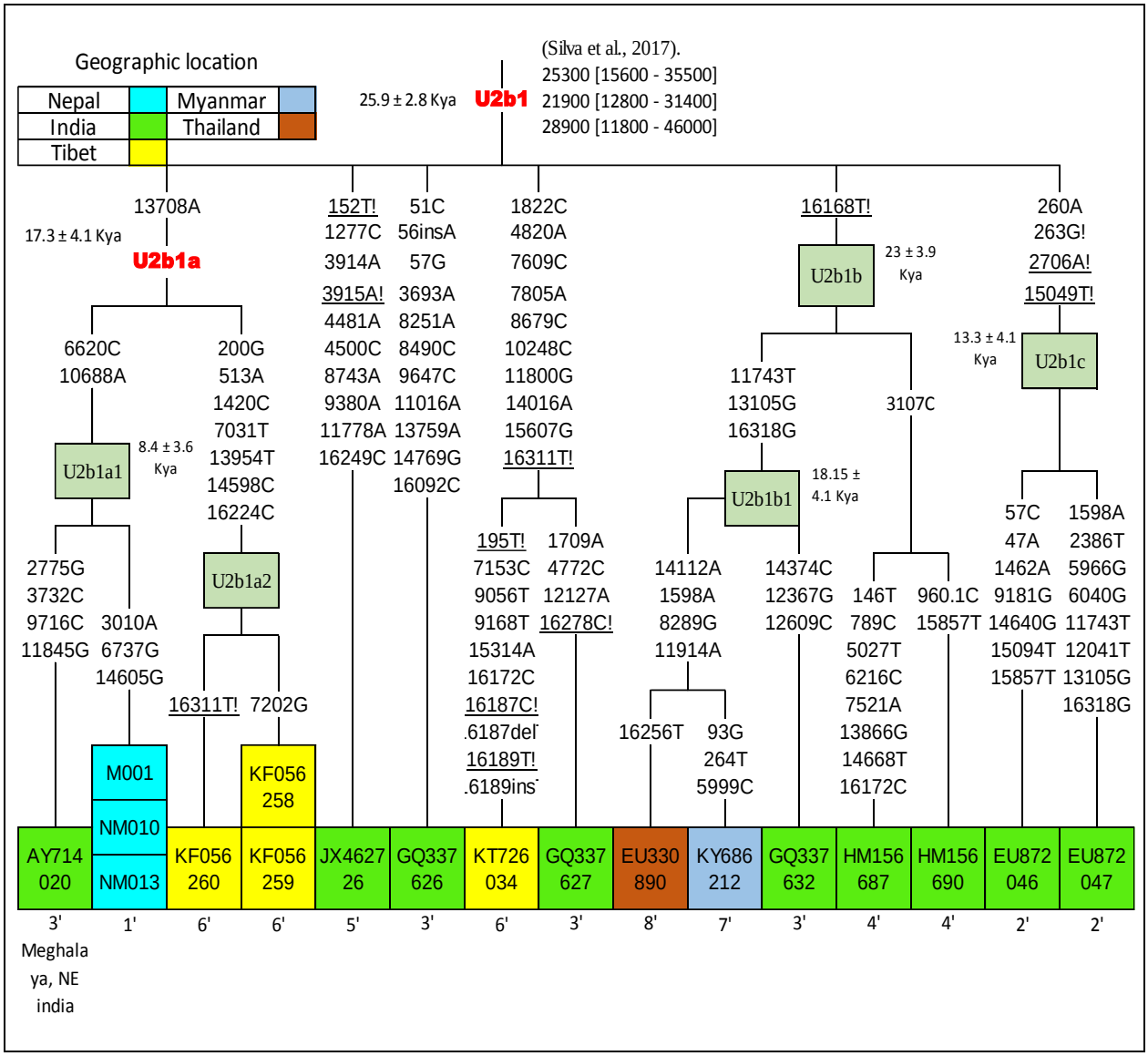


Figure 4.22 | Phylogenetic tree of Hg U2b1 based on complete mitochondrial sequences. Sub lineages U2b1a1, U2b1b and U2b1c that coalesce with U2b1 were newly defined in this study. Mutations relative to the RCRS are indicated on the branches. Sample ID/GenBank accession number for each sequence are shown in the box with geographic color code. The number (1', 2'... n') corresponds to the reference for the analyzed sample. The hg written in black color represents newly defined hg in this study, whereas red color indicates the predefined Hg in PhyloTree (Oven, 2015).

Similarly, complete mtDNA analysis of the samples belonging to the hg U2c1 was also performed as shown in **Figure 4.23**. Sub hg U2c1 shows huge diversity Among the Nepali populations. U2c1 might be one of the oldest lineage to be present in the geographic region occupied by modern day Nepal. Age of hg U2c1 was estimated to be ≈33.41 Kya, in a good agreement to the previous study which estimated the Age of hg U2c1 to be ≈28,800 years (Silva et al., 2017). In present study, new branch that coalesce with the U2c1 was identified and named as U2c1c. Further, four different clades were identified, hereby designated as U2c1c, U2c1c1, U2c1c1a and U2c1c1b. Complete sequence analysis revealed Nepali sample to cluster with Indian and Thai-Laos sample. Further Nepali sample share a close ancestry with the Thai-Laos sample as reflected by high sequence homogeneity of Thai-Laos sample with the Nepali sample compared to Indian sample. Interestingly, Brahmin of Tanahun shared several basal variants with the Maharjan of Kathmandu.

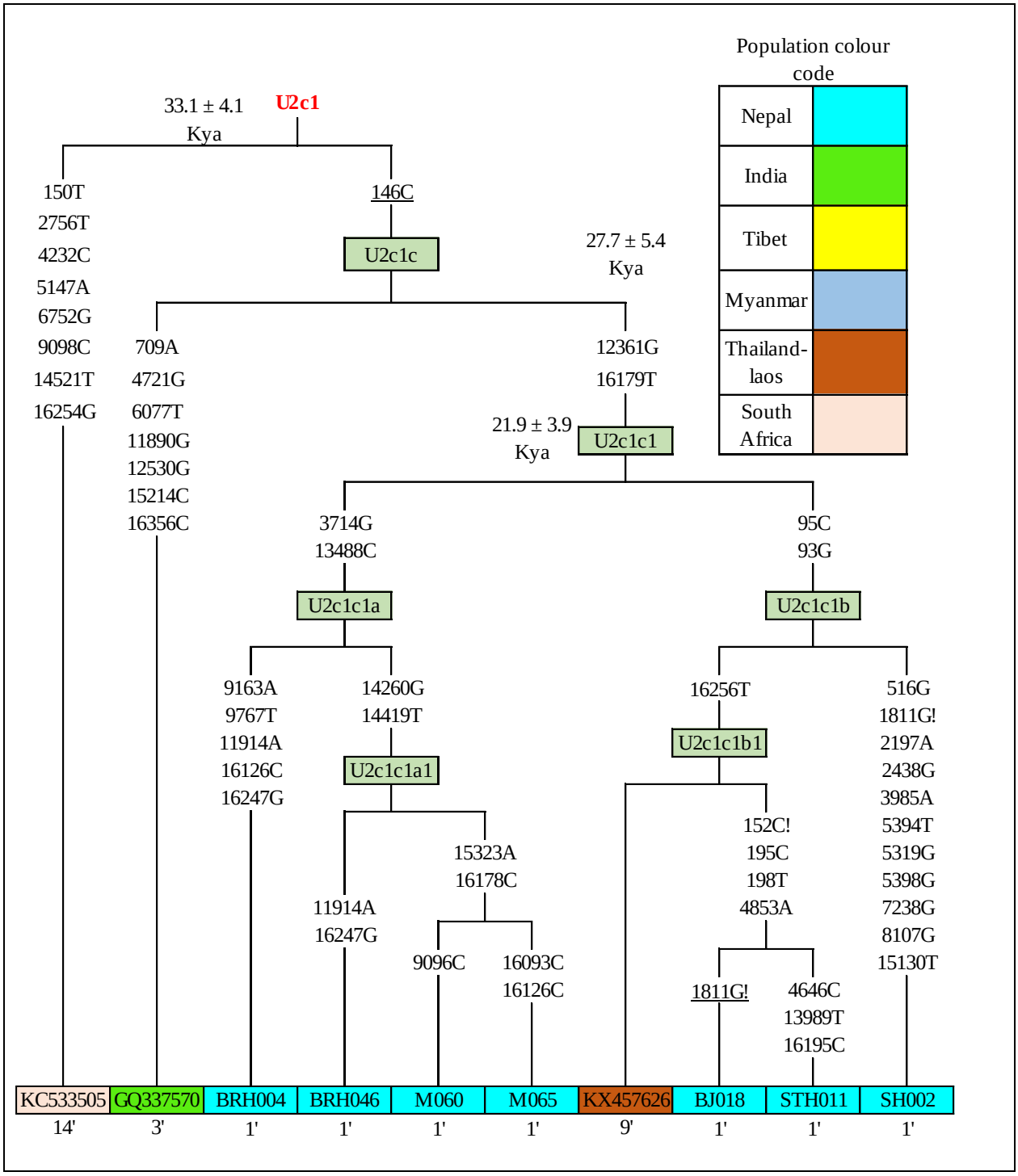


Figure 4.23 | Phylogenetic tree of Hg U2c1 based on complete mitochondrial sequences. Hg U2 is one of the oldest lineage to inhabit South Asia. Mutations relative to the RCRS are indicated on the branches. Sample ID/GenBank accession number for each sequence are shown in the box with geographic color code. The number (1', 2'... n') corresponds to the reference for the analyzed sample. The hg written in black color represents newly defined hg in this study, whereas red color indicates the predefined Hg in PhyloTree (Oven, 2015).

5. Haplogroup U5

Hg U5 is found throughout Europe. Surprisingly only a Single Maharjan sample belonging to subclade U5a1b1f were observed in present study. Previously, subclade U5a1 was observed in low frequency in Kathmandu and Morang (Wang et al., 2012).

4.4.1.9.2 Haplogroup U7

Hg U7 is the major west Eurasian lineage found in Newar populations. Frequency of hg U7 was higher in Shakya (15.78%) followed by Maharjan (10%), Shrestha (8.33%), Newa-mix (2.08%) and Brahmin (2.08%). Surprisingly, Hg U7 was completely absent in Udaya and Magar. Previous study done in Nepali population shows high incidence of U7 among Tamang (15.6%) and completely absent among Tharu population (Fornarino et al., 2009; Gayden et al., 2013). Hg U7 is detected at high frequencies in populations throughout Iran, Pakistan, Northwest India and the Arabian Peninsula (Quintana-Murci et al., 2004). U7a, U7a1a and U7a3 are the major U7 clades present in Indian populations and these lineages have wide spread distribution in India (Palanichamy et al., 2015).

Phylogenetic analysis: Intriguingly, the **U7a3a** lineage observed in Newar population of Nepal were found to cluster exclusively with the Afghanistan sample. Whole sequence analysis shows sharing of three basal variants (5291C-9266A-16223T) between Newar and Afghan sample (Pashtun people). This has led in to the identification of new sub lineage, named herein as U7a3a2. Further, within the Nepali samples; Shrestha, Maharjan and Shakya share two more variants (98545A, 15364T) to form a distinct and possibly a Nepal specific subclade named herein as U7a3a2a. The homogeneity (less genetic diversity) between the Shrestha, Shakya and Maharjan may indicate that the population has been descended from a small number of colonizing ancestors (**Founder effect**). Further the variant 9099T, is shared in between two Shakya samples. Absence of variant 9098C in other sub caste of Newar, shows the population expansion has occurred within the Shakya caste. This can be explained as a result of introduction of caste system the variant is limited within the Shakya clan (Endogamy). Sub Hg U7a3a1 were not found in Nepali population. Only mtDNA D loop sequence of Tamang were available from previous study (Gayden et al., 2013), of which all D-loop variants of Tamang fit to Hg U7a except one variant (16298C) which were not observed in Newar samples. This difference in the lineage between these two populations might suggest expansion of specific lineage within the restricted geographical region.

Sub Hg U7a1a has a wide distribution throughout India and are observed mainly in the Brahmin and Muslim population. Indian U7a1a sequences cluster with the Pakistani and Iran-Persian samples (M. Derenko et al., 2013). Sub Hg U7a1a was found in low frequency (2.78%) among the Shrestha population of Nepal. Complete sequence analysis shows Shrestha sample to cluster with Brahmin from North India (Jammu-Kashmir and Uttar Pradesh) and West Bengal. Shrestha sample share two variants (16172C and 16309G!) with the Indian samples to form a new sub clade name herein as U7a1a1a.

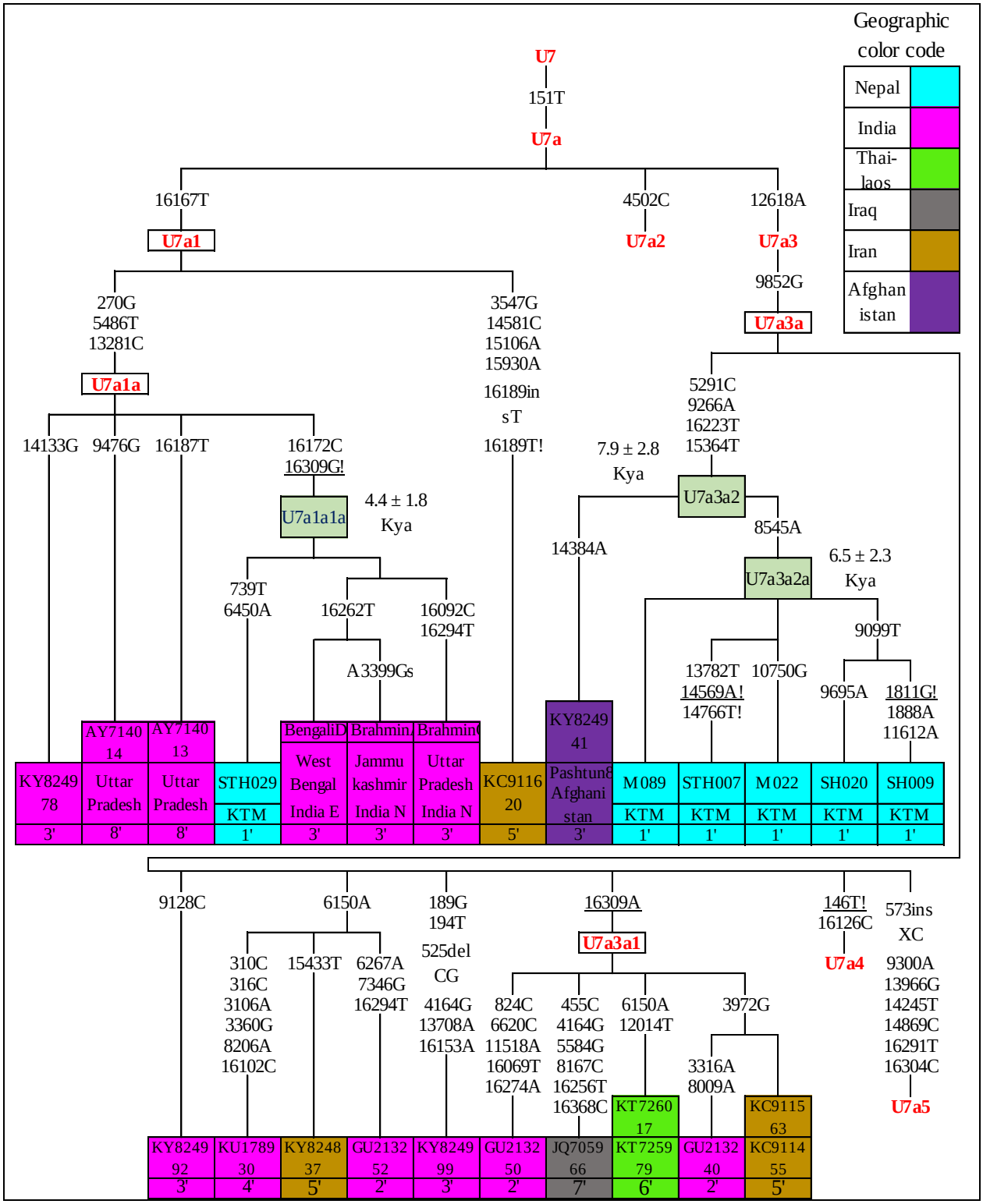


Figure 4.24 | Phylogenetic tree of Hg U7a1 and U7a3a based on 24 complete mitochondrial sequences of which 6 complete mtDNA sequences were generated in this study. U7a3a lineage of Nepal share an ancestry with the Pashtun people from Afghanistan. Pashtun people are predominantly Eastern Iranian people and lives in Pakistan and Afghanistan. Sample ID/GenBank accession number for each sequence are shown in the box with geographic color code. The number (1', 2'... n') corresponds to the reference

for the analyzed sample. The hg written in black color represents newly defined hg in this study, whereas red color indicates the predefined Hg in PhyloTree (Oven, 2015).

Sub Hg U7b was found in very low frequency in Shrestha (2.78%) and Maharjan (1%). Complete sequences for this sub lineage were available from Italy, Croatia, Russia, Israel, and India.

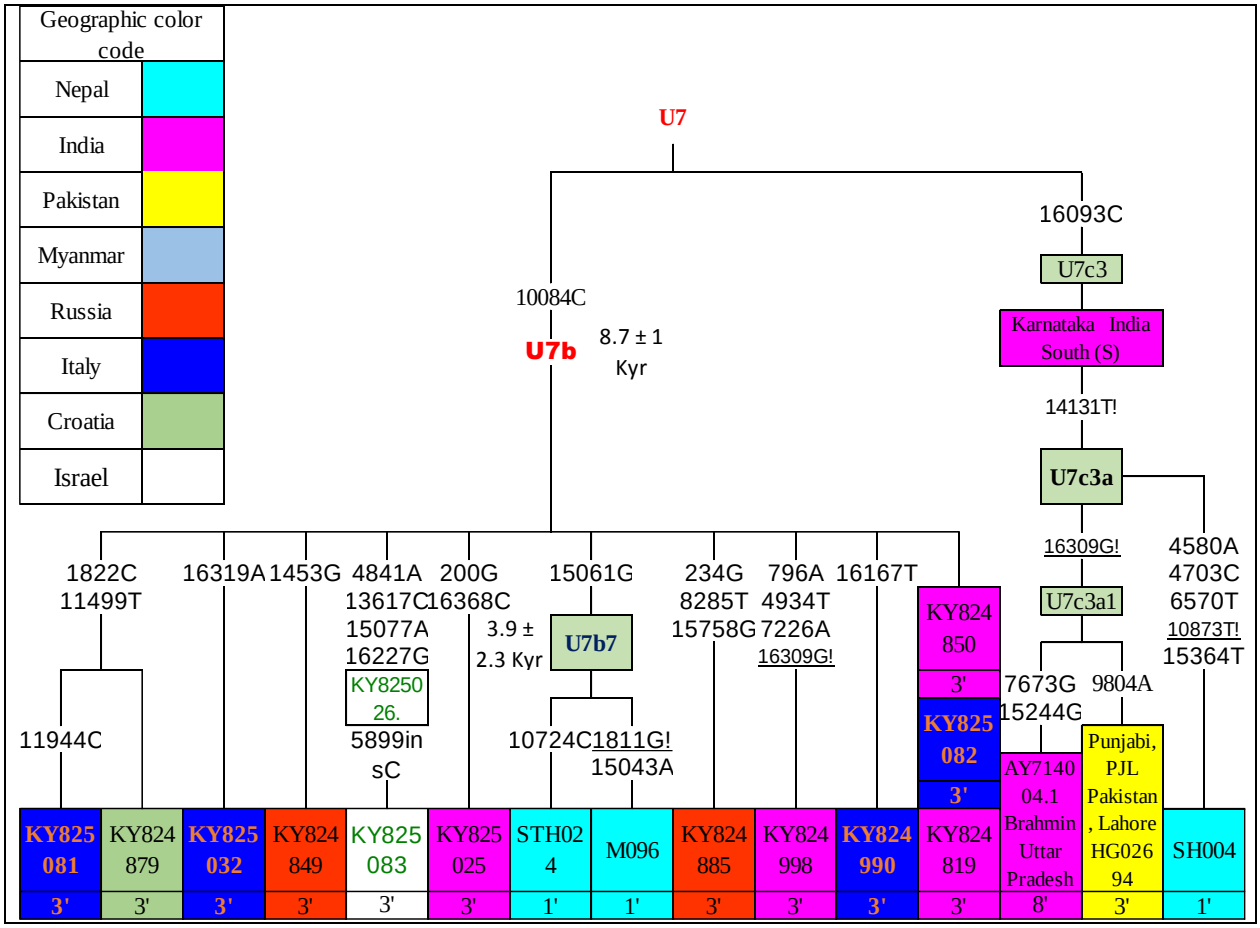


Figure 4.25 | Phylogenetic tree of Hg U7b based on 20 complete mitochondrial sequences of which 3 complete mtDNA sequences were generated in this study. Age of hg U7b is between 5,400 and 15,600 years (Behar et al., 2012b).

One sample belonging to Shakya was found to cluster with the sample from Karnataka (south India), Brahmin of Uttar Pradesh and Punjabi of Lahore (Pakistan). New subclade U7c3 was named accordingly to the recent study (Sahakyan et al., 2017). Further, 14131T variation was found to be present in Uttar Pradesh and Nepali sample were as it was absent in Punjabi sample.

4.4.2 Haplogroup N1a1b

Hg N1a is present in West Eurasia. N1a has also been observed in central Asia and Southern Siberia. Hg N1a was found in a 2,500-year-old Scytho-Siberian burial in the Altai region (Miroslava Derenko et al., 2007). Hg I1e, and N1a1b1 were observed in the present study.

4.4.2.1 Haplogroup I

Hg" I" is a sister clade of Hg N1a1b1. Hg "I" is present across West Asia and Central Asia, and is also found at traces frequencies in South Asia.

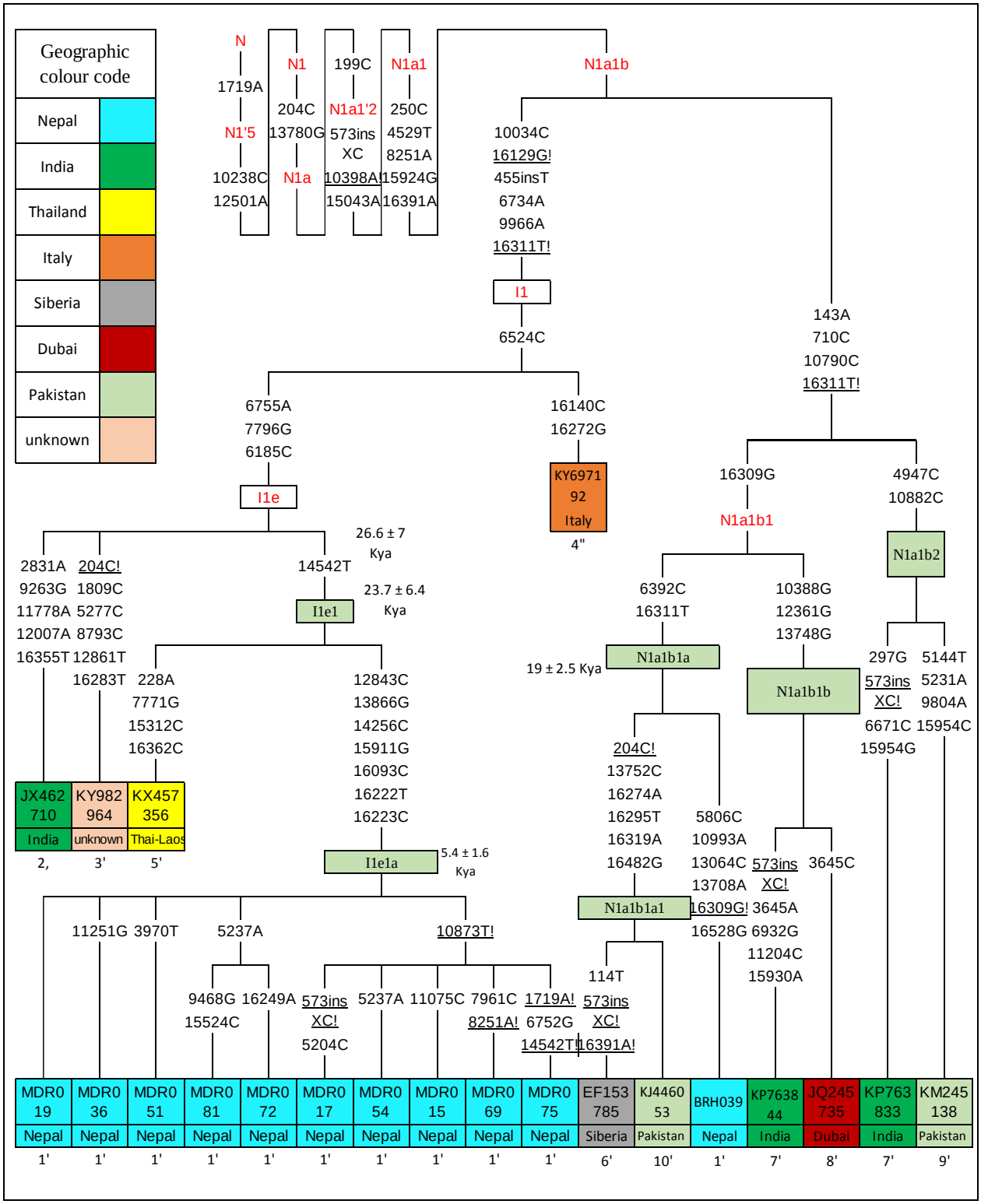


Figure 4.26 | Phylogenetic tree of Hg N1a1b based on 22 complete mitochondrial sequences. Mutations relative to the RCRS are indicated on the branches. Sample ID/GenBank accession number of each sequence are shown in the box with geographic color code. The number (1', 2'... n') corresponds to the reference for the analyzed sample. The hg written in black color represents newly defined hg in this study, whereas red color indicates the predefined Hg in PhyloTree (Oven, 2015).

After the first initial settlement ≈45-65 Kyr ago the earliest movement into the South Asian population occurred from West Eurasia. Genetic evidence for this movement is provided by the

Hg N1a1b (Silva et al., 2017). Hg "I" has been previously reported to be present in Nepali population; Hindu from Morang (4.2%) (Fornarino et al., 2009) and Nepali-other (individual from Kathmandu and Eastern Nepal) (2%)(Wang et al., 2012). It is also present at trace frequencies among the Kshatriya (1.28%) and Arya (2.1%) of Uttarakhand, India (Negi et al., 2016). Hg I was reported to be absent in a previous study done on Newar population (Gayden et al., 2013). In present study, Hg I4e was detected in high frequency (10.75%) exclusively only in Manandhar.

Manandhar sample belonging to Sub Hg I1e shares an ancestry with the population from Thailand and India. Further, Manandhar is relatively closer with Thailand, in which one basal variant (14542T) is shared between Manandhar and Thailand sample (GenBank: KX457356). In the present study, several sub Hgs were newly named as shown in the phylogenetic tree (**Figure 4.26**).

4.4.2.2 Haplogroup N1a1b1

Hg **N1a1b1** has been observed exclusively in Brahmin (4.16%). Previously, **N1a1b1** was reported to be present in Kathmandu (2.6%) (Gayden et al., 2013). Sub Hg **N1** was also present at trace frequency (0.8%) among the individual from Kathmandu and Eastern Nepal (Wang et al., 2012). This Hg is also detected in India, but in very low frequency.

Common ancestor of Lineage N1a1b1a present in Nepal (Brahmin), Siberia and Pakistan dates to ≈19 Kyr (**Figure 4.26**). In addition, Nepali samples shared two variants with the Siberia and Pakistani samples, indicating a deep ancestral relationship.

4.4.3 Haplogroup W

Hg W is predominant in the eastern half of Europe, in the North Caucasus, in Central Asia, in Iran and in India. Though Hg W has wide distribution throughout North-West of the Indian subcontinent it is rarely found in the East Indians. Sub-Hgs of W such as W1c, W1g, W3, W3a1, W3a1b, W3a2, W4 and W4a are observed in the Indian populations (Palanichamy et al., 2015).

Sub Hg W3a1b is found specifically in India and Pakistan (Quintana-Murci et al. 2004; Rakha et al. 2011). Previous study showed presence of Hg "W" only in Sherpa population and was completely absent in other Nepali populations including Newar. In contrary to previous study, W3a1b lineage has been observed in Newar in small frequency**;** in Manandhar (1%), Shrestha (2.2%) and Udaya (3.4%). Complete mtDNA sequence analysis shows W3a1b of Manandhar and Shrestha are different than those of Udaya as reflected by their distinct clusters in the phylogenetic tree (**Figure 4.27**). In present study, Individuals from Manandhar and Shrestha cluster with W3a1b lineage of India and Pakistan, in which single basal variant is shared further between Nepali and Indian sample to form a new sub clade. Whereas, W3a1b of Udaya share ancestry with Thailand-Laos and India, in which one basal variant (225A) is shared between these populations to form a new subclade. Further one more basal variant (11194G) is shared between India and Thailand.

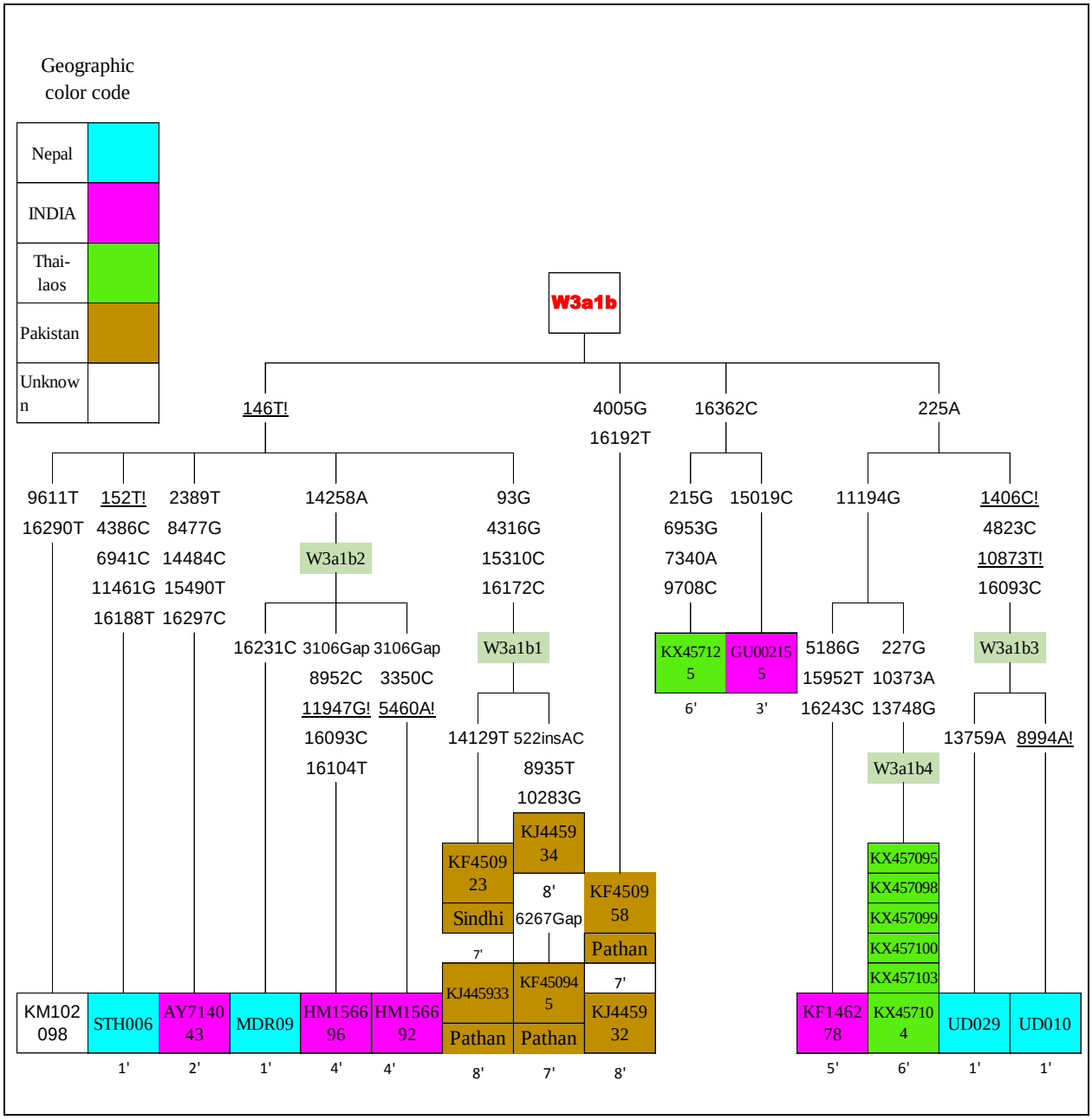


Figure 4.27 | Phylogenetic tree of Hg W3a1b based on 23 complete mitochondrial sequences. Mutations relative to the RCRS are indicated on the branches. Sample ID/GenBank accession number for each sequence are shown in the box with geographic color code. The number (1', 2'... n') corresponds to the reference for the analyzed sample. The hg written in black color represents newly defined hg in this study, whereas red color indicates the predefined Hg in PhyloTree (Oven, 2015).

4.4.4 Haplogroup A

Hg A is a Northeast Asian specific Hg. Highest frequency of this Hg is observed in the indigenous people of the Americas. Sub Hg A2 is one of the five mtDNAs found among the indigenous people of America, the other being B, C, D and X. lineage belonging to A2 is also found among the several populations from Siberia. Its greatest variety and overall population frequency is observed among the East Asians **(Table 8.2)**. Thus, This Hg is believed to be originated and spread from Far East (Fagundes et al., 2008). Frequency of hg "A" observed in Nepali and Indian population are shown in the **Table 4.6**.

Population	Location	Frequency (%)	Population	Location	Frequency (%)
Sherpa	Nepal	24.0	Adi	North-East India	26.7
BJR	Nepal	10.0	Apatani	North-East India	19.0
N-MIX2	Nepal	7.6	Nishi	North-East India	18.2
MDR	Nepal	7.5	Shah	North-India	17.1
Udaya	Nepal	6.9	Changpa	North-India	16.1
Nepali-other	Nepal	6.1	Bhoi	North-East India	13.8
N-MIX	Nepal	6.0	Tamta	North-India	11.8
Kathmandu	Nepal	5.2	Tipperah	North-East India	10.0
Tamang	Nepal	4.5	Goswami	North-India	8.7
STH	Nepal	2.8	Ladhak	North-India	8.3
BR-NP	Nepal	2.1	Arya	North-India	4.3
Maharjan	Nepal	1.0	Kshatriya	North-India	3.8
SHK	Nepal	0.0	Apatani	North-East India	3.8
MGR	Nepal	0.0	Brahmin-UTK	North-India	2.4
H-TE	Nepal	0.0	Tharu-UTKH	North-India	2.2
TH-CI	Nepal	0.0	Lyngnga	North-East India	1.4
TH-CII	Nepal	0.0	Khyn	North-East India	1.2
TH-E	Nepal	0.0	Tharu -UP	North-India	0.6

Table 4.6 | Frequency of Hg A in Nepali and Indian populations. Highest frequency of hg A was observed among the Sherpa (24%) and Tibeto-Burman speaking population from North-East India. Surprisingly Indo-European speaking population Shah from Uttarakhand (India) shows higher frequency for hg A (17.4%). The background color gradient in green, white and red represent high, moderate and low frequency respectively.

In present study, majority of lineage "A" of Newar were unable to assign into predefined sub hg. Each sample branch off form the node A+152+16362, from which also arise hg A1, A2, A6, A15, A16, A17, A18, A19, A20, A21, A22, A23, A24, A25 and A26. Two sample from Manandhar fall under A17 hg, whereas the other sample branch off from A+152+16362 separately and do not cluster with any of these predefined hg. New Hg A27 and their sub clades (A27a and A27b) were newly defined for these samples. As shown in the phylogenetic tree, all the samples belonging to newly defined hg A27 shared similar basal haplotype. This homogeneity in the sequence between the Maharjan, Manandhar, Shakya, Shrestha, Udaya and Newa_mix indicate that the population has been descended from a small number of colonizing ancestors (Founder effect).

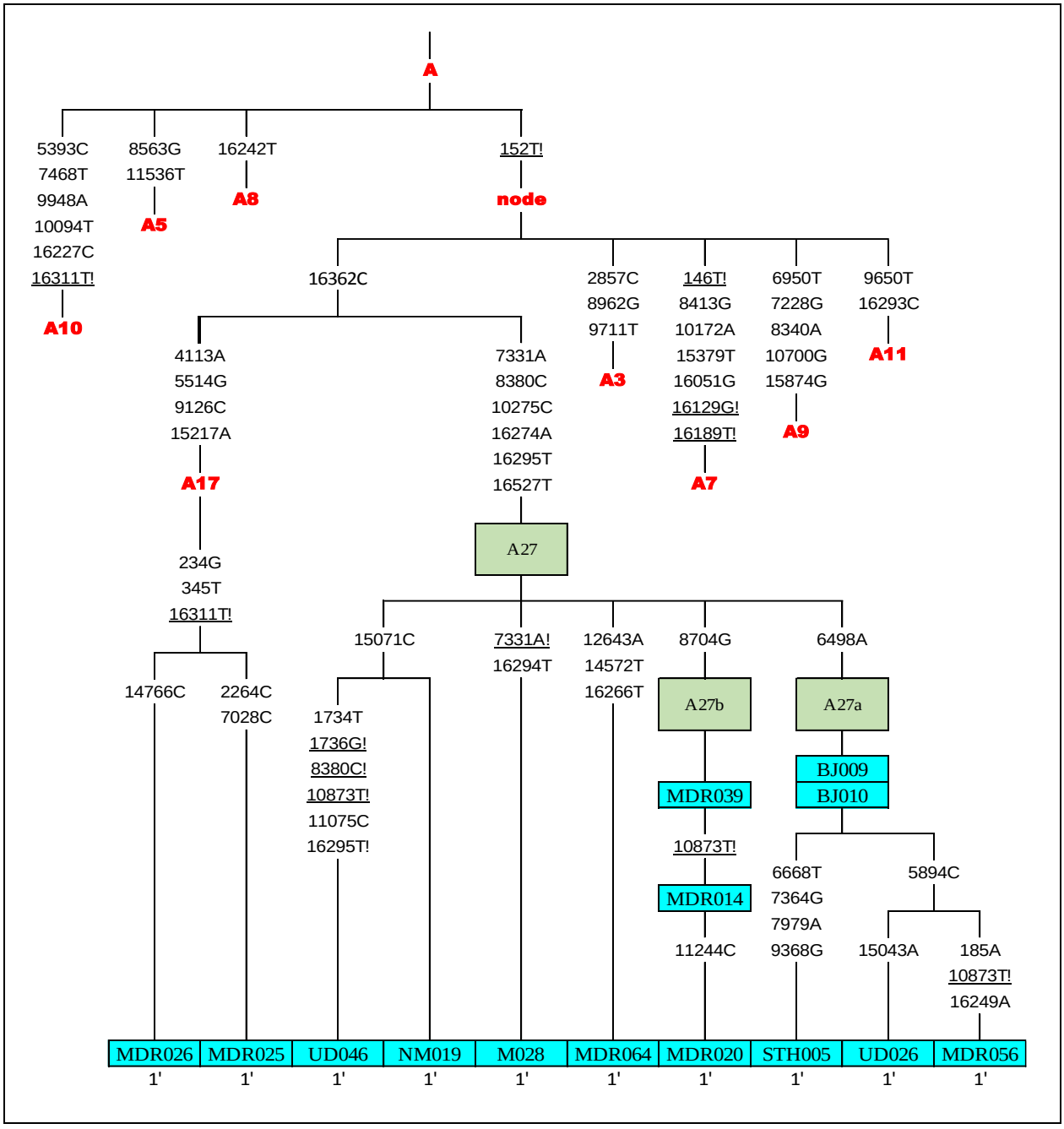


Figure 4.28 | Phylogenetic tree of Hg A based on 14 complete mitochondrial sequences. Mutations relative to the RCRS are indicated on the branches. Sample ID/GenBank accession number for each sequence are shown in the box with geographic color code. The number (1', 2'... n') corresponds to the reference for the analyzed sample. The hg written in black color represents newly defined hg in this study, whereas red color indicates the predefined Hg in PhyloTree (Oven, 2015).

4.5 Phylogeographic pattern of the dominant Hgs within the Macro Hg M.

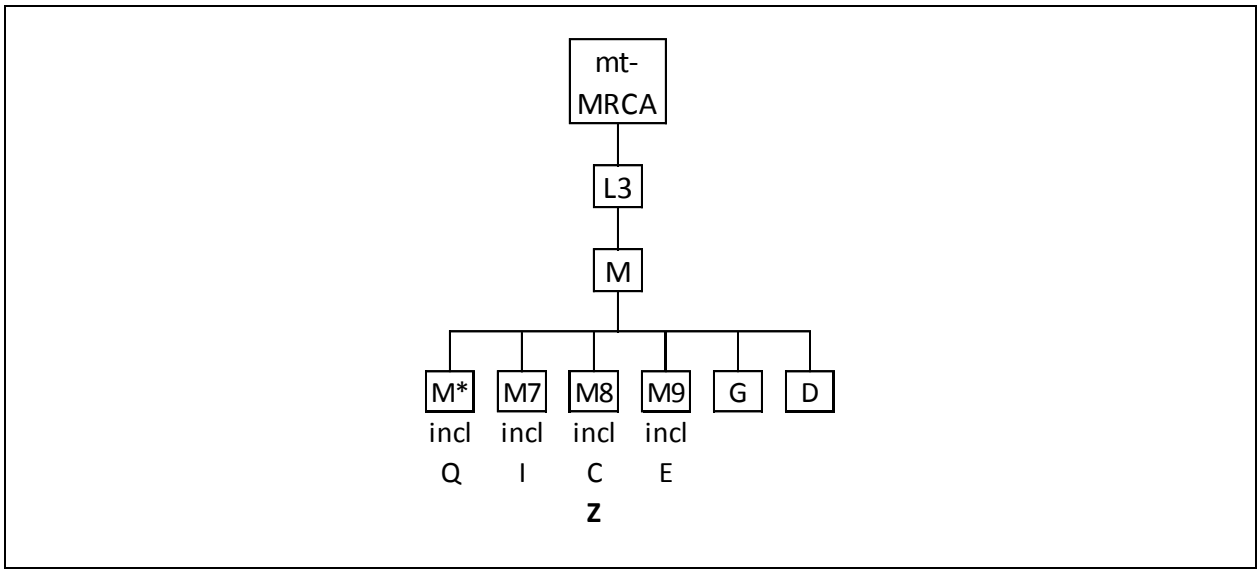


Figure 4.29 | Phylogenetic tree of mtDNA Super Hg M. For convenient, tree of Macro Hg M is divided into 6 subtrees (Phylotree.org).

4.5.1 Haplogroup M2

Hg M2 is the oldest lineage on the Indian subcontinent (Kumar et al., 2008; Rajkumar, Banerjee, Gunturi, Trivedi, & Kashyap, 2005) believed to be originated in South west India (Silva et al., 2017). It is most common among the population from Uttar Pradesh such as Musahar (13.33%), Koli (8.33%), Harijan (6.89%) and South Indian such as Andh (13.63%), Naikpod (11.62%) and Pardhan (9.35%) (Table 8.2 & Table 8.1). In the present study, sample each belonging to Bajracharya, Shakya and Udaya are allocated to sub Hg M2a3a. Similarly, one sample from Maharjan belong to M2a1a+207 sub Hg. Previous study showed presence of hg M2 in Newar (3.03%) and Nepali-other (0.4%) group (Table 8.2 & Table 8.1).

When Nepali and Indian M2a3a lineage were compared, one basal polymorphism (11827C) was found to be shared in between two Indian sample and two Nepali sample. Subsequently two basal polymorphisms (9182A, 16227G) was shared between two Nepali samples to form a new subclade M2a3a4. In present study Hg M2a3a1, M2a3a3, M2a3a3 and M2a3a4 where newly named as shown in the PhyloTree (**Figure 4.30**). The shared basal variants between Nepal and Indian M2a3a lineage provides a direct genetic evidence to the long ancient link between India and Nepal.

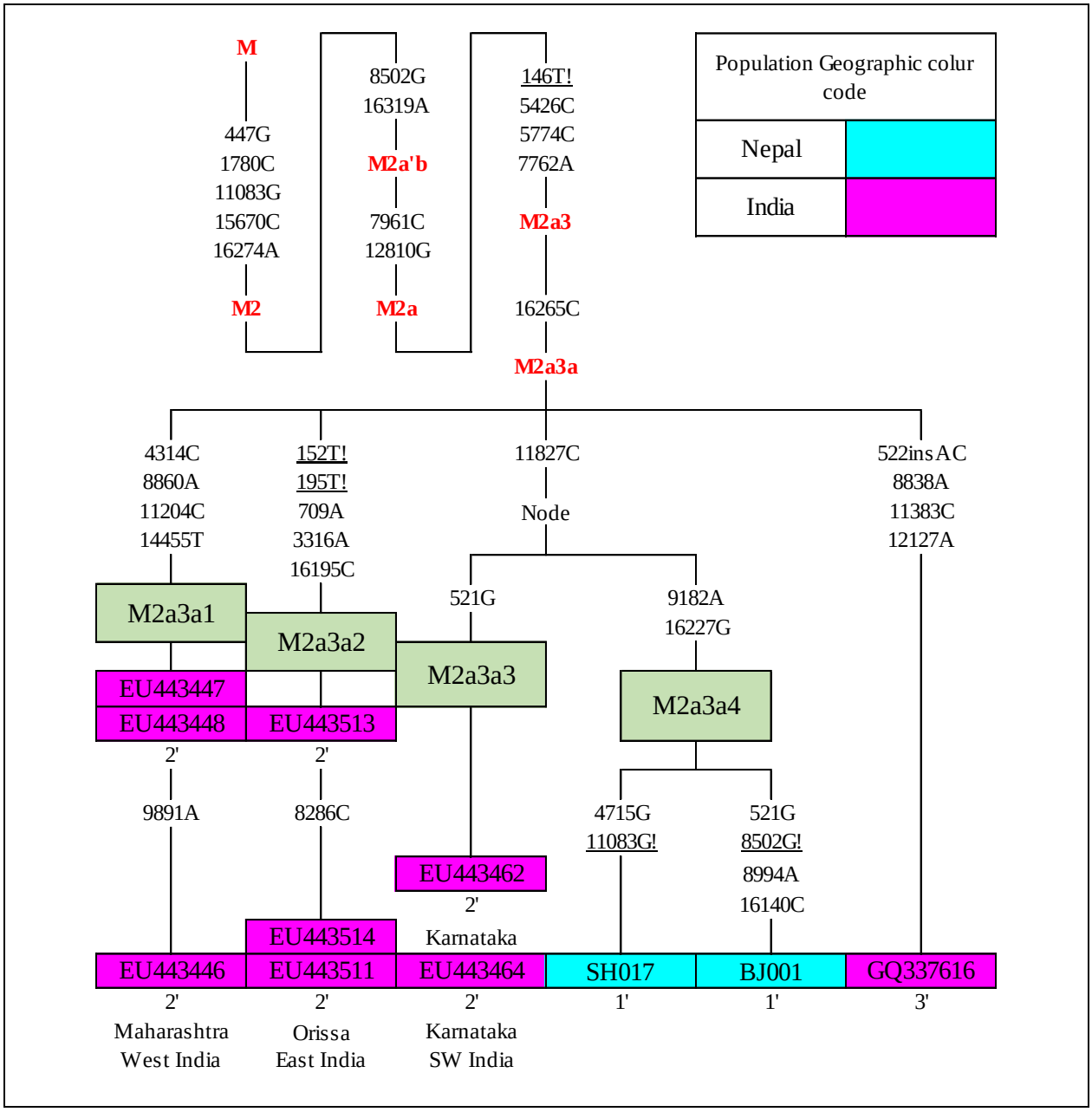


Figure 4.30 | Phylogenetic tree of Hg M2a3a based on 11 complete mitochondrial sequences. Mutations relative to the RCRS are indicated on the branches. Sample ID/GenBank accession number for each sequence are shown in the box with geographic color code. The number (1', 2'... n') corresponds to the reference for the analyzed sample. The hg written in black color represents newly defined hg in this study, whereas red color indicates the predefined Hg in PhyloTree (Oven, 2015). M2a3a1, M2a3a2, M2a3a3 and M2a3a4 are the newly defined sub lineages in this study.

4.5.2 Haplogroup M3

Hg M3 is found mainly in South Asia with highest concentration observed in West and North-West India. In present study, Hg M3 was observed among Newar, with highest frequency in **Shakya (10.52%),** Shrestha (5.55%) and Udaya (5.17%). Similarly, this Hg were also detected in **Brahmin (8.33%)** and Magar (2.7%). According to previous study, Hg M3 were also found among **Hindu-Terai (8.33%**), Kathmandu (5.19%) and Nepal-other (3.6%) population.

Single sample of M3c2a of Nepal clustered with the individual from India and Pakistan.

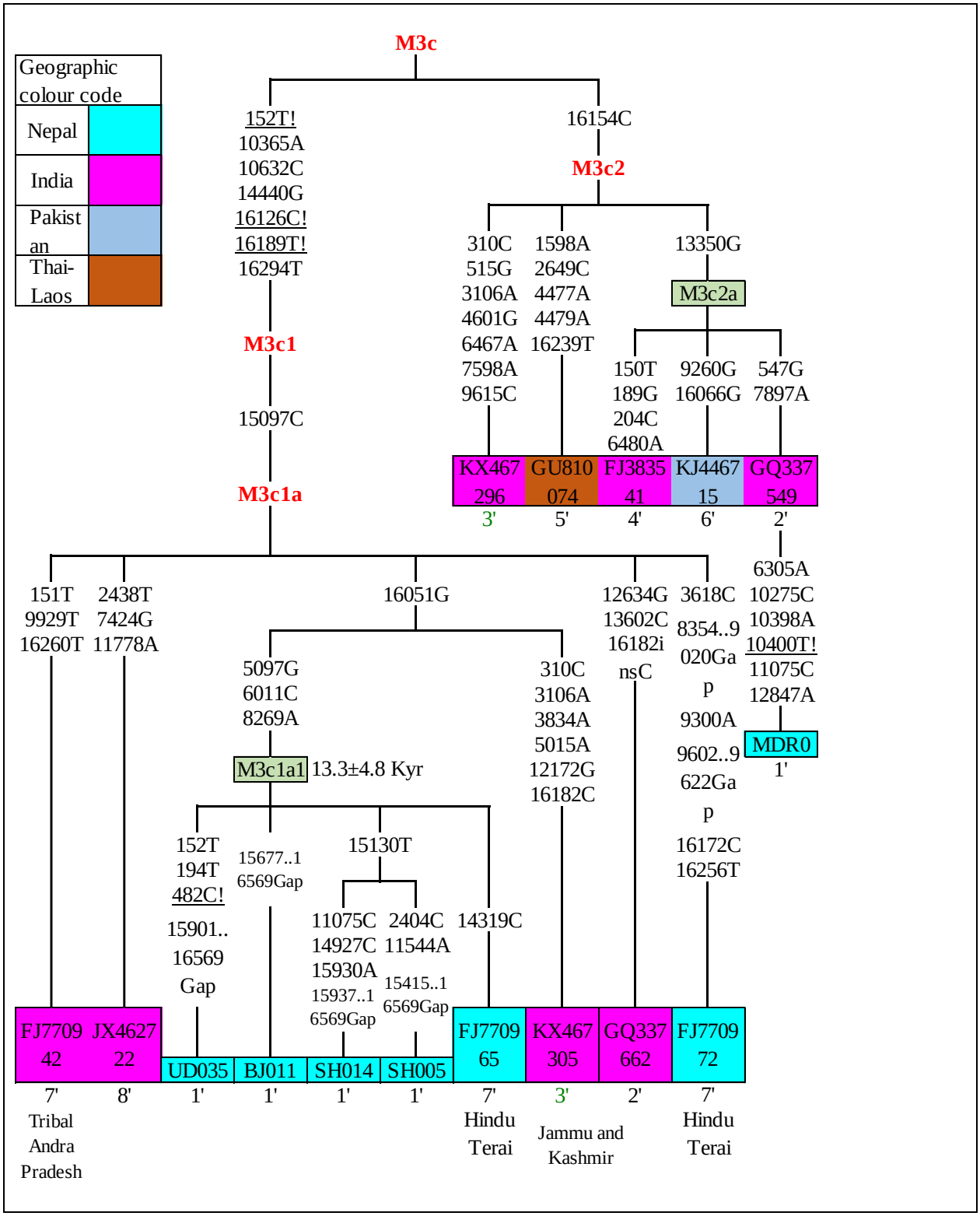


Figure 4.31 | Phylogenetic tree of Hg M3c based on 16 complete mitochondrial sequences. Mutations relative to the RCRS are indicated on the branches. Sample ID/GenBank accession number for each sequences are shown in the box with geographic color code. The number (1', 2'... n') corresponds to the

reference for the analyzed sample. The hg written in black color represents newly defined hg, whereas red color indicates the predefined Hg in PhyloTree.

Sample belonging to Udaya, Bajracharya, Shakya and Hindu-Terai share a basal variant (16051G) with the sample from Jammu and Kashmir, India.

Complete mtDNA analysis of hg M3d shows its recent arrival via migration into Nepal. Bearer of hg M3d1a might have entered recently into Nepal via migration as revealed by the complete mtDNA (**Figure 4.32**).

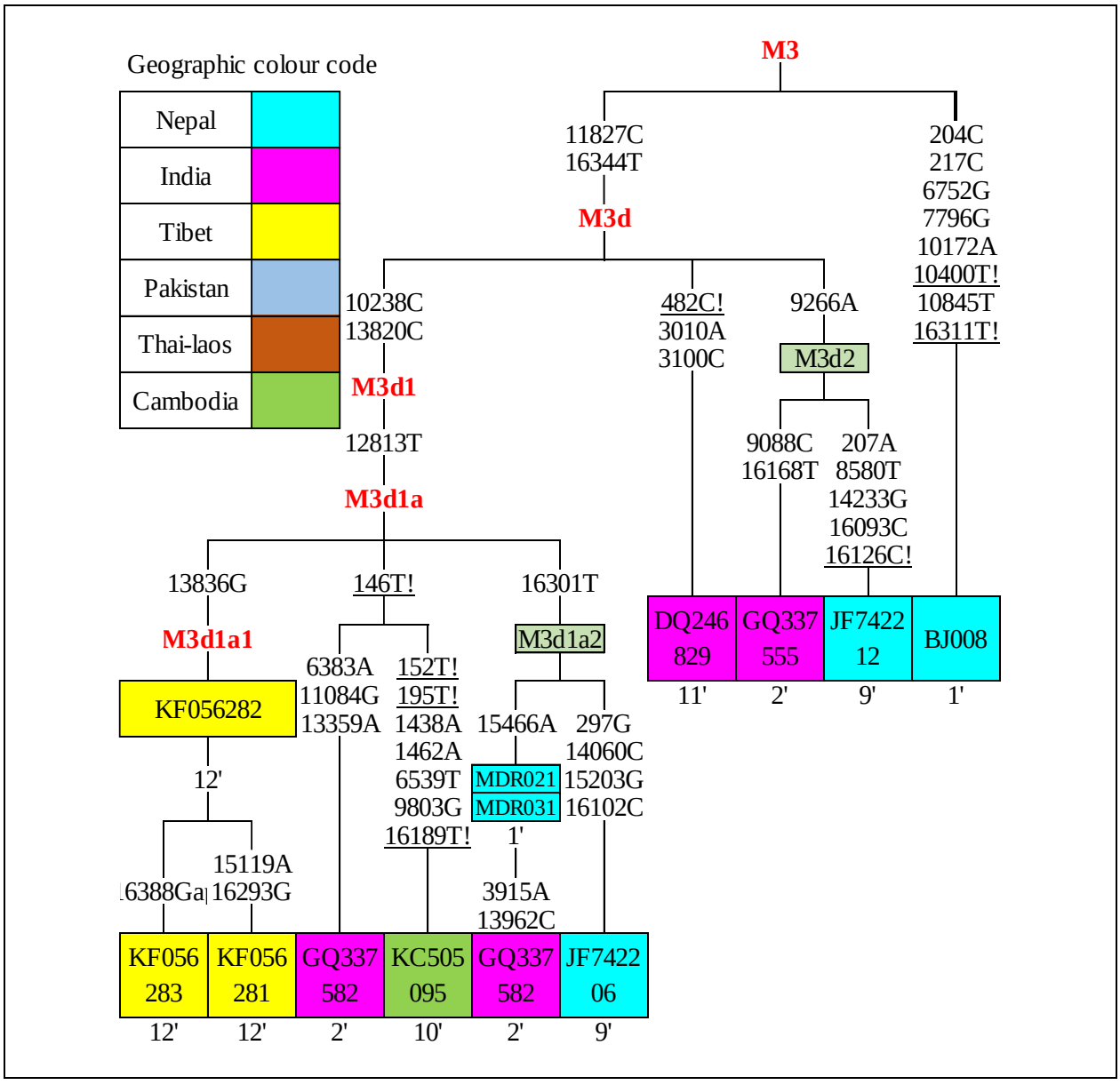


Figure 4.32 | Phylogenetic tree of Hg M3d based on 12 complete mitochondrial sequences. One sample belonging to Bajracharya was unable to locate to the predefined branch of hg M3.

4.5.3 Haplogroup M4"67

Hg M4"67 is assumed to be originated in central India (Silva et al., 2017). This hg shows an extraordinary branching which extends to the wider geographic regions of South Asia. The age of

these lineage is ≈45-35 Kyr (Silva et al., 2017). Few sample from Brahmin (2.083%) in this study and Tamang (4.5%) from previous study was unable to assign to the already defined Hgs belonging to super Hg M4"67.

Super Hg M4”67 encompasses the following Hgs as shown below

• M4'65'67:	❖ M4 M65 M67		• M43
• M18"38	❖ M18 M38		• M45 M63
• M30			• M64
• M37			• M66

Hg M65b was present exclusively only among the Brahmin population in a very low frequency (2.08%). Hg M65 was found in higher frequencies among the population from Uttarakhand, India (Negi et al., 2016). Hg M18 is present in Magar (2.70%) and Shrestha (2.77%). Previously reported to be found in the individual from Kathmandu, East Nepal (Wang et al., 2012) and Tharu population from Chitwan (Fornarino et al., 2009). Hg M38 is present in India, Myanmar and Nepal.

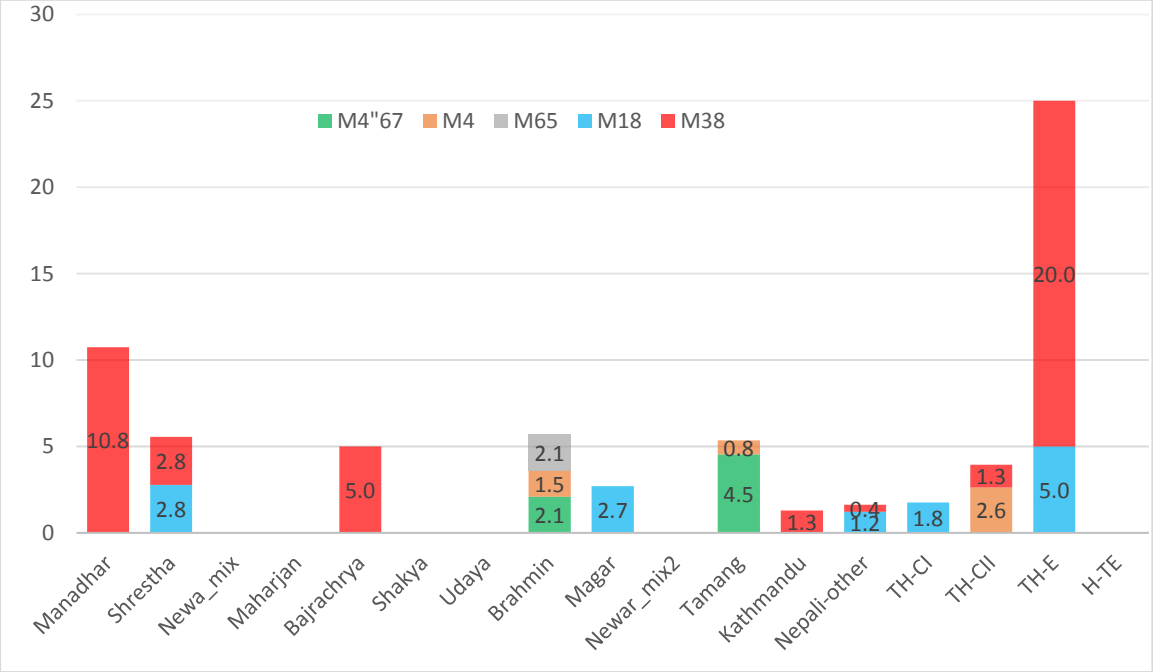


Figure 4.33 | frequencies of Hg M4”67 observed in the Nepali populations. Hg M38 observed in high frequencies in Hindu Terai is also observed in Manandhar, Bajracharya and Shrestha. Whereas this Hg was completely absent in other Newar populations. M30 was also absent in Magar and Brahmin group

4.5.3.1 Haplogroup M38

Hg M38 is one of the most common lineage in Manandhar caste (10.75%) and detected in low frequency in Shrestha (2.78%). As reported by Previous studies, this Hg was also detected in high frequency among Tharu (20%) from Morang District of Southeastern Nepal, Tibeto-Burman speaking Naga people (16.27%) from Nagaland (North East India), Tibeto-Burman speaking people from Rakhine state (4.7%) and Magway state (4.16%) of Myanmar **(Table 8.2).**

One sample from Myanmar share a variant (16111T) with the Indian sample. Similarly sample from India and Nepal belonging to M38a clade, share three basal variants (77149G, 15940C, 16319A) to form a new sub Hg named herein as M38a1. Several basal variants are shared within the different caste of Newar to form sub Hgs newly named as M38d1 and M32d2.

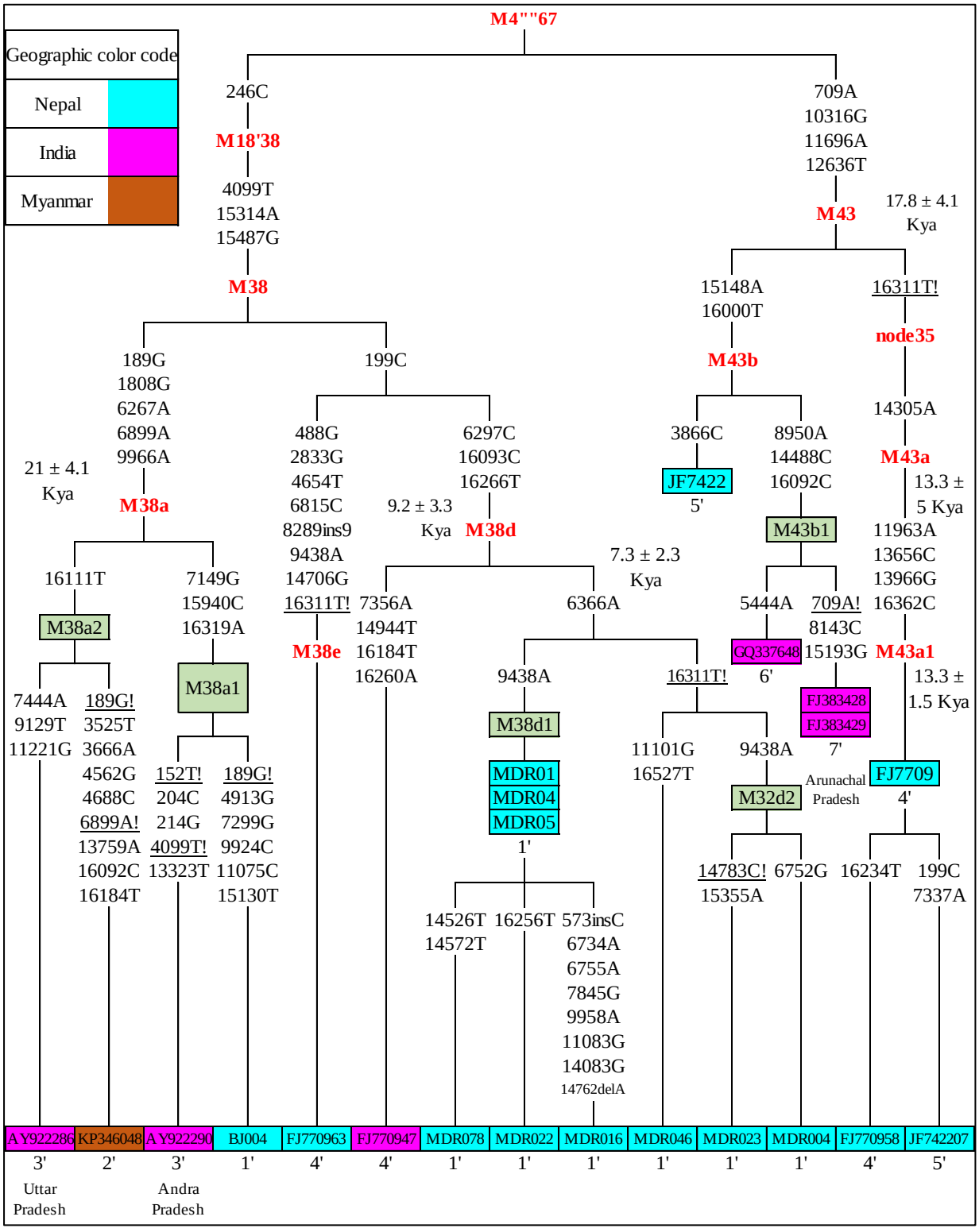


Figure 4.34| Phylogenetic tree of Hg M38 & M43 based on 22 complete mitochondrial sequences. M38a2, M38a1, M38d1, M38d2 and M43b1 are the newly named sub Hgs in this study. Mutations relative to the RCRS are indicated on the branches. Sample ID/GenBank accession number for each sequence are

shown in the box with geographic color code. The number (1', 2'... n') corresponds to the reference for the analyzed sample. The hg written in black color represents newly defined hg in this study, whereas red color indicates the predefined Hg in PhyloTree (Oven, 2015).

4.5.3.2 Haplogroup M43b

Hg M43b was present in low frequency in Shrestha (2.77%) and Magar (2.27%). It is one of the most common lineage among the Tharu of Chitwan (TH-CI: 10%, TH-CII: 9.21%), whereas detected in low frequency among the several other Nepali populations. In present study, whole mtDNA sequencing was not performed for this lineage. Hence the mutations observed in present study were compared with the basal variants observed in the phylogenetic tree of M43b (**Figure 4.34**) generated using all the available published sequences. It was observed that, the Shrestha M43b lineage share a terminal mutation 3866C with the Nepali sample (GenBank: JF742211) and 16092C variation with the samples from India (GenBank: GQ337648, FJ383428, FJ383429).

4.5.3.3 Haplogroup M30

M30 Hg has been detected in Yemen, Egypt, Tunisia, Palestine, Saudi Arabia, South Africa, Iran, India, Nepal and China (**Table 8.2**). Hg M30 and its sub Hg M30a, M30b, M30c and M30d has a broad geographic, ethnic and linguistic distribution in India. It is a South Asian specific Hg. Among the Newar, the highest frequency of M30 has been observed in Shrestha (11.11%), followed by Udaya (10.34%) and Manandhar (3.22%). Its frequency was 4.1% in Brahmin. According to previous study, highest frequency of M30 was observed in Hindu (25%) from Morang, Nepal. Similarly, this Hg was also reported from Kathmandu (5.19%) (Gayden et al., 2013), individual from East-Nepal including Kathmandu (6.5%)(Wang et al., 2012).

- M30+16234T: Udaya (5.17%),
- M30b: **Shrestha (8.33%)**, Newa-mix (6%), Udaya (1.72%), Manandhar (1%),
- M30d/M30d1: **Hindu Terai (25%),** Nepali-other (4.47%), Udaya (3.44%), Shrestha (2.77%), Brahmin (2.08%),
- M30e: Kathmandu (1.29%)
- M30g: Brahmin (2.08%). Age of this hg is ≈8.7kyr (Silva et al., 2017)

Unresolved M30+16234T lineages belonging to Udaya shows, the sharing of terminal polymorphism (8886A, 11437C) between the two samples of Udaya. This led into the identification of a new sub-Hg named as M30h. Newar sample belonging to M30b2 lineage clustered with the Indian sample and one unknown sample.

Complete mtDNA sequencing of sample belonging to M30d1 lineage, shows the sharing of variants between (7419A, 8502G, 13020C, 14518G, 15773A) between two samples of Udaya to form a new sub Hg (m30d1c). Similarly, the samples belonging to Shrestha and Newar-mix matches the terminal variants with the Indian sample (GenBank: GQ337540) indicating the Indian maternal influence on Newar.

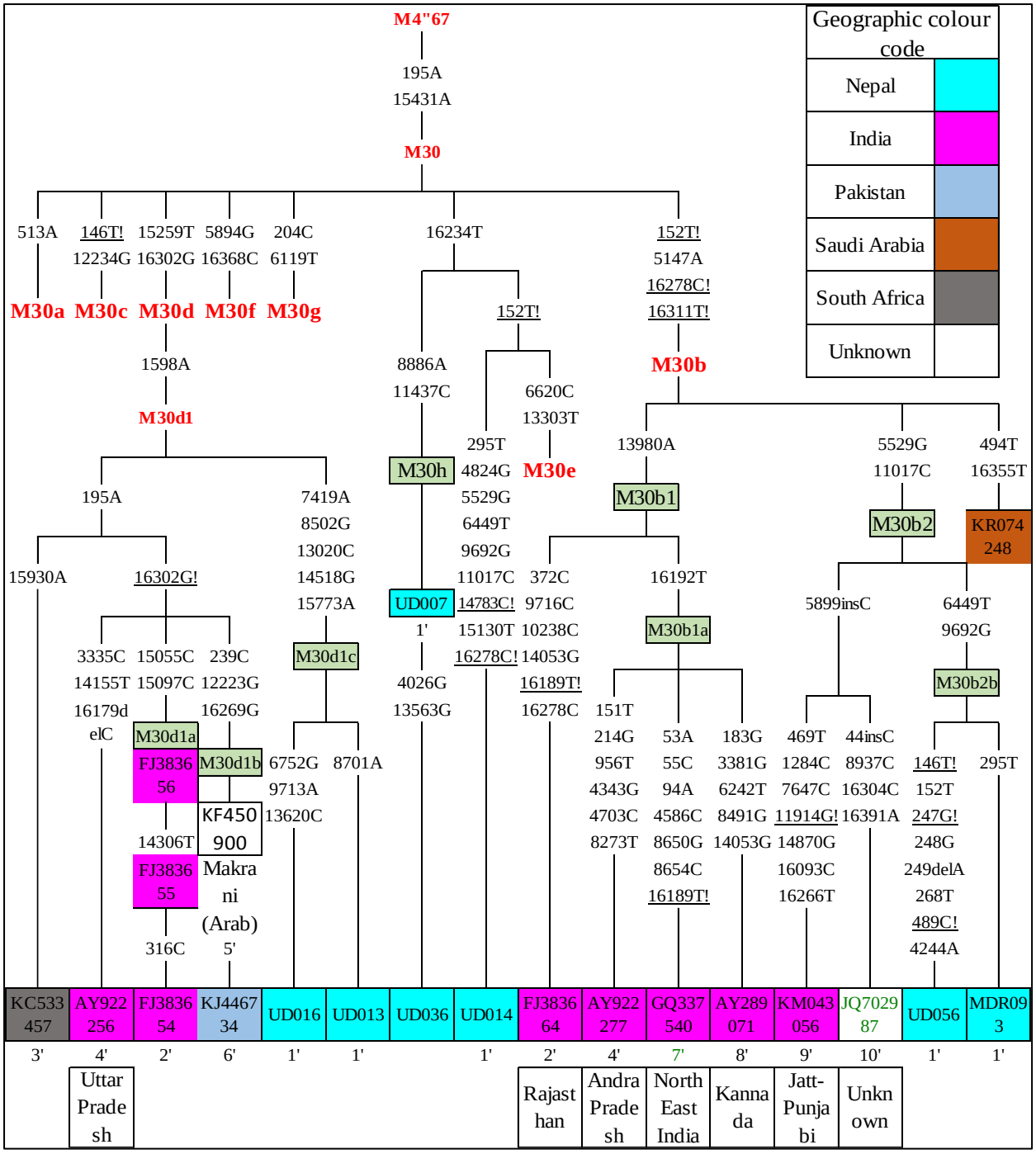


Figure 4.35 | A maximum-parsimony tree of Hg M30 was constructed based on 21 complete mtDNA sequences. Mutations relative to the RCRS are indicated on the branches. Sample ID/GenBank accession number for each sequence are shown in the box with geographic color code. The number (1', 2'... n') corresponds to the reference for the analyzed sample. The hg written in black color represents newly defined hg in this study, whereas red color indicates the predefined Hg in PhyloTree (Oven, 2015).

4.5.4 Haplogroup M5

Genetic evidence for the gene flow from South Asia in the Newar populations are mainly provided by the Hg M5. The frequency and diversity of Hg M5 reveals that it might have originated in central India and spread out to the eastern and western regions of India (Chandrasekar et al., 2009). In India, hg M5 has a wide distribution in both Dravidian as well as Indo-Aryan speaking

population. Interestingly, M5 are also reported among the Austroasiatic speaking population such as Saharia (9.43%) from Madhya Pradesh (**Table 8.1**).

Hg M5 is one of the most common lineage in Newar and Tharu (Chitwan II). The frequency of M5 Hg is higher in Shakya (15.78%), Manandhar (11.82%), Udaya (13.79%), Newar mix (12.12%) and Bajracharya (5%). Low frequency of hg M5 was observed in Brahmin (2.7%) and Magar (2.7%), most likely resulted due to the gene flow from neighboring populations. Higher Frequency of hg M5 was reported among the Uttar Pradesh populations such as Musahar (20%), Lodhi (14.2%), Bhar (10.24%), Harijan (6.89%) and Tharu (6.66%). Frequency of M5 were also higher among the South Indian population from Andra Pradesh, such as Naikpod (15.11%) and Pardhan (9. 94%) (**Table 8.1**). In Uttarakhand it was present in Kshatriya (10.25%) and Tharu (6.67%). Frequency of hg M5 was 8.5% among the Hindu samples from New Delhi (**Table 8.1**).

Interestingly, M5 sub lineage observed among the Newar sub caste, Magar and Brahmin populations where different with each other. Majority of M5 lineages observed in Manandhar (11.82%), belongs to M5a (9.67%). Shakya which contains the highest frequency of M5a (15.78%) includes three majors sub-Hg; M5a2a1a2 (5.2%), M5b2b (5.2%) and M5c2 (5.26%). Maharjan only contains M5a lineage (4%). Detail sub lineage of M5 observed in the studied populations is given below:

- M5a
  - M5a: Manandhar (9.67%), Maharjan (4%), Udaya (3.4%), Newa-mix (2%)
  - M5a2a1a1: Brahmin (2.08%)
  - M5a2a1a2: Shakya (5.26%),
- M5b2b: Shakya (5.26%), Bajracharya (5.26%), Udaya (3.4%), **Magar** (2.7%),
- M5c
  - M5c1: Shrestha (2.7%)
  - M5c2: Udaya (6.89%), Shakya (5.26%), Manandhar (1%)
- M5d: Manandhar (1%)

Hg M5 is detected in higher frequency in Newar and Tharu (Th-CII). Whereas its frequency is > 5% in other Nepali populations. Hg M5 were also observed in very low frequency in Burma, Tibet, Thailand, Iran, Laos etc. To further access the variability within the Hg M5, mtDNA phylogeny based on complete mtDNA sequences of this study and all available published sequences were constructed.

Basal variant 12477C was shared between Manandhar and Udaya population to form a new sub clade named herein as M5a6. Sub Hg M5a2a1a was present only in one individual belonging to Brahmin population. Sub Hg M5b2b were present in Shakya (5.26%), Bajracharya (5%), Udaya (1.72%), Maharjan (3%) and Newa mix (2%). Previous study shows, this Hg were present in small frequency among the Kathmandu and Nepali-other population. Whole sequencing analysis of this lineage led into the identification of three new sub lineage M5b2b1b, M5b2b1c and M5b2b1c1. One sample belonging to Hg M5b2b1a were found in Udaya and Magar.

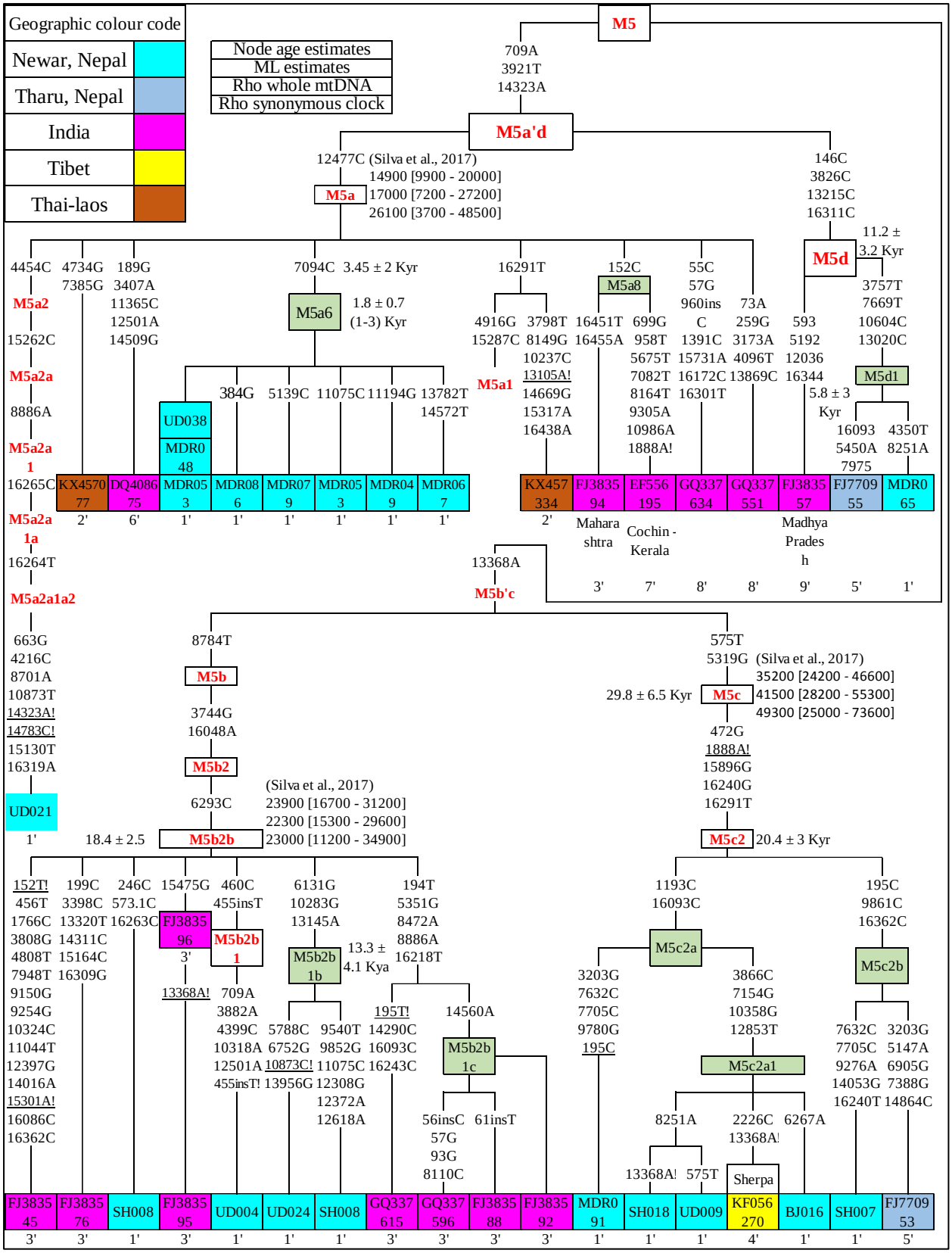


Figure 4.36 | A maximum-parsimony tree of Hg M5 constructed based on 38 complete mtDNA sequences. Sample ID/GenBank accession number for each sequence are shown in the box with geographic color code. The number (1', 2'... n') corresponds to the reference for the analyzed sample. The hg written in black color represents newly defined hg in this study, whereas red color indicates the predefined Hg in PhyloTree (Oven, 2015). In particular clade M5c2 most probably originated in Nepal is detected in Udaya Shakya and Tharu whereas completely absent in Maharjan. The sister clade M5c1

observed in India is detected in Shrestha in low frequency ≈2%. Newly calculated age in present study is given in Kyr. Whereas the other age provided in tree were taken from published literature (Silva et al., 2017).

Sub Hg M5c2 was found in Udaya (6.89%), Shakya (5.2%) and Manandhar (1.07%). Previous study revealed presence of M5c2 in the Tharu population from Chitwan. Complete mtDNA Sequencing of M5c2 individuals from Nepal indicates, clustering of Manandhar, Shakya, Udaya, Bajracharya and Tibetan sample. A series of basal variant were found to be shared between the Nepali and Tibetan population, to form a distinct sub clade named herein as M5c2a1. The presence of Nepal specific branch on a Tibetan sample indicates a recent gene flow from Nepali population to the Tibet. Similarly, Shakya and Tharu share three variants (M5c2b).

M5d was only present in one individual from Manandhar population. Previously, it was also reported to be present among the Tharu population of Nepal (Chaubey et al., 2014). Tharu and Manandhar sequence share four variants, to form a new sub clade named as M5d1.

4.5.5 Haplogroup M7

East Asian specific Hg M7 where not observed among the Newar, Magar and Brahmin populations.

4.5.6 Haplogroup M8

Phylogeny of Hg M8 (Phylotree.org) based on their presence on Nepali populations:

- M8a/M8a2: not observed in Nepali populations. This Hg is found in East Asia, notably in Japan.
- CZ (ancestor of Hg C and Z)
  - C
    - C4 (Unresolved): Tamang (2.7%), Kathmandu (1.2%)
    - C4a1a1a: Magar (2.7%), Nepali-other (0.4%)
    - C4a2c2: Brahmin (2.08%)
    - C4a3b: Sherpa from Nepal and Tibet (21.82%), Kathmandu (2.59%)
    - **C7b**: Magar (5.4%)
  - Z
    - Z1
      - Z1a
    - Z2
    - Z3
      - Z3a/Z3a1a: **Newar and Magar** populations.
    - Z4
    - Z7: Shakya

Hg C and Z are the sister clades of Hg M8. Hg C is found in North-East Asia, Siberia and also among the **Indigenous people of America**. The subclades C1b, C1c, C1d, and C4c are found in the first people of the Americas, whereas C1a is found only in Asia.

4.5.6.1 Haplogroup C

In present study, East Asian specific hg C was present in **Magar (8.1%)** and Brahmin (2.27%), whereas it was completely absent in Newar. The highest frequency of Hg M8 was observed in Sherpa (28%). Sub Hg C7 has a greater diversity in **China** including Thailand. Sub clade C7a has been reported to be present among the individuals from China, Thailand, Taiwan, Korea and North-East India (Chandrasekar et al., 2009; M. Derenko et al., 2010; Jiang et al., 2014; Ko et al., 2014; Kutanan et al., 2017). **Sub clade C7b is present in individual from China and North-East India particularly in Arunanchal Pradesh (Chandrasekar et al., 2009). In present study, sub clade C7b was detected in Magar (5.4%)**. Interestingly Hg C7b has not been reported so far among the Tibetan highlanders.

4.5.6.2 Haplogroup Z

Hg Z is one of the rarest mtDNA Hg (**Table 8.1**). Hg Z is also present in several Siberian populations and among the Saami people of northern Scandinavia. Interestingly, sub Hg Z1 predominates the maternal lineages of the above-mentioned populations, whereas sub Hg Z3a1a has been present in very low frequency among the Tibetans (X. Qi et al., 2013) and North East Indian. Z3a1a has been reported in North East Indian; three samples each belonging to Lepcha, Lachungpa and Dirang Monpa (Chandrasekar et al., 2009). Overall frequency of Hg Z has been reported to be **very low (1.92%) in Tibet** (X. Qi et al., 2013). Similarly, Z3/Z3a/Z3a1a sub clade is also detected in very high frequency among the several Thai populations as listed in the **Table 4.7** (Kutanan et al., 2017). These lineages where detected in both Tai-Kadai and Austroasiatic speaking population from Thailand. Hg Z has high frequency diversity among the populations from southern China (H. Li et al., 2007).

Table 4.7 Huge diversity of Hg Z was observed among the population from Thailand (Kutanan et al., 2017).

Population	Code	Country	Linguistic Family	Sample Size	Frequency (%)	HG
Lawa	LW2	Thailand	Austroasiatic	24	4.16	Z3a
Paluang	PL	Thailand	Austroasiatic	25	4	Z3a
Blang	BL1	Thailand	Austroasiatic	25	4	Z
Phuan	PU5	Thailand	Tai-Kadai	25	16	Z4
Khon-Mueang	KM3	Thailand	Tai-Kadai	24	4.16	Z3a1a

In the present study, Hg Z was observed in high frequency in Newar (18.66%). Within the hg Z, subclade Z3a1a was the most dominant and present in all major Newar caste analyzed in this study. Whereas, sub Hg Z7 was confined within Shakya (10.62%) and Bajracharya (5%). Hg Z7 was either absent in other Newar caste or present in very low frequency. Interestingly Hg Z3a1a was also present in low frequency in **Magar (5.4%)** and was not detected in Brahmin.

To get further insight into the origin and migratory route of lineage Z, contour maps of hg Z, Z3a and Z3a1a were created based on the Hg frequencies (**Table 8.1**).

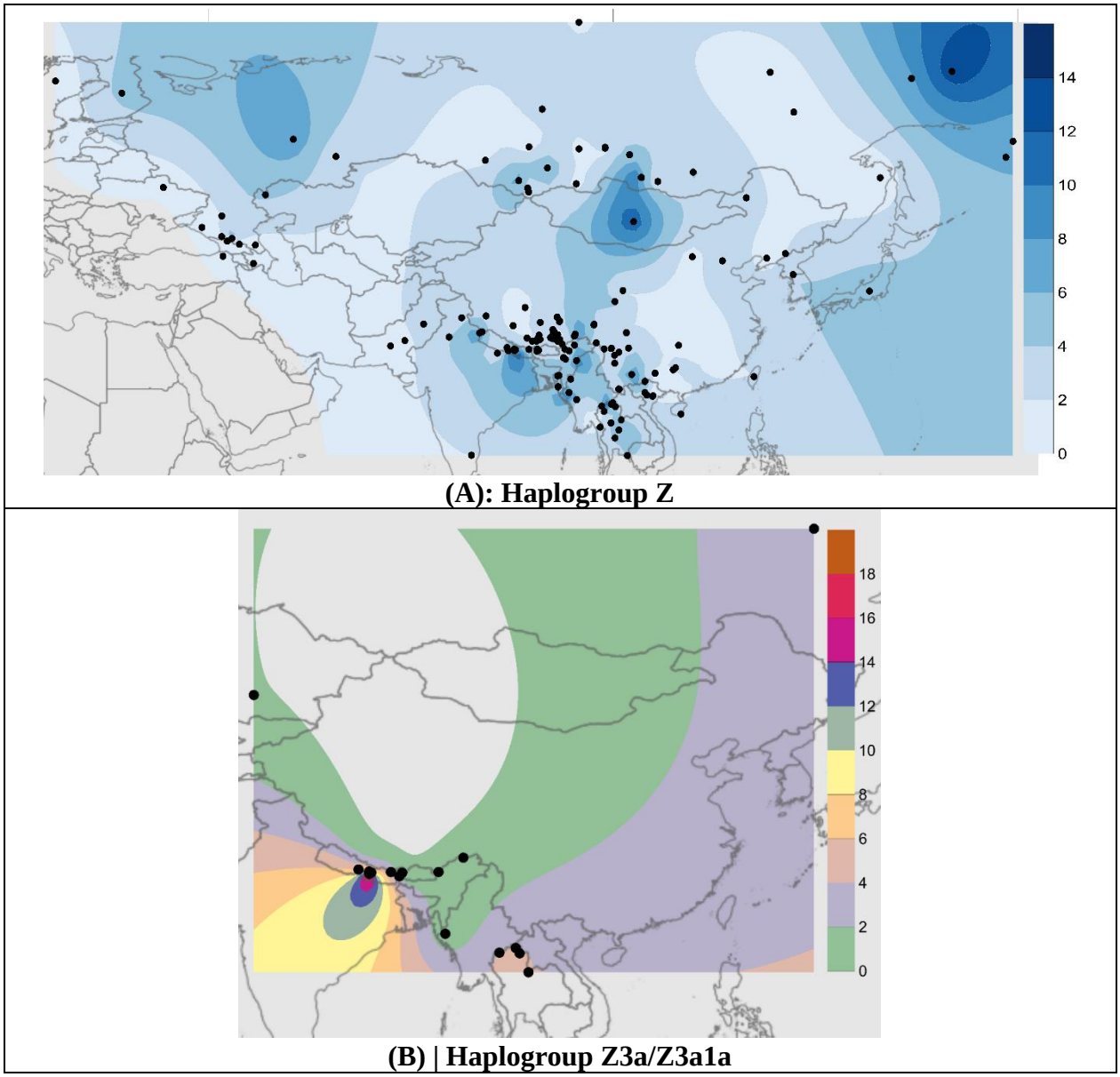


Figure 4.37 | Contour map of (A) Haplogroup Z, (B) hg Z3a/Z3a1a, of Nepali and other Asian populations. The dots in the map indicate the approximate geographic locations of the analyzed populations. These spatial-frequency distributions were created using the Kriging linear model of the Surfer 12.8 package, based on the frequency of each Hg in different populations (Supplementary **Table 8.2**).

Z3a1a has been observed in the Tibetan populations (Xuebin Qi et al., 2013), but their frequency were unavailable to include in the counter map analysis. As shown in the counter map, Z3a1a were also detected in North East India, Burma and Thailand (Chandrasekar et al., 2009; Y. C. Li et al., 2015). Hence, there are two possible routes through which these lineages might have introduced into Nepal;

1. From Tibet (across the Himalayas)
2. Through North East India.

To get more insights into the origin and dispersal of Hg Z in Eurasia Median joining network and a complete phylogenetic tree based on **139 complete mtDNA sequences** were computed of

which 39 complete mtDNA sequences where generated in this study. Due to the large file size the tree was unable to include in word (only available in pdf format on request).

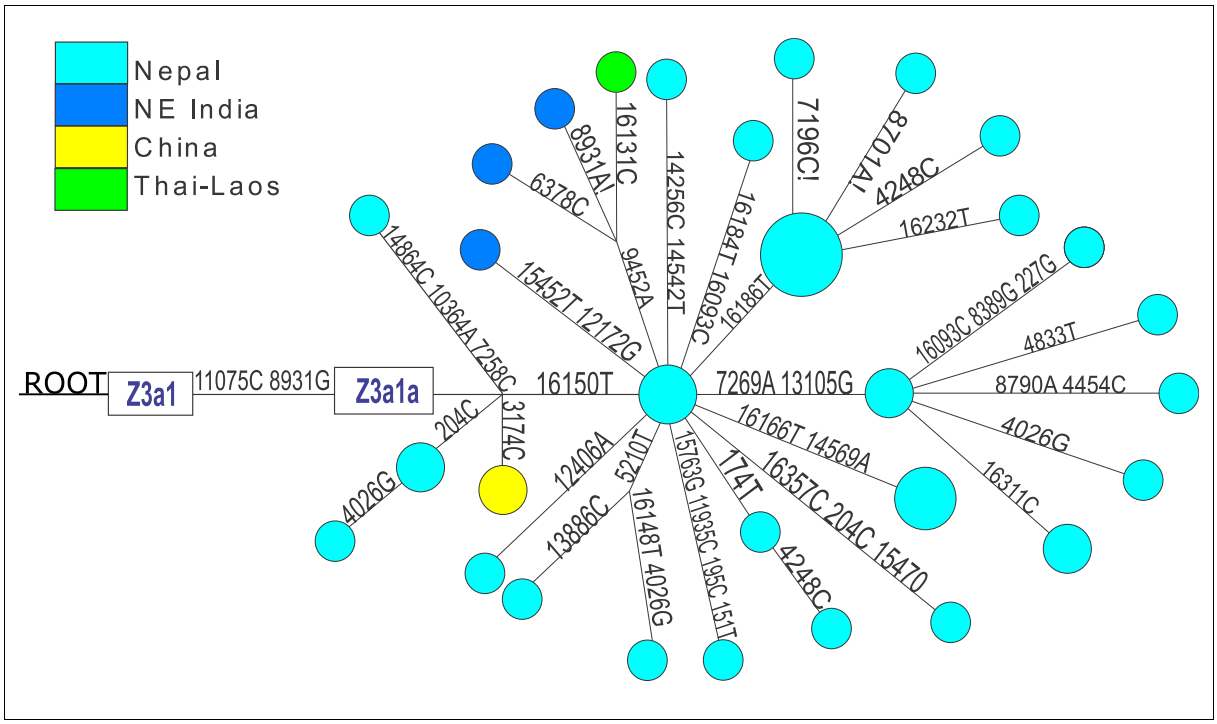


Figure 4.38 | Median Joining Network of Hg Z based on complete mtDNA sequences. The size of the circles is proportional to the number of individual cmtDNA sequences. Nucleotide position number shown in the network are consistent with the Revised Cambridge Reference Sequence (rCRS). The geographic origin of sample is shown by different colors. All the Nepali samples are shown by same colour. The C stretch length polymorphism in regions 303–315, AC indels at 515–522, 16182C, 16183C, 16193.1C(C) and 16519 were disregarded for the network reconstruction. As shown by the Median joining Network, Thai-Laos, North East Indian and Nepali samples (Newar and Magar) belonging to lineage Z3a1a are derived from the ancestral variant 16150T.

Majority of the of Z3a1a lineage of Nepal, India and Thailand-Laos share a **common basal variant 16150T**, suggesting their common maternal origin. Previous study reported Z3a1a to be present in a very small frequency among the Tibetans, unfortunately complete Tibetan mtDNA sequences for this lineage was unavailable. Therefore, median joining network for Hg Z, based on the HVR I sequence was also constructed (**Figure 4.39**), which showed the sharing of basal polymorphism 16150T among the Tibetan, Nepali, Thai-Laos and North East Indian samples to form a New sub lineage named herein as Z3a1a1.

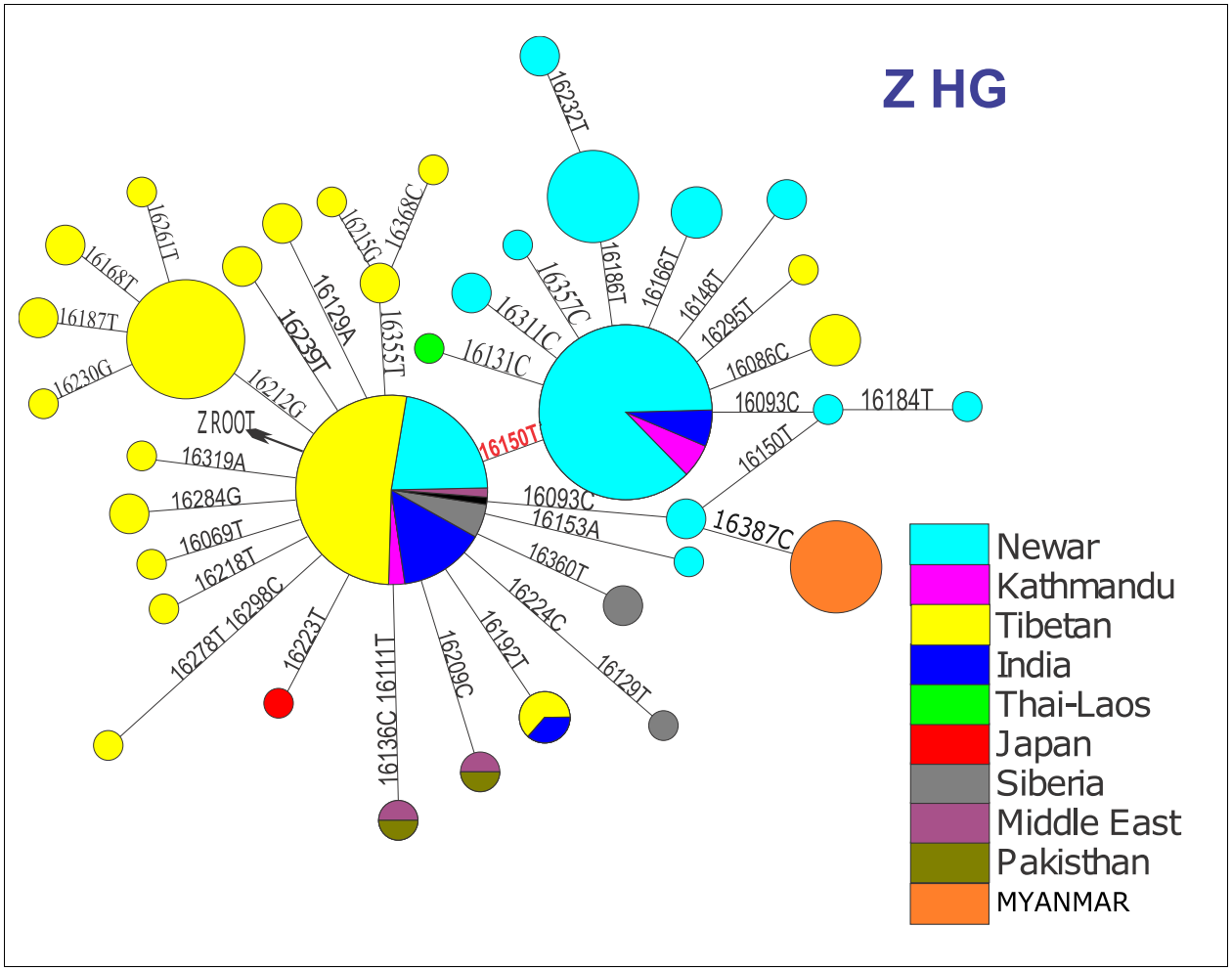


Figure 4.39 | Median Joining Network for Hg Z3a1a, based on mtDNA HVR1sequences. The size of the circles is proportional to the number of sample. Nucleotide position number shown in the network are consistent with the Revised Cambridge Reference Sequence (rCRS). The geographic origin of sample is shown by different colors. The C stretch length polymorphism in regions 303–315, AC indels at 515–522, 16182C, 16183C, 16193.1C(C) and 16519 were disregarded for the network reconstruction. The network suggests a common maternal origin for Z3a1a lineage present in Tibet, Thai-Laos, Nepal and North-East India.

As shown in the **Figure 4.38**, variant 9452A is shared between two North East Indian sample and one Thai-Laos sample indicating a further closeness between Indian and Thailand Z3a1a to form a new sub lineage named herein as Z3a1a1d. On the basis of the constructed networks, we can see that Nepali sample belonging to Z3a1a share some basal haplotypes with the Tibetan samples. Nepali Z3a1a harbor's a number of unique (Nepal specific) haplotypes at the terminal level, most of which branched off directly came off from the nodes occupied by Tibetan Z3a1a lineage. Further, the network analysis suggests that Z3a1a tend to form a star-like shapes, suggesting recent expansion of Z3a1a lineages among the Nepali population. Three new Nepal specific lineage Z3a1a1a (13105G, 7269A), Z3a1a1b (161866T) and Z3a1a1c (A16166T, G14569A) has been newly defined in this study.

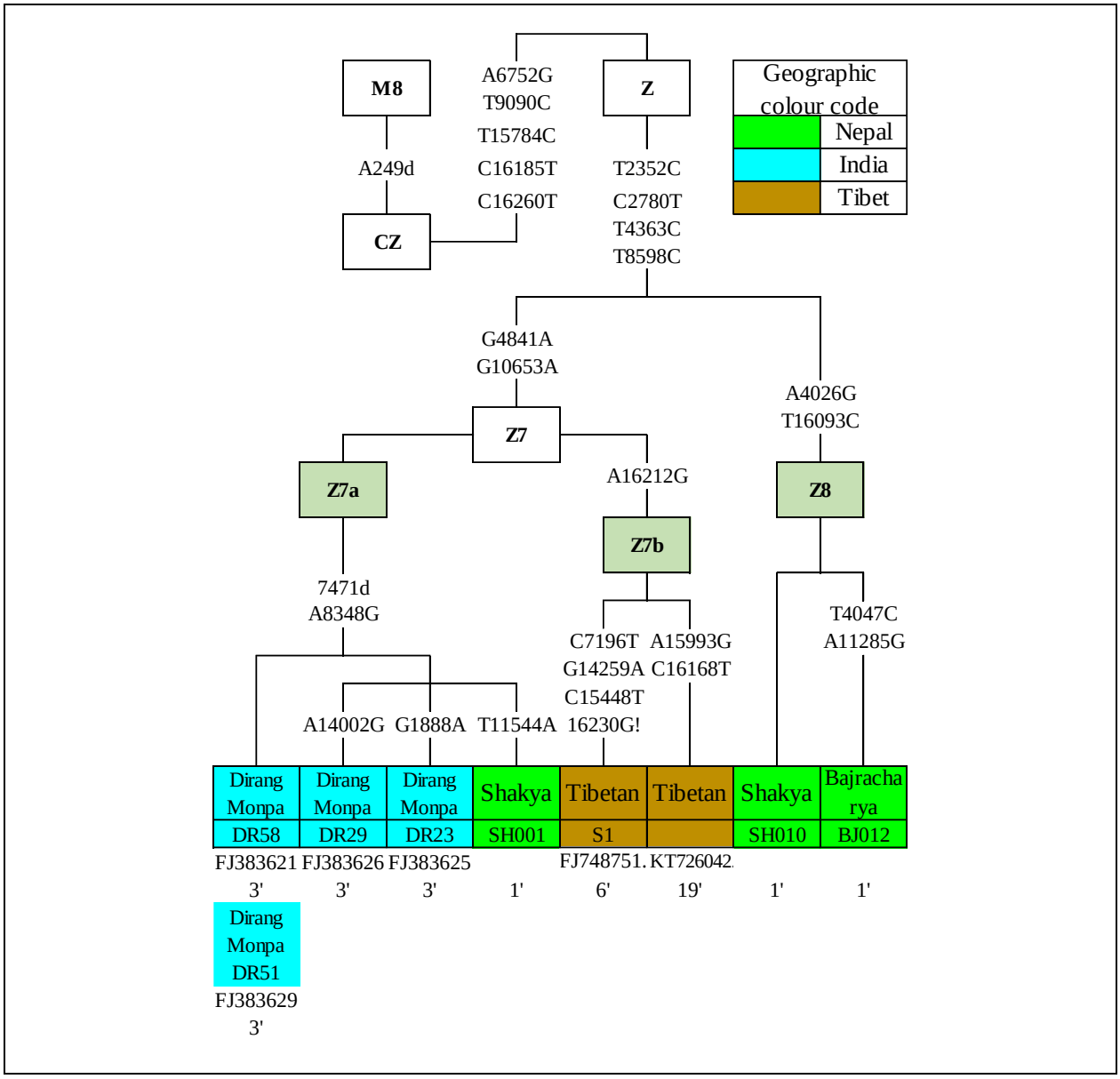


Figure 4.40 | Phylogenetic tree of Hg Z7 based on 7 complete mitochondrial sequences. One Individual belonging to Z7 lineage of Shakya shared a terminal variant with Dirang monpa of North East India, indicating a recent gene flow between these populations. New Hg Z7 (16212G) and Z8 (4047G, 11285G) has been newly defined in this study. Mutations relative to the RCRS are indicated on the branches. Sample ID/GenBank accession number for each sequence are shown in the box with geographic color code. The number (1', 2'... n') corresponds to the reference for the analyzed sample. The hg written in black color represents newly defined hg in this study, whereas red color indicates the predefined Hg in PhyloTree (Oven, 2015).

4.5.7 Haplogroup M9

Hg M9 is an East Asian specific Hg. Its greatest frequency was observed among the Tibetans (22.48%) (X. Qi et al., 2013). In present study, this hg was either absent or present in very low frequency in Newar. Whereas sub Hg M9a1a2 was present in higher frequencies (13.5%) in Magar. This hg was completely absent in Brahmin.

Previous study reported this Hg to be present in higher frequency in Sherpa (24.22%), Tharu Chitwan I (19.29%), Tamang (13.52%) and Tharu Chitwan II (11.84%). Whereas its frequency was less than 2% in Kathmandu (**Table 8.1**). M9a and its four sub-Hgs (M9a1a, M9a1a, M9a1b and M9a1a1c1b1a) are widely distributed in East Asia and Southeast Asia. In the present study, majority of M9 lineage belongs to sub clade M9a1a2 except one sample from Udaya belong to M9a1b1. M9a1a are present in higher frequency among the Tibetans, 16.3% (X. Qi et al., 2013). The Tibetan M9a1a lineage contains the back mutation 16362C (Kang et al., 2013) which was also detected in all Nepali M9a1a lineages. High frequency of M9a1b1 lineage in the Magar group suggest a close maternal relationship with the Tibetans.

4.5.8 Haplogroup M10

Hg M10 is an East Asian specific Hg. In present study, this Hg was either absent or present in very low frequency in different Newar sub caste (Gene flow in Newar is dominated by East Asian (50.8%) haplogroups followed by South Asian (35.9%) haplogroups. Whereas the contribution from West Eurasia (12.8%) is very low in Newar. Two East Asian haplogroups Z and F contribute 35.6% of the total mtDNA gene pool of Newar. Hg Z and F both are also detected in Tibetan population. The other East Asian hgs with appreciable frequency include only hg D (5.9%) and A (4.8%). Hg A and D are found in higher frequency diversity among the Tibetans.

Among the South Asian (35.9%) hg, only two hgs M5 (9.6%) and M3 (5.3%) shows significant frequency contribution on Newar. Whereas the other south Asian hgs are present in very low frequency. West Eurasian hg U7 has 4.3% contribution to the total mtDNA gene pool of Newar.

Table 4.5). Hg M10 was completely absent in Brahmin and Magar.

4.5.9 Haplogroup M12'G

M12'G consists two Hgs (Phylotree.org):

1. M12: Detected in Arunachal Pradesh and Orissa of India. Hg M12 was not observed among the Nepali population in present as well as in previous study (**Table 8.1**).
2. G: It is an East Asian specific Hg.

4.5.9.1 Haplogroup G

Hg G is an East Asian specific Hg. This hg is also present in Central Asia where it makes up to 20% of the mtDNA pool. Hg G is divided into subclades G1, G2, G3, and G4 (Phylotree.org). Sub Hg G2 was only observed in Newar, whereas Hg G3 were observed in Brahmin (6.25%). Interestingly, this Hg were not detected in Magar. Sub Hg G2 were also observed in high frequency among the Tharus population from Nepal (**Table 8.1**). Complete mtDNA analysis shows the clustering of Brahmin G3 lineage with those detected in China, Tibet and Kashmir. Moreover, Brahmin G3 lineage share the most recent common ancestors with G3 lineage detected in Kashmir. Newar sample (Maharjan, Udaya, Manandhar and Newamix) belonging to lineage G2a1d2 share an ancestry with the sample from Belarus (Central Asia). The shared basal variants (4833G and 16227G!) between Belarus and several Newar sample define a new clade named as G2a1d2b. Whereas the two sample from Udaya belonging to Haplogroup G2b2a1 cluster with the sample from India

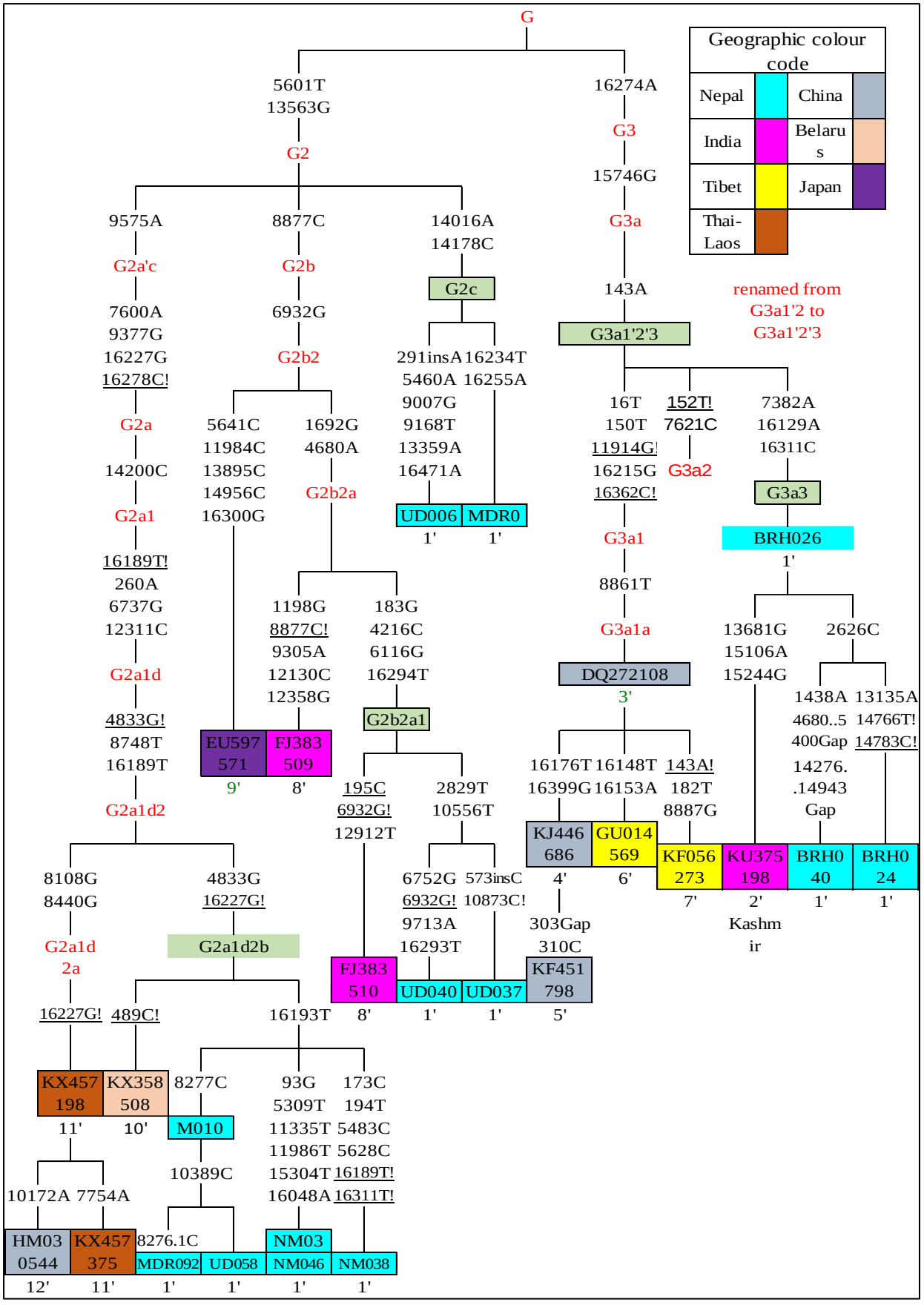


Figure 4.41| A maximum-parsimony tree of Hg G. Tree was constructed using 26 complete mitochondrial sequences of which 13 complete sequences were generated in this study. Mutations relative to the RCRS are indicated on the branches. Sample ID/GenBank accession number for each sequence are shown in the

box with geographic color code. The number (1', 2'... n') corresponds to the reference for the analyzed sample. The hg written in black color represents newly defined hg in this study, whereas red color indicates the predefined Hg in PhyloTree (Oven, 2015).

4.5.10 Haplogroup M31

Haplogroup M31 was observed among several Nepalese populations with a frequency <2.5% (**Table 8.1** & **Table 8.2**). Frequency and distribution of M31 haplogroup is given below:

- M31a
  - M31a1: Andamanese-specific lineage,
  - M31a2: Present in Tamang (2.27%)
- M31b:
  - M31b: Tharu Chitwan I (1.17%), Tharu Chitwan II (1.3%)
  - M31b2: Brahmin (2.08%)
- M31c: Newa_mix (2%), Maharjan (1%), Nepali-other (0.4%)

M31 was reported to be preset in Andamanese Island. This haplogroup was present in high frequency among the **Austroasiatic** speaking tribes from North East (Meghalaya) India (Reddy et al., 2007). Haplogroup M31 is also reported to be present among the Saharia (Austro-Asiatic) from Madhya Pradesh, India. In North India, this haplogroup where reported to be present only in Tharu (2.2%) from Uttarakhand. Small frequency were also reported among the Tibeto-Burman speaking populations from Myanmar; Burmans (1.8%) and Naga (2.3%) (Y. C. Li et al., 2015). Similarly, in Tibet M31 was found only among the Monb (3.9%) group. This haplogroup where reported in very low frequency among Chakma and Marma; Tibeto-Burman speaking population from Bangladesh (Gazi et al., 2013).

4.5.11 Haplogroup M33

Haplogroup M33 is a South Asian specific haplogroup and one of the oldest lineage present in South Asia. In present study, this haplogroup was found in very low frequency in Udaya (3.44%), Manandhar (1.07) and Magar. Previous study reported this haplogroup to be present in higher frequency in Newar (6.06%) (Gayden et al., 2013). Similarly, this haplogroup were also reported to be present in several other Nepali populations; Tharu CII (7.01%), Eastern Tharu (7.5%), Kathmandu (6.4%) and Nepali -other (2.8%) (**Table 8.1**). Haplogroup M33 has a wider distribution throughout India and has been reported to present among Tibeto-Burman, Austroasiatic, Dravidian as well as Indo European speaking population of India. M33a is reported to be present in Gujrat India (Thangaraj et al., 2006). This haplogroup has also been reported to present in high frequency in Meghalaya (North East India) and Uttarakhand (North India). Whereas, small frequency where observed in South India.

Complete sequence analysis shows M33a linage detected in Udaya cluster with the Tharu and individual from Kathmandu/Eastern Nepal (Wang et al., 2012) to form a Nepal specific sub haplogroup named herein as M33a1a2. Similarly, another Udaya sample share a basal variant (16192T) with the samples from Sikkim to form a new sub clade named newly as M33a1a1.

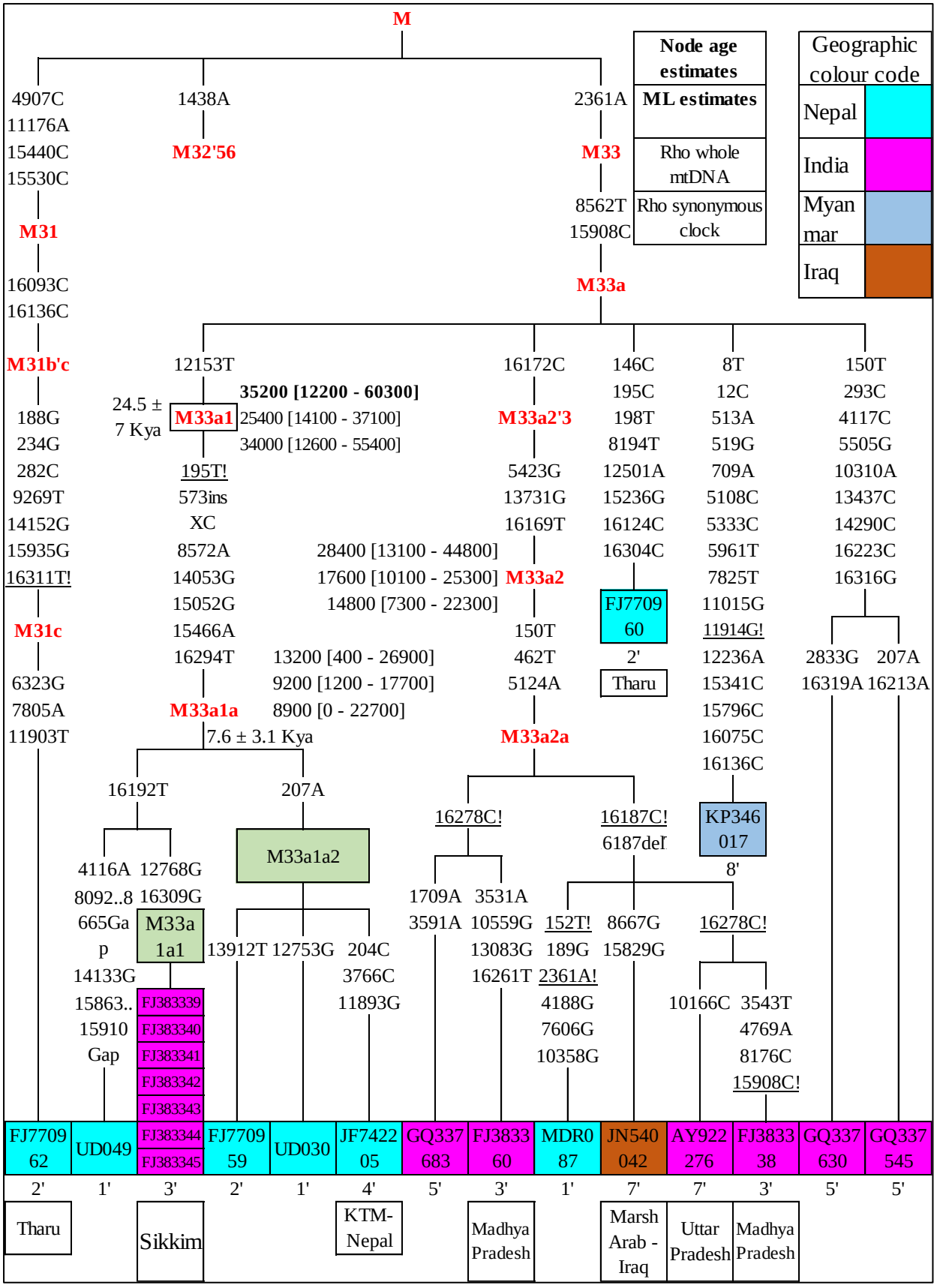


Figure 4.42 | A maximum-parsimony tree of haplogroup M31 and M33 based on 22 complete mitochondrial sequences. Mutations relative to the RCRS are indicated on the branches. Sample ID/GenBank accession number for each sequence are shown in the box with geographic color code. The number (1', 2'... n') corresponds to the reference for the analyzed sample. The hg written in black color

represents newly defined hg in this study, whereas red color indicates the predefined Hg in PhyloTree (Oven, 2015).

Haplogroup M33a1 was observed only in Nepal and Sikkim. Within the M33a1a clade high diversity was observed among the Nepali group compared to Sikkim, suggesting a gene flow from Nepal to Sikkim at ≈2Kyr ago. Age of haplogroup M33a1 is estimated to be ≈25.4 Kyr old (Silva et al., 2017). Hence, Nepali populations harbors ancient and indigenous south Asian specific mtDNA haplogroups.

4.5.12 Haplogroup M34

Small clades of haplogroup M34 are present in South Asia. M34a are reported to be present in **Karnataka** of India (Thangaraj et al., 2006). Among the Newar, haplogroup M34 was found only in Shrestha (5.5%) and Udaya (1.7%). This haplogroup was also found among the Brahmin (2.08%) and Magar (2.08%) from Tanahun district. Interestingly the sub-haplogroups found among Brahmin, Shrestha and Udaya where different with each other. Frequency and distribution of Haplogroup M34 is given below (**Table 8.1**):

- **M34"57** (unresolved): Magar (2.08%)
- **M34** (Unresolved): Kathmandu (1.29%), Nepali-other (0.4%),
  - **M34a**: Newar_mix2 (1.5%),
    - **M34a1a:** Udaya (1.72%), Nepali-other (1.29%),
    - **M34a2:** Shrestha (5.55%)
  - **M34b:** Brahmin (2.08%), Nepali-other (2.4%)

4.5.13 Haplogroup M35

Haplogroup M35a is found in India. Sub haplogroup **M35b** has been reported to present in **Karnataka India**, Nepal, Myanmar and Slovakia. Among the Newar, M35b is only present in Manandhar (4.30%) and Udaya (3.44%). Haplogroup M34b2 were also observed in Brahmin (4.1%). Hg M34b were also detected in the previous studies done on the Nepali populations; Newar (1.5%), Kathmandu (5.1%), Nepali-other (3.2%) and Tharu CII (1.3%) (**Table 8.1**). M35 lineage detected in Newar belongs to sub haplogroup M35b4. M35b4 observed among the Newar were not detected in other Nepalese populations studied so far. Lineage M35b2 detected in Brahmin from Tanahun district, were also observed in Kathmandu (Gayden et al., 2013). Frequency and distribution of haplogroup M35 in Nepali populations (**Table 8.1**) studied so far is given below:

- **M35** (Unresolved): Newar_mix2 (1.5%), Tharu CII (1.31%)
  - M35b
    - **M35b**: Nepali-other (3.25%)
    - **M35b2**: Brahmin (4.16%), Kathmandu (5.19%)
    - **M35b4**: Manandhar (4.3%), Udaya (3.44%)

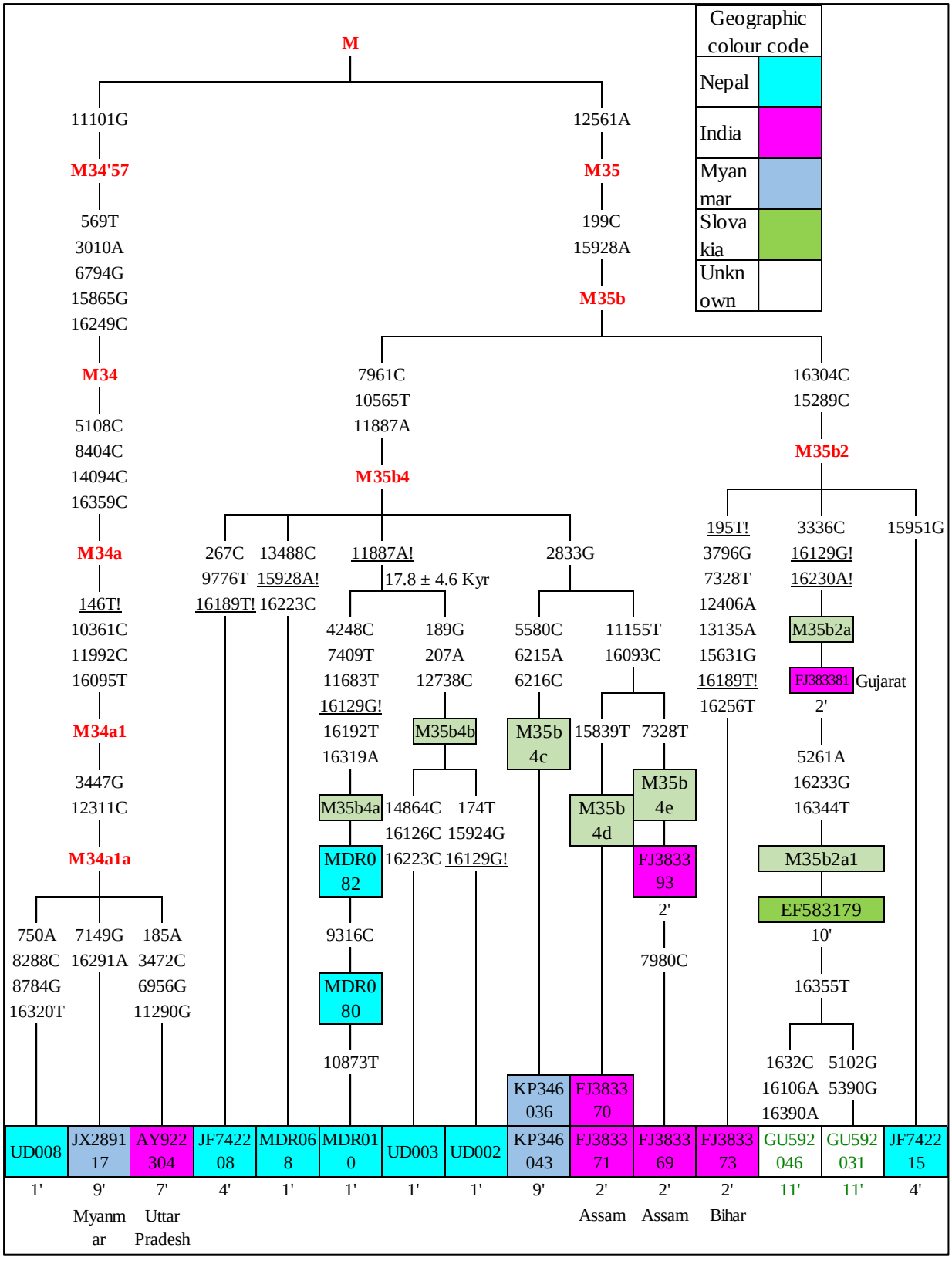


Figure 4.43 | A maximum-parsimony tree of haplogroup M34 and M35 based on 20 complete mitochondrial sequences. Mutations relative to the RCRS are indicated on the branches. Sample ID/GenBank accession number for each sequence are shown in the box with geographic color code. The number (1', 2'... n') corresponds to the reference for the analyzed sample. The hg written in black color represents newly defined hg in this study, whereas red color indicates the predefined Hg in PhyloTree (Oven, 2015).Two Myanmar sample clusters with Indian sample by sharing basal variant (2833G),

indicating a recent gene flow from India to Myanmar. In present study, several sub haplogroup were newly defined as shown in the tree. Age of clade M35b4 is assumed to be ≈8.6 Kyr old (Silva et al., 2017).

4.5.14 Haplogroup M52

M52 is a south Asia specific haplogroup. M52 haplogroup contain two sub haplogroups M52a and M52b. M52a haplogroup has been previously detected in Indian Muslim populations (Eaaswarkhanth et al., 2010), Tharu from Uttarakhand (India), Tharu from Nepal (Chaubey et al., 2014) and Yemen. In the present study, Haplogroup M52b was present in high frequency in Magar **(10.8%)**, whereas its frequency was lower in Newar and Brahmin (2.98%). Previous study revealed this sub clade was present in Tharu (2.5%) from Morang District of Southeastern Nepal and Nepali-other (2.84%) population (**Table 8.1**).

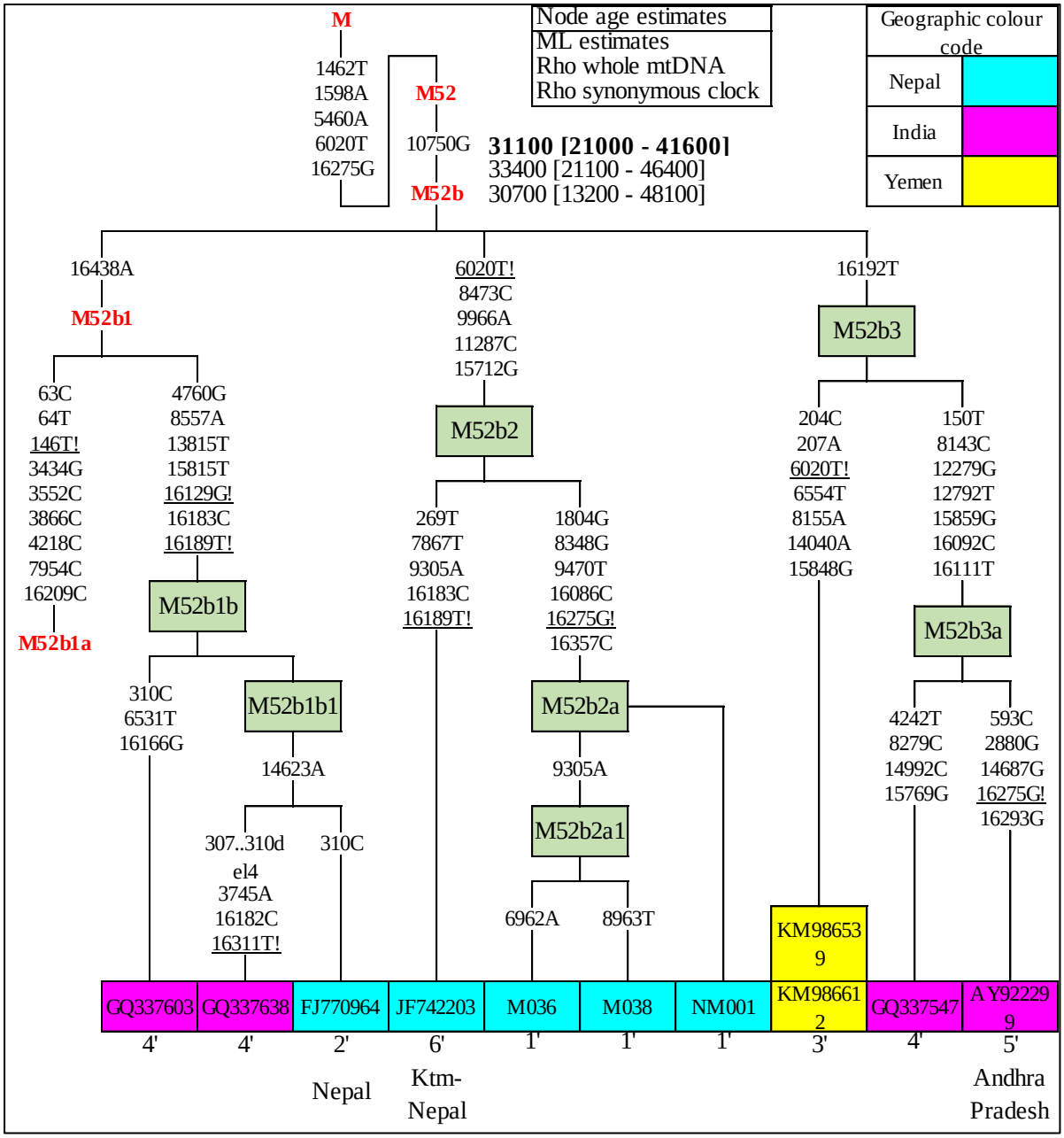


Figure 4.44 | Phylogenetic tree of haplogroup M52b based on 11 complete mitochondrial sequences of which three complete sequences were generated in this study. Mutations relative to the RCRS are

indicated on the branches. Sample ID/GenBank accession number for each sequence are shown in the box with geographic color code. The number (1', 2'... n') corresponds to the reference for the analyzed sample. The hg written in black color represents newly defined hg in this study, whereas red color indicates the predefined Hg in PhyloTree (Oven, 2015). The Homogeneity in the sequences among the populations from Kathmandu (Gene Bank: JF742203), Maharjan and Newamix has led into the formation of new Nepal specific branches M52b2.

Complete sequence analysis shows the North East Indian sample and Tharu sample share a series of basal variants to form a new subclade named herein as M52b1b and M52b1b1. Similarly, Maharjan and Newamix cluster with the sample belonging to Kathmandu to form a new distinct sub group name herein as M52b2, M52b2a1 and M52b2a1. M52b2 lineage of Magar and Brahmin (only HVR sequence available) contain two basal variation (269T and 16189C) which was also present in Kathmandu sample from previous study (GenBank: JF742203) (Wang et al., 2012). Age of M52b is estimated to be ≈33.4 Kyr old. Presence of ancient south Asian lineage M52b2 only in Nepali populations (Brahmin, Maharjan, Magar and Newamix) suggest an in-situ differentiation of this lineage in Nepal and might represent one of the oldest lineage present in the geographic territory of the present-day Nepal. Similarly, the Arabian samples from the Yemen cluster with the Indian samples to form a new clade named herein as M52b3 and M52b3a.

4.5.15 Haplogroup M58

Haplogroup M58 is an ancient South Asian specific lineage having Deep time depth of ≈29,000 years. This haplogroup has been reported to be present in Munda of North East India and Burmese of Myanmar. In the present study, Haplogroup M58 were present **in Magar (2.7%),** whereas completely absent in other Nepali populations studied so far.

4.5.16 Haplogroup M80'D

4.5.16.1 Haplogroup D

Haplogroup D is found in East, Central and North Asia. Sub hg D4 and D5 are the two principle branches of hg D. Sub hg D4 is mostly found in Japan, spread all over China and Korea. This haplogroup is also detected in Russian and Native Americans. Sub hg D1 is the basal branch of hg D4, which is widespread and diverse in Americas. Overall frequency of Haplogroup D in Tibet is 16.53% of which D4 (12.72%) is the predominant sub haplogroup (Supplementary **Table 8.2)**.

Among Newar, haplogroup D is present in Shakya (15.78%), Bajracharya (10%), Newar-mix2 (10%), Udaya (6.8%), Maharjan (6%) and Shrestha (5.5%). Whereas it was completely absent in Manandhar. In present study, Haplogroup D was present in high frequency in **Magar (24.32%)** and present in traces frequency in Brahmin (2.08%). As reported by the Previous study, this haplogroup was present among the several population from Nepal; Tamang (27.5%), Tharu from Chitwan II (11.84%), Tharu from Morang district (10%), Kathmandu (11.6%), Tharu from Chitwan I (5.2%), individuals from Kathmandu and Eastern Nepal (8.5%) and Sherpa (4.57%). This is one of the most important hg in Tamang and Magar followed by Shakya.

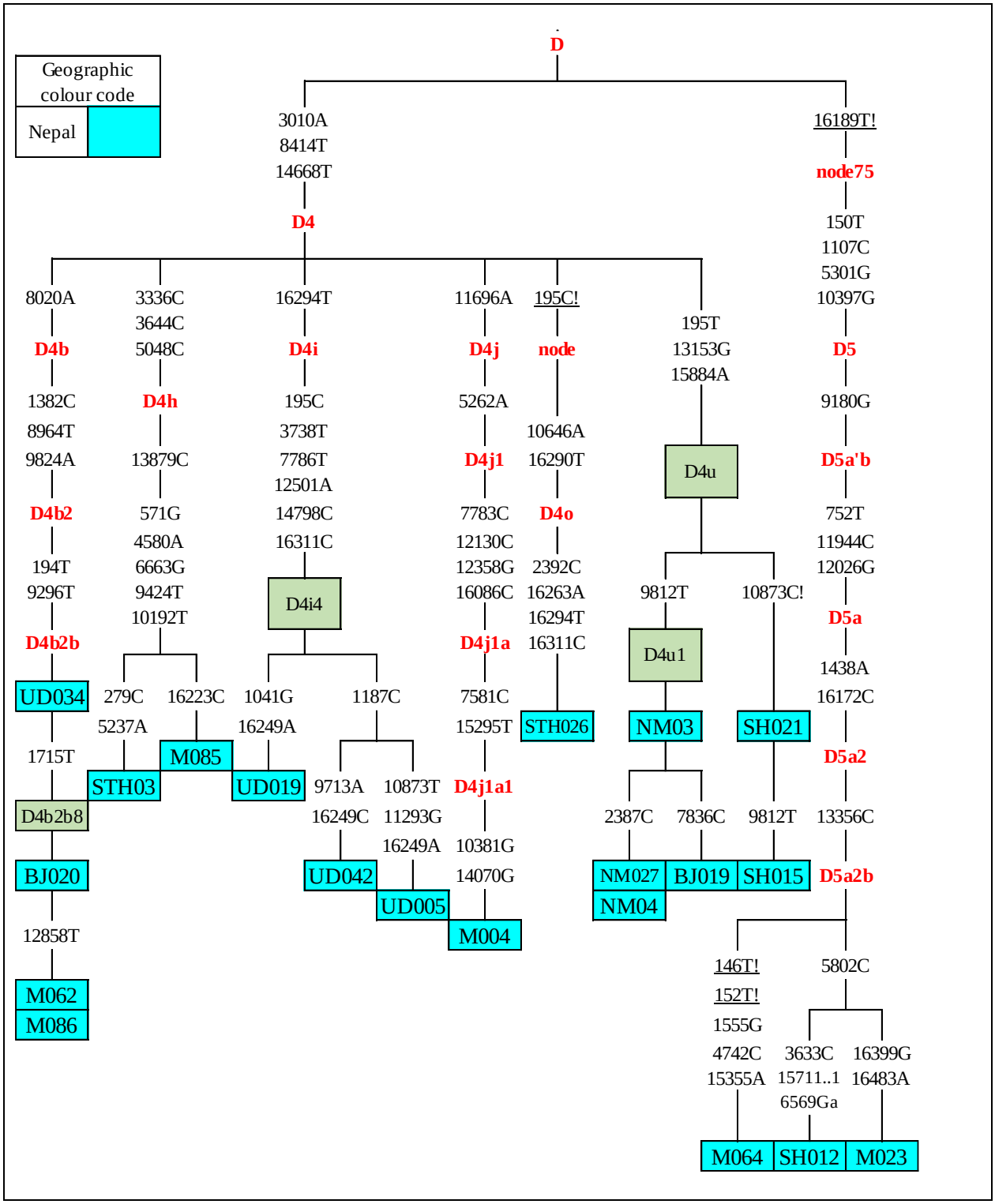


Figure 4.45 | Phylogenetic tree of haplogroup D based on the 20 complete mitochondrial sequences generated in this study. Mutations relative to the RCRS are indicated on the branches. Sample ID/GenBank accession number for each sequence are shown in the box with geographic color code. The hg written in black color represents newly defined hg in this study, whereas red color indicates the predefined Hg in PhyloTree (Oven, 2015).

4.6 Time Estimates

Table 4.8 | Estimated time to the most recent common ancestor of the Hgs observed in the Nepali populations based on the complete mtDNA (see material and methods section 3.9.7 for more info).

Haplogroup	Complete genome substitutions		Haplogroup	Complete genome substitutions	
	Ro(Sigma) ρ (σ)	Calculated age in Kyr		Ro(Sigma) ρ (σ)	Calculated age in Kyr
M2a3a	4.3 [1]	11.4 ± 2.5	M43a	5(2.1)	13.3 ± 5
M2a3a + 11827C	2.3 [0.9]	6 ± 2.3	M43a1	1(0.6)	2.5 ± 1.5
M3c1a1	5 [1.86]	13.3 ± 4.8	M52+110750	11.7(1.9)	32.8 ± 4.9
M3d	5.7(1.5)	15.3 ± 3.8	M52b	10.7(1.6)	29.8 ± 4.1
M3d1	4.3(1.3)	11.4 ± 3.3	M52b2	6.5(2)	17.5 ± 5.2
M3d1a	3.3 [0.8]	8.7 ± 2	M52b3	9.2(2)	25.3 ± 5.2
M3d1a1	0.7(0.5)	1.8 ±1.2	Z3a1a + 16150T	2.77 [0.954]	7 ± 2.3
M3d1a+16301	2.2(1)	5.7 ± 2.5	Z3a1a1	1.97 [0.46]	5.1 ± 1.1
M3d1a+16301+15466	0.7(0.5)	1.8 ± 1.2	Z3a1a1a (13105G+7269A)	2.27 [1.07]	5.9 ± 2.7
M5a + 7094C	1.33 [0.8]	3.45 ± 2	Z3a1a1b (16186T)	1.076 [0.78]	2.7 ± 2
M5a6	0.7 [0.3]	1.8 ± 0.7	Z3a1a1+9452A	1 [0.57]	2.5 ± 1.4
M5b2b	6.8(1)	18.4 ± 2.5	G3a1'2'3	6.8(1.7)	18.4 ± 4.4
M5b2b1b	5(1.6)	13.3 ± 4.6	G3a3	2.2 [0.8]	5.7 ± 2
M5c+472G+1888A!+15896A+16240G+16291T	10.7 [2.49]	29.8 ± 6.5	F1d	7.2(2.1)	19.5 ± 5.4
M5c2	7.5 [1.86]	20.4 ± 4.8	F1d1	3.14 [1.18]	8.2 ± 3
M5d	4.25 [1.25]	11.2 ± 3.2	F1d1a1	1.90 [0.73]	4.9 ± 1.8
M5d1	2.5 [1.18]	5.8 ± 3	F1d1a1b	0.78 [0.25]	2 ± 0.6
M4"67	12(1.4)	33.7 ± 3.6	F1c1a2	3.44 [1.16]	9 ± 3
M30+16234T	6.3 (1.7)	17 ± 4.4	F1c1a2a+15388C	1.8 [0.95]	4.6 ± 2.5
M30d1	4.2(1.1)	11± 2.8	F1c1a2a1	1 [0.33]	2.5 ± 0.8
M30b2	5.7(1.3)	15.3 ± 3.3	F1c1a2a+15680+16058	3.33 [1.52]	8.7 ± 3.9
M30b2b	4(1.4)	10.6 ± 3.6	F1c1a2a2	2 [1.16]	5.2 ± 3
M33a1	8.9(2.7)	24.5 ± 7	U2	18.9(1.5)	55.4 ± 3.9
M33a1a	2.9(1.2)	7.6 ± 3.1	U2b	14.7(1.9)	42 ± 4.9
M35b4+11877	6.6(1.8)	17.8 ± 4.6	U2b1	9.4(1.1)	25.9 ± 2.8
M38	9.5(1.8)	26.2 ± 4.6	U2b1a	6.4(1.6)	17.3 ± 4.1
M38a	7.7(1.6)	21 ± 4.1	U2b1a1	3.2(1.4)	8.4 ± 3.6
M38d	3.5(1.3)	9.2 ± 3.3	U2b1b	8.4(1.5)	23 ± 3.9
M38 + 6366	2.8(0.9)	7.3 ± 2.3	U2b1b1	6.7(1.6)	18.15 ± 4.1
M38d+16311!	2.3(1)	6 ± 2.5	U2b1c	5(1.6)	13.3 ± 4.1
M38d+ 9438	1.5(0.9)	3.9 ± 2.3	U2c1	11.18[1.63]	33.1 ± 4.1
M43	6.6(1.6)	17.8 ± 4.1	U2c1+146C	10.62 [2.35]	29.6 ± 6.1
U2c1c+12361GG+16179T	10[2.08]	27.7 ± 5.4	R6b1a	2.2(1)	2.2 ± 2.5
U2c1c1	8[1.53]	21.9 ± 3.9	R30b2a1	2.2(1)	5.7 ± 2.5
U7a3a2	3(1.1)	7.9 ± 2.8	HV12b1b	4.41 [1.32]	11.7 ± 3.4
U7a3a2a	2.5(0.9)	6.5 ± 2.3	HV12b1b1	3.92[1.25]	10.3 ± 3.2
U7a1a1a	1.7(0.7)	4.4 ± 1.8	1284C+ HV12b1b1	2.4 [1.16]	6.2 ± 3
U7b	3.3(0.4)	8.7 ± 1	HV12b1b1b1a	1.4 [0.6]	3.6 ± 1.5
U7b7	1.5(0.9)	3.9 ± 2.3	H13a1d	2.5[0.8]	6.5 ± 2
R5a2b+16309G!	4.75 [1.67]	12.7 ± 4.3	I1e+14542T	9.63[2.67]	26.6 ± 7
R5a2b6	3.75 [1.34]	9.9 ± 3.4	I1e1	8.63[2.48]	23.7 ± 6.4
R6b	7.5(1.8)	20.4 ± 4.7	I1e1a	2.1[0.65]	5.4 ± 1.6
R6b1	4.2(1.5)	11.1 ± 3.9	N1a1b1a	7 [1]	19 ± 2.5

4.7 Mitochondrial DNA diversity in Ladhak and Changpa

The mtDNA Hgs/sub-Hgs observed among the studied populations were classified into three main groups: East Eurasian, South Asian and West Eurasian. The East Eurasian prevalent Hgs were grouped into East Eurasian ancestries (Xuebin Qi et al., 2013). Accordingly, south Asian and west Eurasian prevalent Hgs were also allocated into South Asian and West Eurasian ancestries (Silva et al., 2017). East Eurasian ancestries where higher among Changpa (**69%**) followed by Ladhak (**37.1%**). Surprisingly, the mtDNA gene pool of Ladhak were dominated by the West Eurasian (42.7%) specific haplogroups. Substantial genetic contribution from South Asian specific lineages were detected in both Changpa (24.1%) and Ladhak (20.2%).

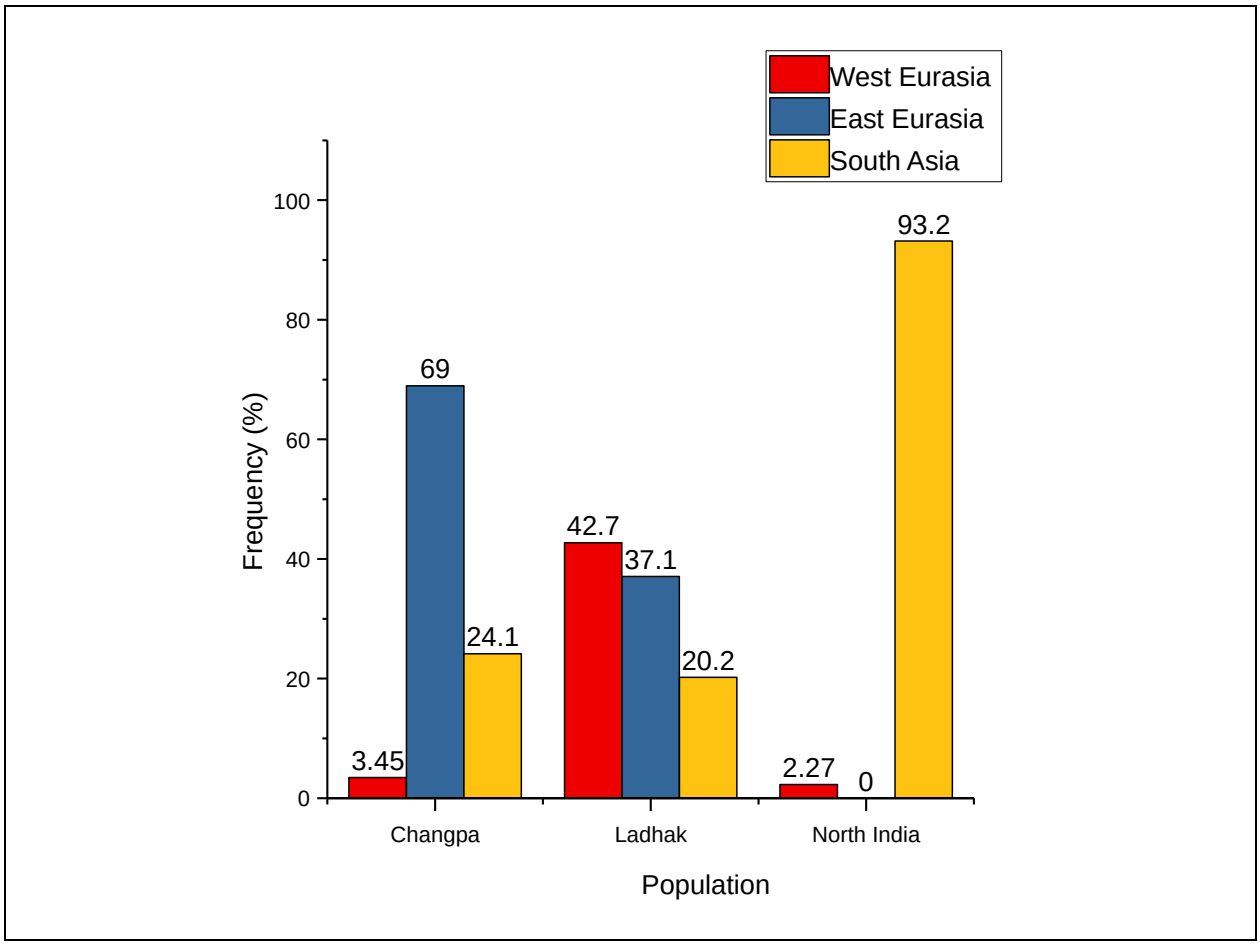


Figure 4.46 | Histograms of mtDNA components observed in the studied populations. mtDNA data for these populations were computed based on the mtDNA Hg frequencies (**Table 4.9**).

4.7.1 Classification of mtDNA sequences

The dominant haplogroups observed in Changpa are hg M9a (18.4%), A (16.1%), C (12.6%), G2 (9.2%) and M34 (8%). These hgs altogether cover 64.36% of the total Changpa sample. All of these haplogroups belongs to East Eurasian ancestries, except M34 which belong to South Asian ancestries. Hg M9a (16.9%), A (12.4%), I (11.2%), HV (8.9%), U2 (8.9%), W (6%) and U7 (5.6%) are the major Hgs observed in general individuals from the Ladhak. Hg M9 and Hg A observed in Changpa belongs to Tibetan (East Asian) ancestries, whereas hg I, HV, W all belongs to the Near Eastern clades. Likewise, the samples from North India are dominated by the south Asian specific

lineage's such as M6 (13.6%), U2 (11.3%), M5 (11.3%), M2 (11.3%), M18 (11.3%), M3 (9%) and M35 (6.8%).

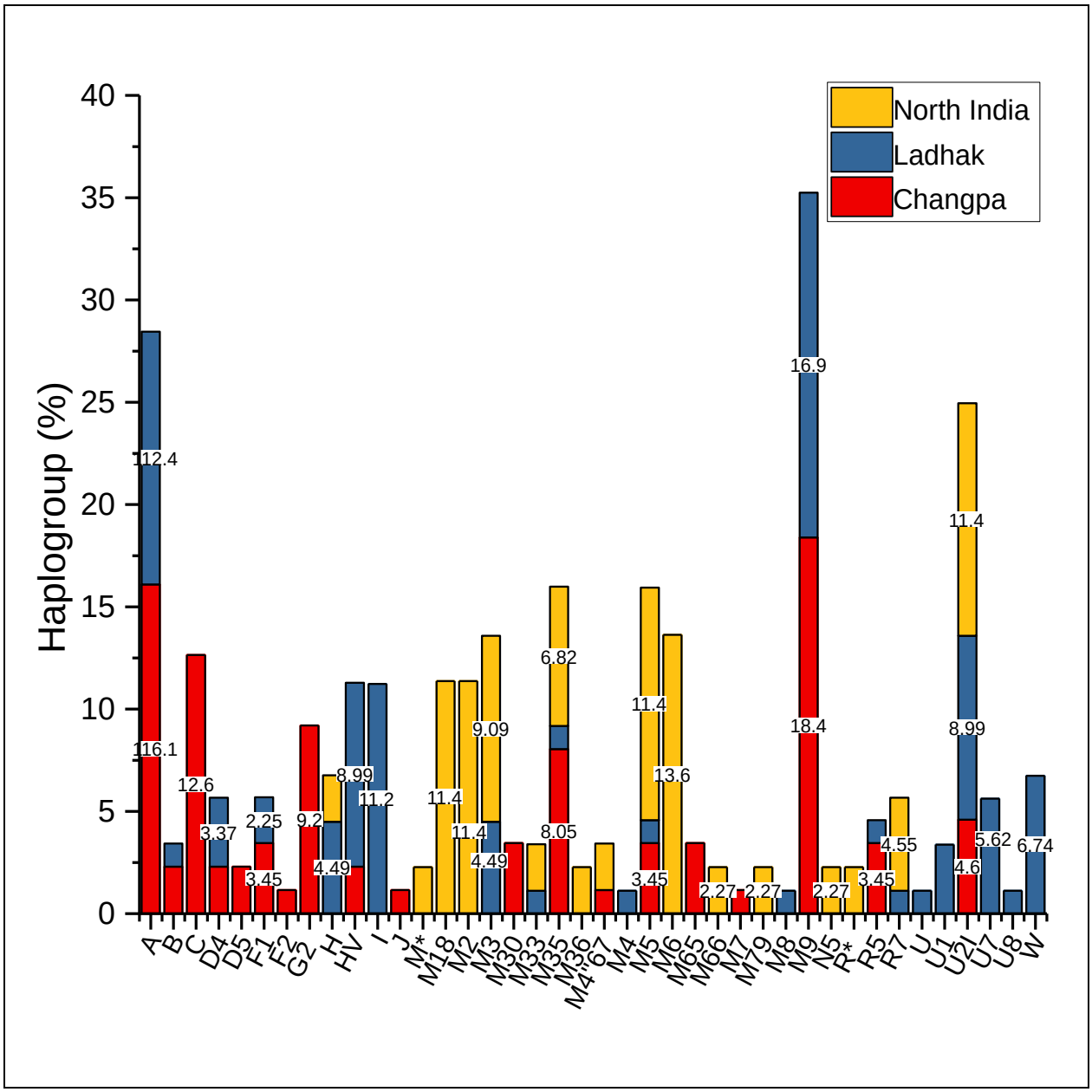


Figure 4.47 | mtDNA Hg frequencies observed in the studied populations. More detail information regarding the mtDNA Hg frequencies is provided in the **Table 4.9**.

In overall, the dominant Hgs observed in Changpa and Ladhak are the bearer of the East Eurasian specific matrilineal lineages; M9a and A. These same haplogroups M9a (22.48%) and A (14.63%) constituted the major Hgs among the Tibetans (X. Qi et al., 2013). The four sub haplogroups of M9a (M9a1a, M9a1a2, M9a1b1and M9a1a1c1b1a) are widely distributed in East Asia and Southeast Asia. Surprisingly, in Ladhak all of the individuals of M9a belong to lineage **M9a1a1c1b1a**, whereas Changpa constitute all of these sub haplogroups of M9a except M9a1a2. Hg M9a1a1c1b1a has been detected in very low frequency among most Asian populations, but highly prevalent in Ladhak (100% of the total M9a individuals), Sherpa (58.16% of the total M9a individuals) (Bhandari et al., 2015), Tibetans (58.06% of the total M9a individuals) (X. Qi et al.,

2013) and Changpa (37.5% of the total M9a). The shared components of sub haplogroup **M9a1a1c1b1a** between Ladhak, Changpa, Sherpa and Tibetans suggests the close maternal relationship between these populations. M9a1c1a1b lineages observed in Changpa and Ladhak are expanded from the Tibetans as suggested by the star like network in which the lineages M9a1c1ab is surrounded by the other haplotypes with few mutation (one, two or three) steps away from the core haplotype (**Figure 4.48**).

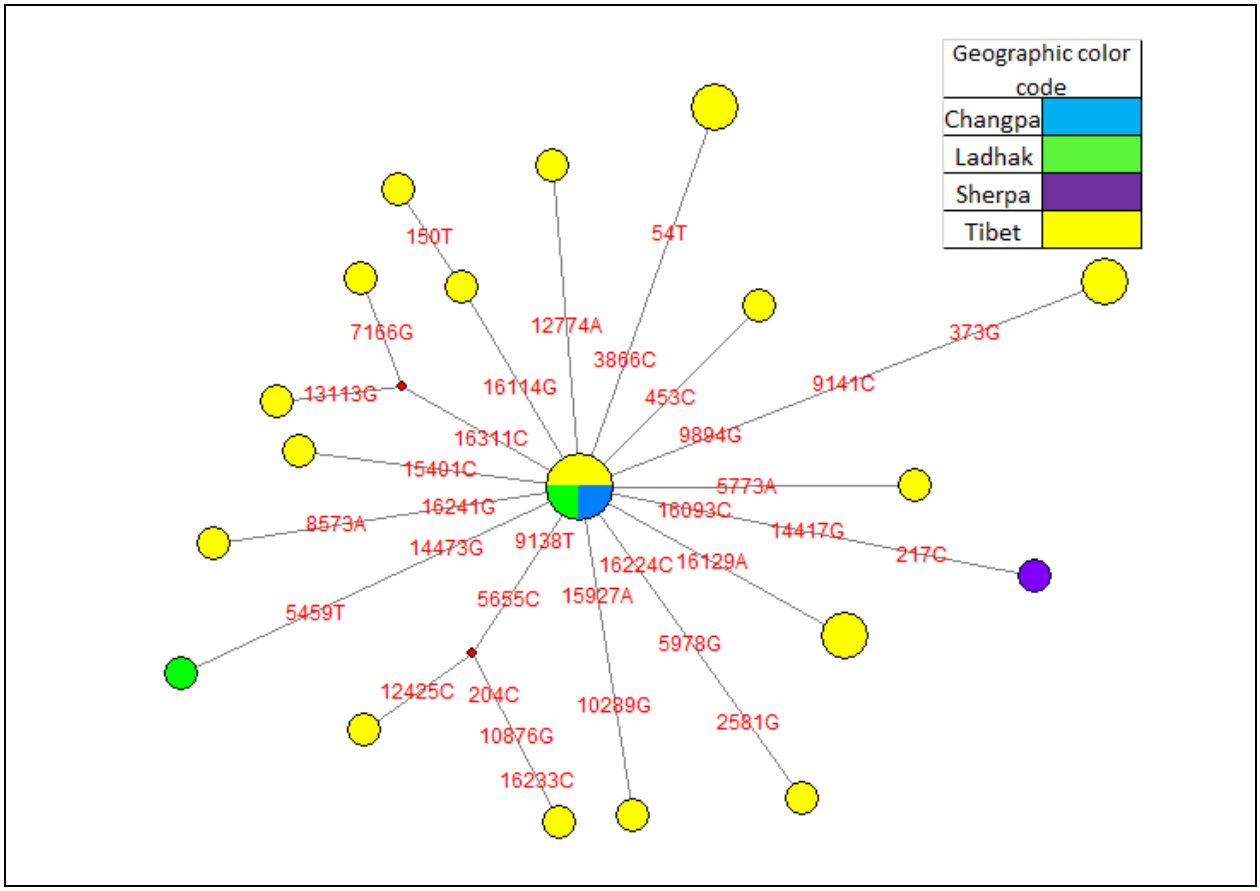


Figure 4.48 | Median Joining Network of Hg M9a1c1ab based on complete mtDNA sequence. The size of the circles is proportional to the number of individual cmtDNA sequences. Nucleotide position number shown in the network are consistent with the Revised Cambridge Reference Sequence (rCRS). The geographic origin of sample is shown by different colors. The C stretch length polymorphism in regions 303–315, AC indels at 515–522, 16182C, 16183C, 16193.1C(C) and 16519 were disregarded for the network reconstruction

The genetic affinity is further supported by the presence of Hg A in high frequency in Changpa (16.1%), Ladhak (12.4%) and Tibetans (14.63%). Network analysis showed that A11a lineage observed in Ladhak, Changpa and Tibetans are derived from the ancestral variant (16093C). Similarly, A21 lineages observed among the Changpa and Ladhak share some Internal Haplotypes with the Tibetans. Further, Changpa and Ladhak belonging to lineage A21 and A11a, shares basal haplotypes with the Tibetan samples, in which most individuals from Changpa and Ladhak lies at the tips, suggesting a relatively recent lineage expansion among Changpa and Ladhak. East Eurasian specific haplogroup G2a1 was also observed in Changpa (9.19%), but was completely absent in Ladhak.

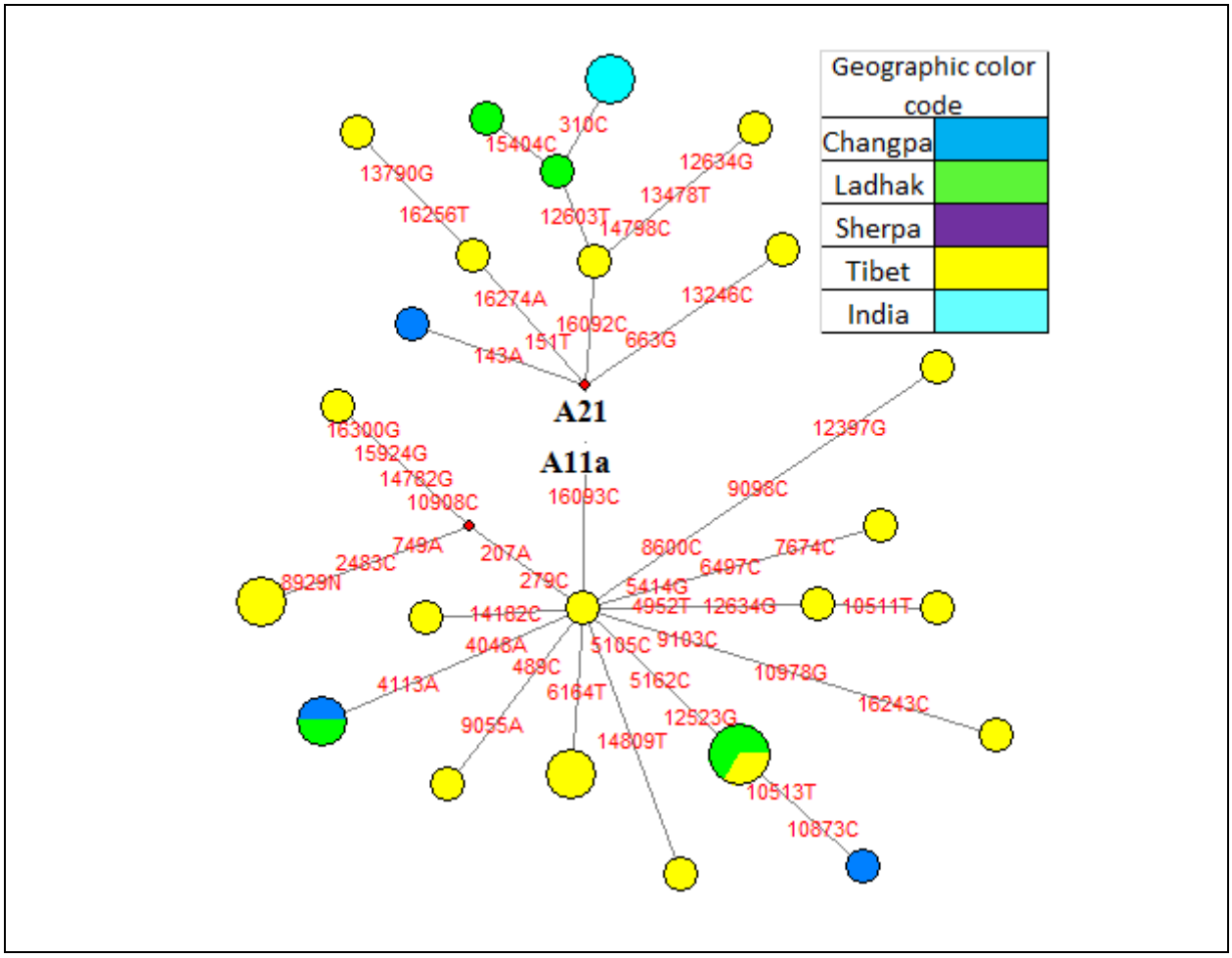


Figure 4.49 | Median Joining Network of Hg A11a and A21 based on complete mtDNA sequence. The size of the circles is proportional to the number of individual cmtDNA sequences. Nucleotide position number shown in the network are consistent with the Revised Cambridge Reference Sequence (rCRS). The geographic origin of sample is shown by different colors. The C stretch length polymorphism in regions 303–315, AC indels at 515–522, 16182C, 16183C, 16193.1C(C) and 16519 were disregarded for the network reconstruction

Intriguingly, Changpa/Ladhak specific haplogroup C4a2c2a were not detected in Tibetans. This sub-haplogroup accounts 12.6% of the maternal lineage of Changpa. As shown in the phylogenetic tree, expansion of lineage C4a2c2a has occurred very recently. The coalescence age of the clade A21 and newly defined clade A11a1 was estimated to be ≈2.5 Kyr and ≈1.8 Kyr respectively. Network analysis (ancestral variant 16093C) along with the younger age (≈1.8Kyr) of A11a suggests a recent bottleneck during migration and latter population expansion in Tibet followed by Ladhak. Similarly, age of the Changpa specific clade C4a2c2a was estimated to be ≈3.1Kyr. The estimated younger age of these basal lineages is in consistent with the proposed recent expansions of the East Eurasian components in Ladhak from Tibet.

Other haplogroups detected in low frequency in Changpa are

1. East Asian: D (4.59%), F (4.5%) and B (2.3%).
2. West Eurasian: HV (2.29%) and J (1.1%),
3. South Asian: U2 (4.5%), M30 (3.4%), M5 (3.44%), R5 (3.4%) and M4"67 (1.1%).

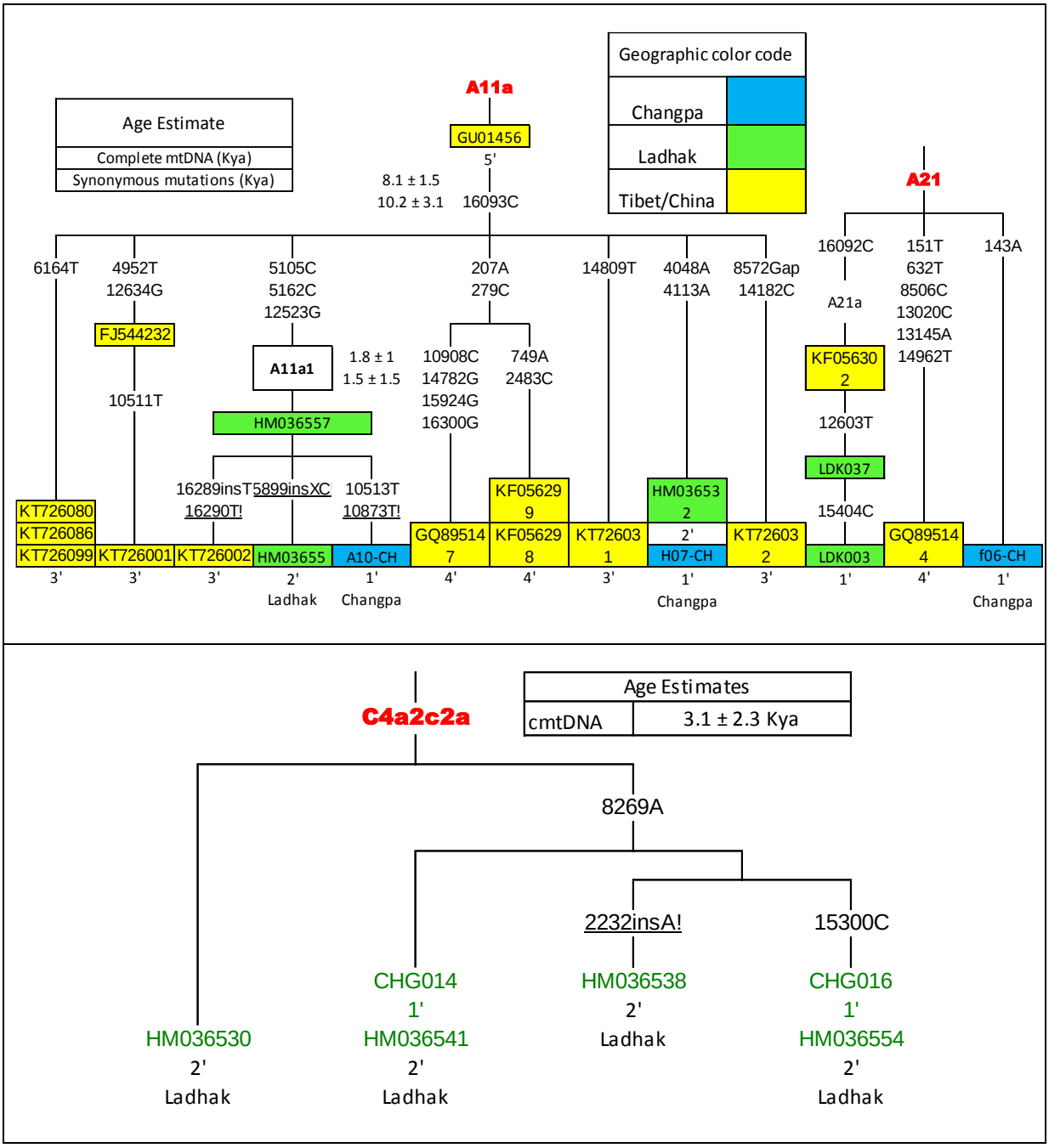


Figure 4.50 | Phylogenetic tree of Hg A21, A11a and Ladhak specific sub Hg C4a2c2a based on the complete mtDNA sequences. Mutations relative to the RCRS are indicated on the branches. Sample ID/GenBank accession number for each sequence are shown in the box with geographic color code. The number (1', 2'... n') corresponds to the reference for the analyzed sample. The hg written in black color represents newly defined hg in this study, whereas red color indicates the predefined Hg in PhyloTree (Oven, 2015).

Haplogroup I4b shows the signal of dispersion from West Eurasia (mainly from Near East) into South Asia, most probably occurred during the Bronze Age. However, mtDNA study showed the genetic influx might have restricted into Pakistan, Afghanistan and few reached up to North India. In present study high frequency of hg I4b was observed in Ladhak (11.2%) but completely absent

in Changpa. Further complete mtDNA analysis revealed the sharing of basal Variant 8768T and 11530G between the individual from Pakistan, Afghanistan and Ladhak to form a new haplogroup newly named herein as I4b1. The coalescence age of I4b1 was estimated to be ≈3.1 Kyr.

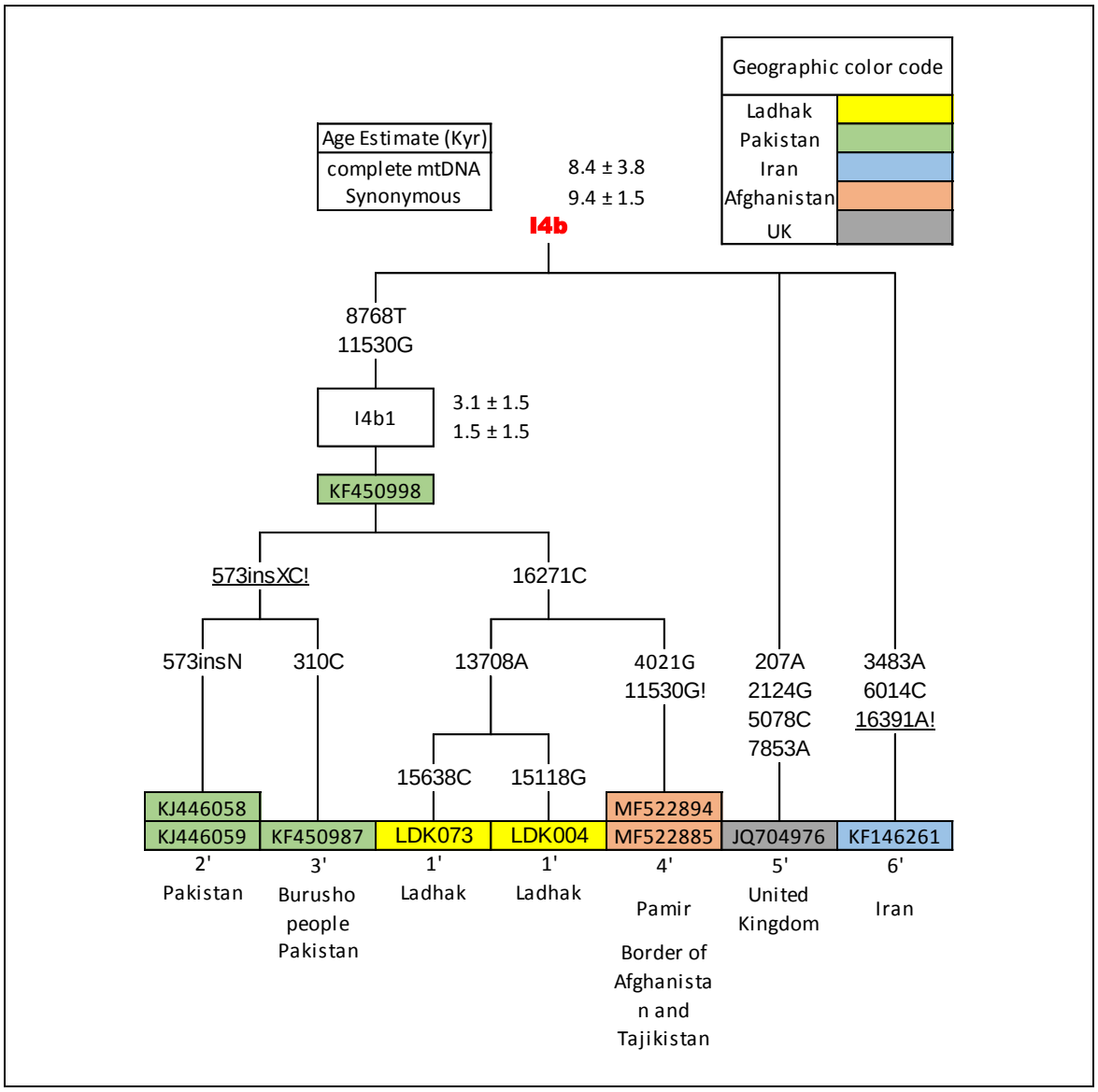


Figure 4.51| Phylogenetic tree of haplogroup I4b based on the 10 complete mitochondrial sequences of which 2 cmtDNA sequences were generated in this study. Mutations relative to the RCRS are indicated on the branches. Sample ID/GenBank accession number for each sequence are shown in the box with geographic color code. The number (1', 2'... n') corresponds to the reference for the analyzed sample. The hg written in black color represents newly defined hg in this study, whereas red color indicates the predefined Hg in PhyloTree (Oven, 2015).

Hg HV12 is mainly found in Iran (M. Derenko et al., 2013). HV12a and HV12b, subclades of Hg HV12 have been reported in the Near Eastern populations **(Table 8.2)**. Only the lineage of HV12b (HV12b1 and HV12b1a) were observed in India and Nepal. In the present study, hg HV12b1 of Ladhak traces its ancestry to the populations from Iran and Armenia (Near East) as shown in the **Figure 4.18**. South Asian specific hg U2b detected in Ladhak (6.7%) are mainly present in India. Except India, Sub hg U2b1 is also reported to be present in Nepal, Myanmar, Thailand and Tibet. Other haplogroups detected in Ladhak are

1. West Eurasian: W (6.7%), U7a (5.6%), H (4.49%), U1 (3.1%) and U8 (1.1%),
2. South Asian: M3 (4.5%) and those with ≈1% frequency (M35, M33, M4, M5, R5 and R7)
3. East Eurasian: B (1.1%), D3 (3.3%), F (2.2%) and M8 (1.1%)

Table 4.9 | mtDNA haplogroup frequencies (%) observed in Changpa, Ladhak and North India

Haplogroup	Ladhak	North India	Changpa
A+152+16362+200	0.0	0.0	2.3
A11a	1.1	0.0	6.9
A17	0.0	0.0	2.3
A21	11.2	0.0	1.1
A6	0.0	0.0	1.1
A6b	0.0	0.0	2.3
B4a4	0.0	0.0	2.3
B4b1a3a	1.1	0.0	0.0
C4a1a5	0.0	0.0	2.3
C4a2	0.0	0.0	1.1
C4a2c2a	0.0	0.0	8.0
C5	0.0	0.0	1.1
D4j1a1	1.1	0.0	0.0
D4j1a2	0.0	0.0	2.3
D4j1b	1.1	0.0	0.0
D4j3a	1.1	0.0	0.0
D5a2a1	0.0	0.0	2.3
F1b1+@152	0.0	0.0	2.3
F1c1a1a	2.2	0.0	1.1
F2+16291	0.0	0.0	1.1
G2a1	0.0	0.0	6.9
G2a1h	0.0	0.0	2.3
H6a	1.1	0.0	0.0
H6a1b	2.2	0.0	0.0
H7b	0.0	2.0	0.0
HV	1.1	0.0	0.0
HV12b1	4.5	0.0	1.1
HV14	2.2	0.0	0.0
HV14a	1.1	0.0	0.0
HV2	1.1	0.0	0.0
HV20	0.0	0.0	1.1
I4b	11.2	0.0	0.0
J1b1a1	0.0	0.0	1.1
M	0.0	2.0	0.0
M18	0.0	10.2	0.0
M2	0.0	10.2	0.0
M30	0.0	4.1	0.0
M30b	0.0	4.1	0.0
M30c1a	0.0	0.0	3.4
M33	0.0	2.0	0.0
M33+16362	1.1	0.0	0.0
M35b3	1.1	0.0	8.0
M35	0.0	6.1	0.0
M36b	0.0	2.0	0.0
M3a2	2.2	0.0	0.0
M3d1	2.2	0.0	0.0

Haplogroup	Ladhak	North India	Changpa
M4	1.1	0.0	0.0
M4"67	0.0	2.0	1.1
M5a	0.0	10.2	0.0
M5a2a	1.1	0.0	0.0
M5b2b1a	0.0	0.0	3.4
M65a+@16311	0.0	0.0	2.3
M65a2	0.0	0.0	1.1
M66	0.0	2.0	0.0
M6a1a	0.0	12.2	0.0
M79	0.0	2.0	0.0
M7c1a1a	0.0	0.0	1.1
M8a	1.1	0.0	0.0
M9	0.0	0.0	3.4
M9a1a	0.0	0.0	1.1
M9a1a1a	0.0	0.0	2.3
M9a1a1c1b	0.0	0.0	1.1
M9a1a1c1b1a	16.9	0.0	2.3
M9a1a1c1b1a2	0.0	0.0	3.4
M9a1b	0.0	0.0	1.1
M9a1b1c	0.0	0.0	3.4
N5	0.0	2.0	0.0
R	0.0	2.0	0.0
R5a2	1.1	0.0	1.1
R5a2b	0.0	0.0	2.3
R7	1.1	0.0	0.0
R7b	0.0	4.1	0.0
U	1.1	0.0	0.0
U1a	1.1	10.2	0.0
U1a1c1d	2.2	0.0	0.0
U2a	2.2	2.0	0.0
U2a1	0.0	2.0	0.0
U2a2	0.0	2.0	0.0
U2b	6.7	0.0	0.0
U2b1	0.0	2.0	1.1
U2c1	0.0	2.0	0.0
U2c1a	0.0	0.0	1.1
U2c1b	0.0	0.0	2.3
U7a	3.4	0.0	0.0
U7a3b	2.2	0.0	0.0
U8b1a1	1.1	0.0	0.0
W	1.1	0.0	0.0
W3a	1.1	0.0	0.0
W3b	1.1	0.0	0.0
W4	2.2	0.0	0.0
W4a	1.1	0.0	0.0

4.8 Ectodysplasin Receptor (EDAR).

Adaptive non-synonymous Allele rs3827760, also known as 1540T/C, 370A or Val370Ala, is a Derived coding variant (SNP) in the Ectodysplasin A (EDA) receptor (EDAR) gene on chromosome 2. Several studies have shown the high frequencies and restricted distribution of the derived variant 1540C allele especially in the East Asians and the Americans suggesting that the allele 370A was positively selected somewhere in Central China approximately 30,000 years ago (Kamberov et al., 2013). This variant is completely absent in the African and the European populations. This variant was either completely absent or present in very low frequency among the Indo-European speaking populations from India, Pakistan and central Asia. The EDAR profile of the studied populations in the present study is given below (**Figure 4.52**).

Population		Shrestha & newa mix	Maharjan	Shakya	Bajrachary a	Udaya	Magar	Brahmin
Gneotype Frequency (%)	**GG / GA / TC AA / TT**	6	8	5	10	7	54	4
		42	40	42	25	44	35	33
		52	52	53	65	49	12	63
frequency (%)	**A / T**	73	72	67	72	65	29	79
	G / C	27	28	33	28	35	71	21

Figure 4.52 | Frequency of 1540C Allele of EDAR Gene in Nepal. In Nepal derived variant 1540C/T has been observed in the populations of the East Asian origin. The allele frequency was estimated by using the simple gene counting method.

In Nepal, 1540C allele was observed mainly among the Tibeto-Burman speaking populations. However its frequency varied greatly between the Magar and Newar. In the present study, Magar show the highest, Brahmin shows the least whereas Newar shows the intermediate frequency of the derived allele 1540C (**Figure 4.52**). For Detail information on the EDAR profile of the Nepali and other populations see **Table 8.7** (Appendix I).

4.9 Endothelial PAS domain protein 1 (EPAS1)

4.9.1 Signal of selection

Changpa are genetically closer to Tibetans. Non-Tibetan CHB/JPT, which reside in a low altitude region are on average closely related to Tibetans. Hence any specific region in the gene which shows a deep divergence between the CHB/JPT (Han Chinese in Beijing, China/ Japanese in Tokyo, Japan) and Tibetan (including Changpa) possibly indicate that those genetic loci might have been positively selected for the high-altitude adaptation.

Table 4.10 | Signal of natural selection identified in Changpa and Tibetan population using cross-population composite likelihood ratio (XPCLR). Symbol "†" indicates number of SNPs in windows. Genes with strongest frequency changes in Changpa population. Data derived from whole genome SNP.

Chr	#†	Physical position	XP-CLR score	Gene	Chr	#†	Physical position	XP-CLR score	Gene
2	15	46591833	53.25637	EPAS1	17	6	29246888	67.224235	ADAP2
2	16	46593833	72.43459	EPAS1	17	6	29248888	69.279027	ADAP2
2	15	46595833	52.2641	EPAS1	17	6	29250888	65.084368	ADAP2
2	14	46597833	69.86909	EPAS1	17	6	29252888	60.307282	ADAP2
2	15	46599833	97.33317	EPAS1	17	6	29254888	59.648904	ADAP2
2	14	46601833	71.93769	EPAS1	17	6	29256888	52.133302	ADAP2
2	14	46605833	80.12476	EPAS1	17	9	62276888	51.788352	TEX2
2	13	46607833	105.7098	EPAS1	18	9	34337591	56.812145	FHOD3
2	10	46609833	92.05422	EPAS1	18	8	34341591	74.391311	FHOD3
2	8	46611833	64.09962	EPAS1	18	8	34343591	76.72874	FHOD3
2	8	46663833	49.73342	TMEM247	18	7	34345591	63.681075	FHOD3
2	8	46705833	52.43413	TMEM247	18	7	34347591	61.98495	FHOD3
13	4	114469878	64.58926	TMEM255B	18	22	56719591	52.178008	OACYLP
13	4	114471878	88.37826	TMEM255B	18	5	76995591	54.683312	ATP9B
13	4	114473878	74.87649	TMEM255B	20	11	50252347	56.372605	ATP9A
13	4	114475878	72.16733	TMEM255B	22	8	28426858	54.591591	TTC28
					22	7	28428858	50.189638	TTC28

In present study, 725,634 SNPs were utilized to explore signals of selection. Since, F_{st} value between Tibetan refugees and Changpa population was 0±0.0007, both populations were merged, to increase power to detect signal of selection. Before calculation of F_{st}, pruning was performed and excluded 193,014 SNPs, with r2>0.75.

In the present study, only those regions under natural selection were considered, which scores top 99.995 percentile of XPCLR (score >49.71) and identified 10 candidate genes for high-altitude adaptation in Ladhak (**Table 4.10** and **Figure 4.53**). Of which, *EPAS1* was already reported with maximum XPCLR score equal to 105. The other candidate genes for high altitude adaptation has been newly reported in this study.

The other genes are: Transmembrane protein (TMEM) 255B and 247, ArfGAP with dual PH domains 2 (ADAP2), ATPase Phospholipid Transporting 9A (ATP9A), and Tetratricopeptide Repeat Domain 28 (TTC28). All of these genes are protein coding genes, however function of these gene on high altitude adaptation are unknown. Function of TMEM gene is unknown, whereas protein encoded by ADAP2 and TTC28 genes binds beta tubulin and increase the stability of microtubule. Likewise ATP9A helps in ion channel transport and cardiac conduction (Gene card: http://www.genecards.org; NCBI Gene: www.ncbi.nlm.nih.gov/gene/).

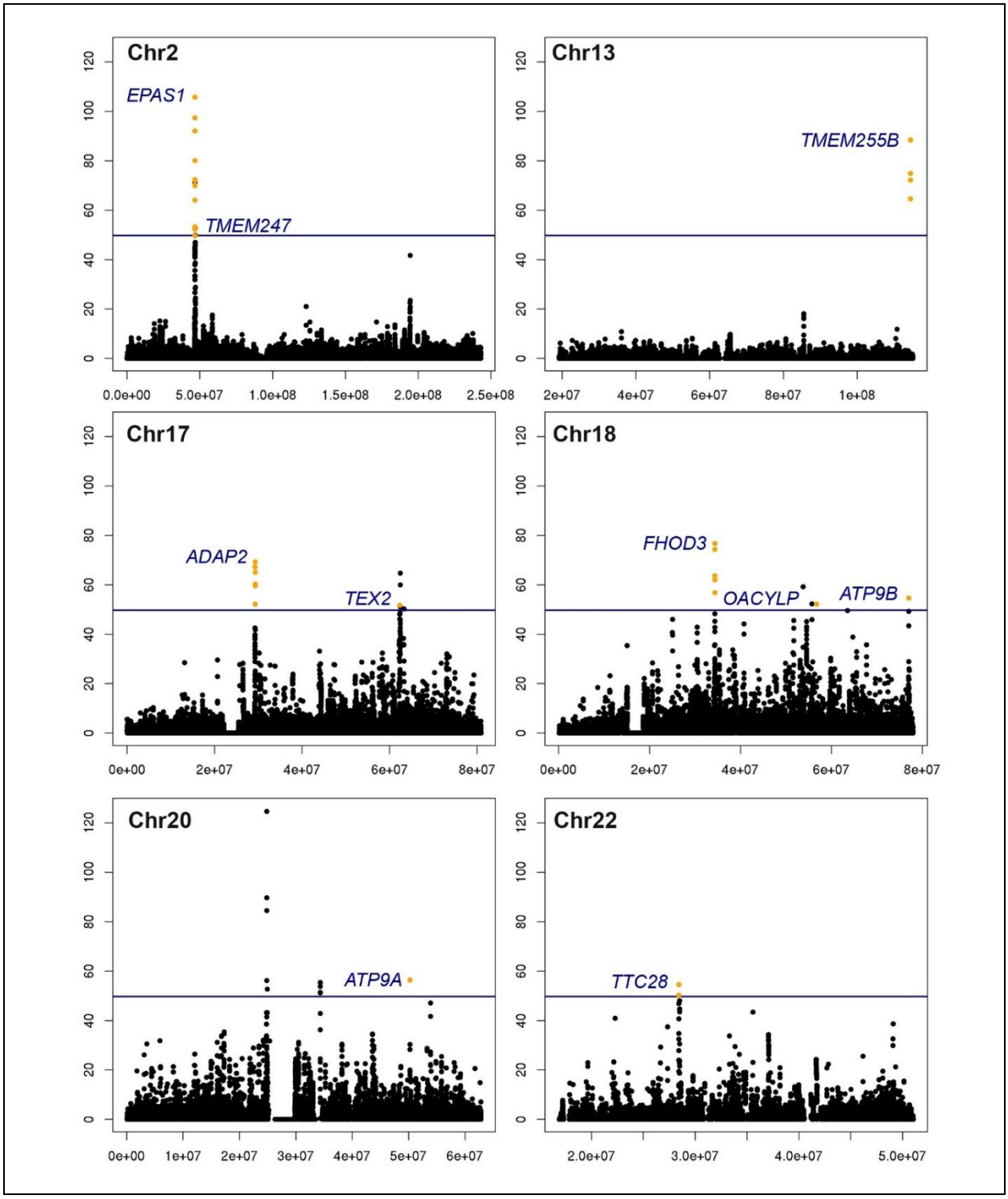


Figure 4.53 | Signal of natural selection identified in Changpa and Tibetan population using XPCLR. The orange dots represent the genetic region which are under natural selection and within gene. In brief, orange dots represent the Specific region in the gene which shows a deep divergence between the CHB/JPT (Han Chinese in Beijing, China/ Japanese in Tokyo, Japan) and Tibetan (including Changpa) strongly indicating that those genetic region might have been positively selected for the high-altitude adaptation. The strong signal of positive selection came from EPAS1 gene with maximum XPCLR score equal to 105.

4.9.2 Association between genetic variants of EPAS1 gene and altitude

Since, Changpa population analyzed collectively with Tibetan refugees, 4 SNPs of EPAS1 gene were also genotyped, rs150877473 (C>G), rs13419896 (G>A), rs1868092 (G>A) and rs1868093 (G>A) in 695 subjects from Nepal and Ladhak, who were living from 1300-4000-meter altitude.

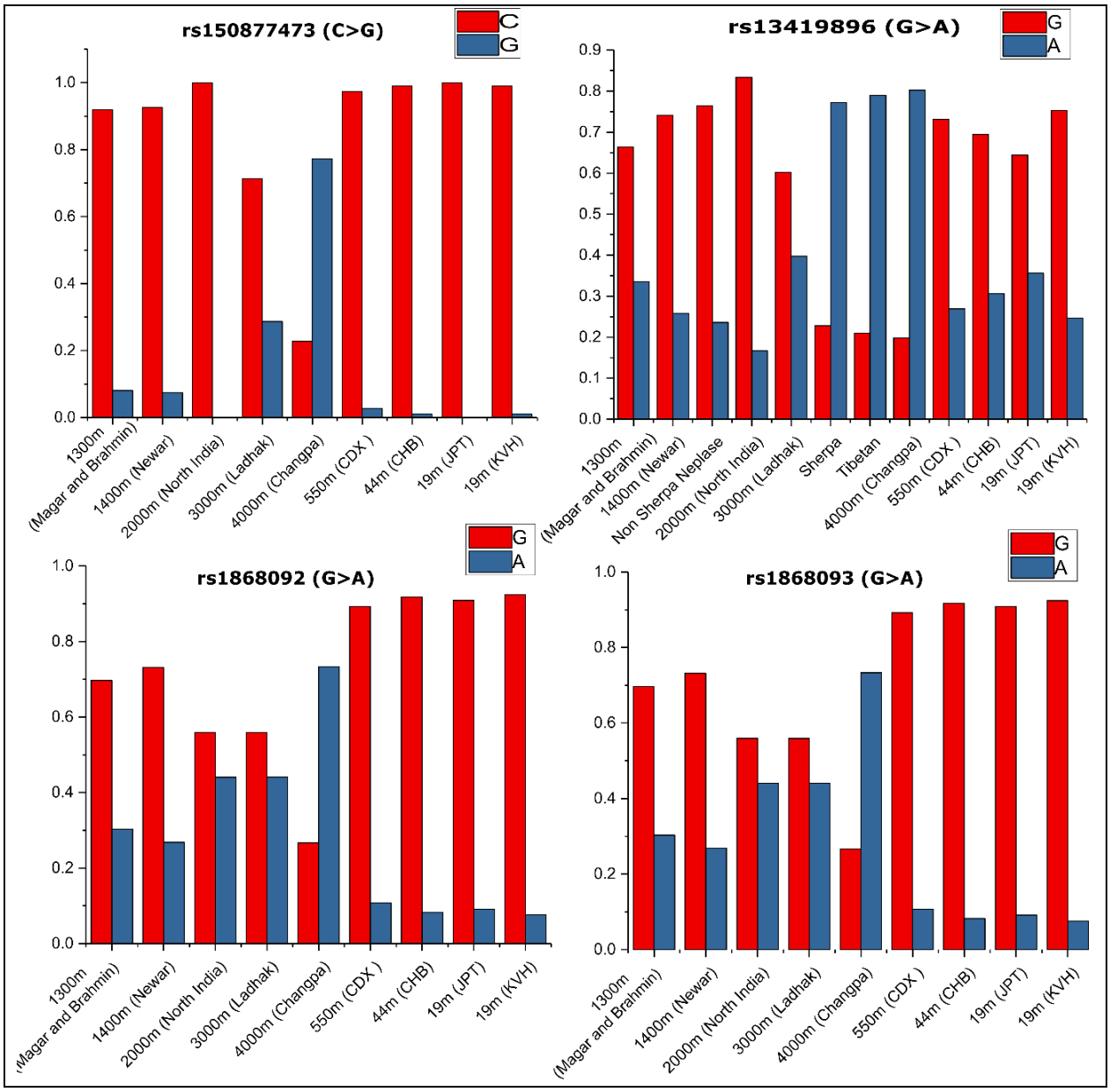


Figure 4.54 | Frequency of four SNPs analyzed in this study. Red colour indicate the ancestral allele, whereas blue color indicates the derived allele. Genetic data for Tibetan, Sherpa and Non Sherpa Nepalese were obtained from reference (Hanaoka et al., 2012).

SNP rs150877473 (C>G) shows **78 %** higher frequency in Tibetans than in Han Chinese (Yi et al., 2010). Given that the Han and Tibetan diverged 2,750 years ago, the author described it to be the fastest change in allele frequency observed at any human gene to date. In a good agreement to the previous study, Changpa population also shows **76%** higher frequency of rs150877473 (C>G). Whereas rs150877473 (C>G) was observed in low frequency in non-Changpa populations from Ladhak (3000m). SNP rs13419896(G>A) observed in higher frequency in Sherpa (77%) and Tibetan

(79%) (Hanaoka et al., 2012) were also observed in higher frequency in Changpa (80%). Similarly, 39% of derived SNP rs13419896 (G>A) were observed in Ladhak (3000m). Frequency of the haplotype of alleles rs1868092 (G>A) and rs1868093 (G>A) were higher in Changpa (73%) and Ladhak (44%). Hence the frequency of Tag SNPs in Changpa, paralleled with those observed in Tibetan as well as Sherpa. Further, positive correlation of derived allele with altitude for each SNP was identified using the Pearson correlation analysis as shown in **Figure 4.55**.

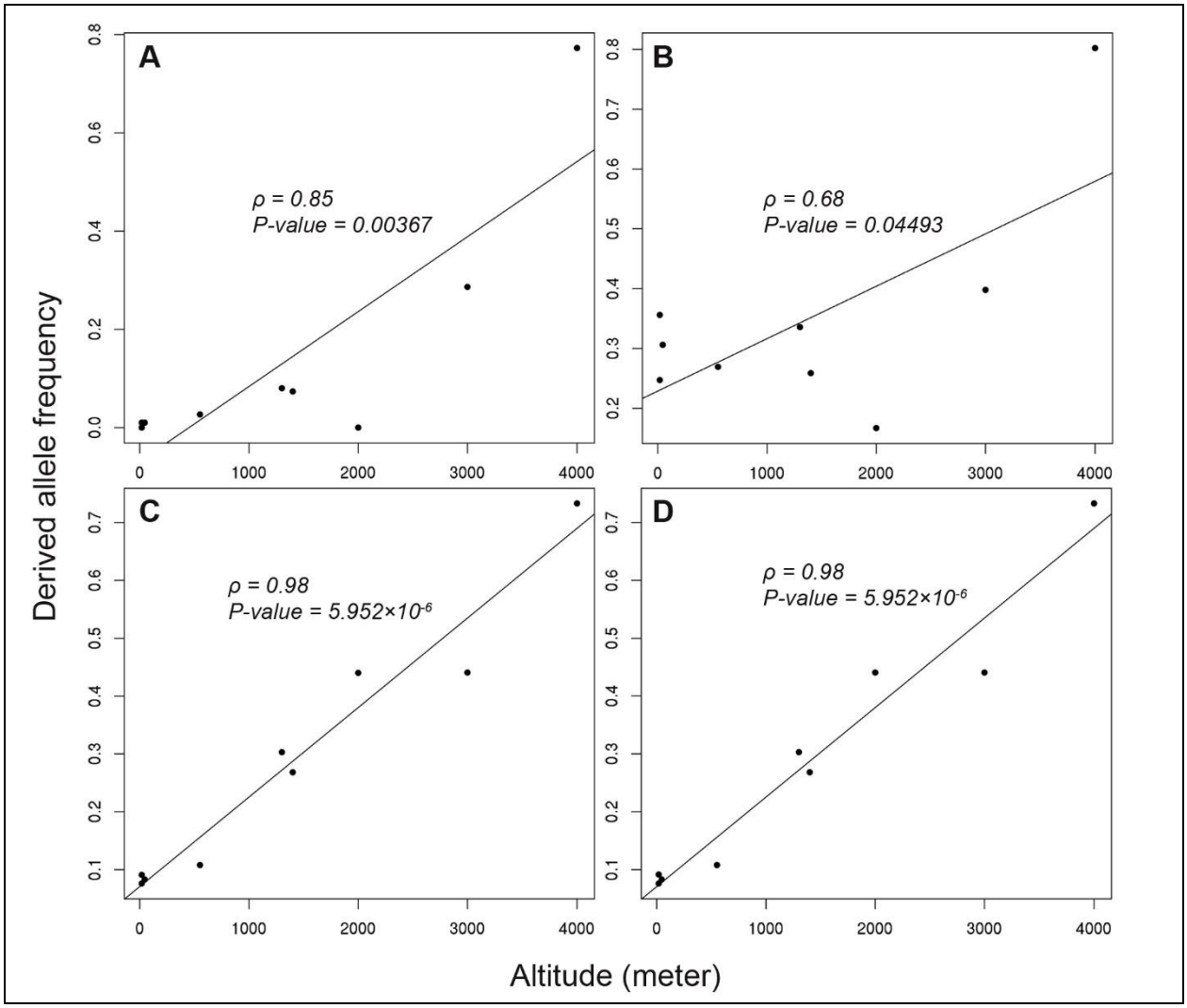


Figure 4.55 | Correlation of (A) rs150877473G, (B) rs13419896A, (c) rs1868092A and (D) rs1868093 A, with altitude. The scatter plot shows the measure of strength between the two variables; altitude and Frequency of the genotyped SNPs. Positive correlation (p) between the genotype allele and altitude was observed in the present study, indicating that both variables increase or decrease together. The p-value demonstrates the significance of correlation using Pearson correlation analysis. Letter A, B, C and D represents the correlation between altitude and the four different genotyped SNPS, rs150877473 (C>G), rs13419896 (G>A), rs1868092 (G>A) and rs1868093 (G>A) respectively. Populations analyzed in the present study and their corresponding altitude is given as: 1300m (Magar and Brahmin), 1400m (Newar), 2000m (North India), 3000m (Ladhak) and 4000m (Changpa). Likewise, the population at altitude 550m, 44m, 19m and 19m belongs to CDX (Chinese Dai in Xishuangbanna), CHB (Han Chinese in Beijing, China), JPT (Japanese in Tokyo, Japan) and KVH (Kinhin Ho Chi Minh City) respectively (http://hapmap.ncbi.nlm.nih.gov/index.html.en).

5 Discussion

Indigenous South Asian mtDNA lineages

Hg M and R are the two major founder clades for the south Asian specific Hgs observed in Nepali populations. It is well accepted fact that the early out of Africa migrants colonized South Asia in between 45-65 Kyr ago after the Mount Toba eruption. This route is considered to be the major route of dispersal of AMH to the other geographical regions mainly to the South-East Asia and Oceania. The presence of high genetic diversity among the Indian populations, the second highest after the sub-Saharan populations suggests the initial colonization of the AMH in South Asia occurred mainly in West India (Behar et al., 2008; Kivisild et al., 2004; Macaulay et al., 2005; Mishmar et al., 2003; Thangaraj et al., 2005). Macro Hg M and R is assumed to have arrived in the West of the Indian sub-continent. After the initial settlement, Hg M might have undergone re-expansion in the other regions of South Asia, during the drier period which occurred between 30-35 Kyr ago. The drier period might have triggered the population movements followed by the expansion of the lineage in the South Asia including Nepal. The evidence for this movement is provided by the presence of oldest AMH fossils found in Sri Lanka which date back to ≈30 Kyr BP (Silva et al., 2017). This is also the earliest dated record of AMHs in South Asia. Finding of Neolithic tools in the ancient city Kathmandu also suggests that people were living in the Himalayan region in the distant past (Desai & Sharma, 2003). These microlithic technology can be traced to ≈45 Kyr ago in Central India. This suggest the Central India to be the likely source for the re-expansion of South Asian specific lineage. **M4'67, M35, M52** are the ancient and indigenous mtDNA lineages originated in **Central India** (Silva et al., 2017). M4'67 which dates back to ≈40 Kyr ago shows an extraordinary branching giving rise to several Hgs including M4, M18, **M30, M38 and M43**. Majority of these ancient South Asian specific Hgs were observed in the Nepali populations including Newar.

Table 5.1 | Frequency of ancient and indigenous South Asian Hgs present among the Nepali populations. All the Newar sub caste are included in a single group as Newa total. Likewise all Tharu from Chitwan and Eastern Nepal are included in a single group and named as Tharu (**Table 8.1**).

Hg	Newar Total	Brahmin	Magar	Tamang	KTM	Nepali-other	Tharu
U2c	1.8	10.4	2.7	0.0	2.6	5.3	0.6
M52	0.7	2.1	10.8	0.0	0.0	2.8	0.6
M33	1.8	0.0	2.7	2.3	6.5	2.8	4.6

Hg M52b and U2 are the most ancient South Asian lineages observed in the Nepali populations, with average coalescence time estimates of ≈29,000 and ≈27,000 years respectively. **In overall, these haplogroups mentioned in the Table 5.1 contribute ≈5% of the total Newar gene pool**. List of ancient putative south Asian Hgs with their corresponding estimated ages is given below:

1. M52b [29.8 ±4.1 Kya]: **Magar (10.8%),** Maharjan (2%), Newar-mix (2%), Brahmin (2.98%), East Tharu (2.5%) and Nepali-other (2.84%).
2. U2c1c+12361G+16179T [27.7±5.4]: **Brahmin (10.4%),** Shrestha (5.55%), Shakya (5.26%), Maharjan (4%) and Magar (2.7%).
3. M33a1 [24.5±3.1]

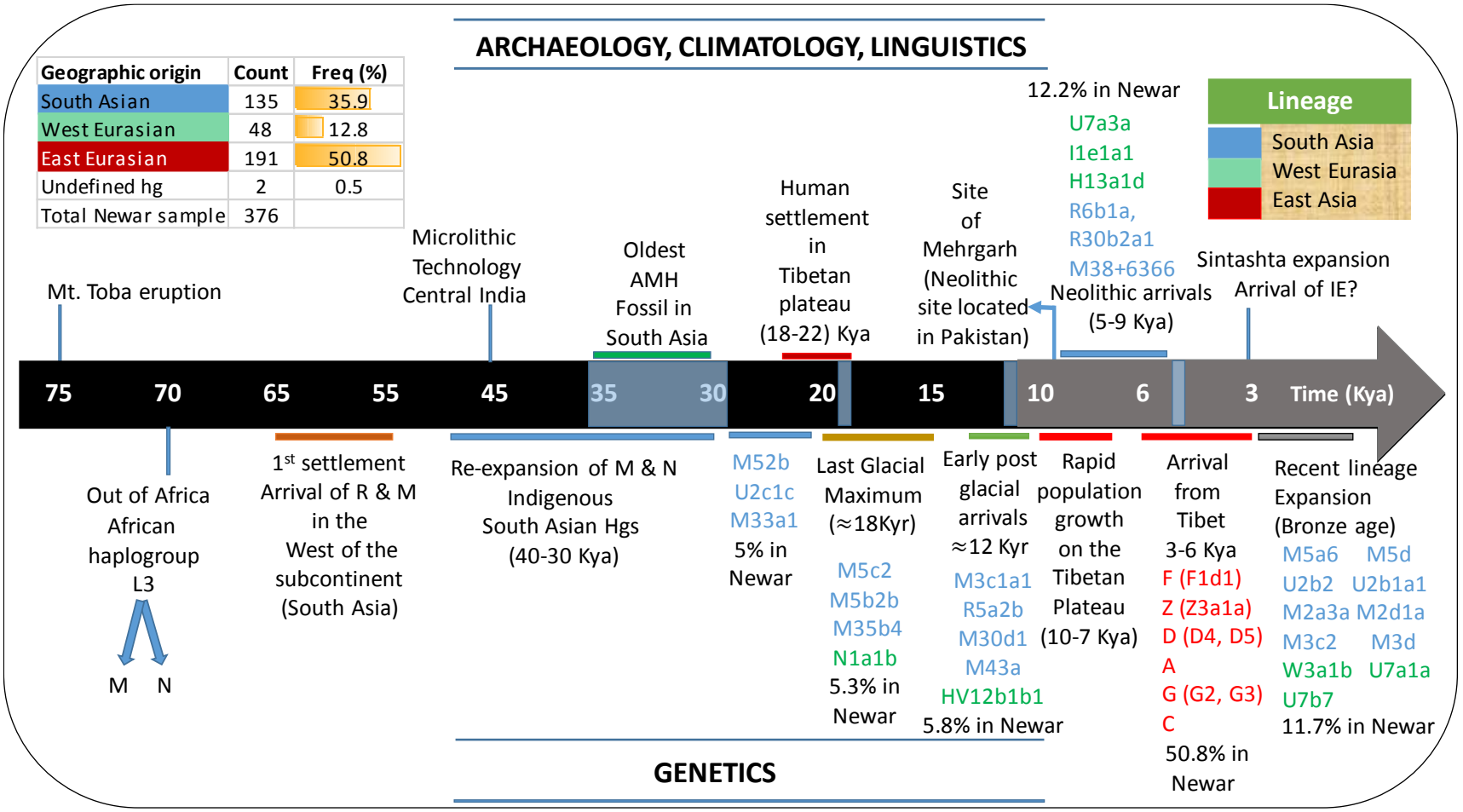


Figure 5.1 | Timeline for AMH evolution in South Asia, Tibet and Nepal based on genetic, archaeological, climatological and linguistic evidence. Black and grey portions of the arrow represent Pleistocene and Holocene, respectively. Blue sections correspond to periods of climate changes: dryer periods between 35 and 30 Kya, Last Glacial Maximum ~18 Kya, Younger Dryas ~12 Kya and the "4.2 Kya" event (Mellars et al., 2013; Silva et al., 2017; Qi, X et al. 2014). Lineages shown in the timeline belongs to Newar, Brahmin and Magar. Lineages in red stand for the genetic influx from East Eurasia; green for West Eurasia and blue from South Asia. Recent lineage expansion indicates the genetic events in the last 4 Kya. The major admixture event in Newar might have occurred from East Asia (50.8%) as shown by the red colored lineages.

Ancient Hg M52b detected in High frequency **(10.8%)** in Magar were observed in lower frequency in other Nepali populations. Sub lineages belonging to hg R (R30, R7, R31, and U2) are the ancient mtDNA lineages present in South Asia. Hg U2 is considered to be the oldest lineage in Europe and South West Asia (region at the crossroads between Asia, Africa and Europe). Two oldest Homo sapiens DNA samples of old Cro-Magnons from the Kostenki site on the Don River in the Russia, dated to ≈37,000 and 33,000-year belonged to Hg U2 (Fu et al., 2016). U2c1 lineage detected in Thailand branched off directly from the node occupied exclusively by the Nepali sample suggesting a gene flow from Nepal to Thailand. In present study, incidence of U2c were higher in Brahmin (10.4%) followed by Shrestha (5.55%), Shakya (5.26%), Maharjan (4%) and Magar (2.7%). These Hgs were also detected in Hindu from Terai, Kathmandu and Nepali-other group (**Figure 4.21**). M33a1 lineage detected in Nepal and Sikkim shows a deep coalescence age estimated to be ≈24.5 Kyr old. Presence of Nepal specific South Asian lineages M33a1 in Sikkim reflected the ancient connection between Nepal and Sikkim.

Last Glacial Maximum (LGM) and Late Glacial arrivals

After the first initial settlement ≈45-66 Kya, the evidence for the movement from West Eurasia to the South Asia is provided by the lineage N1a1b (Silva et al., 2017). LGM corresponds with the time of ≈ (22–18) Kya BP. The Putative Glacial/postglacial genetic flux from West Eurasia (with a probable source in the near East) into South Asia are strengthened by the detection of Lineage N1a1b (N1a1b1 and I1e) in Brahmin and Manandhar group of Nepal

Other similar Hgs include Hg T2. Hg T2 has been observed in Bajracharya, Brahmin and Kathmandu. Hg I1e, the Sister clade of N1a1b1 shows an incredible differentiation of branch confined within the Manandhar sub caste of Newar suggesting either deep coalescence age of ≈25Kyr BP or a Late arrival ≈5Kya (**Figure 4.26**).

Similarly, indigenous South Asian lineages M5 (M5c2 and M5b2b subclades) and M35b4 shows age estimates of ≈17-20 Kya corresponding with the population movement during the Last Glacial Maximum (LGM). The change in temperature during this time period is believed to have triggered the lineage expansion.

1. **M5c2** [`20±4.8 Kya]: Udaya (6.89%), Shakya (5.2%) and Manandhar (1.07%)
2. M5b2b [18.4±2.5] Kya: Shakya (5.26%), Bajracharya (5.26%), Udaya (3.4%), Magar (2.7%)
3. M35b4 [17.8±4.6] Kya: Manandhar (4.3%) and Udaya (3.44%)

The frequency and diversity of Hg M5 revealed that it might have originated in Central India and spread out to the eastern and western regions of India (Chandrasekar et al., 2009). Ancient Gene flow from South Asia into the Kathmandu valley are further strengthened by the restricted distribution of Hg M5c and M5b in Newar**. The Nepali sequences for Hg M5c2 and M5b are hugely diverged from the non-Nepalese sequences. Therefore, it is highly unlikely that these ancient mtDNA lineage were brought into Nepal only recently via migration**.

Early post glacial arrivals

Early post glacial (≈12 Kyr BP) arrival corresponds with the expansion of indigenous (autochthonous) lineage across the south Asia. At the end of the younger Dryas or the Pleistocene/ Holocene transition, the movement of the population increased/ intensified across the South Asia as suggested by the presence of many indigenous south Asian lineages in the studied populations. Over the time distinctive nested Nepal specific branch are present in these lineages.

1. M3c1a1 ≈13.3±4.8 Kya. Detected mainly in Shakya (10.5%) and Bajracharya (5%).
2. R5a2b ≈9.9±3.4 Kya. Detected in Brahmin, Manandhar and Hindu Terai (**Figure 6.2**).
3. M30d1 [11±2.8]: Detected in **Hindu Terai (25%),** Nepali-other (4.47%), Udaya (3.44%), Shrestha (2.77%), Brahmin (2.08%).
4. M43a [13±5.4]: Detected in **Tharu** populations from Chitwan (TH-CI: 10%, TH-CII: 9.21%), Kathmandu (2.59%), Tamang (2.27%), Nepali-other (1.21%), Eastern Tharu (2.5%).
5. HV12b1b1 [10.3±3.2 Kya]: Detected only in Manandhar.

In India, Hg R5 has a wide distribution, especially in North India, Madhya Pradesh and West India. Sub Hg **R5a1** are reported to be present in high frequencies among the Indo-European speaking population, whereas R5a2 is present among the Dravidic groups of India and Sri Lanka (Chaubey et al., 2008). Age of R5a2b was estimated to be ≈9.5 Kyr old (Silva et al., 2017). In a good agreement to the previous study, this study estimated the age corresponding to the similar time frame of ≈9.9 Kyr old.

Neolithic arrivals

Neolithic (≈5-9 Kya) is associated with the spread of agriculture and animal domestication along with significant migration with genetic consequences and language replacement in South Asia. The first animals to be domesticated were dog, cattle, sheep and goat and the first plants to be cultivated were wheat and barley. Rice and Pig's domestication occurred in East Asia. The earliest Neolithic sites, on the Indus Valley around Mehrgarh in Baluchistan (Pakistan), date to before ≈9 Kyr BP (Jarrige J-F, 2006; Petrie, 2015). Hence, the putative Neolithic lineages might have entered to the South Asia at ≈9.5Kyr from Anatolia, the Caucasus and Iran (Silva et al., 2017). The genetic evidence for the Neolithic entry of West Eurasian Hg in Nepal is provided by the lineage U7a3a in which Newa specific U7a3a shows genetic affinity with the populations (Pashtun people) from Afghanistan. Hg U7a is the major West Eurasian lineage found in Newar populations. Hg U7 is detected at high frequencies in populations throughout Iran, Pakistan, Northwest India and the Arabian Peninsula (Quintana-Murci et al., 2004). Nepal specific clade Hv12b1b1, which dates back to ≈10.3 Kya is mainly detected in Manandhar. Sub Hg HV12b1 from Manandhar population traces its ancestry back to the populations from Armenia and Iran. Back mutation at position @150T has been observed in all sequences belonging to Iran, Myanmar, Armenian, Ladhak and Nepalese sample. The most recent ancestral variant 16242T was shared between Iran, Nepal (Manandhar) and Myanmar. I1e linage observed in Nepal and Myanmar share the most recent common ancestor suggesting a gene flow from Nepal to Myanmar. Restricted distribution of Hg I1e and HV12b1 within the Manandhar caste might be due the endogamy nature of the

Manandhar caste. The finding of two new west Eurasian specific lineage in high frequency in Manandhar, clearly indicates the importance of sample size and sub caste specific study of Newar group, which was lacking in previous study (Gayden et al., 2013).

Many South Asian specific Hgs also arrived at the similar time frame. List of the mtDNA linages estimated to have arrived in Nepal in the similar time frame are: R6b1a [5.7 Kya], R30b2a1 [5.7 Kya], I1e1a1 [5.4 Kya], M38+6366 [7.3±2.5 Kya], and H13a1d [6.5 Kya]. Hg R6b has been observed in Nepal and Uttar Pradesh. However, majority of basal variant observed in Nepali sample (R6b1) shows huge divergence from those observed in Uttar Pradesh (R6b2). Hence the Nepali specific lineage R6b1 shows the ancient origin. Hg R6b1 observed in Myanmar shares a recent ancestry with those found in Nepal.

Arrival from East Eurasia

Tibetan plateau and the great mountain ranges were among the last places colonized by humans in prehistory (Jeong et al., 2016). While most of the part of the Asia was already populated, Tibetan plateau remained an uninhabitable place due to its high altitude at around 4,000 m above sea level, resource scarcity, cold stress, hypoxia, hyper arid and cold desert environment. At the end of the Late Glacial Maximum temperatures rose and glaciers began retreating, making migration onto the plateau from the north more feasible (Aldenderfer and Yinong, 2004). Early human settlement in Tibetan plateau is estimated to have occurred at the end of the LGM (≈18-22 Kyr). Rapid population growth on the Plateau is estimated to have occurred in between 10–7 Kya during the early Neolithic. It is assumed that the major migration(s) of Neolithic agriculturalists into the Plateau have occurred from the Northwestern China, eventually leading to the establishment of agriculture and pastoralism on the Plateau, which provided a sufficient and stable food supply necessary for the rapid population growth that occurred on the Plateau around 10–7 Kya during the early Neolithic (Qi, Cui, Ouzhuluobu, Wu, & Su, 2014). Modern populations residing at the high-altitude transverse valleys (>3,000 m above sea level) of the Himalayan arc from Arunachal Pradesh to Ladhak are genetically closer to the East Asians. The study of the ancient genome (Whole genome sequencing, mtDNA and Epas1) of the 8 ancient samples (dental) from Annapurna Conservation Area (ACA) of Nepal (Mustang District), spanning 3,150–1,250 year before present (YBP) yielded affinities with the peoples living today on the Tibetan plateau (Jeong et al., 2016). All of the Ancient ACA sample were classified in to the East Asian specific Hgs (D4j1b, M9a, Z3a1a, F1c1a and F1d1). Majority of these Hgs detected in ACA samples were also present in Newar group. Incredibly, Hgs Z3a1a and F1d1 detected in ACA samples makeup 36% of the total Newar sample. This confirms that the descendant of the earliest inhabitant of the Himalayan arc also populated the Kathmandu valley located at low elevation. In a good agreement with the above statement, analysis of the major East Eurasian lineages observed in the present study suggests that major genetic influx from East Asia to Nepal have occurred in between **3-6 Kyr BP**. Similarly, Median joining network for Hg Z showed the sharing of basal variant (16150T) among the Tibetan, Nepali, Thai-Laos and North East Indian, suggesting a common ancestor for these populations. The age of the Continental lineage Z3a1a+16150T is estimated to be 7 Kyr old. Nepali harbor's a number of unique Nepal specific haplotypes at the terminal branch**,** majority of which branched off directly from the nodes occupied by almost exclusively by Tibetan lineage.

Similarly, Median joining network of Hg F1d1 indicates that the core haplotype is shared between the Tibet, Nepal, Myanmar and Thailand-Laos. This suggests these populations have been descendent from the common Ancestor and spread to the different geographic regions. Majority of the contemporary populations of Tibet, Thailand, Nepal and Myanmar (only HVR data available) belonging to the lineage F1d1 are derived from the basal variant 16284G. Hence, the most recent common ancestor for this sub lineage were originated some were in East Asia. Basal Variant 15204C is the major ancestral variant of the contemporary Nepali populations belonging to lineage F1d1 (Founder effect) (**Figure 4.13**).

Other East Eurasian specific Hgs such as G and D has relatively low distribution frequencies in Newar. Hg D is one of the most important hg among the Tamang (27.5%) and Magar (24.32%) followed by Shakya (15.78%), Tharu from Chitwan II (11.84%), Tharu from Morang district (10%) and Bajracharya (10%). This hg was also found in high frequency in Kathmandu (11.6%). Although G3a (6.2%) observed in Brahmin cluster with the sample from Tibet, china and Thailand, it shares closer ancestry with the G3a detected in Kashmir. Similarly, G2b2a observed in Newar, cluster with Thailand, whereas G2a1d2 cluster with Belarus (Central Asia). Hg C present in **Magar (8.1%)** was completely absent in Newar. Tibetan specific hg A+152 was also found in appreciable frequencies in Newar.

Intriguingly, Hg F1d1 and Z3a1a which accounts **65.8%** of the total East Eurasian maternal components of Newar, were observed in very low frequency with relict distribution among the Tibetans (Kang et al., 2016). Hence, majority of Newar group are placed away from the Tibetan cluster as shown in PCA map. Whereas, other Nepali groups Sherpa, Tamang, Tharu from Chitwan (CI & CII) and Magar contains high frequency of Tibetan specific Hg and cluster closely with the Tibetans. Z3a1a and F1d1 both contains Nepal specific haplotypes at their terminal branch or basal nodes. Previous study estimated the age of the clade **Z3a1a** to be ≈7.9 Kyr old (Fedorova et al., 2013). Likewise, this study estimated the age of the subclade **Z3a1a1** to be ≈5.1 Kyr old. Similarly, a recent study on the Tibetan populations estimated the age of the clade F1d1a to be ≈4.5 Kya (Kang et al., 2016). In a good agreement to the previous study, our study also estimated the age of the clade F1d1a to be 4.9 Kyr old. Hence the carrier of hg Z3a1a and F1d1a might have entered Nepal across the great Himalayas as early as 3-6 Kyr BP. Other Tibetan specific lineage such as Hg A and D observed in Newar are most likely resulted due to gene flow gene flow from neighboring populations including Tibet. In overall East Asian lineages represent a frequency of ≈50% in the total Newar data set.

Bronze Age arrivals

This admixture event was followed by recent admixture events in the last 4Kya and their assimilation into Newar. List of Hgs estimated to have arrived recently into Nepal via migration in the last ≈2-4 Kya includes: M5a6, M5d, U2b2, U2b1a1, U7a1a, U7b7, M2a3a, M2d1a, M3c2, M3d and W3a1b.

Although, mtDNA analysis shows recent Genetic Influx from Tibet, bearer of the East Eurasian genetic components of Newar shows lower signal of selection in the genetic variants of EPAS1 gene suggesting a short stay of majority of these lineages in the Tibetan plateau. This was further supported by the analysis of the 1540C variants of the **EDAR gene**. Variant 1540C>T of EDAR gene

originated in central china is highly informative for assessing East and Southeast Asian ancestries. Despite higher proportion of East Eurasian specific maternal components observed in the Newar, the frequency of variant 1540C>T was comparatively low, suggesting that the lineage Z3a1a or F1d1 might have arrived into the Tibetan plateau via different migratory route than the central China. Another possibility is that, the lineage Z originated in Central Asia might have arrived into central china and then into Tibet very recently form other geographic region. Indeed abundant Z3a lineages have been detected in the native populations of Sakha (Siberia) as discussed in the PhyloTree of Hg Z. Whereas, higher frequency of variant 1540C>T in Magar reflect their affinity with the populations from central china.

High-altitude adaptation study in Ladhak also identified positive signal of selection in genetic variants of EPAS1 gene. Further, a strong linear correlation between genotyped alleles OF EPAS1 gene and different level of altitude were detected in the Himalayan populations residing in Ladhak and Nepal, suggesting that extremely high-altitude hypoxia environment exert a selective effect on Epas1 variants. Tibetan specific mtDNA hg M9a was present in higher frequency in both Changpa and general individuals of Ladhak. The mtDNA pool of Changpa was dominated by East Asian specific hgs which were also observed in higher frequency diversity among the Tibetans. Hence Changpa group shares most of their matrilineal lineages with the indigenous Tibetans suggesting that the Tibetans are the ancestral populations of Changpa and the adaptive traits for the high-altitude adaptation has been recently inherited form their ancestors in Tibet. Whereas the general individuals from Ladhak shows higher proportion of West Eurasian maternal components with substantial genetic contribution from Tibet and India. mtDNA pool of general individuals from Ladhak contains significant proportion of East Eurasian hgs and majority of these hgs were shared with the Changpa tribes.

The frequency of the four tag EPAS1 SNPs in Changpa, paralleled with the Tibetan as well as Sherpa, strongly suggesting that the EPAS1 gene is under selection in people of Tibetan ethnicity. Hence, the change in the genetic variants in EPAS1 gene might be responsible for resulting in an efficient hemoglobin-oxygen transport system in Changpa at high altitude hypoxia environment.

6 Summary

In present study, samples belonging to Newa, Magar and Brahmin were classified into known mtDNA sub Hgs belonging to three main origin: **East Eurasian**, **South Asian** and **West Eurasian**. Investigation was performed based on the high resolution mtDNA sequencing.

Gene flow from East Eurasia

Majority of the East Eurasian components observed in Newar are present in all of the Major Newar sub caste, except Shakya which lacks F1d Hg. Further, most of the East Eurasian components observed in Newar populations were detected mainly in Tibet, Thailand and Myanmar. Magar contain East Asian specific hg M9 and D. Gene flow from East Eurasia was also detected in Brahmin in lower frequency.

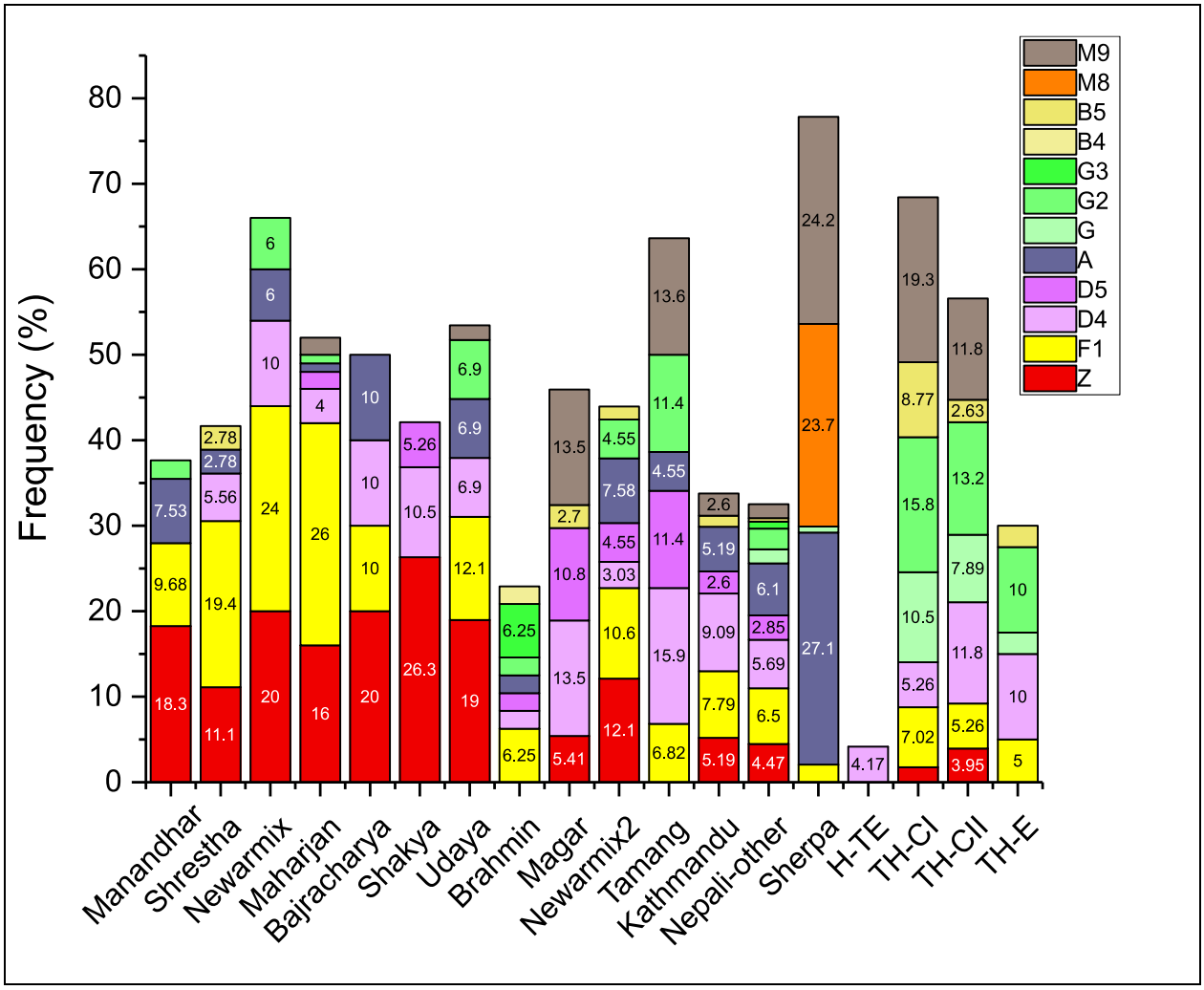


Figure 6.1 | Frequencies of East Eurasian Maternal components observed in the several Nepali populations. Subclade Z3a1a and F1d1 are highly prevalent in Newar followed by small frequency of D4 and G2. Brahmin (F1d1, D4, D5, G3) population shows relatively lower proportion, but higher proportion of East Asian specific lineages (M9, D4, D5, C) were observed in Magar.

Previous study showed higher proportion of Tibetan specific lineages among the Sherpa (A, M9 & M8), Tamang (D4, D5, M9 and G2), Central Tharu I (M9 & G2) and Central Tharu II (G2, M9 and

D4) (**Table 8.1**). Hg Z originated some were in Central Asia are found with the relict distributions among the East Asians (**Figure 4.37**). Whereas Hg F is mainly present in East and South-East Asia (**Figure 4.11**). Hg F1d1 observed in High frequencies in Newar (especially in Maharjan) were also detected in Thailand, Myanmar, Tibet and North India (Uttarakhand). Hg F1d (4.2%) and F1d1 (4.3%) were present among the Mon (MO2) people of North East Thailand and West Thailand (border between Thailand and Myanmar). Mon are considered to be one of the earliest group to be present along the South-East Asia (Kutanan et al., 2017). The Mon were responsible for the spread of Theravada Buddhism in South East Asia (Indochina). Likewise they are considered to have a major source of influence on the culture of Burma. They speak the Mon language, an Austroasiatic language (Kutanan et al., 2017). The presence of both lineage F1d & F1d1 in the Mon group of Thailand/Myanmar and the distribution pattern of Hg F1d1 might provide the genetic evidence for the involvement of the Mon people in the cultural transformation across the South-East Asia.

Gene flow from South Asia

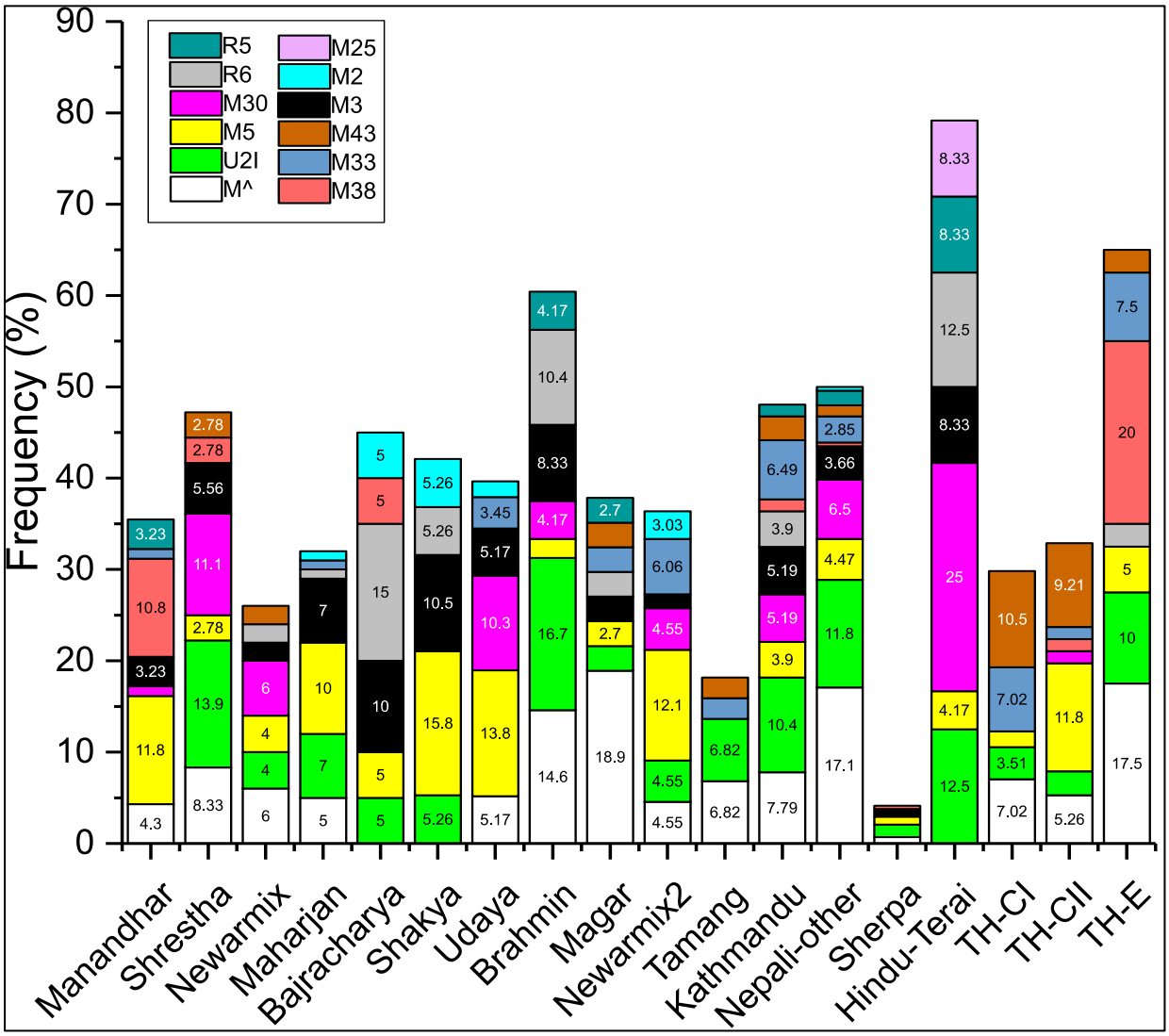


Figure 6.2 | Frequencies of the major South Asian specific Maternal components observed in the Nepali populations. Symbol "^" includes all the sub branches of M haplogroup detected in low frequency**.** Except Sherpa and Tamang all other Nepali populations studied so far has significant proportion of South Asian

specific maternal components (Bhandari et al., 2015; Fornarino et al., 2009; Gayden et al., 2013; Wang et al., 2012).

The proportion of south Asian specific maternal components varies among the Nepali populations. In Newar, comparatively Shrestha and Shakya has highest, Maharjan and Newamix has the least, whereas other shows intermediate level of South Asian maternal components. As expected, Brahmin (Indo European) and Magar (Tibeto-Burman) shows 60%, and 37.8% of South Asian maternal components respectively. Limited and low distribution frequencies of South Asian lineage such as M18, M31, M32, M4, M58, M61, M62, M65, R30 and R8 indicates a limited contribution in Newar mtDNA gene pool and other studied Nepali popualtions.

Gene flow from West Eurasia.

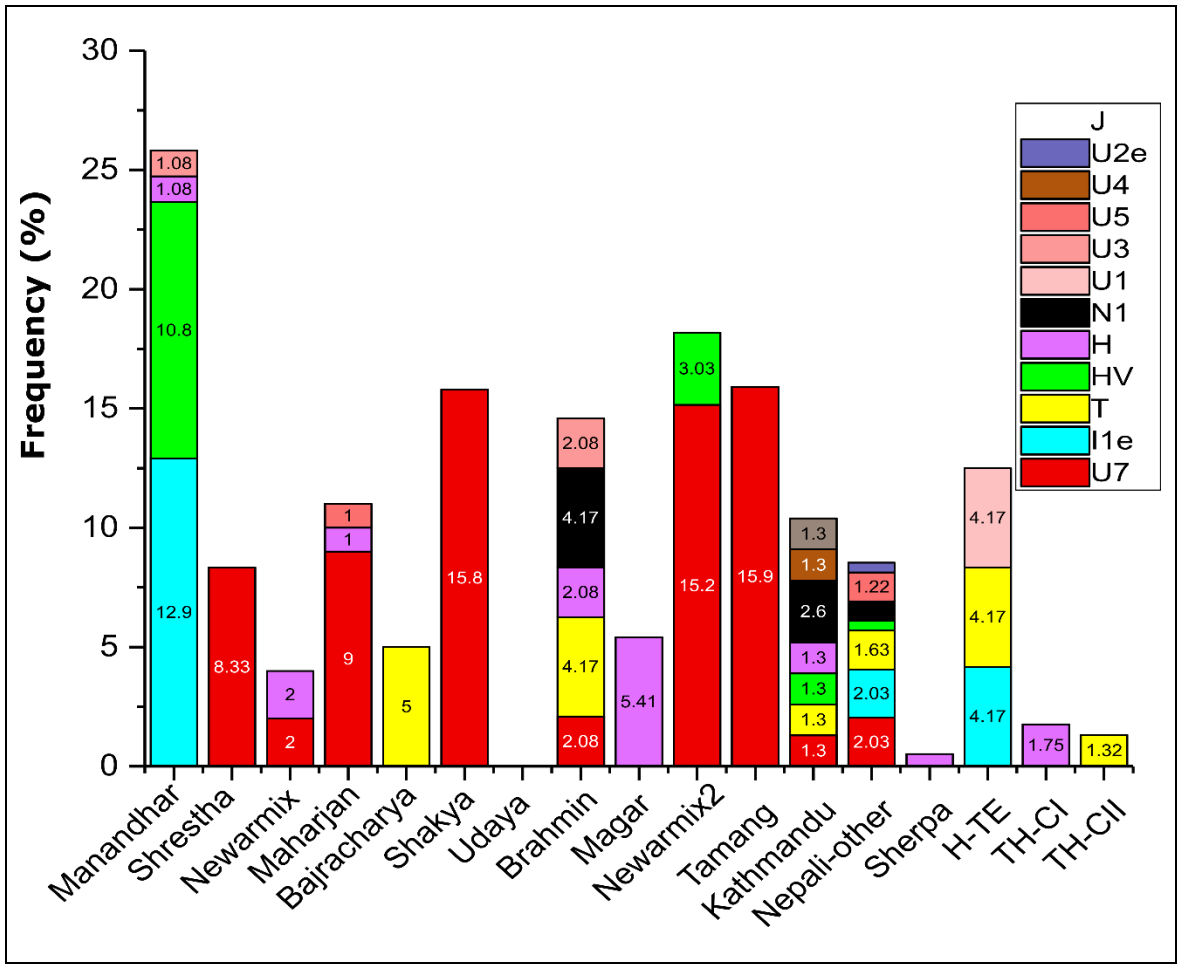


Figure 6.3 | Frequencies of the major West Eurasian specific Maternal components observed in the Nepali populations. Genetic influences from west Eurasia in Newar are represented predominantly by Hg U7 and its subclade U7a3a, U7b and U7a1a. This is followed by other Hgs: I, HV, U5, H, N1a and T.

Manandhar shows higher proportion of West Eurasian (Near Eastern clade) hg than any other Newa sub caste. Surprisingly these Major West Eurasian hg where confined within the Manandhar and was unshared among the other Newa sub caste. Hg U7 was the major founding west Eurasian clade present among the Newa sub caste except Manandhar. Brahmin shows significant proportion of west Eurasian hg whereas its frequency were very lower in Magar.

Phylogeographic pattern of the Major Haplogroups

Table 6.1: Phylogeographic pattern of the Major Haplogroups. Summary about the Major hgs, frequency, ancestries, possible route of migration and TMRCA of the major haplogroups is given in the table. The term "Recent arrival" includes the genetic events in the last 4Kya.

HG	Frequency	Ancestries	Major sub Haplogroup	Ancestral variant	Most recent ancestor	Major Nepal specefic variant	Major Sub caste specefic variant	Major subcaste	Migratory Wave	Estimated age arrived in Nepal (Kyr)
Z	17.80%	East Eurasian	**Z3a1a**	16150T	Tibet, North East India, Burma, Thailand, Nepal, Siberia, China	16186T, 7296A, 1305G	16186T	MDR	From Tibet across the Himalayas	5.1 ± 1.1
							7296A 1305G	Mhj, Udaya, BJR, SHK		
			Z7	4841A, 10653A	North East India, Nepal,Tibet,			Shakya	From Tibet across the Himalayas	Recent arrival (~2.5-5 Kya)
F	17.80%	East Eurasian	**F1d1**	16284G	Tibet, Burma, Thailand, Nepal, Kyrgyzstan, Siberia, North India (Uttarakhand), China	15204C	268T	Udaya, STH (1)	From Tibet across the Himalayas	4.9 ± 1.8
							12840T	MHJ, STH, Newa-mix		
D	5.90%	East Eurasian	D4, D5	Huge diversity	Tibet, China, Japan, Korea, North East India	Huge Diversity	-	Shakya (15.7%) BJR (10%), MGR (24%), TMG (27.5%)	From Tibet across the Himalayas	Recent Arrival
M5	9.60%	South Asian	M5a	12477C	India, Nepal, Thailand	472G, 1888A, 15896G, 16240G, 16291T	7094C	MDR, MHJ, Udaya, Newa-mix	-	1.8 ± 0.7
			M5b2b	6293C	India, Nepal		6131G, 10283G, 13145A	SHK, BJR, Udaya		13.3 ± 4.1
			M5c2	16240G	Nepal		1193C, 16093C	Udaya, SHK, MDR, BJR, Sherpa		20.4 ± 3
							195C, 9861C, 16362C	Shakya, Tharu		
			M5d	16311C	India, Nepal	3757T, 7699T, 10604C, 13020C	-	Tharu, MDR		11.2 ± 3.2
M3	5.30%	South Asian	M3c1a	16051G	India (Jammu and Kashmir), Nepal,	5097G, 6011C, 8269A	-	SHK, BJR, Udaya, Hindu Terai	-	13.3 ± 4.8
			M3d1a	16301T	India, Nepal	-	-	MHJ, MDR, Newa-mix, Tharu	-	Recent arrival
U2I (U2a, U2b, U2c)	4.30%	South Asian	U2b1	6620, 10688a	India, Nepal			MHJ, Newa-mix		Recent arrival
			U2c1	146C	India, Nepal, Thailand	12361G, 16179T	3714G, 13488C	BRH, MHJ	Geneflow Nepal to Thailand	21.9 ± 3.9
							95C, 93G	SHK, BJR, STH, Thai		
U7	4.30%	West Eurasian	U7a3a	5291C-9266A-16223T-15364T	Afghanistan, Nepal	8545A	9099T	SHK	west Eurasia to Nepal through Afghanistan	6.5 ± 2.3
							-	MHJ, STH		
I1	3.20%	West Eurasian/ Near East	I1e	14542T	India, Nepal, Thailand		12843C, 13866G, 14256C	Manadhar	Through Iran	5.4 ± 1.6
Hv12	2.70%	West Eurasian/ Near East	Hv12b1	150T!	Iran, Armenia, Nepal, Myanmar, India (Ladhak)	1284C		Manadhar	Through Iran	3.6 ± 1.5

Haplogroup M52b, U2c1 and M33a1 are the ancient and deep time depth South Asian lineages observed in the Nepali populations suggesting that this region might been inhabited by the earliest settlers during the initial peopling of the South Central Asia. This provides a direct evidence to the long ancient connection between India and Nepal. This was further added by the entry of the putative Glacial/postglacial genetic influx from West Eurasia to the South Asia, estimated to have occurred in between 15-21 Kyr BP. Indicator for this genetic influx was provided by lineage N1a1b (Silva et al., 2017). This claim was further strengthened by the presence of deep time depth west Eurasian specific Hg N1a1b (N1a1b1 and I1e) in Brahmin and Manandhar group of Nepal. Considering the close vicinity between Nepal and the North India in geography, a possibility that West Eurasian genetic components might have spread into Nepal at the same time as they initially spread in North India could not be ruled out. During the Holocene several East Asian, South Asian as well as West Eurasian specific lineages were added into contemporary Newar via multiple dispersal from several distinct sources in different time, rather than just one or two major admixture events in the Neolithic/Bronze age. However, Newar and Magar experiences huge gene flow during the Neolithic/Bronze age mainly from East Asia. The study of the ancient genome (Whole genome sequencing, mtDNA and Epas1) of the 8 ancient samples (dental) from Annapurna Conservation Area (ACA) of Nepal (Mustang District), spanning 3,150–1,250 year before present (YBP) yielded affinities with the peoples living today on the Tibetan plateau. Majority of the East Asian specific Haplogroups (D4j1b, M9a, Z3a1a, F1c1a and F1d1) detected in ACA samples were also detected in Newar and Magar. However Hgs Z3a1a and F1d1 detected in ACA samples makeup 36% of the total Newar sample. This confirms that the descendant of the earliest inhabitant of the Himalayan arc and the Kathmandu valley shares a most recent common ancestor.

Most of the East Eurasian haplotype observed in Newar, Magar and Brahmin branched off directly from the nodes occupied by the Tibetan lineage, albeit a few haplotypes were shared in between Nepali and Northeast Indian. Hence this study suggest that Tibet might be the possible homeland through which most of the East Eurasian maternal components has entered into Nepal. Massive genetic Influx from East Asia (50% of the total Newar sample) most probably occurred in between 3-6 Kyr, has made a hefty contribution to the modern gene pool of the Newar as compared to South Asia (35.9% of the total Newar sample) or West Eurasia (12.8% of the total Newar sample). Hence, our investigation of the Newars origin via extensive, high resolution mtDNA diversity analysis supports a close genetic relationship between Newar and East Asia with substantial genetic contribution from South Asia and West Eurasia. The clustering pattern in the PCA clearly shows the gene flow into the Newar from East/Southeast Asia and South Asia. Hence, this clustering pattern supports the proposed antiquity of Newar, who are postulated to be the mixture of Dravidian, Mongoloid and Aryan origins (Regmi, 1969). Mitochondrial gene pool of Brahmin were dominated by South Asian hg, whereas Magar were dominated by East Eurasian hg. Further, this study performed the most precise identification of East/Southeast Asian, South Asian and West Eurasian Hgs based on the large scale complete mtDNA analysis of the Nepali populations and their thorough whole mtDNA sequence comparison with the other Asian populations.

Frequency of variant 1540C>T of EDAR gene was comparatively lower in Brahmin and Newar. Whereas its frequency was higher in Magar.

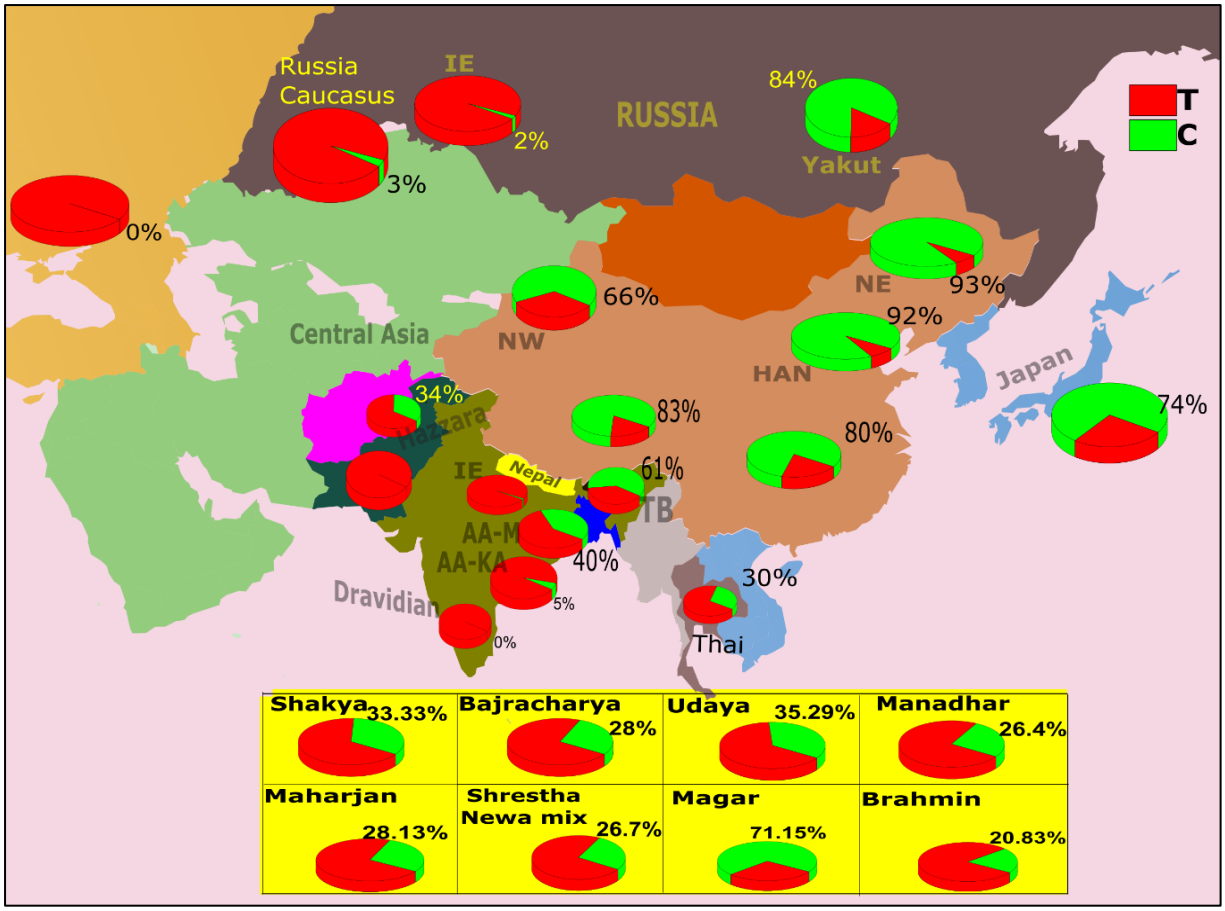


Figure 6.4 | Comparison of EDAR profile of Newar, Magar and Brahmin with the other populations. Distribution of 1540C allele of the EDAR gene in association with the geographic and linguistic affiliations. The background yellow color represents the populations analyzed in present study. The Austroasiatic (AA) speaking population of India is divided into Khasi-Aslian (KA) and Munda (M). Tibeto-Burman and Indo-European speaking populations are abbreviated as TB and IE respectively.

Variant 1540C>T of EDAR gene are detected in higher frequency in those populations who share ancestry with the central china populations. Higher frequency of variant 1540C>T in Magar reflect their affinity with the populations from central china.

Newar shows lower signal of selection in the genetic variants of EPAS1 gene suggesting a short stay of majority of these lineages in the Tibetan plateau. In a good agreement with the previous studies on Tibetans and Sherpa, high-altitude adaptation in Ladhak also identified positive signal of selection in genetic variants of EPAS1 gene. Further, a strong linear correlation between genotyped alleles of EPAS1 gene and different level of altitude were detected in the Himalayan populations residing in Ladhak and Nepal, suggesting that extremely high-altitude hypoxia environment exert a selective effect on Epas1 variants. Changpa, the semi nomadic tribes living around the Ladhak regions shares most of their matrilineal lineages (mtDNA) with the indigenous Tibetans suggesting that the Tibetans are the ancestral populations of Changpa and the adaptive traits for the high-altitude adaptation has been recently inherited form their ancestors in Tibet. Whereas the general individuals from Ladhak shows higher proportion of West Eurasian maternal components with substantial genetic contribution from Tibet and India.

7 Conclusion

Newar, Magar and Brahmin harbors albeit limited, ancient and deep time depth South Asian lineages suggesting that this region might been inhabited in between 20-30 Kya by the earliest settlers during the initial peopling of the South Central Asia. The presence of ancient South Asian hg in the studied populations provides a direct genetic evidence to the long ancient link between India and Nepal. Further several gene flow from South Asia and West Eurasia were observed during Glacial/postglacial period. Over the time distinctive nested Nepal specific branch were present in these lineages Newar experiences several waves of migration during the Holocene adding several East Asian, South Asian as well as West Eurasian specific lineages into contemporary Newar via multiple dispersal from several distinct sources in different time, rather than just one or two major admixture events in the Neolithic/Bronze age. The study of the ancient genome of the 8 ancient samples from Annapurna Conservation Area (ACA), spanning 3,150–1,250 YBP yielded affinities with the peoples living today on the Tibetan plateau. Hgs Z3a1a and F1d1 detected in ACA samples makeup 36% of the total Newar sample. Despite the presence of low frequency of hg F and Z in Tibet, this study suggest that Tibet might be the possible homeland through which most of the East Eurasian maternal components has entered into Nepal. Hg Z and F detected in Newar is phylogenetically closer to those found in Tibet, Myanmar and Thailand suggesting a common ancestor for these lineages. Most of the East Eurasian haplotype, observed in Newar, Magar and Brahmin branched off directly from the nodes occupied by the Tibetan lineage, albeit a few haplotypes were shared in between Nepali and Northeast Indian. Massive genetic Influx from East Asia (50.8%) most probably occurred in between 3-6 Kyr, has made a hefty contribution to the modern gene pool of the Newar as compared to South Asia (35.9%) or West Eurasia (12.8%). Hence, our investigation of the Newars origin via extensive, high resolution mtDNA diversity analysis supports a close genetic relationship between Newar and East Asia with substantial genetic contribution from South Asia and West Eurasia.

In the past few years, several studies of high-altitude adaptation in Tibetan populations including Sherpa of Nepal, identified positive selection signals in genetic variants of EPAS1 gene. Present study of high-altitude adaptation in the Himalayan population's form Ladhak also identified positive signal of selection in genetic variants of EPAS1 gene with maximum XPCLR score equal to 105. Further, a strong linear correlation between genotyped alleles of EPAS1 gene and different level of altitude was detected in the Himalayan populations suggesting that extremely high-altitude hypoxia environment exert a selective effect on HI2A variants among the populations from Ladhak. The frequency of the four tag EPAS1 SNPs in Changpa, paralleled with the Tibetan as well as Sherpa, strongly suggesting that the EPAS1 gene is under selection in people of Tibetan ethnicity. Hence, the change in the genetic variants in EPAS1 gene might be responsible for resulting in an efficient hemoglobin-oxygen transport system in Changpa at high altitude hypoxia environment. Changpa, the semi nomadic tribes living around the Ladhak regions shares most of their matrilineal lineages (mtDNA) with the indigenous Tibetans suggesting that the Tibetans are the ancestral populations of Changpa and the adaptive traits for the high-altitude adaptation has been recently inherited form their ancestors in Tibet. Whereas the mtDNA pool of general individuals from Ladhak were dominated by West Eurasian Hgs with the substantial genetic contribution from Tibet and south Asia.

8 Appendix

8.1 Appendix A

Table 8.1 | Geographic location, linguistic affiliation, sample size and references of the 289 populations analyzed in the present study. More information in detail regarding the mitochondrial frequencies are displayed in the **Table 8.2** (Appendix A). Samples analyzed in this study are abbreviated as PS (present study).The number 1, 2…n provided in the reference column corresponds with the reference of the analyzed populations as given in the Appendix B.

Id	Population	State/Location	Country	Longitude	Latitude	Language	Size	Reference
1	Maharjan	Kirtipur	Nepal	85.32	27.72	Tibeto-Burman	100	PS
2	Shrestha	Bhaktapur	Nepal	85.4	27.7	Tibeto-Burman	36	PS
		Banepa	Nepal	85.5	27.6			
		Kathmandu	Nepal	85.3	27.7			
3	Newa_mix	Bhaktapur	Nepal	85.4	27.7	Tibeto-Burman	50	PS
		Sankhu	Nepal	85.3	27.5			
		Kathmandu	Nepal	85.3	27.7			
4	Manandhar	Kathmandu	Nepal	85.3	27.7	Tibeto-Burman	93	PS
		Patan	Nepal	85.3	27.7			
5	Bajracharya	Kathmandu	Nepal	85.3	27.7	Tibeto-Burman	20	PS
6	Shakya	Kathmandu	Nepal	85.3	27.7	Tibeto-Burman	19	PS
7	Udaya	Kathmandu	Nepal	85.3	27.7	Tibeto-Burman	58	PS
		Patan	Nepal	85.3	27.7			
8	Brahmin	Tanahun	Nepal	84.4	27.9	Indo-European	48	PS
9	Magar	Tanahun	Nepal	84.3	28.0	Tibeto-Burman	36	PS
10	Changpa	Jammu and Kashmir	India	78.6	33.9	Tibeto-Burman	90	PS
11	Ladhak	Jammu and Kashmir	India	78.1	34.0	Mixed	89	PS
12	North India		India	80.5	26.9	Indo-European	44	PS
13	Newar_mix2	Kathmandu	Nepal	85.3	27.8	Tibeto-Burman	65	1
14	Tamang	Kathmandu	Nepal	84.3	27.7	Tibeto-Burman	45	1
15	Kathmandu	Kathmandu	Nepal	85.4	27.7	Mixed	77	1
16	Nepali-other	Kathmandu and East Nepal	Nepal	87.5	27.7	Mixed	262	2
17	Sherpa	Khumbu valley	Nepal	86.8	27.8	Tibeto-Burman	582	3
		Zhangmuzhen	Tibet	86.0	28.2			
18	Hindu_Terai	Nepal	Nepal	85.2	27.1	Indo-European	23	4
19	Tharu_Chitwan	Chitwan	Nepal	84.5	27.5	Indo-European	133	4
20	Tharu_Chitwan	Chitwan	Nepal	84.4	27.5			4
21	Tharu_Morang	Morang	Nepal	87.5	26.7	Indo-European	40	4
22	Hindu_New Delhi	New Delhi	India	77.2	28.6	Indo-European	83	4
23	Tribal Ap	Andra Pradesh	India	78.7	14.9	Dravidian	78	4
24	Tharu_Haldwani	Uttarakhand	India	79.5	29.2	Indo-European	22	5
25	Tharu_Nainital	Uttarakhand	India	79.5	29.4	Indo-European	23	5
26	Tharu_Gorakhpur	Uttar Pradesh	India	75.7	29.4	Indo-European	38	5
27	Tharu_Allahabad	Uttar Pradesh	India	81.8	25.4	Indo-European	37	5
28	Tharu_Deoria	Uttar Pradesh	India	83.8	26.5	Indo-European	47	5
29	Tharu_Mahrajganj	Uttar Pradesh	India	83.1	26.3	Indo-European	42	5

30	Brahmin	Uttarakhand	India	80.6	30.2	Indo-European	35	6
31	Kshatriya	Uttarakhand	India	80.4	30.1	Indo-European	46	6
32	Shah	Uttarakhand	India	79.9	29.2	Indo-European	46	6
33	Go swami	Uttarakhand	India	80.2	30.0	Indo-European	34	6
34	Arya	Uttarakhand	India	79.9	29.3	Indo-European	24	6
35	Tamta	Uttarakhand	India	79.9	29.8	Indo-European	29	6
36	Koli	Uttar Pradesh	India	80.1	27.9	Indo-European	24	7
37	Harijan	Uttar Pradesh	India	80.2	27.3	Indo-European	29	8
38	Lodhi	Uttar Pradesh	India	81.8	26.4	Indo-European	49	8
39	Bhar	Uttar Pradesh	India	83.5	25.9	Indo-European	22	9
40	Mushar	Uttar Pradesh	India	83.2	26.1	Indo-European	49	9
41	Pardhan	Adilabad, Andhra Pradesh	India	78.5	19.66 4688,	Dravidian	193	10
42	Naikpod	Adilabad, Andhra Pradesh	India	78.5	19.65 6605,	Dravidian	88	10
43	Andh	Adilabad, Andhra Pradesh	India	78.5	19.7	Indo-European	66	10
44	Tibetans	Tibet	China	86.9	33.6	Tibeto-Burman	111	11
45	DEngs	Nyingchi, Tibet	China	94.4	29.9	Tibeto-Burman	76	11
46	Lhobas	Tibet	China	93.8	29.5	Tibeto-Burman	17	11
47	Sherpa’s	Tibet	China	85.9	28.3	Tibeto-Burman	668	11
48	Monpas	Tibet	China	86.0	28.0	Tibeto-Burman	151	11
49	Japanese	Tokyo	Japan	138.4	35.8	Mixed	199	12
50	ALCQ	Nagri	China	85.2	31.1	Tibeto-Burman	151	13
51	ALGJ	Nagri	China	81.1	32.5	Tibeto-Burman	199	13
52	ALGZ	Nagri	China	84.1	32.3	Tibeto-Burman	162	13
53	CDBB	Chamdu	China	94.7	30.9	Tibeto-Burman	30	13
54	CDCD	Chamdu	China	97.1	31.2	Tibeto-Burman	120	13
55	CDDQ	Chamdu	China	95.6	31.4	Tibeto-Burman	62	13
56	CDJD	Chamdu	China	89.2	31.5	Tibeto-Burman	28	13
57	CDLW	Chamdu	China	96.6	31.2	Tibeto-Burman	80	13
58	CDMK	Chamdu	China	98.7	29.6	Tibeto-Burman	76	13
59	LSDQ	Lhasa	China	91.1	30.0	Tibeto-Burman	625	13
60	LSDX	Lhasa	China	91.1	30.5	Tibeto-Burman	792	13
61	LSGK	Lhasa	China	91.8	29.8	Tibeto-Burman	325	13
62	LSLZ	Lhasa	China	91.2	30.2	Tibeto-Burman	567	13
63	LSQS	Lhasa	China	90.7	29.4	Tibeto-Burman	427	13
64	LZBM	Nyingchi	China	95.8	29.9	Tibeto-Burman	163	13
65	LZCY	Nyingchi	China	97.5	28.6	Tibeto-Burman	68	13
66	LZDQ	Nyingchi	China	94.3	29.6	Tibeto-Burman	30	13
67	NQAD	Nagqu	China	91.7	32.3	Tibeto-Burman	493	13
68	NQDB	Nagqu	China	94.1	32.0	Tibeto-Burman	17	13
69	NQXB	Nagqu	China	90.1	31.4	Tibeto-Burman	16	13
70	RKAR	Xigaze	China	87.2	29.3	Tibeto-Burman	49	13
71	RKBL	Xigaze	China	89.2	29.1	Tibeto-Burman	113	13
72	RKDJ	Xigaze	China	87.8	28.4	Tibeto-Burman	47	13
73	RKDQ	Xigaze	China	88.8	29.3	Tibeto-Burman	461	13
74	RKDR	Xigaze	China	87.1	28.6	Tibeto-Burman	22	13
75	RKJZ	Xigaze	China	89.6	28.9	Tibeto-Burman	56	13

76	RKNM	Xigaze	China	89.0	29.7	Tibeto-Burman	81	13
77	RKSJ	Xigaze	China	88.0	28.9	Tibeto-Burman	67	13
78	RKXB	Xigaze	China	85.3	29.4	Tibeto-Burman	40	13
79	RKXT	Xigaze	China	88.3	29.4	Tibeto-Burman	49	13
80	RKYD	Xigaze	China	88.9	27.6	Tibeto-Burman	65	13
81	SNCN	Shannan	China	91.9	28.0	Tibeto-Burman	19	13
82	SNDQ	Shannan	China	91.8	29.2	Tibeto-Burman	177	13
83	SNGG	Shannan	China	91.0	29.3	Tibeto-Burman	32	13
84	SNLZI	Shannan	China	92.4	28.5	Tibeto-Burman	44	13
85	SNND	Shannan	China	91.7	29.0	Tibeto-Burman	47	13
86	SNQS	Shannan	China	92.1	29.1	Tibeto-Burman	66	13
87	SNZL	Shannan	China	91.3	29.2	Tibeto-Burman	27	13
88	QHMQ	Qinghai	China	100.3	34.5	Tibeto-Burman	130	13
89	Menba	Shannan	China	91.9	28.0	Tibeto-Burman	57	13
90	Sherpa	Xigaze	China	85.3	29.4	Tibeto-Burman	29	13
91	Adi	Assam	India	92.9	26.3	Tibeto-Burman	45	14
92	Apatani	Arunachal Pradesh	India	94.1	28.3	Tibeto-Burman	26	14
93	Apatani	Tripura	India	92.0	24.0	Tibeto-Burman	21	14
94	Naga	Nagaland	India	94.7	26.2	Tibeto-Burman	43	14
95	Nishi	Tripura	India	92.0	24.0	Tibeto-Burman	44	14
96	Tipperah	Tripura	India	92.0	24.0	Tibeto-Burman	20	15
97	Garo	Meghalaya	India	90.6	25.6	Tibeto-Burman	76	16
98	Lyngngam	Meghalaya	India	90.9	25.4	Austro-Asiatic	74	16
99	Nongtrai	Meghalaya	India	91.1	25.9	Austro-Asiatic	27	16
100	Maram	Meghalaya	India	91.2	25.5	Austro-Asiatic	60	16
101	Bhoi	Meghalaya	India	91.8	26.0	Austro-Asiatic	29	16
102	Khynriam	Meghalaya	India	91.8	25.5	Austro-Asiatic	82	16
103	War_Khas	Meghalaya	India	91.9	25.2	Austro-Asiatic	29	16
104	Pnar	Meghalaya	India	92.3	25.5	Austro-Asiatic	51	16
105	War_Jaint	Meghalaya	India	92.3	25.2	Austro-Asiatic	17	16
106	Tibetan	Nagqu, Tibet	China	92.0	31.6	Tibeto-Burman	168	17
107	Tibetan	Rikaze, Tibet	China	88.9	29.4	Tibeto-Burman	220	17
108	Tibetan	Liangshan, Sichuan	China	102.3	27.9	Tibeto-Burman	62	17
109	Tibetan	Guide, Qinghai	China	101.4	36.1	Tibeto-Burman	76	17
110	Tibetan	Chamdo, Tibet	China	97.2	31.3	Tibeto-Burman	61	18
111	Tibetan	Garze, Sichuan	China	102.0	30.1	Tibeto-Burman	55	18
112	Tibetan	Lhasa, Tibet	China	91.1	29.7	Tibeto-Burman	59	18
113	Tibetan	Nagqu, Tibet	China	92.0	31.8	Tibeto-Burman	58	18
114	Monba	Nyingchi, Tibet	China	94.4	29.7	Tibeto-Burman	51	18
115	Tibetan	Nyingchi, Tibet	China	94.3	29.4	Tibeto-Burman	53	18
116	Lhoba	Shannan, Tibet	China	91.8	29.3	Tibeto-Burman	20	18
117	Tibetan	Shannan, Tibet	China	91.8	28.8	Tibeto-Burman	56	18
118	Tibetan	Rikaze, Tibet	China	88.6	28.9	Tibeto-Burman	59	18
119	Hani-YN	Xishuangbanna, Yunnan	China	100.8	22.0	Tibeto-Burman	80	19
120	Bai-YN1	Dali	China	100.2	25.7	Tibeto-Burman	69	19
121	Bai-YN2	Xishuangbanna	China	100.8	22.1	Tibeto-Burman	19	19
122	Yi-YN1	Xishuangbanna	China	100.8	22.1	Tibeto-Burman	16	19
123	Yi-YN2	Chuxiong	China	101.5	25.1	Tibeto-Burman	40	19
124	Jino-YN	Xishuangbanna	China	100.8	22.1	Tibeto-Burman	18	19
125	Lahu-YN	Simao, Xishuangbanna	China	101.0	22.8	Tibeto-Burman	37	19

126	Pumi-YN	Ninglang	China	100.9	27.3	Tibeto-Burman	35	19
127	Naxi-YN	Lijiang	China	100.2	26.9	Tibeto-Burman	45	19
128	Zang-YN1	Diqing	China	99.7	27.8	Tibeto-Burman	88	17,20
129	Zang-YN2	Zhongdian	China	99.7	27.9	Tibeto-Burman	35	19
130	Lisu-YN	Gongshan	China	98.7	27.8	Tibeto-Burman	30	21
131	Va-1	Simao	China	101.0	22.8	Austro-Asiatic	22	20
132	Va-1	Gengma,Ximeng	China	99.4	23.6	Austro-Asiatic	36	22
133	Bugan	Xichou	China	104.7	23.5	Austro-Asiatic	32	23
134	Dai-YN1	Jinghong	China	100.8	22.1	Tai-Kadai	81	21, 23
135	Dai-YN2	Xishuangbanna	China	100.8	22.0	Tai-Kadai	21	20
136	Buyang-YN	Guangnan	China	105.1	24.1	Tai-Kadai	31	23
137	Lachi-YN	Maguan	China	104.4	23.0	Tai-Kadai	30	23
138	OT-YN	Malipo	China	104.7	23.1	Tai-Kadai	25	23
139	GL-YN	Malipo	China	104.7	23.2	Tai-Kadai	14	23
140	Miao-YN	Wenshan	China	104.3	23.4	Hmong-Mien	39	24
141	Yao-YN1	Mengla	China	101.6	21.5	Hmong-Mien	37	24
142	Yao-YN2	Malipo	China	104.7	23.2	Hmong-Mien	40	24
143	BY-GZ1	Libo	China	107.9	25.4	Tai-Kadai	33	23
144	BY-GZ2	Pingtang	China	107.3	25.9	Tai-Kadai	30	23
145	GL-GZ1	Bijie	China	105.3	27.3	Tai-Kadai	12	23
146	GL-GZ2	Majiang	China	107.6	26.5	Tai-Kadai	29	23
147	GL-GZ3	Dafang	China	105.6	27.2	Tai-Kadai	31	23
148	Yao-HN1	Jishou	China	109.7	28.3	Hmong-Mien	103	24
149	Yao-HN2	Jianghua	China	111.6	25.2	Hmong-Mien	24	24
150	ML-GX	Luocheng	China	108.9	24.8	Tai-Kadai	66	23
151	Maonan-GX	Huanjiang	China	108.3	24.8	Tai-Kadai	32	23
152	Caolan-GX	Fangcheng	China	108.4	21.8	Tai-Kadai	30	23
153	Sui-GX	Rongshui	China	109.3	25.1	Tai-Kadai	30	23
154	Zhuang-GX1	Tianlin	China	106.2	24.3	Tai-Kadai	25	23
155	Zhuang-GX2	Hezhou	China	111.6	24.4	Tai-Kadai	55	25
156	WS-GX	Rongshui	China	109.3	25.1	Tai-Kadai	33	23
157	Yerong-GX	Napo	China	105.8	23.4	Tai-Kadai	15	23
158	GL-GX	Longlin	China	105.4	24.9	Tai-Kadai	30	23
159	Dong-GX	Sanjiang	China	109.6	25.8	Tai-Kadai	72	25
160	Yao-GX1	Dahua	China	108.0	23.8	Hmong-Mien	19	24
161	Yao-GX2	Tianlin	China	106.2	24.3	Hmong-Mien	64	24
162	Yao-GX3	Fuchuan	China	111.3	24.8	Hmong-Mien	102	24
163	Yao-GX4	Fangcheng	China	108.4	21.8	Hmong-Mien	30	24
164	Yao-GX5	Hezhou	China	111.6	24.4	Hmong-Mien	41	24
165	Yao-GX6	Shangsi	China	108.0	22.2	Hmong-Mien	32	24
166	Yao-GX7	Jinxiu	China	110.2	24.1	Hmong-Mien	67	25
167	Yao-GD	Liannan	China	108.0	22.2	Hmong-Mien	35	25
168	Danga-HI	Lingshui	China	110.0	18.5	Tai-Kadai	40	23
169	Lingao-HI	Lingao	China	109.7	19.9	Tai-Kadai	31	23
170	Hlai-Qi-HI	Tongza	China	109.5	18.8	Tai-Kadai	34	23
171	Jiamao-HI	Baoting	China	109.7	18.6	Tai-Kadai	27	23
172	Cun-HI	Dongfang	China	108.7	19.1	Tai-Kadai	30	23
173	Kinh	Hanoi	Vietnam	105.9	21.0	Austro-Asiatic	139	26
174	Middle Viet	Middle Vietnam	Vietnam	107.6	16.5	Austro-Asiatic	62	23
175	Northern VIE1	Hanoi	Vietnam	105.9	21.1	Austro-Asiatic	187	27,28

176	Northern VIE2	Vietnam	Vietnam	105.0	21.2	Austro-Asiatic	42	28
177	Viet-South	South Vietnamese from California	Vietnam	105.0	10.5	Austro-Asiatic	35	29
178	Thai-KK	Khon Kaen	Thailand	102.6	16.0	Tai-Kadai	44	30
179	Phuthai-THA	Nakhon Pathom	Thailand	100.2	14.0	Tai-Kadai	25	30
180	LSg-THA	Suphan Buri	Thailand	100.0	14.6	Tai-Kadai	25	30
181	Thai-CM	Chiang Mai	Thailand	98.7	18.9	Tai-Kadai	220	30,31
182	Thai-Jin	Thailand	Thailand	99.7	17.3	Tai-Kadai	40	31
183	Thai-Yao	Northern Thailand	Thailand	99.3	18.9	Tai-Kadai	34	21
184	Thai-Korat	Thailand	Thailand	99.7	17.3	Tai-Kadai	32	32
185	Akha	Chiang Rai	Thailand	99.8	19.9	Tibeto-Burman	91	33
186	Lahu	Chiang Mai	Thailand	99.0	18.8	Tibeto-Burman	39	33
187	Lisu	Chiang Mai	Thailand	98.8	19.2	Tibeto-Burman	54	33
188	Lisu	Chiang Rai	Thailand	99.7	19.9	Tibeto-Burman	41	33
189	Lisu	Mae Hong Son	Thailand	99.0	18.9	Tibeto-Burman	25	33
190	Mussur	Chiang Mai	Thailand	99.7	19.9	Tibeto-Burman	21	30
191	Lisu_4	Chiang Mai	Thailand	97.9	18.8	Tibeto-Burman	25	30
192	Sakai	Trang	Thailand	99.6	7.6	Austro-Asiatic	20	30
193	Chong	Chanthaburi	Thailand	102.1	12.6	Austro-Asiatic	25	30
194	Khm	Northeast Thailand	Thailand	102.2	12.8	Austro-Asiatic	22	32
195	ChB	Northeast Thailand	Thailand	102.3	14.9	Austro-Asiatic	20	32
196	Cambodia	Siem Reap, NW Cambodia	Cambodia	103.9	13.4	Austro-Asiatic	31	34
197	Semang	West Malaysia	Malaysia	101.8	4.8	Austro-Asiatic	112	35
198	Senoi	West Malaysia	Malaysia	101.9	5.1	Austro-Asiatic	52	6
199	Semelai	West Malaysia	Malaysia	102.3	2.9	Austro-Asiatic	61	6
200	Nicobarese	Eastern India	India	92.5	10.6	Austro-Asiatic	46	36, 37
201	Barma	Myanmar	Myanmar	98.1	16.7	Tibeto-Burman	116	38
202	Karen	Myanmar	Myanmar	98.0	17.1	Tibeto-Burman	155	38
203	Chakma	Chittagong hill tract	Bangladesh	91.8	22.4	Tibeto-Burman	108	39
204	Marma	Chittagong hill tract	Bangladesh	91.8	22.4	Tibeto-Burman	97	39
205	Tripura	Chittagong hill tract	Bangladesh	91.8	22.3	Tibeto-Burman	97	39
206	Burmans_1	Sagaing	Myanmar	95.4	21.8	Tibeto-Burman	32	40
207	Burmans_2	Sagaing	Myanmar	94.1	23.4	Tibeto-Burman	53	40
208	Burmans_3	Sagaing	Myanmar	94.6	23.8	Tibeto-Burman	51	40
209	Burmans_4	Magway	Myanmar	94.6	20.6	Tibeto-Burman	122	40
210	Burmans_5	Bago	Myanmar	95.7	17.8	Tibeto-Burman	69	40
211	Burmans_6	Ayeyarwady	Myanmar	95.3	18.4	Tibeto-Burman	72	40
212	Naga_1	Sagaing	Myanmar	95.3	25.9	Tibeto-Burman	32	40
213	Naga_2	Sagaing	Myanmar	94.9	25.5	Tibeto-Burman	30	40
214	Naga_3	Sagaing	Myanmar	95.0	25.2	Tibeto-Burman	39	40
215	Chin_1	Chin	Myanmar	93.7	23.5	Tibeto-Burman	58	40
216	Chin_2	Chin	Myanmar	93.5	21.6	Tibeto-Burman	187	40
217	Chin_3	Magway	Myanmar	94.3	20.2	Tibeto-Burman	13	40
218	Rakhine_1	Magway	Myanmar	94.1	21.7	Tibeto-Burman	24	40
219	Rakhine_2	Rakhine	Myanmar	94.8	18.7	Tibeto-Burman	63	40
220	Bharia	Chhindwara, MP	Inida	78.9	22.1	Dravidian	65	42
221	Bhil	sehore, MP	Inida	77.1	23.2	Indo-European	49	42
222	Saharia	Shivuri, MP	Inida	77.7	25.4	Austro-Asiatic	95	42

223	Khon Mueang	North	Thailand	99.2	19.7	Tai-Kadai	25	43
224	Khon Mueang	North	Thailand	99.9	20.1	Tai-Kadai	25	43
225	Khon Mueang	North	Thailand	100.3	19.6	Tai-Kadai	24	43
226	Khon Mueang	North	Thailand	98.3	19.6	Tai-Kadai	25	43
227	Khon Mueang	North	Thailand	99.7	18.7	Tai-Kadai	23	43
228	Khon Mueang	North	Thailand	98.1	18.7	Tai-Kadai	25	43
229	Khon Mueang	North	Thailand	99.2	19.7	Tai-Kadai	25	43
230	Khon Mueang	North	Thailand	100.3	19.6	Tai-Kadai	24	43
231	Khon Mueang	North	Thailand	98.3	19.6	Tai-Kadai	25	43
232	Yuan	North	Thailand	99.6	19.4	Tai-Kadai	17	43
233	Yuan	Central	Thailand	100.6	14.3	Tai-Kadai	25	43
234	Shan	North	Thailand	99.9	20.1	Tai-Kadai	25	43
235	Lao Isan	Northeast	Thailand	102.1	17.3	Tai-Kadai	25	43
236	Lao Isan	Northeast	Thailand	102.0	17.7	Tai-Kadai	25	43
237	Lao	North	Laos	102.2	21.1	Tai-Kadai	25	43
238	Lao	Central	Laos	102.3	19.9	Tai-Kadai	24	43
239	Phutai	Northeast	Thailand	102.1	17.3	Tai-Kadai	25	43
240	Seak	Northeast	Thailand	104.5	15.7	Tai-Kadai	26	43
241	Nyaw	Northeast	Thaialnd	103.2	15.8	Tai-Kadai	25	43
242	Black Tai	Central	Thailand	100.6	14.3	Tai-Kadai	25	43
243	Phuan	North	Thailand	99.2	19.7	Tai-Kadai	25	43
244	Phuan	North	Thailand	99.9	20.1	Tai-Kadai	25	43
245	Phuan	North	Thailand	100.3	19.6	Tai-Kadai	25	43
246	Phuan	Central	Thailand	102.0	14.3	Tai-Kadai	25	43
247	Phuan	Central	Thailand	100.3	15.1	Tai-Kadai	25	43
248	Mon	Northeast	Thailand	102.0	17.7	Austroasiatic	25	43
249	Mon	West	Thailand	99.4	13.7	Austroasiatic	23	43
250	Mon	Central	Thailand	100.6	14.3	Austroasiatic	25	43
251	Mon	Central	Thailand	100.6	14.3	Austroasiatic	22	43
252	Khmer	Northeast	Thailand	104.5	15.7	Austroasiatic	19	43
253	Khmer	Northeast	Thailand	103.2	15.8	Austroasiatic	25	43
254	Bru	Northeast	Thailand	103.2	15.8	Austroasiatic	24	43
255	Blang	North	Thailand	99.2	19.7	Austroasiatic	25	43
256	Paluang	North	Thailand	99.9	20.1	Austroasiatic	25	43
257	Lawa	North	Thailand	100.3	19.6	Austroasiatic	22	43
258	Lawa	North	Thailand	98.3	19.6	Austroasiatic	24	43
259	Lawa	North	Thailand	99.7	18.7	Austroasiatic	24	43
260	Makrani		Pakistan	69.1	29.0	Arabic	81	44
261	Caucasus		Caucasus	46.6	39.9	Indo-European	147	45
262	Altaians		Siberia	102.4	55.4	Altaic	110	49
263	Khakassians		Siberia	90.3	53.5	Altaic	53	49
264	Buryats		Siberia	98.8	56.3	Altaic	90	49
265	Sojots		Siberia	93.0	55.6	Altaic	30	49
266	Todjins		Siberia	94.9	74.0	Altaic	49	49
267	Tuvinians		Siberia	94.5	51.2	Altaic	90	49
268	Tofalar		Siberia	104.2	52.1	Altaic	58	49
269	Tubalar	East	Siberia	94.9	56.2	Turkic	72	46
270	Tuvan	East	Siberia	94.5	51.2	Turkic	96	46
271	Buryat	East	Siberia	98.8	56.4	Mongolic	25	46
272	Tofalar	East	Siberia	104.2	52.1	Turkic	46	46
273	Evenki	East	Siberia	102.7	53.8	Tungusic	71	46
274	Negidal	East	Siberia	130.9	51.3	Tungusic	33	46

275	Ulchi	East	Siberia	139.1	52.1	Tungusic	87	46
276	Nivkhi	East	Siberia	37.9	56.4	unknown	56	46
277	Udegey	East	Siberia	135.1	48.5	Tungusic	46	46
278	Afghanistan		Uzbekistan	66.4	38.8	Unknown	85	47
279	Kyrgyzstan		Uzbekistan	70.0	41.4	Unknown	226	47
280	Kazakhstan		Uzbekistan	61.0	41.4	Unknown	214	47
281	Russia		Uzbekistan	57.5	43.7	Unknown	150	47
282	Tajikistan		Uzbekistan	67.5	38.9	Unknown	238	47
283	Turkmenistan		Uzbekistan	63.6	39.8	Unknown	212	47
284	Fergana		Uzbekistan	71.8	40.4	Unknown	49	47
285	Karakalpakstan		Uzbekistan	59.2	43.7	Unknown	40	47
286	Qashkadarya		Uzbekistan	66.1	38.8	Unknown	62	47
287	Tashkent		Uzbekistan	69.3	41.3	Unknown	55	47
288	Khorezm		Uzbekistan	60.9	41.4	Unknown	94	47
289	Minnan	Kaohsiung	Taiwan	120.9	23.8	Unknown	50	48
						Total (N)	20917	

Table 8.2 | Hg frequency (%) of 289 populations analyzed in this study. The value 0 indicates the frequency (%) range between 0.1-0.49. Due to the large size of the file, the complete data for **Table 8.2** is available only in online (excel format) on request. Here the data shown in the table are the representative data included from the first and last segment of the original table

		A	B	B4	B5	B6	C	CZ	D	D/G	D1	D4	D5	D6	E	F	F1	F2	F3	F4	G	G1	G2	G3	G4	H	HV	I	J	K	M*
1	MDR	8	-	-	-	-	-	-	-	-	-	-	-	-	-	-	10	-	-	-	-	-	2	-	-	1	11	13	-	-	-
2	STH	3	-	-	3	-	-	-	-	-	-	6	-	-	-	-	19	-	-	-	-	-	-	-	-	-	-	-	-	-	-
3	N-MIX	6	-	-	-	-	-	-	-	-	-	10	-	-	-	-	24	4	-	-	-	-	6	-	-	2	-	-	-	-	-
4	Maharjan	1	-	-	-	-	1	-	-	-	-	4	2	-	-	-	26	-	-	-	-	-	1	-	-	1	-	-	-	-	1
5	BJR	10	-	-	-	-	-	-	-	-	-	10	-	-	-	-	10	-	-	-	-	-	-	-	-	-	-	-	-	-	-
6	SHK	-	-	-	-	-	-	-	-	-	-	11	5	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
7	Udaya	7	-	-	-	-	-	-	-	-	-	7	-	-	-	-	12	3	-	-	-	-	7	-	-	-	-	-	-	-	-
8	BR-NP	2	-	2	-	-	2	-	-	-	-	2	2	-	-	-	6	-	-	-	-	-	2	6	-	2	-	-	-	-	-
9	MGR	-	-	-	3	-	8	-	-	-	-	14	11	-	-	-	-	-	-	-	-	-	-	-	-	5	-	-	-	-	-
10	N-MIX2	8	-	-	2	-	-	-	-	-	-	3	5	-	-	-	11	2	-	-	-	-	5	-	-	-	3	-	-	-	-
11	Tamang	5	-	-	-	-	2	-	-	-	-	16	11	-	-	-	7	-	-	-	-	-	11	-	-	-	-	-	-	-	-
12	Kathmandu	5	-	-	1	-	4	-	-	-	-	9	3	-	-	-	8	-	-	-	-	-	-	-	-	1	1	-	1	-	-
13	Nepali-other	6	-	-	0	-	1	-	1	-	-	6	3	-	-	-	7	-	-	-	2	-	2	1	-	-	0	2	-	-	-
14	Sherpa	27	-	-	-	-	-	-	6	-	-	-	-	-	-	-	2	1	-	-	1	-	-	-	-	1	-	-	-	-	-
15	H-TE	-	-	-	-	-	-	-	-	-	-	4	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	4	-	-	-
16	TH-CI	-	-	-	9	-	-	-	-	-	-	5	-	-	-	-	7	-	-	-	11	-	16	-	-	2	-	-	-	-	-
17	TH-CII	-	-	-	3	-	5	-	-	-	-	12	-	-	-	-	5	-	-	-	8	-	13	-	-	-	-	-	-	-	1
18	TH-E	-	-	-	3	-	3	-	-	-	-	10	-	-	-	-	5	-	-	-	3	-	10	-	-	-	-	-	-	-	-
19	H-ND	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	2	-	-	-	-	2	-	-	-	-	2
20	Tribal-AP	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	10
21	Tharu-UTK	2	-	-	2	-	-	-	-	-	-	9	-	-	-	-	2	-	-	-	14	-	9	-	-	-	-	-	-	-	7
22	Tharu -UP	1	-	-	4	-	3	-	-	-	-	4	-	-	-	1	-	-	-	-	3	-	5	-	-	-	-	-	1	1	1
23	Brahmin-UTK	2	-	-	-	-	2	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	1	-	-	1	-	-	-	2
24	Kshatriya	4	-	1	-	-	-	-	-	-	-	12	4	-	-	-	13	1	-	-	-	-	-	5	-	-	-	1	-	-	4
25	Shah	17	-	6	-	-	6	-	-	-	-	11	-	-	-	6	6	-	-	-	-	3	-	6	-	-	3	-	-	-	-
26	Goswami	9	-	-	-	-	4	-	-	-	-	4	-	-	-	-	2	-	-	-	-	-	9	-	-	-	-	-	-	-	2
27	Arya	4	-	-	-	-	-	-	-	-	-	2	2	-	-	-	-	-	-	-	-	-	-	2	-	-	-	2	-	-	-
28	Tamta	12	-	-	-	-	3	-	-	-	-	6	-	-	-	-	-	-	-	-	-	-	-	3	-	-	-	-	-	-	-

		R2	R21	R22	R3	R30	R31	R32	R5	R6	R7	R8	R9	T	U	U1	U2I	U2	U2	u3	U4	U5	U7	U8	U9	V	W	X	Y	Z
281	TUR	1	-	-	-	-	-	-	0	-	-	-	-	6	-	1	1	-	2	1	4	2	1	-	-	-	2	3	1	-
282	FER	-	-	-	-	-	-	-	8	-	-	-	-	4	-	-	-	-	-	-	-	10	-	-	-	-	2	-	2	-
283	KAR	-	-	-	-	-	-	-	-	-	-	-	-	5	-	3	-	-	5	-	5	3	-	-	-	-	3	5	-	-
284	QAS	2	-	-	-	2	-	-	-	-	-	-	-	6	-	-	-	-	2	2	-	-	5	-	-	-	3	-	-	-
285	TAS	-	-	-	-	-	-	-	-	-	-	-	-	7	-	-	-	-	9	2	-	7	2	-	-	-	-	2	-	-
286	XOR	2	-	-	-	-	-	-	-	-	-	-	-	12	-	2	-	-	1	1	3	3	1	-	-	-	2	3	-	-
287	Minnan (TAIW	-	-	-	-	-	-	-	-	-	-	-	-	4	-	-	-	-	-	-	-	-	-	-	-	-	-	-	2	2

8.2 Appendix B

References for **Table 8.1** (population information) and **Table 8.2** (Haplogroup frequency of the 289 analyzed populations) is provided below:

1. Gayden, T. et al. The Himalayas: barrier and conduit for gene flow. Am J Phys Anthropol 151, 169-182, doi:10.1002/ajpa.22240 (2013).
2. Wang, H. W. et al. Revisiting the role of the Himalayas in peopling Nepal: insights from mitochondrial genomes. J Hum Genet 57, 228-234, doi:10.1038/jhg.2012.8 (2012).
3. Bhandari, S. et al. Genetic evidence of a recent Tibetan ancestry to Sherpas in the Himalayan region. Sci Rep 5, 16249, doi:10.1038/srep16249 (2015).
4. Fornarino, S. et al. Mitochondrial and Y-chromosome diversity of the Tharus (Nepal): a reservoir of genetic variation. BMC Evol Biol 9, 154, doi:10.1186/1471-2148-9-154 (2009).
5. Chaubey, G. et al. Unravelling the distinct strains of Tharu ancestry. Eur J Hum Genet 22, 1404-1412, doi:10.1038/ejhg.2014.36 (2014).
6. Negi, N. et al. The paternal ancestry of Uttarakhand does not imitate the classical caste system of India. J Hum Genet 61, 167-172, doi:10.1038/jhg.2015.121 (2016).
7. Thangaraj, K. et al. Different population histories of the Mundari- and Mon-Khmer-speaking Austro-Asiatic tribes inferred from the mtDNA 9-bp deletion/insertion polymorphism in Indian populations. Hum Genet 116, 507-517, doi:10.1007/s00439-005-1271-6 (2005).
8. Metspalu, M. et al. Most of the extant mtDNA boundaries in south and southwest Asia were likely shaped during the initial settlement of Eurasia by anatomically modern humans. BMC Genet 5, 26, doi:10.1186/1471-2156-5-26 (2004).
9. Chaubey G, Metspalu M, Karmin M et al.: Language shift by indigenous population: a model genetic study in South Asia. International Journal of Human Genetics 2008; 8: 41.
10. Thanseem, I. et al. Genetic affinities among the lower castes and tribal groups of India: inference from Y chromosome and mitochondrial DNA. BMC Genet 7, 42, doi:10.1186/1471-2156-7-42 (2006).
11. Kang, L. et al. MtDNA analysis reveals enriched pathogenic mutations in Tibetan highlanders. Sci Rep 6, 31083, doi:10.1038/srep31083 (2016).
12. Peng, M. S. & Zhang, Y. P. Inferring the population expansions in peopling of Japan. PLoS One 6, e21509, doi:10.1371/journal.pone.0021509 (2011).
13. Qi, X. et al. Genetic evidence of paleolithic colonization and neolithic expansion of modern humans on the tibetan plateau. Mol Biol Evol 30, 1761-1778, doi:10.1093/molbev/mst093 (2013).
14. Cordaux, R. et al. Mitochondrial DNA analysis reveals diverse histories of tribal populations from India. Eur J Hum Genet 11, 253-264, doi:10.1038/sj.ejhg.5200949 (2003).
15. Roychoudhury, S. et al. Genomic structures and population histories of linguistically distinct tribal groups of India. Hum Genet 109, 339-350, doi:10.1007/s004390100577 (2001).
16. Reddy, B. M. et al. Austro-Asiatic tribes of Northeast India provide hitherto missing genetic link between South and Southeast Asia. PLoS One 2, e1141, doi:10.1371/journal.pone.0001141 (2007).
17. Zhao, M. et al. Mitochondrial genome evidence reveals successful Late Paleolithic settlement on the Tibetan Plateau. Proc Natl Acad Sci U S A 106, 21230-21235, doi:10.1073/pnas.0907844106 (2009).
18. Qin, Z. et al. A mitochondrial revelation of early human migrations to the Tibetan Plateau before and after the last glacial maximum. Am J Phys Anthropol 143, 555-569, doi:10.1002/ajpa.21350 (2010).
19. Wen, B. et al. Analyses of genetic structure of Tibeto-Burman populations reveals sex-biased admixture in southern Tibeto-Burmans. Am J Hum Genet 74, 856-865, doi:10.1086/386292 (2004).
20. Qian, Y. P. et al. Mitochondrial DNA polymorphisms in Yunnan nationalities in China. J Hum Genet 46, 211-220, doi:10.1007/s100380170091 (2001).
21. Yao, Y. G. et al. Genetic relationship of Chinese ethnic populations revealed by mtDNA sequence diversity. Am J Phys Anthropol 118, 63-76, doi:10.1002/ajpa.10052 (2002).
22. Yao, Y. G. & Zhang, Y. P. Phylogeographic analysis of mtDNA variation in four ethnic populations from Yunnan Province: new data and a reappraisal. J Hum Genet 47, 311-318, doi:10.1007/s100380200042 (2002).
23. Li, H. et al. Mitochondrial DNA diversity and population differentiation in southern East Asia. Am J Phys Anthropol 134, 481-488, doi:10.1002/ajpa.20690 (2007).
24. Wen, B. et al. Genetic structure of Hmong-Mien speaking populations in East Asia as revealed by mtDNA lineages. Mol Biol Evol 22, 725-734, doi:10.1093/molbev/msi055 (2005).

25. Gan, R. J. et al. Pinghua population as an exception of Han Chinese's coherent genetic structure. J Hum Genet 53, 303-313, doi:10.1007/s10038-008-0250-x (2008).
26. Peng, M. S. et al. Tracing the Austronesian footprint in Mainland Southeast Asia: a perspective from mitochondrial DNA. Mol Biol Evol 27, 2417-2430, doi:10.1093/molbev/msq131 (2010).
27. Irwin, J. A. et al. Mitochondrial control region sequences from a Vietnamese population sample. Int J Legal Med 122, 257-259, doi:10.1007/s00414-007-0205-3 (2008).
28. Jin, H. J., Tyler-Smith, C. & Kim, W. The peopling of Korea revealed by analyses of mitochondrial DNA and Y-chromosomal markers. PLoS One 4, e4210, doi:10.1371/journal.pone.0004210 (2009).
29. Oota, H. et al. Extreme mtDNA homogeneity in continental Asian populations. Am J Phys Anthropol 118, 146-153, doi:10.1002/ajpa.10056 (2002).
30. Fucharoen, G., Fucharoen, S. & Horai, S. Mitochondrial DNA polymorphisms in Thailand. J Hum Genet 46, 115-125, doi:10.1007/s100380170098 (2001).
31. Zimmermann, B. et al. Forensic and phylogeographic characterization of mtDNA lineages from northern Thailand (Chiang Mai). Int J Legal Med 123, 495-501, doi:10.1007/s00414-009-0373-4 (2009).
32. Lertrit, P. et al. Genetic history of Southeast Asian populations as revealed by ancient and modern human mitochondrial DNA analysis. Am J Phys Anthropol 137, 425-440, doi:10.1002/ajpa.20884 (2008).
33. Oota, H., Settheetham-Ishida, W., Tiwawech, D., Ishida, T. & Stoneking, M. Human mtDNA and Y-chromosome variation is correlated with matrilocal versus patrilocal residence. Nat Genet 29, 20-21, doi:10.1038/ng711 (2001).
34. Black, M. L., Dufall, K., Wise, C., Sullivan, S. & Bittles, A. H. Genetic ancestries in northwest Cambodia. Ann Hum Biol 33, 620-627 (2006).
35. Hill, C. et al. Phylogeography and ethnogenesis of aboriginal Southeast Asians. Mol Biol Evol 23, 2480-2491, doi:10.1093/molbev/msl124 (2006).
36. Prasad, B. V. et al. Mitochondrial DNA variation in Nicobarese Islanders. Hum Biol 73, 715-725 (2001).
37. Thangaraj, K. et al. Genetic affinities of the Andaman Islanders, a vanishing human population. Curr Biol 13, 86-93 (2003).
38. Summerer, M. et al. Large-scale mitochondrial DNA analysis in Southeast Asia reveals evolutionary effects of cultural isolation in the multi-ethnic population of Myanmar. BMC Evol Biol 14, 17, doi:10.1186/1471-2148-14-17 (2014).
39. Gazi, N. N. et al. Genetic structure of Tibeto-Burman populations of Bangladesh: evaluating the gene flow along the sides of Bay-of-Bengal. PLoS One 8, e75064, doi:10.1371/journal.pone.0075064 (2013).
40. Li, Y. C. et al. Ancient inland human dispersals from Myanmar into interior East Asia since the Late Pleistocene. Sci Rep 5, 9473, doi:10.1038/srep09473 (2015).
42. Sharma, G., Tamang, R., Chaudhary, R., Singh, V. K., Shah, A. M., Anugula, S., . . . Thangaraj, K. (2012). Genetic affinities of the central Indian tribal populations. PLoS One, 7(2), e32546-e32546. doi:10.1371/journal.pone.0032546
43. Kutanan, W., Kampuansai, J., Srikummool, M., Kangwanpong, D., Ghirotto, S., Brunelli, A., & Stoneking, M. (2017). Complete mitochondrial genomes of Thai and Lao populations indicate an ancient origin of Austroasiatic groups and demic diffusion in the spread of Tai-Kadai languages. Hum Genet, 136(1), 85-98. doi:10.1007/s00439-016-1742-y
44. Siddiqi, M. H., Akhtar, T., Rakha, A., Abbas, G., Ali, A., Haider, N., . . . Ahmad, J. (2015). Genetic characterization of the Makrani people of Pakistan from mitochondrial DNA control-region data. Legal Medicine, 17(2), 134-139.
45. Schönberg, A., Theunert, C., Li, M., Stoneking, M., & Nasidze, I. (2011). High-throughput sequencing of complete human mtDNA genomes from the Caucasus and West Asia: high diversity and demographic inferences. European Journal of Human Genetics, 19(9), 988-994.
46. Starikovskaya, E. B., Sukernik, R. I., Derbeneva, O. A., Volodko, N. V., Ruiz- Pesini, E., Torroni, A., . . . Huoponen, K. (2005). Mitochondrial DNA diversity in indigenous populations of the southern extent of Siberia, and the origins of Native American Hgs. Annals of human genetics, 69(1), 67-89.
47. Irwin, J. A., Ikramov, A., Saunier, J., Bodner, M., Amory, S., Röck, A., . . . Mukhamedov, R. (2010). The mtDNA composition of Uzbekistan: a microcosm of Central Asian patterns. International Journal of Legal Medicine, 124(3), 195-204.
48. Ko, A. M.-S., Chen, C.-Y., Fu, Q., Delfin, F., Li, M., Chiu, H.-L., . . . Ko, Y.-C. (2014). Early Austronesians: into and out of Taiwan. The American Journal of Human Genetics, 94(3), 426-436.

8.3 Appendix C

Table 8.3 | mtDNA sequence variations of Newar, Magar and Brahmin. Polymorphism (Poly) not associated with the Hgs are segregated into two groups 1) Not_found_polys & 2) Remaining_Polys. If the Hg associated polymorphisms (Polys) are missing, they are indicated as Not_found_polys. Whereas extra variants not associated with particular Hg are represented by Remaining_polys. Remaining polys may also contain mutational hotspots such as 309.XC, 16519, 523-524d etc. For more detail see Materials and Methods (Haplogroup Assignment). Due to the large size of the file, the complete data for **Table 8.3** is available only in online format (excel format) on request. Here the data shown in the table are the few representative data included from the original table.

Maharjan (M001-M100)					
Sample ID	**HG**	Range	Input_sample	N0t_found_polys	Remaining_Polys
M001	U2b1a	1-16569 ;	73G 146C 263G 309.1C 315.1C 523d 524d 750G 1438G 1811G 3010A 3915A 4093G 4769G 5186T 6620C 6737G 7028T 8860G 10688A 11467G 11719A 12106T 12308G 12372A 13194A 13708A 14605G 14766T 15049T 15326G 16051G 16168T		309.1C 315.1C 523d 524d 3010A 6620C 6737G 10688A 14605G
M002	Z3a1a	1-16569 ;	73G 152C 207A 249d 263G 309.1CCC 489C 750G 1438G 2706G 4715G 4769G 6752G 7028T 7196A 8584A 8701G 8860G 8931G 9090C 9540C 9713A 10208C 10398G 10400T 10873C 11075C 11719A 12705T 13620C 14766T 14783C 15043A 15301A 15487T 15784C 15928A 16150T 16185T 16223T 16260T 16298C		309.1CCC 16150T
M003	M9a1a2	10025-10818 ; 3266-3957 ; 12040-12547 ; 7196-7916 ; 12611-13446 ; 13432-14248 ; 14894-15629 ; 3831-4661 ; 15831-16569 ;	3394C 4491A 7256T 7468T 10398G 10400T 12705T 13731G 15043A 15301A 15326G 16145A 16223T 16234T 16316G	16362C	7468T 13731G
M004	D4j1a1	1-16569 ;	73G 263G 309.1C 315.1C 489C 750G 1438G 2706G 3010A 4769G 4883T 5178A 5262A 7028T 7581C 7783C 8414T 8701G 8860G 9540C 10381G 10398G 10400T 10873C 11696A 11719A 12130C 12358G 12705T 14070G 14668T 14766T 14783C 15043A 15295T 15301A 15326G 16086C 16223T 16362C		309.1C 315.1C 10381G 14070G
M005	U7a3a	10074-10848 ; 15014-15746 ; 3268-3875 ; 15818-16544 ; 12036-12777 ; 7966-8806 ; 9317-10117 ; 12610-13383 ; 13371-14175 ; 1-749 ; 3831-4661 ;	73G 151T 152C 263G 267C 309.1C 315.1C 523d 524d 3741T 8137T 8545A 8684T 9852G 10142T 12308G 12372A 12618A 13500C 15326G 15364T 16223T 16309G 16318T 16519C		267C 309.1C 315.1C 523d 524d 8545A 15364T 16223T 16519C
M006	F1d1	10025-10842 ; 14943-15710 ; 3199-3967 ; 16185-16543 ; 11971-12725 ; 12596-13343 ; 13361-14270 ; 3831-4650 ; 5298-6038 ; 5347-6039 ;	3970T 5628C 10310A 10609C 12406A 12882T 13135A 13928C 15204C 15326G 15402T 16189C 16263C 16284G 16304C 16519C		15204C 16263C 16284G 16519C
M007	F1d1	15039-15856 ; 3195-3964 ; 11980-12766 ; 16184-16568 ; 10027-10852 ; 3832-4667 ; 5350-6042 ; 12604-13154 ;	3970T 5628C 10310A 10609C 12406A 12840T 12882T 13135A 15204C 15326G 15402T 16189C 16284G 16304C 16519C		12840T 15204C 16284G 16519C
M008	M5a	14971-15709 ; 3225-3931 ; 15833-16564 ; 11974-12622 ; 10026-10758 ; 10026-10651 ; 12605-13474 ; 3832-4657 ; 1288-1991 ;	1438G 1888A 3398C 3921T 10398G 10400T 12477C 12705T 15043A 15301A 15326G 16129A 16223T 16224C 16311C 16519C		3398C 16224C 16311C 16519C
M009	F1g	10025-10846 ; 15037-15776 ; 3201-3932 ; 11980-12637 ; 16184-16544 ; 12619-13473 ; 3833-4666 ; 1983-2415 ;	2389T 3398C 3630T 3970T 10310A 10609C 12406A 12882T 15326G 16189C 16304C 16519C		3630T 16519C

8.4 Appendix D

Table 8.4 | mtDNA sequence variations of Changpa, Ladhak and North India. Due to the large size of the file, the complete data for Table 7.6 is available only in online format (excel format) on request. Here the data shown in the table are the representative data included from original table.

Ladhak (LDK001-LDK089)					
Sample ID	**Haplo group**	**Range**	**Input_Sample**	**Not_Found_Polys**	**Remaining_Polys**
LDK001	**U2a**	1471-2005 ; 11961-12601 ; 3889-4657 ; 5-507 ; 508-716 ; 10087-10848 ; 15913-16569 ; 12050-12776 ;	1811G 4171T 44.1C 524.1A 12034T 16051G 200G 16291T 309.2C 146C 16206C 12696C 309.1C 73G 524.2C 12308G 16129A 12372A 263G 16519C 315.1C		44.1C 146C 200G 524.1AC 4171T 12034T 12696C 16129A 16291T
LDK002	**U2b**	6570-7288 ; 713-1418 ; 1401-2011 ; 11970-12532 ; 3890-4667 ; 15189-15681 ; 389-712 ; 10075-10846 ; 15911-16569 ; 12051-12778 ;	750G 1438G 4053G 16051G 1888A 7028T 1811G 12106T 15326G 709A 12397G 16124C 12308G 16239T 12372A		709A 1888A 4053G 12397G 16124C 16239T
LDK003	**A21**	1-16569 ;	11719A 12603T 12705T 14364A 1438G 14766T 15326G 15404C 16092C 16223T 16290T 16319A 16362C 1736G 2706G 4248C 4769G 4824G 663G 7028T 750G 8794T 8860G 263G 309.1C 309.2C 315.1C 235G 152C 73G		12603T 15404C 16092C
LDK004	**I4b**	1-16569 ;	10034C 10238C 10398G 11530G 11719A 12501A 12705T 13662T 13708A 13780G 1438G 14766T 15118G 15326G 15924G 16129A 16223T 16271C 16391A 16519C 1719A 2308G 2706G 4113A 4529T 4769G 7028T 750G 8251A 8519A 8768T 8860G 73G 199C 204C 250C 263G 315.1C 573.1C 573.2C 573.3C 573.4C 573.5C 573.6C 15043A		8768T 11530G 13708A 15118G 16271C
LDK005	**A21**	11966-12664 ; 3891-4657 ; 15078-15710 ; 1-355 ; 444-729 ; 10087-10840 ; 15913-16569 ; 12124-12775 ; 14011-14550 ;	4248C 15326G 14364A 663G 12705T 16092C 73G 152C 235G 263G 309.1C 309.2C 309.3C 315.1C 315.2C 12603T 16223T 16290T 16319A 16362C 16519C		12603T 16092C
LDK006	**HV12b1**	12000-12778 ; 3936-4603 ; 16105-16400 ; 13560-13900 ; 15148-15740 ; 440-715 ; 10088-10847 ; 12125-12780 ;	15326G 15682G 16356C 13889A 15300C 249d 10873C		249d 10873C 15300C
LDK007	**R7**	6597-6958 ; 7247-8042 ; 8025-8692 ; 8682-9396 ; 9330-10114 ; 11380-12016 ; 12711-13407 ; 671-1413 ; 13402-14277 ; 14277-14423 ; 1912-2681 ; 2666-3295 ; 3268-3912 ; 4644-5434 ; 5434-5946 ; 5911-6621 ; 1339-1919 ; 11972-12777 ; 3890-4665 ; 15108-15704 ; 341-714 ; 16043-16162 ;	1007A 11719A 12867T 13105G 13647T 1438G 15326G 15416C 16085T 2706G 3633C 4769G 5601T 750G 7660C 8697A 8860G 9531G		1007A 3633C 5601T 7660C 8697A 9531G 12867T 13647T 15416C 16085T
LDK008	**M9a1a1c1b1a**	1341-1960 ; 1985-2658 ; 2632-3275 ; 3265-3874 ; 4622-5155 ; 5395-5945 ; 5942-6638 ; 708-1417 ; 11996-12664 ; 3826-4655 ; 15005-15687 ; 5-710 ; 10087-10858 ; 12051-12782 ; 8008-8613 ;	1438G 2706G 4769G 5899.1C 750G 4491A 15043A 73G 10398G 12705T 3394C 15301A 315.1C 15326G 489C 263G 10400T 1005C 711C	1041G	1005C

8.5 Appendix E

References for the accession number of the sequences used to construct PhyloTree are given below. The number (1, 2…n) provided here in the reference corresponds with the reference number (1', 2'... n') for the Sample ID/GenBank accession number for each sequence as shown in the corresponding Figure (PhyloTree).

- **Figure 4.8** (hg R5), **Figure 4.9** (hg R6) and **Figure 4.10** (hg R30)

1) Present study
2) Behar, D. M. et al. counting the founders: the matrilineal genetic ancestry of the Jewish Diaspora. PLoS One 3, e2062, doi:10.1371/journal.pone.0002062 (2008).
3) Khan, N. A. Unpublished
4) Silva, M. et al. A genetic chronology for the Indian Subcontinent points to heavily sex-biased dispersals. BMC Evol Biol 17, 88, doi:10.1186/s12862-017-0936-9 (2017).
5) Palanichamy, M. G. et al. Phylogeny of mitochondrial DNA macrohaplogroup N in India, based on complete sequencing: implications for the peopling of South Asia. Am J Hum Genet 75, 966-978, doi:10.1086/425871 (2004).
6) Chaubey, G. et al. Phylogeography of mtDNA haplogroup R7 in the Indian peninsula. BMC Evol Biol 8, 227, doi:10.1186/1471-2148-8-227 (2008).
7) Derenko, M. et al. Complete mitochondrial DNA diversity in Iranians. PLoS One 8, e80673, doi: 10.1371/journal.pone.0080673 (2013).
8) Sharma,V., Phylogenomic study of a concealed Ladhak tribe of the Great Himalayas
9) Summerer, M. et al. Large-scale mitochondrial DNA analysis in Southeast Asia reveals evolutionary effects of cultural isolation in the multi-ethnic population of Myanmar. BMC Evol Biol 14, 17, doi:10.1186/1471-2148-14-17 (2014).

- **Figure 4.13** (hg F1d1)

1) Present study
2) Kang, L. et al. mtDNA lineage expansions in Sherpa population suggest adaptive evolution in Tibetan highlands. Mol Biol Evol 30, 2579-2587, doi:10.1093/molbev/mst147 (2013).
3) Kutanan, W. et al. Complete mitochondrial genomes of Thai and Lao populations indicate an ancient origin of Austroasiatic groups and demic diffusion in the spread of Tai-Kadai languages. Hum Genet 136, 85-98, doi:10.1007/s00439-016-1742-y (2017).
4) Fornarino, S. et al. Mitochondrial and Y-chromosome diversity of the Tharus (Nepal): a reservoir of genetic variation. BMC Evol Biol 9, 154, doi:10.1186/1471-2148-9-154
5) Ji, F. et al. Mitochondrial DNA variant associated with Leber hereditary optic neuropathy and high-altitude Tibetans. Proc Natl Acad Sci U S A 109, 7391-7396, doi:10.1073/pnas.1202484109 (2012).
6) Tanaka, M. et al. Mitochondrial genome variation in eastern Asia and the peopling of Japan. Genome Res 14, 1832-1850, doi:10.1101/gr.2286304 (2004).

- **Figure 4.15** (hg F1c1a2)

1) Present study
2) Kang, L. et al. mtDNA lineage expansions in Sherpa population suggest adaptive evolution in Tibetan highlands. Mol Biol Evol 30, 2579-2587, doi:10.1093/molbev/mst147 (2013).
3) Kutanan, W. et al. Complete mitochondrial genomes of Thai and Lao populations indicate an ancient origin of Austroasiatic groups and demic diffusion in the spread of Tai-Kadai languages.
4) Ji, F. et al. Mitochondrial DNA variant associated with Leber hereditary optic neuropathy and high-altitude Tibetans. Proc Natl Acad Sci U S A 109, 7391-7396, doi:10.1073/pnas.1202484109 (2012).
5) Greenspan, B. Direct submission to GenBank, Unpublished

6) Khan, N.A., Govindaraj, P. and Thangaraj, K. Direct submission to GenBank

- **Figure 4.16** (hg F1g)

1) Present study
2) Kang, L. et al. mtDNA lineage expansions in Sherpa population suggest adaptive evolution in Tibetan highlands. Mol Biol Evol 30, 2579-2587, doi:10.1093/molbev/mst147 (2013).
3) Zheng2016 unpublished
4) Kutanan, W. et al. Complete mitochondrial genomes of Thai and Lao populations indicate an ancient origin of Austroasiatic groups and demic diffusion in the spread of Tai-Kadai languages. Hum Genet 136, 85-98, doi:10.1007/s00439-016-1742-y (2017).
5) Guo, H. et al. Presence of mutation m.14484T>C in a Chinese family with maternally inherited essential hypertension but no expression of LHON. Biochim Biophys Acta 1822, 1535-1543, doi: 10.1016/j.bbadis.2012.06.010 (2012).
6) Liu, J. et al. Deciphering the signature of selective constraints on cancerous mitochondrial genome. Mol Biol Evol 29, 1255-1261, doi:10.1093/molbev/msr290 (2012).
7) Wang, H. W. et al. Revisiting the role of the Himalayas in peopling Nepal: insights from mitochondrial genomes. J Hum Genet 57, 228-234, doi:10.1038/jhg.2012.8 (2012).

- **Figure 4.17** (hg F2b1)

1) Present study
2) Zheng, H. The mitochondrial DNA diversity of World populations, Unpublished.
3) Kutanan, W. et al. Complete mitochondrial genomes of Thai and Lao populations indicate an ancient origin of Austroasiatic groups and demic diffusion in the spread of Tai-Kadai languages. Hum Genet 136, 85-98, doi:10.1007/s00439-016-1742-y (2017).
4) Liu, J. et al. Deciphering the signature of selective constraints on cancerous mitochondrial genome. Mol Biol Evol 29, 1255-1261, doi:10.1093/molbev/msr290 (2012).

- **Figure 4.18** (hg HV12b1)

1) Present study
2) Palanichamy, M. G. et al. Phylogeny of mitochondrial DNA macrohaplogroup N in India, based on complete sequencing: implications for the peopling of South Asia. Am J Hum Genet 75, 966-978, doi:10.1086/425871 (2004).
3) van der Walt, E. M. et al. Characterization of mtDNA variation in a cohort of South African pediatric patients with mitochondrial disease. Eur J Hum Genet 20, 650-656, doi:10.1038/ejhg.2011.262 (2012).
4) Derenko, M. et al. Complete mitochondrial DNA diversity in Iranians. PLoS One 8, e80673, doi:10.1371/journal.pone.0080673 (2013).
5) Li, Y. C. et al. Ancient inland human dispersals from Myanmar into interior East Asia since the Late Pleistocene. Sci Rep 5, 9473, doi:10.1038/srep09473 (2015).
6) Schonberg, A., Theunert, C., Li, M., Stoneking, M. & Nasidze, I. High-throughput sequencing of complete human mtDNA genomes from the Caucasus and West Asia: high diversity and demographic inferences. Eur J Hum Genet 19, 988-994, doi:10.1038/ejhg.2011.62 (2011)

- **Figure 4.19** & **Figure 4.20** (hg HV12b1 & H13)

1) Present study
2) Greenspan, B. Unpublished.
3) Roostalu, U. et al. Origin and expansion of haplogroup H, the dominant human mitochondrial DNA lineage in West Eurasia: the Near Eastern and Caucasian perspective. Mol Biol Evol 24, 436-448, doi:10.1093/molbev/msl173 (2007).

4) Raule, N. et al. The co-occurrence of mtDNA mutations on different oxidative phosphorylation subunits, not detected by haplogroup analysis, affects human longevity and is population specific. Aging Cell 13, 401-407, doi:10.1111/acel.12186 (2014).
5) Derenko, M. et al. Complete mitochondrial DNA diversity in Iranians. PLoS One 8, e80673, doi:10.1371/journal.pone.0080673 (2013).
6) Fregel, R. et al. Multiple ethnic origins of mitochondrial DNA lineages for the population of Mauritius. PLoS One 9, e93294, doi:10.1371/journal.pone.0093294 (2014).
7) Margaryan, A. et al. Eight Millennia of Matrilineal Genetic Continuity in the South Caucasus. Curr Biol 27, 2023-2028 e2027, doi:10.1016/j.cub.2017.05.087 (2017).
8) Palanichamy, M. G. et al. Phylogeny of mitochondrial DNA macrohaplogroup N in India, based on complete sequencing: implications for the peopling of South Asia. Am J Hum Genet 75, 966-978, doi:10.1086/425871 (2004).
9) Behar, D. M. et al. Counting the founders: the matrilineal genetic ancestry of the Jewish Diaspora. PLoS One 3, e2062, doi:10.1371/journal.pone.0002062 (2008).
10) Khan, N. A. Unpublished. Direct submission to GenBank
11) Wang, H. W. et al. Revisiting the role of the Himalayas in peopling Nepal: insights from mitochondrial genomes. J Hum Genet 57, 228-234, doi:10.1038/jhg.2012.8 (2012).

- **Figure 4.22 and Figure 4.23 (Haplogroup U2)**

1) Present Study
2) Bhat, A. Direct submission on GenBank
3) Palanichamy, M. Direct submission on GenBank
4) Govindaraj, P. et al. Mitochondrial dysfunction and genetic heterogeneity in chronic periodontitis. Mitochondrion 11, 504-512, doi:10.1016/j.mito.2011.01.009 (2011).
5) Khan, N. A. Unpublished. Direct submission to GenBank
6) Kang, L. et al. MtDNA analysis reveals enriched pathogenic mutations in Tibetan highlanders. Sci Rep 6, 31083, doi:10.1038/srep31083 (2016).
7) Silva, M. et al. A genetic chronology for the Indian Subcontinent points to heavily sex-biased dispersals. BMC Evol Biol 17, 88, doi:10.1186/s12862-017-0936-9 (2017).
8) Direct submission by FTDNA (Family Tree DNA)
9) Kutanan, W. et al. Complete mitochondrial genomes of Thai and Lao populations indicate an ancient origin of Austroasiatic groups and demic diffusion in the spread of Tai-Kadai languages. Hum Genet 136, 85-98, doi:10.1007/s00439-016-1742-y (2017).
10) Sharma, I. Direct submission
11) Illeperuma, R. J. and Bamshad, M. J. unpublished
12) Zheng,H.-X. Direct submission to GenBank
13) Lippold, S. et al. Human paternal and maternal demographic histories: insights from high-resolution Y chromosome and mtDNA sequences. Investig Genet 5, 13, doi:10.1186/2041-2223-5-13 (2014).
14) 14' van der Walt, E. M. et al. Characterization of mtDNA variation in a cohort of South African pediatric patients with mitochondrial disease. Eur J Hum Genet 20, 650-656, doi:10.1038/ejhg.2011.262 (2012).
15) Greenspan, B. Direct submission to GenBank
16) Achilli, A. et al. Saami and Berbers--an unexpected mitochondrial DNA link. Am J Hum Genet 76, 883-886, doi:10.1086/430073 (2005).

- **Figure 4.24 and Figure 4.25 (Haplogroup U7**

1) Present Study
2) Palanichamy, M. G. et al. West Eurasian mtDNA lineages in India: an insight into the spread of the Dravidian language and the origins of the caste system. Hum Genet 134, 637-647,
3) Sahakyan, H. et al. Origin and spread of human mitochondrial DNA haplogroup U7. Sci Rep 7, 46044, doi:10.1038/srep46044 (2017).

4) Sharma, V. unpublished
5) Derenko, M. et al. Complete mitochondrial DNA diversity in Iranians. PLoS One 8, e80673, doi:10.1371/journal.pone.0080673 (2013).
6) Kang, L. et al. MtDNA analysis reveals enriched pathogenic mutations in Tibetan highlanders. Sci Rep 6, 31083, doi:10.1038/srep31083 (2016).
7) Behar, D. M. et al. A "Copernican" reassessment of the human mitochondrial DNA tree from its root. Am J Hum Genet 90, 675-684, doi:10.1016/j.ajhg.2012.03.002 (2012).
8) Palanichamy, M. G. et al. Phylogeny of mitochondrial DNA macrohaplogroup N in India, based on complete sequencing: implications for the peopling of South Asia. Am J Hum Genet 75, 966-978, doi:10.1086/425871 (2004).

- **Figure 4.26 (Hg N1a1b)**

1. Present Study
2. A Direct Submission to GenBank by Khan et al., 2014. unpublished
3. A Direct Submission to GenBank by; Family Tree DNA - Genealogy by Genetics,
4. A Direct Submission to GenBank by; Family Tree DNA - Genealogy by Genetics,
5. Kutanan, W. et al. Complete mitochondrial genomes of Thai and Lao populations indicate an ancient origin of Austroasiatic groups and demic diffusion in the spread of Tai-Kadai languages. Hum Genet 136, 85-98, doi:10.1007/s00439-016-1742-y (2017).
6. Derenko, M. et al. Phylogeographic analysis of mitochondrial DNA in northern Asian populations. Am J Hum Genet 81, 1025-1041, doi:10.1086/522933 (2007).
7. Palanichamy, M. G. et al. West Eurasian mtDNA lineages in India: an insight into the spread of the Dravidian language and the origins of the caste system. Hum Genet 134, 637-647, doi:10.1007/s00439-015-1547-4 (2015).
8. Fernandes, V. et al. The Arabian cradle: mitochondrial relicts of the first steps along the southern route out of Africa. Am J Hum Genet 90, 347-355, doi:10.1016/j.ajhg.2011.12.010 (2012).
9. Fregel, R., Cabrera, V., Larruga, J. M., Abu-Amero, K. K. & Gonzalez, A. M. Carriers of Mitochondrial DNA Macrohaplogroup N Lineages Reached Australia around 50,000 Years Ago following a Northern Asian Route. PLoS One 10, e0129839, doi:10.1371/journal.pone.0129839 (2015).
10. Derenko, M. et al. Western Eurasian ancestry in modern Siberians based on mitogenomic data. BMC Evol Biol 14, 217, doi:10.1186/s12862-014-0217-9 (2014).

- **Figure 4.27 (Hg W3a1b)**

1. Present study
2. Palanichamy, M. G. et al. Phylogeny of mitochondrial DNA macrohaplogroup N in India, based on complete sequencing: implications for the peopling of South Asia. Am J Hum Genet 75, 966-978, doi:10.1086/425871 (2004).
3. Greenspan, B. Direct submission
4. Govindaraj, P. et al. Mitochondrial dysfunction and genetic heterogeneity in chronic periodontitis. Mitochondrion 11, 504-512, doi:10.1016/j.mito.2011.01.009 (2011).
5. Olivieri, A. et al. Mitogenomes from two uncommon haplogroups mark late glacial/postglacial expansions from the near east and neolithic dispersals within Europe. PLoS One 8, e70492, doi:10.1371/journal.pone.0070492 (2013).
6. Kutanan, W. et al. Complete mitochondrial genomes of Thai and Lao populations indicate an ancient origin of Austroasiatic groups and demic diffusion in the spread of Tai-Kadai languages. Hum Genet 136, 85-98, doi:10.1007/s00439-016-1742-y (2017).
7. Lippold, S. et al. Human paternal and maternal demographic histories: insights from high-resolution Y chromosome and mtDNA sequences. Investig Genet 5, 13, doi:10.1186/2041-2223-5-13 (2014).
8. Zheng,H.-X. The mitochondrial DNA diversity of HGDP populations, Unpublished

- **Figure 4.28 (Hg A)**

1. Present study

- **Figure 4.31 and Figure 4.32 (Hg M3)**

1. Present study
2. Palanichamy, M., Northeast Indo-China corridor is the cradle for Asian origin after the Africa exodus. Unpublished
3. Sharma, I. Complete Mitochondrial sequences of individuals belonging to maternal haplogroup M from Jammu and Kashmir, India. Unpublished
4. Chandrasekar, A. et al. Updating phylogeny of mitochondrial DNA macrohaplogroup m in India: dispersal of modern human in South Asian corridor. PLoS One 4, e7447,
5. Pradutkanchana, S. Unpublished
6. Zheng,H.-X. Unpublished
7. Fornarino, S. et al. Mitochondrial and Y-chromosome diversity of the Tharus (Nepal): a reservoir of genetic variation. BMC Evol Biol 9, 154, doi:10.1186/1471-2148-9-154 (2009).
8. Khan, N. A. Occurrence of G11778A LHON mutation in diverse mitochondrial haplogroups in Indian population
9. Wang, H. W. et al. Revisiting the role of the Himalayas in peopling Nepal: insights from mitochondrial genomes. J Hum Genet 57, 228-234, doi:10.1038/jhg.2012.8 (2012).
10. Zhang, X. et al. Analysis of mitochondrial genome diversity identifies new and ancient maternal lineages in Cambodian aborigines. Nat Commun 4, 2599, doi:10.1038/ncomms3599 (2013).
11. Rajkumar, R., Banerjee, J., Gunturi, H. B., Trivedi, R. & Kashyap, V. K. Phylogeny and antiquity of M macrohaplogroup inferred from complete mt DNA sequence of Indian specific lineages. BMC Evol Biol 5, 26, doi:10.1186/1471-2148-5-26 (2005).
12. Kang, L. et al. MtDNA analysis reveals enriched pathogenic mutations in Tibetan highlanders. Sci Rep 6, 31083, doi:10.1038/srep31083 (2016).

- **Figure 4.34** (Hg M38 & M43)

1. Present study
2. Li, Y. C. et al. Ancient inland human dispersals from Myanmar into interior East Asia since the Late Pleistocene. Sci Rep 5, 9473, doi:10.1038/srep09473 (2015).
3. Sun, C. et al. The dazzling array of basal branches in the mtDNA macrohaplogroup M from India as inferred from complete genomes. Mol Biol Evol 23, 683-690, doi:10.1093/molbev/msj078 (2006).
4. Fornarino, S. et al. Mitochondrial and Y-chromosome diversity of the Tharus (Nepal): a reservoir of genetic variation. BMC Evol Biol 9, 154, doi:10.1186/1471-2148-9-154 (2009).
5. Wang, H. W. et al. Revisiting the role of the Himalayas in peopling Nepal: insights from mitochondrial genomes. J Hum Genet 57, 228-234, doi:10.1038/jhg.2012.8 (2012).
6. Palanichamy, M. Northeast Indo-China corridor is the cradle for Asian origin after the Africa exodus. Unpublished.
7. Chandrasekar, A. et al. Updating phylogeny of mitochondrial DNA macrohaplogroup m in India: dispersal of modern human in South Asian corridor. PLoS One 4, e7447,

- **Figure 4.35** (Hg M30)

1. This study
2. Chandrasekar, A. et al. Updating phylogeny of mitochondrial DNA macrohaplogroup m in India: dispersal of modern human in South Asian corridor. PLoS One 4, e7447,
3. van der Walt, E. M. et al. Characterization of mtDNA variation in a cohort of South African pediatric patients with mitochondrial disease. Eur J Hum Genet 20, 650-656, doi:10.1038/ejhg.2011.262 (2012).
4. Sun, C. et al. The dazzling array of basal branches in the mtDNA macrohaplogroup M from India as inferred from complete genomes. Mol Biol Evol 23, 683-690, doi:10.1093/molbev/msj078 (2006).

5. Lippold, S. et al. Human paternal and maternal demographic histories: insights from high-resolution Y chromosome and mtDNA sequences. Investig Genet 5, 13, doi:10.1186/2041-2223-5-13 (2014).
6. Zheng, H Direct submission to GenBank, Unpublished
7. Palanichamy, M. Direct submission to GenBank
8. Ingman, M. & Gyllensten, U. Mitochondrial genome variation and evolutionary history of Australian and New Guinean aborigines. Genome Res 13, 1600-1606, doi:10.1101/gr.686603 (2003).
9. Greenspan, B. Direct Submission
10. Behar, D. M. et al. A "Copernican" reassessment of the human mitochondrial DNA tree from its root. Am J Hum Genet 90, 675-684, doi:10.1016/j.ajhg.2012.03.002 (2012).

- **Figure 4.36 (Hg M5)**

1. Present study
2. Kutanan, W. et al. Complete mitochondrial genomes of Thai and Lao populations indicate an ancient origin of Austroasiatic groups and demic diffusion in the spread of Tai-Kadai languages. Hum Genet 136, 85-98, doi:10.1007/s00439-016-1742-y (2017).
3. Kumar, S. et al. Reconstructing Indian-Australian phylogenetic link. BMC Evol Biol 9, 173, doi:10.1186/1471-2148-9-173 (2009).
4. Kang, L. et al. mtDNA lineage expansions in Sherpa population suggest adaptive evolution in Tibetan highlands. Mol Biol Evol 30, 2579-2587, doi:10.1093/molbev/mst147 (2013).
5. Fornarino, S. et al. Mitochondrial and Y-chromosome diversity of the Tharus (Nepal): a reservoir of genetic variation. BMC Evol Biol 9, 154, doi:10.1186/1471-2148-9-154 (2009).
6. Thangaraj, K. et al. In situ origin of deep rooting lineages of mitochondrial Macrohaplogroup 'M' in India. BMC Genomics 7, 151, doi:10.1186/1471-2164-7-151 (2006).
7. Behar, D. M. et al. Counting the founders: the matrilineal genetic ancestry of the Jewish Diaspora. PLoS One 3, e2062, doi:10.1371/journal.pone.0002062 (2008).
8. Unpublished, direct submission by Palanichamy to GenBank
9. Chandrasekar, A. et al. Updating phylogeny of mitochondrial DNA macrohaplogroup m in India: dispersal of modern human in South Asian corridor. PLoS One 4, e7447, doi:10.1371/journal.pone.0007447 (2009).

- **Figure 4.41** (Hg G)

1. Present Study
2. Greenspan, B., Unpublished
3. Kong, Q. P. et al. Updating the East Asian mtDNA phylogeny: a prerequisite for the identification of pathogenic mutations. Hum Mol Genet 15, 2076-2086, doi:10.1093/hmg/ddl130 (2006).
4. Zheng, H.-X. Direct submission to GenBank.
5. Lippold, S. et al. Human paternal and maternal demographic histories: insights from high-resolution Y chromosome and mtDNA sequences. Investig Genet 5, 13, doi:10.1186/2041-2223-5-13 (2014).
6. Zhao, M. et al. Mitochondrial genome evidence reveals successful Late Paleolithic settlement on the Tibetan Plateau. Proc Natl Acad Sci U S A 106, 21230-21235, doi:10.1073/pnas.0907844106 (2009).
7. Kang, L. et al. mtDNA lineage expansions in Sherpa population suggest adaptive evolution in Tibetan highlands. Mol Biol Evol 30, 2579-2587, doi:10.1093/molbev/mst147 (2013).
8. Kumar, S. et al. Reconstructing Indian-Australian phylogenetic link. BMC Evol Biol 9, 173, doi:10.1186/1471-2148-9-173 (2009).
9. Hartmann, A., Unpublished
10. Pankratov, V. et al. East Eurasian ancestry in the middle of Europe: genetic footprints of Steppe nomads in the genomes of Belarusian Lipka Tatars. Sci Rep 6, 30197, doi:10.1038/srep30197 (2016).
11. Kutanan, W. et al. Complete mitochondrial genomes of Thai and Lao populations indicate an ancient origin of Austroasiatic groups and demic diffusion in the spread of Tai-Kadai languages. Hum Genet 136, 85-98, doi:10.1007/s00439-016-1742-y (2017).

12. Kong, Q. P. et al. Large-scale mtDNA screening reveals a surprising matrilineal complexity in east Asia and its implications to the peopling of the region. Mol Biol Evol 28, 513-522, doi:10.1093/molbev/msq219 (2011).

- **Figure 4.42 and Figure 4.43 (Hg M31, M33, M34 & M35)**

1. Present study
2. Fornarino, S. et al. Mitochondrial and Y-chromosome diversity of the Tharus (Nepal): a reservoir of genetic variation. BMC Evol Biol 9, 154, doi:10.1186/1471-2148-9-154 (2009).
3. Chandrasekar, A. et al. Updating phylogeny of mitochondrial DNA macrohaplogroup m in India: dispersal of modern human in South Asian corridor. PLoS One 4, e7447,
4. Wang, H. W. et al. Revisiting the role of the Himalayas in peopling Nepal: insights from mitochondrial genomes. J Hum Genet 57, 228-234, doi:10.1038/jhg.2012.8 (2012).
5. Palanichamy, M. Northeast Indo-China corridor is the cradle for Asian origin after the Africa exodus. UNPUBLISHED
6. Al-Zahery, N. et al. In search of the genetic footprints of Sumerians: a survey of Y-chromosome and mtDNA variation in the Marsh Arabs of Iraq. BMC Evol Biol 11, 288, doi:10.1186/1471-2148-11-288 (2011).
7. Sun, C. et al. The dazzling array of basal branches in the mtDNA macrohaplogroup M from India as inferred from complete genomes. Mol Biol Evol 23, 683-690, doi:10.1093/molbev/msj078 (2006).
8. Li, Y. C. et al. Ancient inland human dispersals from Myanmar into interior East Asia since the Late Pleistocene. Sci Rep 5, 9473, doi:10.1038/srep09473 (2015).
9. Summerer, M. et al. Large-scale mitochondrial DNA analysis in Southeast Asia reveals evolutionary effects of cultural isolation in the multi-ethnic population of Myanmar. BMC Evol Biol 14, 17, doi:10.1186/1471-2148-14-17 (2014).
10. Malyarchuk, B. A. et al. Mitochondrial DNA variability in Slovaks, with application to the Roma origin. Ann Hum Genet 72, 228-240, doi:10.1111/j.1469-1809.2007.00410.x (2008).
11. Fendt, L. et al. Accumulation of mutations over the entire mitochondrial genome of breast cancer cells obtained by tissue microdissection. Breast Cancer Res Treat 128, 327-336, doi:10.1007/s10549-010-1092-8 (2011).

- **Figure 4.44 (Hg M52)**

1. This study
2. Fornarino, S. et al. Mitochondrial and Y-chromosome diversity of the Tharus (Nepal): a reservoir of genetic variation. BMC Evol Biol 9, 154, doi:10.1186/1471-2148-9-154 (2009).
3. Vyas, D. N. et al. Bayesian analyses of Yemeni mitochondrial genomes suggest multiple migration events with Africa and Western Eurasia. Am J Phys Anthropol 159, 382-393, doi:10.1002/ajpa.22890 (2016).
4. Palanichamy, M. unpublished
5. Sun, C. et al. The dazzling array of basal branches in the mtDNA macrohaplogroup M from India as inferred from complete genomes. Mol Biol Evol 23, 683-690, doi:10.1093/molbev/msj078 (2006).
6. Wang, H. W. et al. Revisiting the role of the Himalayas in peopling Nepal: insights from mitochondrial genomes. J Hum Genet 57, 228-234, doi:10.1038/jhg.2012.8 (2012).

- **Figure 4.45 (Hg D)**

1. This study

8.6 Appendix F

Table 8.5 | List of Newa sub caste included in Newa mix group. Total number of sample included in Newa mix group is 50.

code	NEWAR CASTE	LOCATION						Total
		Bhaktapur	Nala	Banepa	Sankhu	KTM	Kirtipur	
Mhj	Maharjan	-	-	-	-	-	100	100
STH	Shrestha	6	15	3	6	6	-	36
Newa mix	Malla	1	-	-	5	-	-	6
	Ranjitkar	1	-	-	5	-	-	6
	Kharbuja	3	-	-	-	-	-	3
	shahi	-	-	-	3	-	-	3
	Dangol	-	-	-	3	-	-	3
	Vaidhya	-	-	1	2	-	-	3
	Pradhan	1	-	-	2	-	-	3
	Mununkarmi	2	-	-	2	-	-	4
	Karmacharya	-	3	-	1	-	-	4
	Kapali	-	-	-	2	-	-	2
	Mali	-	-	-	2	-	-	2
	Napit	1	1	-	-	-	-	2
	Hada	1	-	-	1	-	-	2
	Prajapati	-	1	-	-	-	-	1
	Balami	-	-	-	-	1 (Jorpati)	-	1
	Phoju	1	-	-	-	-	-	1
	Kayastha	1	-	-	-	-	-	1
	Rajlawat	1	-	-	-	-	-	1
	Madhikarmi	1	-	-	-	-	-	1
	Lakhe	1	-	-	-	-	-	1
	pode	-	-	-	1	-	-	1
	singh	-	-	-	1	-	-	1
	basi	-	-	-	1	-	-	1
	phutwal	-	-	-	1	-	-	1
	Gorkhali	-	-	-	1	-	-	1

8.7 Appendix G

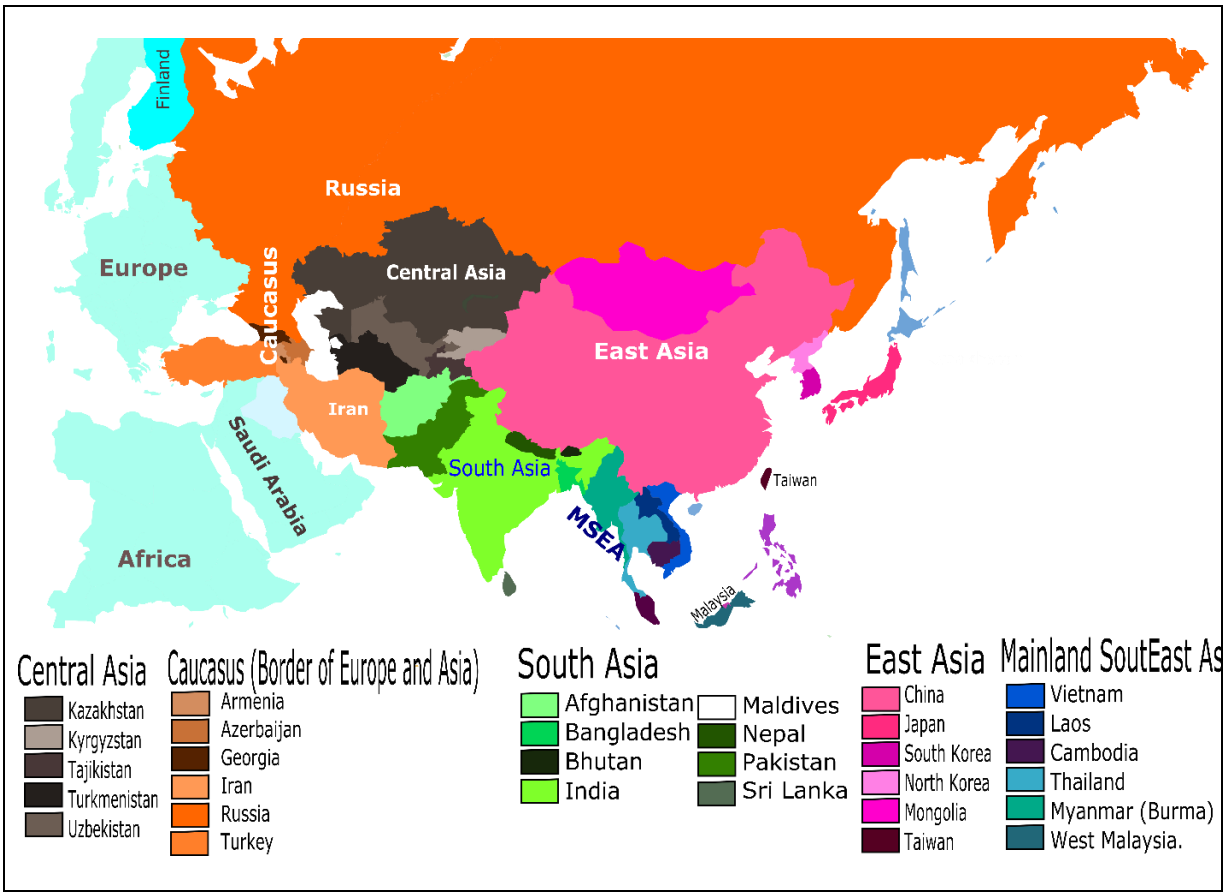


Figure 8.1| Map showing the geographic territory and respective countries. Geographic region including the respective countries of South Asia, Central Asia, Caucasus, East Asia and Mainland South East Asia are shown in the map. Map created using Corel Draw Suite X8 version 18. All colored boundaries in the Map are approximate.

8.8 Appendix H

Table 8.6| Genotyped data for Changpa, Ladhak, and other lowlander populations from Nepal and India. The p value demonstrates significance of correlation analysis

	rs150877473 (C>G)		rs13419896 (G>A)		rs1868092 (G>A)		rs1868093 (G>A)	
Alt	C	G	G	A	G	A	G	A
1300	0.91954	0.08046	0.664179	0.335821	0.69697	0.30303	0.69697	0.30303
1400	0.925926	0.074074	0.741279	0.258721	0.731579	0.268421	0.731579	0.268421
2000	1	0	0.833333	0.166667	0.559524	0.440476	0.559524	0.440476
3000	0.713415	0.286585	0.602273	0.397727	0.559211	0.44079	0.559211	0.44079
4000	0.227273	0.772727	0.197674	0.802326	0.266667	0.733333	0.266667	0.733333
550	0.973	0.027	0.731	0.269	0.892	0.108	0.892473	0.107527
44	0.99	0.01	0.694	0.306	0.917	0.083	0.917476	0.082524
19	1	0	0.644	0.356	0.909	0.091	0.908654	0.091346
19	0.99	0.01	0.753	0.247	0.924	0.076	0.924242	0.075758
Correlation P-value	0.003668		0.04493		5.95E-06		5.99E-06	

8.9 Appendix I

Table 8.7 | EDAR profile of the Nepali and other populations.

Population	State/City	Location	n	Language	1540C (%)	Reference
Shrestha & newa mix	Kathmandu	Nepal	86	Tibeto-Burman	27	Present Study
Maharjan	Kathmandu	Nepal	100	Tibeto-Burman	28	Present Study
Shakya	Kathmandu	Nepal	19	Tibeto-Burman	33	Lab data
Bajracharya	Kathmandu	Nepal	20	Tibeto-Burman	28	Lab data
Udaya	Kathmandu	Nepal	58	Tibeto-Burman	35	Lab data
Magar	Tanahun	Nepal	36	Tibeto-Burman	71	Present Study
Brahmin	Tanahun	Nepal	48	Indo-European	21	Present Study
Biaka Pygmies	Central African	Africa	32	Bantu	0	Xue et al. 2009
Mbuti Pygmies	Democratic Republic of Congo	Africa	13	Bantu	0	Xue et al. 2009
Bantu N.E.	Kenya	Africa	10	Bantu	0	Xue et al. 2009
San	Namibia	Africa	4	Khosian	0	Xue et al. 2009
Yoruba	Nigeria	Africa	17	Niger-Congo	0	Xue et al. 2009
South Africa		Africa	8	Africans	0	Xue et al. 2009
Mozabite	Algeria (Mzab)	Algeria	25	Afro-Asiatic	0	Xue et al. 2009
Maya	Mexico	America-N	24	Maya	88	Xue et al. 2009
Pima	Mexico	America-N	24	Pima	100	Xue et al. 2009
Karitiana	Brazil	Brazil	24	Tupi	100	Xue et al. 2009
Surui	Brazil	Brazil	21	Akwáwa	74	Xue et al. 2009
Northeast	China	East Asia	40	Sinitic	93	Xue et al. 2009
Northwest	China	East Asia	19	Sinitic	66	Xue et al. 2009
Han	China	East Asia	43	Sinitic	92	Xue et al. 2009
Central	China	East Asia	30	Sinitic	80	Xue et al. 2009
Southwest	China	East Asia	50	Sinitic	83	Xue et al. 2009
Japanese	Japan	East Asia	31	Japonic	74	Xue et al. 2009
France	France	Europe	52	Indo-European	0	Xue et al. 2009
Italy	Italy	Europe	46	Indo-European	0	Xue et al. 2009
Orcadian	Orkney Islands	Europe	13	Indo-European	0	Xue et al. 2009
Nyshi	Arunanchal-Pradesh	India	20	Tibeto-Burman	65	Chaubey et al. 2011
Garo	Meghalaya	India	20	Tibeto-Burman	53	Chaubey et al. 2011
Khasi	Meghalaya	India	20	Khasi-Aslian/Austroasiatic	40	Chaubey et al. 2011
Naga	Nagaland	India	17	Tibeto-Burman	65	Chaubey et al. 2011
Kanjar	Uttar Pradesh	India	10	Indo-European	0	Chaubey et al. 2011
Harijan	Uttar Pradesh	India	10	Indo-European	0	Chaubey et al. 2011
Dharkar	Uttar Pradesh	India	10	Indo-European	0	Chaubey et al. 2011
Dusadh	Uttar Pradesh	India	10	Indo-European	0	Chaubey et al. 2011
Kol	Uttar Pradesh	India	10	Indo-European	0	Chaubey et al. 2011
Low Caste	Uttar Pradesh	India	10	Indo-European	0	Chaubey et al. 2011
Brahmin	Uttar Pradesh	India	22	Indo-European	0	Chaubey et al. 2011
Middle Caste	Uttar Pradesh	India	10	Indo-European	0	Chaubey et al. 2011
Musahar	Uttar Pradesh	India	12	Indo-European	4	Chaubey et al. 2011
Tharu	Uttar-Pradesh Mixed	India	24	Indo-European	8	Chaubey et al. 2011
Tharu	Uttarakhand (Haldwani)	India	22	Indo-European	10	Chaubey et al., 2014
Tharu	Uttarakhand (Nainital)	India	23	Indo-European	17	Chaubey et al., 2014
Tharu	Uttar Pradesh (Gorakhpur)	India	38	Indo-European	13	Chaubey et al., 2014
Tharu	Uttar Pradesh (Allahabad)	India	37	Indo-European	3	Chaubey et al., 2014
Tharu	Uttar Pradesh (Deoria)	India	47	Indo-European	10	Chaubey et al., 2014
Tharu	Uttar Pradesh (Mahrajganj)	India	42	Indo-European	12	Chaubey et al., 2014
Chenchu	Andhra-Pradesh	India	20	Dravidian	0	Chaubey et al., 2014
Velamas	Andhra-Pradesh	India	25	Dravidian	0	Chaubey et al., 2014
Brahmin	Andhra-Pradesh	India	19	Dravidian	0	Chaubey et al., 2014
Oraon	Bihar	India	20	Dravidian	0	Chaubey et al., 2014
Kharia	Chattishgarh	India	20	Munda/Austroasiatic	15	Chaubey et al., 2014
Bhunjiya	Chattishgarh	India	20	Indo-European	0	Chaubey et al., 2014
Birhor	Chattishgarh	India	35	Munda/Austroasiatic	3	Chaubey et al., 2014
Kanwar	Chattishgarh	India	11	Indo-European	0	Chaubey et al., 2014
Kolcha	Gujarat	India	15	Indo-European	0	Chaubey et al., 2014
Ho	Jharkhand	India	32	Munda/Austroasiatic	9	Chaubey et al., 2014

Population	State/City	Location	n	Language	1540C (%)	Reference
Mawasi	Jharkhand	India	29	Munda/Austroasiatic	8	Chaubey et al., 2014
Asur	Jharkhand	India	35	Munda/Austroasiatic	6	Chaubey et al., 2014
Santhal	Jharkhand	India	19	Munda/Austroasiatic	2	Chaubey et al., 2014
Halakipikki	Karnatka	India	20	Dravidian	0	Chaubey et al., 2014
Brahmin	Kashmir	India	20	Indo-European	0	Chaubey et al., 2014
Sakkili	Kerla	India	15	Dravidian	0	Chaubey et al., 2014
Paniya	Kerla	India	20	Dravidian	0	Chaubey et al., 2014
Gond	Madhya-Pradesh	India	20	Dravidian	0	Chaubey et al., 2014
Mawasi	Madhya-Pradesh	India	10	Munda/Austroasiatic	0	Chaubey et al., 2014
Baiga	Madhya-Pradesh	India	19	Munda/Austroasiatic	0	Chaubey et al., 2014
Brahmin	Madhya-Pradesh	India	10	Indo-European	0	Chaubey et al., 2014
Birhor	Maharashtra	India	15	Munda/Austroasiatic	3	Chaubey et al., 2014
Tribes	Maharashtra	India	10	Indo-European	0	Chaubey et al., 2014
Tharu	Nepal	India	21	Indo-European	2	Chaubey et al., 2014
Brahmin	Nepal	India	8	Indo-European	0	Chaubey et al., 2014
Bonda	Orissa	India	38	Munda/Austroasiatic	7	Chaubey et al., 2014
Gadaba	Orissa	India	28	Munda/Austroasiatic	0	Chaubey et al., 2014
Savara	Orissa	India	38	Munda/Austroasiatic	7	Chaubey et al., 2014
Juang	Orissa	India	20	Munda/Austroasiatic	3	Chaubey et al., 2014
Oraon	Orissa	India	20	Dravidian	0	Chaubey et al., 2014
Dhurva	Orissa	India	10	Indo-European	0	Chaubey et al., 2014
Baiga	Orissa	India	21	Munda/Austroasiatic	5	Chaubey et al., 2014
Meena	Rajasthan	India	10	Indo-European	0	Chaubey et al., 2014
Toda	Tamilnadu	India	14	Dravidian	0	Chaubey et al., 2014
Kurumba	Tamilnadu	India	16	Dravidian	0	Chaubey et al., 2014
Pulliyar	Tamilnadu	India	16	Dravidian	0	Chaubey et al., 2014
Tharu	Uttar-Pradesh	India	24	Indo-European	8	Chaubey et al., 2014
Chamar	Uttar-Pradesh	India	10	Indo-European	0	Chaubey et al., 2014
Bengali	West-Bengal	India	30	Indo-European	0	Chaubey et al., 2014
Brahmin	West-Bengal	India	10	Indo-European	0	Chaubey et al., 2014
Dusadh	Bihar	India	5	Indo-European	0	Chaubey et al., 2014
Mahali	Jharkhand	India	20	Munda/Austroasiatic	0	Chaubey et al., 2014
Malayan	Kerla	India	25	Dravidian	0	Chaubey et al., 2014
Oorali	Kerla	India	17	Dravidian	0	Chaubey et al., 2014
Chakilliyan	Kerla	India	16	Dravidian	0	Chaubey et al., 2014
Koli	Maharashtra	India	18	Indo-European	0	Chaubey et al., 2014
Meghwal	Rajasthan	India	12	Indo-European	0	Chaubey et al., 2014
Druze	Israel (Carmel)	Middle East	43	Arabic	1	Xue et al. 2009
Palestinian	Israel (Central)	Middle East	46	Arabic	0	Xue et al. 2009
Bedouin	Israel (Negev)	Middle East	36	Arabic	0	Xue et al. 2009
Brahmin	Tanahun	Nepal	8	Indo-European	0	Chaubey et al. 2011
Tharu	Nepal Mixed	Nepal	21	Indo-European	2	Chaubey et al. 2011
Tharu	Nepal-Butwal	Nepal	10	Indo-European	10	Present Study
Tharu	Nepal-Tulsipur	Nepal	21	Indo-European	19	Present Study
Papuan	New Guinea	Oceania	15	Papuan	0	Xue et al. 2009
Brahui	Pakistan	Pakistan	23	Dravidian	0	Xue et al. 2009
Balochi	Pakistan	Pakistan	25	Indo-European	0	Xue et al. 2009
Hazara	Pakistan	Pakistan	24	Indo-European	38	Xue et al. 2009
Makrani	Pakistan	Pakistan	22	Indo-European	0	Xue et al. 2009
Sindhi	Pakistan	Pakistan	22	Indo-European	0	Xue et al. 2009
Pathan	Pakistan	Pakistan	21	Indo-European	0	Xue et al. 2009
Kalash	Pakistan	Pakistan	20	Indo-European	0	Xue et al. 2009
Burusho	Pakistan	Pakistan	24	Language Isolate	0	Xue et al. 2009
NAN Melanesian	Bougainville	Papua New Guinea	21	Austronesian	10	Xue et al. 2009
Russian	Russia	Russia	25	Indo-European	2	Xue et al. 2009
Adygei	Russia Caucasus	Russia	16	Caucasian	3	Xue et al. 2009
Yakut	Siberia	Russia	25	Turkik	84	Xue et al. 2009
Mandenka	Senegal	Senegal	24	Niger-Congo	0	Xue et al. 2009
Cambodian	Cambodia	South East Asia	10	Khasi-Aslian/Austroasiatic	55	Xue et al. 2009
Colombian	Colombia	South East Asia	13	Sign	81	Xue et al. 2009
Java, Indonesia	South East Asia	South East Asia	121	Austronesian	34	Fujimoto et al. 2008
Thailand	South East Asia	South East Asia	65	Tai-Kdai	30	Fujimoto et al. 2008

8.10 Appendix J

- **Genomic DNA from blood sample**

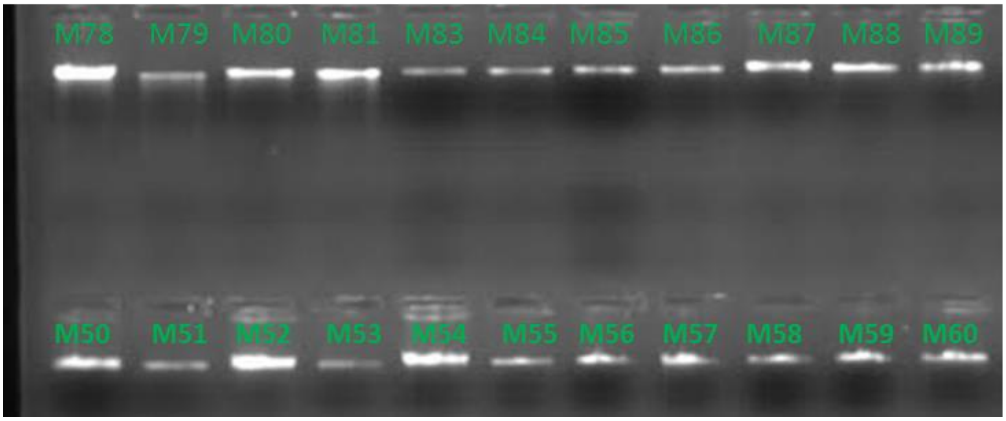


Figure 8.2: Genomic DNA from blood samples. Working concentration of DNA was made to 10ng/µl for all samples and was run on 1% Agarose gel for final confirmation. Genomic bands were visualized under Gel Documentation System.

- **PCR products**: PCR products were electrophoresed at 120V in 2% agarose gel. The PCR products were then visualized under Gel Doc. On obtaining a single band devoid of any primer-dimer bands the PCR products were proceed for purification using ExoSAP enzyme.

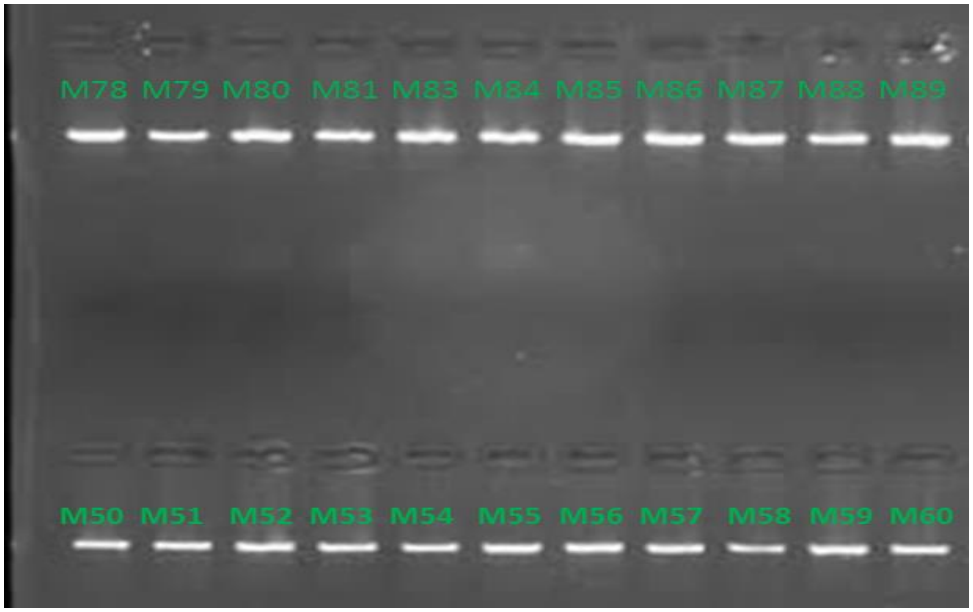


Figure 8.3: Agarose gel electrophoresis of the PCR products (≈650 bp) amplified with the mitochondrial DNA primers.

8.11 Appendix k

Complete mtDNA sequence of one representative sample belonging to Prof. Dr. Tilak R. Shrestha (STH007) in Fasta format is given below. In the sequence, HVR II (1-400) and HVR I (16000-16450) are underlined respectively.

>STH007

gatcacaggtctatcaccctattaaccactcacgggagctctccatgcatttggtattttcgtctgggggtgtgcacgcgatagcattgcgagacgctgga
gccggagcaccctatgtcgcagtatctgtctttgattcctgcctcatctcattatttatcgcacctacgttcaatattacaggcgaacatacttactaaagtgtgtt
aattaattaatgcttgtaggacataataataacaattgaatgtctgcacagccgctttccacacagacatcataacaaaaaatttccaccaaacccccccctcc
ccccgcttctggccacagcacttaaacacatctctgccaaaccccaaaaacaaagaaccctaacaccagcctaaccagatttcaaattttatctttggcggt
atgcacttttaacagtcaccccccaactaacacattattttccctcccactcccatactactaatctcatcaatacaacccccgcccatcctacccagcacac
acacaccgctgctaaccccataccccgaaccaaccaaaccccaaagacacccccccacagtttatgtagcttacctcctcaaagcaatacactgaaaatgtt
tagacgggctcacatcaccccataaacaaataggtttggtcctagcctttctattagctcttagtaagattacacatgcaagcatccccgttccagtgagttca
ccctctaaatcaccacgatcaaaagggacaagcatcaagcacgcagcaatgcagctcaaaacgcttagcctagccacacccccacgggaaacagcag
tgattaacctttagcaataaacgaaagtttaactaagctatactaaccccagggttggtcaatttcgtgccagccaccgcggtcacacgattaacccaagtca
atagaagccggcgtaaagagtgttttagatcacccctccccaataaagctaaaacccacctgagttgtaaaaaactccagttgacacaaaatagactacg
aaagtggctttaacatatctgaacacacaatagctaagacccaaactgggattagataccccactatgcttagccctaaacctcaacagttaaatcaacaaa
actgctcgccagaacactacgagccacagcttaaaactcaaaggacctggcggtgcttcatatccctctagaggagcctgttctgtaatcgataaaccccg
atcaacctcaccacctcttgctcagcctatataccgccatcttcagcaaaccctgatgaaggctacaaagtaagcgcaagtacccacgtaaagacgttagg
tcaaggtgtagcccatgaggtggcaagaaatgggctacattttctaccccagaaaactacgatagcccttatgaaacttaagggtcgaaggtggatttagc
agtaaactgagagtagagtgcttagttgaacagggccctgaagcgcgtacacaccgcccgtcacactcctcaagtatacttcaaaggacatttaactaaaa
ccctacgcatttatatagaggagacaagtcgtaacatggtaagtgtactggaaagtgcacttggacgaaccagagtgtagcttaacacaaagcacccaa
cttacacttaggagatttcaacttaacttgaccgctctgagctaaacctagccccaaacccactccaccttactaccagacaaccttagccaaaccatttacc
caaataaagtataggcgatagaaattgaaacctggcgcaatagatatagtaccgcaagggaaagatgaaaaattatagccaagcataatatagcaaggac
taacccctataccttctgcataatgaattaactagaaataactttgcaaggagagccaaagctaagacccccgaaaccagacgagctacctaagaacagct
aaaagagcacacccgtctatgtagcaaaatagtgggaagatttataggtagaggcgacaaacctaccgagcctggtgatagctggttgtccaagatagaa
tcttagttcaacttaaatttgcccacagaaccctctaaatccccttgtaaatttaactgttagtccaaagaggaacagctctttggacactaggaaaaaacctt
gtagagagagtaaaaaatttaacacccatagtaggcctaaaagcagccaccaattaagaaagcgttcaagctcaacacccactacctaaaaaatcccaaa
catataactgaactcctcacacccaattggaccaatctatcaccctatagaagaactaatgttagtataagtaacatgaaaacattctcctccgcataagcctg
cgtcagattaaaacactgaactgacaattaacagcccaatatctacaatcaaccaacaagtcattattaccctcactgtcaacccaacacaggcatgctcata
aggaaaggttaaaaaaagtaaaaggaactcggcaaatcttaccccgcctgtttaccaaaaacatcacctctagcatcaccagtattagaggcaccgcctg
cccagtgacacatgtttaacggccgcggtaccctaaccgtgcaaaggtagcataatcacttgttccttaaatagggacctgtatgaatggctccacgaggg
ttcagctgtctcttacttttaaccagtgaaattgacctgcccgtgaagaggcgggcatgacacagcaagacgagaagaccctatggagctttaatttattaat
gcaaacagtacctaacaaacccacaggtcctaaactaccaaacctgcattaaaaatttcggttggggcgacctcggagcagaacccaacctccgagcag
tacatgctaagacttcaccagtcaaagcgaactactatactcaattgatccaataacttgaccaacggaacaagttaccctagggataacagcgcaatccta
ttctagagtccatatcaacaatagggtttacgacctcgatgttggatcaggacatcccgatggtgcagccgctattaaaggttcgtttgttcaacgattaaagt
cctacgtgatctgagttcagaccggagtaatccaggtcggtttctatctacnttcaaattcctccctgtacgaaaggacaagagaaataaggcctacttcaca
aagcgccttcccccgtaaatgatatcatctcaacttagtattatacccacacccacccaagaacagggtttgttaagatggcagagcccggtaatcgcataa
aacttaaaactttacagtcagaggttcaattcctcttcttaacaacatacccatggccaacctcctactcctcattgtacccattctaatcgcaatggcattccta
atgcttaccgaacgaaaaattctaggctatatacaactacgcaaaggcccaacgttgtaggcccctacgggctactacaacccttcgctgacgccataaa
actcttcaccaaagagcccctaaaacccgccacatctaccatcaccctctacatcaccgccccgaccttagctctcaccatcgctcttctactatgaacccc
cctccccataccaaccccctggtcaacctcaacctaggcctcctatttattctagccacctctagcctagccgtttactcaatcctctgatcagggtgagcat
caaactcaaactacgccctgatcggcgcactgcgagcagtagcccaaacaatctcatatgaagtcactctagccatcattctactatcaacattactaataa
gtggctcctttaacctctccacccttatcacaacacaagaacacctctgattactcctgccatcatgacccttggccataatatgatttatctccacactagcag
agaccaaccgaaccccccttcgaccttgccgaaggggagtccgaactagtctcaggcttcaacatcgaatacgccgcaggccccttcgccctattcttcata
gccgaatacacaaacattattataataaacaccctcaccactacaatcttcctaggaacaacatatgacgcactctcccctgaactctacacaacatattttgt
caccaagaccctacttctaacctccctgttcttatgaattcgaacagcatacccccgattccgctacgaccaactcatacacctcctatgaaaaaacttcctac
cactcaccctagcattacttatatgatatgtctccatacccattacaatctccagcattccccctcaaacctaagaaatatgtctgataaaagagttactttgata
gagtaaataataggagcttaaacccccttatttctaggactatgagaatcgaacccatccctgagaatccaaaattctccgtgccacctatcacaccccatcc
taaagtaaggtcagctaaataagctatcgggcccataccccgaaaatgttggttatacccttcccgtactaattaatcccctggcccaacccgtcatctactct
accatctttgcaggcacactcatcacagcgctaagctcgcactgatttttacctgagtaggcctagaaataaacatgctagcttttattccagttctaaccaaa
aaaataaaccctcgttccacagaagctgccatcaagtatttcctcacgcaagcaaccgcatccataatccttctaatagctatcctcttcaacaatatactctcc
ggacaatgaaccataaccaatactaccaatcaatactcatcattaataatcataatggctatagcaataaaactaggaatagcccccttttcacttctgagtccc
agaggttacccaaggcacccctctgacatccggcctgcttcttctcacatgacaaaaactagcccccatctcaatcatataccaaatctctccctcactaaac
gtaagccttctcctcactctctcaatcttatccatcatagcaggcagttgaggtggattaaaccaaacccagctacgcaaaatcttagcatactcctcaattac

ccacataggatgaataatagcagttctaccgtacaaccctaacataaccattcttaatttaactatttatattatcctaactactaccgcattcctactactcaactt
aaactccagcaccacgaccctactactatctcgcacctgaaacaagctaacatgactaacacccttaattccatccaccctcctctccctaggaggcctgcc
cccgctaaccggctttttgcccaaatgggccattatcgaagaattcacaaaaaacaacagcctcatcatccccaccatcatagccaccatcaccctccttaa
cctctacttctacctacgcctaatttactccacctcaatcacactactccccatatctaacaacgtaaaaataaaatgacagtttgaacatacaaaacccaccc
cattcctccccacactcatcgcccttaccacgctactcctacctatctccccttttatactaataatcttatagaaatttaggttaaatacagaccaagagccttca
aagccctcagtaagttgcaatacttaatttctgtaacagctaaggactgcaaaaccccactctgcatcaactgaacgcaaatcagccactttaattaagctaa
gcccttactagaccaatgggacttaaacccacaaacacttagttaacagctaagcaccctaatcaactggcttcaatctacttctcccgccgccgggaaaaa
aggcgggagaagccccggcaggtttgaagctgcttcttcgaatttgcaattcaatatgaaaatcacctcggagctggtaaaaagaggcctaacccctgtct
ttagatttacagtccaatgcttcactcagccattttacctcacccccactgatgttcgccgaccgttgactattctctacaaaccacaaagacattggaacacta
tacctattattcggcgcatgagctggagtcctaggcacagctctaagcctccttattcgagccgagctgggccagccaggcaaccttctaggtaacgacca
catctacaacgttatcgtcacagcccatgcatttgtaataatcttcttcatagtaatacccatcataatcggaggctttggcaactgactagttcccctaataatc
ggtgcccccgatatggcgtttcccgcataaacaacataagcttctgactcttacctccctctctcctactcctgctcgcatctgctatagtggaggccggag
caggaacaggttgaacagtctaccctcccttagcagggaactactcccaccctggagcctccgtagacctaaccatcttctccttacacctagcaggtgtct
cctctatcttaggggccatcaatttcatcacaacaattatcaatataaaaccccctgccataacccaataccaaacgcccctcttcgtctgatccgtcctaatca
cagcagtcctacttctcctatctctcccagtcctagctgctggcatcactatactactaacagaccgcaacctcaacaccaccttcttcgaccccgccggag
gaggagaccccattctataccaacacctattctgatttttcggtcaccctgaagtttatattcttatcctaccaggcttcggaataatctcccatattgtaacttact
actccggaaaaaaagaaccatttggatacataggtatggtctgagctatgatatcaattggcttcctagggtttatcgtgtgagcacaccatatatttacagta
ggaatagacgtagacacacgagcatatttcacctccgctaccataatcatcgctatccccaccggcgtcaaagtatttagctgactcgccacactccacgg
aagcaatatgaaatgatctgctgcagtgctctgagccctaggattcatctttcttttcaccgtaggtggcctgactggcattgtattagcaaactcatcactaga
catcgtactacacgacacgtactacgttgtagctcacttccactatgtcctatcaataggagctgtatttgccatcataggaggcttcattcactgatttccccta
ttctcaggctacaccctagaccaaacctacgccaaaatccatttcactatcatattcatcggcgtaaatctaactttcttcccacaacactttctcggcctatccg
gaatgccccgacgttactcggactaccccgatgcatacaccacatgaaacatcctatcatctgtaggctcattcatttctctaacagcagtaatattaataattt
tcatgatttgagaagccttcgcttcgaagcgaaaagtcctaatagtagaagaaccctccataaacctggagtgactatatggatgccccccaccctaccac
acattcgaagaacccgtatacataaaatctagacaaaaaaggaaggaatcgaaccccccaaagctggtttcaagccaaccccatggcctccatgacttttt
caaaaaggtattagaaaaaccatttcataactttgtcaaagttaaattataggctaaatcctatatatcttaatggcacatgcagcgcaagtaggtctacaaga
cgctacttcccctatcatagaagagcttatcaccttttcatgatcacgccctcataatcattttccttatctgcttcctagtcctgtatgccctttttcctaacactcaca
acaaaactaactaatactaacatctcagacgctcaggaaatagaaaccgtctgaactatcctgcccgccatcatcctagtcctcatcgcccctcccatcccta
cgcatcctttacataacagacgaggtcaacgatccctcccttaccatcaaatcaattggccaccaatggtactgaacctacgagtacaccgactacggcgg
actaatcttcaactcctacatacttcccccattattcctagaaccaggcgacctgcgactccttgacgttgacaatcgagtagtactcccgattgaagccccc
attcgtataataattacatcacaagacgtcttgcactcatgagctgtccccacattaggcttaaaaacagatgcaattcccggacgtctaaaccaaaccacttt
taccgctacacgaccgggggtatactacggtcaatgctctgaaatctgtggagcaaaccacagtttcatgcccatcgtcctagaattaattcccctaaaaatc
tttgaaatagggcccgtatttaccctatagcacccccctctaccccctctagagcccactgtaaagctaacttagcattaacctttaagttaaagattaagagaa
ccaacacctctttacagtgaaatgccccaactaaatactaccgtatggcccaccataattacccccatactccttacactattcctcatcacccaactaaaaat
attaaacacaaactaccacctacctccctcaccaaagcccataaaaataaaaaattataacaaaccctgagaaccaaaatgaacgaaaatctgttcacttca
ttcattgccccacaatcctaggcctacccgccgcagtactgatcattctatttccccctctattgatccccacctccaaatatctcatcaacaaccgactaatc
accacccaacaatgactaatcaaactaatctcaaaacaaatgataaccatacacaacactaaaggacgaacctgatctcttatactagtatccttaatcatttt
attgccacaactaacctcctcggactcctgcctcactcatttacaccaaccacccaactatctataaacctagccatggccatccccttatgagcgggcgca
gtgattataggctttcgctctaagattaaaaatgccctagcccacttcttaccacaaggcacacctacacccctatccccatactagttattatcgaaaccatc
agcctactcattcaaccaatagccctggccgtacgcctaaccgctaacattactgcaggccacctactcatgcacctaattggaagcgccaccctagcaat
atcaaccattaaccttccctctacacttatcatcttcacaattctaattctactgactatcctagaaatcgctgtcgccttaatccaagcctacgttttcacacttcta
gtaagcctctacctgcacgacaacacataatgacccaccaatcacatgcctatcatatagtaaaacccagcccatgacccctaacaggagccctctcagc
cctcctaatgacctccggcctagccatgtgatttcacttccactccataacgctcctcatactaggcctactaaccaacacactaaccatataccaatgatgg
cgcgatgtaacacgagaaagcacataccaaggccaccacacaccacctgtccaaaaaggccttcgatacgggataatcctatttattacctcagaagttttt
ttcttcgcaggatttttctgagccttttaccactccagcctagcccctacccccaattaggagggcactggccccaacaggcatcaccccgctaaatccc
ctagaagtcccactcctaaacacatccgtattactcgcatcaggagtatcaatcacctgagctcaccatagtctaatagaaaacaaccgaaaccaaataatt
caagcactgcttattacaattttactgggtctctattttaccctcctacaagcctcagagtacttcgagtctcccttcaccatttccgacggcatctacggctcaa
catttttgtagccacaggcttccacggacttcacgtcattattggctcaactttcctcgctatctgcttcatccgccaactaatatttcactttacatccaaacatc
actttggcttcgaagccgccgcctgatactggcattttgtagatgtggtttgactatttctgtatgtctccatctattgatgagggtcttactcttttagtataaatag
taccgttaacttccaattaactagttttgacaacattcaaaaaagagtaataaacttcgccttaattttaataatcaacaccctcctagccttactactaataattatt
acattttgactaccacaactcaatggctacatagaaaaatccacccccttacgagtgcggcttcgaccctatatcccccgcccgcgtccctttctccataaaatt
cttcttagtagctattaccttcttattatttgatctagaaattgccctccttttaccccctaccatgagccctacaaacaactaacctgccactaatagttatgtcatcc
ctcttattaatcatcatcctagccctaagtctggcctatgagtgactacaaaaaggattagactgaaccgaattggtatatagtttaaacaaaacgaatgatttc
gactcattaaattatgataatcatatttaccaaatgcccctcatttacataaatattatactagcatttaccatctcacttctaggaatactagtatatcgctcacac
ctcatatcctccctactatgcctagaaggaataatactatcgctgttcattatagctactctcataccctcaacacccactccctcttagccaatattgtgcctat
tgccatactagtctttgccgcctgcgaagcagcggtgggcctagccctactagtctcaatctccaacacatatggcctagactacgtacataacctaaacct
actccaatgctaaaactaatcgtcccaacaattatattactaccactgacatgactttccaaaaaacacataatttgaatcaacacaaccacccacagcctaat

tattagcatcatccctctactatttttttaaccaaatcaacaacaacctatttagctgttccccaacctttttcctccgaccccctaacaaccccccctcctaatactaa
ctacctgactcctacccctcacaatcatggcaagccaacgccacttatccagtgaaccactatcacgaaaaaaactctacctctctatactaatctccctaca
aatctccttaattataacattcacagccacagaactaatcatattttatatcttcttcgaaaccacacttatccccaccttggctatcatcacccgatgaggcaac
cagccagaacgcctgaacgcaggcacatacttcctattctacaccctagtaggctcccttcccctactcatcgcactaatttacactcacaacaccctaggct
cactaaacattctactactcactctcactgcccaagaactatcaaactcctgagccaacaacttaatatgactagcttacacaatagcttttatagtaaagatac
ctctttacggactccacttatgactccctaaagcccatgtcgaagcccccatcgctgggtcaatagtacttgccgcagtactcttgaaactaggcggctatgg
tataatacgcctcacactcattctcaaccccctgacaaaacacatagcctacccctccttgtactatccctatgaggcataattataacaagctccatctgcct
acgacaaacagacctaaaatcgctcattgcatactcttcaatcagccacatagccctcgtagtaacagccattctcatccaaaccccctgaagcttcaccgg
cgcagtcattctcataatcgcccacggacttacatcctcattactattctgcctagcaaactcaaactacgaacgcactcacagtcgcatcataatcctctctc
aaggacttcaaactctactcccactaatagctttttgatgacttctagcaagcctcgctaacctcgccttacccccccactattaacctactgggagaactctctg
tgctagtaaccacgttctcctgatcaaatatcactctcctacttacaggactcaacatactagtcacagccctatactccctctacatatttaccacaacacaat
ggggctcactcacccaccacattaacaacataaaaccctcattcacacgagaaaacaccctcatgttcatacacctatccccccattctcctcctatccctcaa
ccccgacatcattaccgggttttcctcttgtaaatatagtttaaccaaaacatcagattgtgaatctgacaacagaggcttacgaccccttatttaccgagaaa
gctcacaagaactgctaactcatgcccccatgtctaacaacatggctttctcaacttttaaaggataacagctatccattggtcttaggccccaagaattttggt
gcaactccaaataaaagtaataaccatgcacactactataaccaccctaaccctaacttccctaattcccccccatccttaccaccctcgttaaccctaacaaa
aaaaactcataccccccattatgtaaaatccattgtcgcatccaccttttattatcagtctcttccccacaacaatattcatgtgcctagaccaagaagttattatctc
gaactgacactgagccacaacccaaacaacccagctctccctaagcttcaaactagactacttctccataatattcatccctgtagcattattcgttacatggt
ccatcatagaattctcactgtgatatataaactcagacccaaacattaatcagttcttcaaatatctactcatcttcctaattaccatactaatcttagttaccgcta
acaacctattccaactgttcatcggctgagagggcgtaggaattatatccttcttgctcatcagttgatgatacgcccgagcagatgccaacacagcagcca
ttcaagcaatcctatacaaccgtatcggcgatatcggtttcatcctcgccttagcatgatttatcctacactccaactcatgagacccacaacaaatagcccttt
ctaaacgctaatccaagcctcaccccactactaggcctcctcctagcagcagcaggcaaatcagcccaattaggtctccacccctgactcccctcagccat
agaaggccccaccccagtctcagccctactccactcaagcactatagttgtagcaggaatcttcttactcatccgcttccacccctagcagaaaatagccc
actaatccaaactctaacactatgcttaggcgctatcaccactctgttcgcagcagtctgcgcccttacacaaaatgacatcaaaaaaatcgtagccttctcc
acttcaagtcaactaggactcataatagttacaatcggcatcaaccaaccacacctagcattcctgcacatctgtacccacgccttcttcaaagccatactatt
tatgtgctccgggtccatcatccacaaccttaacaatgaacaagatattcgaaaaataggaggactactcaaaaccatacctctcacttcaacctccctcacc
attggcagcctagcattagcaggaatacctttcctcacaggcttctactccaaagaccacatcatcgaaaccgcaaacatatcatacacaaacgcctgagc
cctatctattactctcatcgctacctccctgacaagcgcctatagcactcgaataattcttctcaccctaacaggtcaacctcgcttccccacccttactaacatt
aacgaaaataacccccaccctactaaaccccattaaacgcctggcagccggaagcctattcgcaggatttctcattactaacaacatttcccccgcatccccc
ttccaaacaacaattcccctctacctaaaactcacagccctcgctgtcactttcctaggacttctaacagccctagacctcaactacctaaccaacaaacttaa
aataaaatccccactatgcacattttatttctccaacatactcggattctaccctagcatcacacaccgcacaatcccctatctaggccttcttacgagccaaaa
cctgcccctactcctcctagacctaacctgactagaaaagctattacctaaaacaatttcacagcaccaaatctccacctccatcatcacctcaacccaaaaa
ggcataattaaacttttacttcctctctttcttcttcccactcatcctaaccctactcctaatcacataacctattccccgagcaatctcaattacaatatatacacc
aacaaacaatgttcaaccagtaactactactaatcaacgcccataatcatacaaagcccccgcaccaataggatcctcccgaatcaaccctgacccctctc
cttcataaattattcagcttcctacactattaaagtttaccacaaccaccaccccatcatactctttcacccacagcaccaatcctacctccatcgctaacccca
ctaaaacactcaccaagacctcaacccctgacccccatgcctcaggatactcctcaatagccatcgctgtagtatatccaaagacaaccatcattccccta
aataaattaaaaaaactattaaacccatataacctcccccaaaattcagaataataacacacccgaccacaccgctaacaatcaatactaaaccccccataaa
taggagaaggcttagaagaaaaccccacaaaccccattactaaacccacactcaacagaaacaaagcatacatcattattctcgcacggactacaaccac
gaccaatgatatgaaaaaccatcgttgtatttcaactacaagaacaccaatgaccccaatacgcaaaactaacccccctaataaaattaattaaccactcattc
atcgacctccccaccccatccaacatctccgcatgatgaaacttcggctcactccttggcgcctgcctgatcctccaaatcaccacaggactattcctagcc
atgcactactcaccagacgcctcaaccgccttttcatcaatcgcccacatcactcgagacgtaaattatggctgaatcatccgctaccttcacgccaatggc
gcctcaatattctttatctgcctcttcctacacatcgggcgaggcctatattacggatcatttctctactcagaaacctgaaacatcggcattatcctcctgcttg
caactatagcaacagccttcataggctatgtcctcccgtgaggccaaatatcattctgaggggccacagtaattacaaacttactatccgccatcccatacat
tgggacagacctagttcaatgaatctgaggaggctactcagtagacagtcccaccctcacacgattctttacctttcacttcatcttgcccttcattattgcagc
cctagcagcactccacctcctattcttgcacgaaacgggatcaaataaccccctaggaatcacctcccattccgataaaatcaccttccacccttactacaca
atcaaagacgccctcggcttacttctcttccttctctccttaatgacattaacactattctcaccagacctcctaggcgacccagacaattataccctagccaac
cccttaaacacccctccccacatcaagcccgaatgatatttcctattcgcctacacaattctccgatccgtccctaacaaactaggaggcgtccttgccctatt
actatccatcctcatcctagcaataatccccatcctccatatatccaaacaacaaagcataatatttcgcccactaagccaatcactttattgactcctagccgc
agacctcctcattctaacctgaatcggaggacaaccagtaagctacccttttaccatcattggacaagtagcatccgtactatacttcacaacaatcctaatcc
taataccaactatctccctaattgaaaacaaaatactcaaatgggcctgtccttgtagtataaactaatacaccagtcttgtaaaccggagatgaaaacctttt
ccaaggacaaatcagagaaaaagtctttaactccaccattagcacccaaagctaagattctaatttaaactattctctgttctttcatggggaagcagatttgg
gtaccacccaagtattgactcacccatcaacaaccgctatgtatttcgtacattactgccagccaccatgaatattgtacggtaccataaatacttgaccacct
gtagtacataaaaacccaatccacatcaaaacccccctccccatgcttacaagcaagtacagcaatcaaccttcaactatcacacatcaactgcaactccaa
agccacccctcacccactaggataccaacaaacctacccacccttaacagtacatggtacataatgccatttaccgtacatagcacattacagtcaaatccc
ttctcgtccccatggatgacccccctcagataggggtcccttgaccaccatcctccgtgaaatcaatatcccgcacaagagtgctactctcctcgctccggg
cccataacacttgggggtagctaaagtgaactgtatccgacatctggttcctacttcagggccataaagcctaaatagcccacacgttccccttaaataaga
catcacgatg

8.12 Appendix L

Letter of approval of present research by the Ethical Review Board of Nepal Health Research Council.

Government of Nepal

Nepal Health Research Council (NHRC)

Estd. 1991

Ref. No.: 2070

02 June 2016

Mr. Rajdip Basnet
Principal Investigator
Central Department of Biotechnology
Tribhuvan University, Kirtipur

Ref: **Approval of Research Proposal** entitled **Genetic Affinity of Newa Population of Kathmandu Valley**

Dear Mr. Basnet,

It is my pleasure to inform you that the above-mentioned proposal submitted on 29 April 2016 (**Reg.no. 108/2016** please use this Reg. No. during further correspondence) has been approved by NHRC Ethical Review Board on 01 June 2016.

As per NHRC rules and regulations, the investigator has to strictly follow the protocol stipulated in the proposal. Any change in objective(s), problem statement, research question or hypothesis, methodology, implementation procedure, data management and budget that may be necessary in course of the implementation of the research proposal can only be made so and implemented after prior approval from this council. Thus, it is compulsory to submit the detail of such changes intended or desired with justification prior to actual change in the protocol.

If the researcher requires transfer of the bio samples to other countries, the investigator should apply to the NHRC for the permission. The researchers will not be allowed to ship any raw/crude human biomaterial outside the country; only extracted and amplified samples can be taken to labs outside of Nepal for further study, as per the protocol submitted and approved by the NHRC. The remaining samples of the lab should be destroyed as per standard operating procedure, the process documented, and the NHRC informed.

Further, the researchers are directed to strictly abide by the National Ethical Guidelines published by NHRC during the implementation of their research proposal and submit progress report and full or summary report upon completion.

As per your research proposal, the total research amount is **Self- Funded** and accordingly the processing fee amount to **NRs.1000.00.** It is acknowledged that the above-mentioned processing fee has been received at NHRC.

If you have any questions, please contact the Ethical Review M & E section of NHRC.

Thanking you,

Dr. Khem Bahadur Karki
Member-Secretary

Prof. Dr. K. Thangaraj Group
Centre for Cellular and Molecular Biology, Hyderabad, India

9 References

Allentoft ME, Sikora M, Sjogren KG, Rasmussen S, Rasmussen M, Stenderup J, et al. (2015). Population genomics of Bronze Age Eurasia. *Nature, 522*(7555), 167-172. doi:10.1038/nature14507

Armitage SJ, Jasim SA, Marks AE, Parker AG, Usik VI, & Uerpmann HP. (2011). The southern route "out of Africa": evidence for an early expansion of modern humans into Arabia. *Science, 331*(6016), 453-456. doi:10.1126/science.1199113

Aubert M, Pike AW, Stringer C, Bartsiokas A, Kinsley L, Eggins S, et al. (2012). Confirmation of a late middle Pleistocene age for the Omo Kibish 1 cranium by direct uranium-series dating. *J Hum Evol, 63*(5), 704-710. doi:10.1016/j.jhevol.2012.07.006

Awasthi N, Manandhar KD and Shrestha TR (2014). Genetic structure of sub-ethnic groups of Newar population of Kathmandu valley. M.Sc. thesis submitted in Central Department of Biotechnology, Tribhuvan University, Nepal.

Batini C, Hallast P, Zadik D, Delser PM, Benazzo A, & Ghirotto S. (2015). Large-scale recent expansion of European patrilineages shown by population resequencing. *Nat Commun, 6*. doi:10.1038/ncomms8152

Beall CM, Cavalleri GL, Deng L, Elston RC, Gao Y, Knight J, et al. (2010). Natural selection on EPAS1 (HIF2α) associated with low hemoglobin concentration in Tibetan highlanders. *Proceedings of the National Academy of Sciences of the United States of America, 107*(25), 11459-11464. doi:10.1073/pnas.1002443107

Behar DM, Villems R, Soodyall H, Blue-Smith J, Pereira L, Metspalu E, et al. (2008). The dawn of human matrilineal diversity. *Am J Hum Genet, 82*(5), 1130-1140. doi:10.1016/j.ajhg.2008.04.002

Bersaglieri T, Sabeti PC, Patterson N, Vanderploeg T, Schaffner SF, Drake JA, et al. (2004). Genetic signatures of strong recent positive selection at the lactase gene. *Am J Hum Genet, 74*(6), 1111-1120. doi:10.1086/421051

Bhandari S, Zhang X, Cui C, Bianba, Liao S, Peng Y, et al. (2015). Genetic evidence of a recent Tibetan ancestry to Sherpas in the Himalayan region. *Sci Rep, 5*, 16249. doi:10.1038/srep16249

Bigham A, Bauchet M, Pinto D, Mao X, Akey JM, Mei R, et al. (2010). Identifying signatures of natural selection in Tibetan and Andean populations using dense genome scan data. *PLoS Genet, 6*(9), e1001116. doi:10.1371/journal.pgen.1001116

Bista DB. (1972). People of Nepal

Browning SR, & Browning BL. (2007). Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. *The American Journal of Human Genetics, 81*(5), 1084-1097.

Caramelli D, Lalueza-Fox C, Vernesi C, Lari M, Casoli A, Mallegni F, et al. (2003). Evidence for a genetic discontinuity between Neandertals and 24,000-year-old anatomically modern Europeans. *Proc Natl Acad Sci U S A, 100*(11), 6593-6597. doi:10.1073/pnas.1130343100

Central Bureau of Statistics of Nepal: Major Highlights. (2011). 1-13.

Chandrasekar A, Kumar S, Sreenath J, Sarkar BN, Urade BP, Mallick S, et al. (2009). Updating phylogeny of mitochondrial DNA macrohaplogroup m in India: dispersal of modern human in South Asian corridor. *PLoS One, 4*(10), e7447. doi:10.1371/journal.pone.0007447

Chaubey G, Karmin M, Metspalu E, Metspalu M, Selvi-Rani D, Singh VK, et al. (2008). Phylogeography of mtDNA haplogroup R7 in the Indian peninsula. *BMC Evol Biol, 8*, 227. doi:10.1186/1471-2148-8-227

Chaubey G, Singh M, Crivellaro F, Tamang R, Nandan A, Singh K, et al. (2014). Unravelling the distinct strains of Tharu ancestry. *Eur J Hum Genet, 22*(12), 1404-1412. doi:10.1038/ejhg.2014.36

Chen H, Patterson N, & Reich D. (2010). Population differentiation as a test for selective sweeps. *Genome Res, 20*(3), 393-402. doi:10.1101/gr.100545.109

Cook AL, Chen W, Thurber AE, Smit DJ, Smith AG, Bladen TG, et al. (2009). Analysis of cultured human melanocytes based on polymorphisms within the SLC45A2/MATP, SLC24A5/NCKX5, and OCA2/P loci. *J Invest Dermatol, 129*(2), 392-405. doi:10.1038/jid.2008.211

Degener A. (2005). Languages of the Himalayas. An Ethnolinguistic Handbook of the Greater Himalayan Region containing an Introduction to the Symbiotic Theory of Language. Vol. 1. 2.(Handbook of Oriental Studies. Abt. 2. Bd. 10): JSTOR.

Delfin F, Min-Shan Ko A, Li M, Gunnarsdottir ED, Tabbada KA, Salvador JM, et al. (2014). Complete mtDNA genomes of Filipino ethnolinguistic groups: a melting pot of recent and ancient lineages in the Asia-Pacific region. *Eur J Hum Genet, 22*(2), 228-237. doi:10.1038/ejhg.2013.122

Derenko M, Malyarchuk B, Bahmanimehr A, Denisova G, Perkova M, Farjadian S, et al. (2013). Complete mitochondrial DNA diversity in Iranians. *PLoS One, 8*(11), e80673. doi:10.1371/journal.pone.0080673

Derenko M, Malyarchuk B, Grzybowski T, Denisova G, Dambueva I, Perkova M, et al. (2007). Phylogeographic analysis of mitochondrial DNA in northern Asian populations. *The American Journal of Human Genetics, 81*(5), 1025-1041.

Derenko M, Malyarchuk B, Grzybowski T, Denisova G, Rogalla U, Perkova M, et al. (2010). Origin and post-glacial dispersal of mitochondrial DNA haplogroups C and D in northern Asia. *PLoS One, 5*(12), e15214. doi:10.1371/journal.pone.0015214

Desai Z-DA, & Sharma AK. (2003). *Purā-prakāśa : recent researches in epigraphy, numismatics, manuscriptology, Persian literature, art, architecture, archaeology, history and conservation : Dr. Z.A. Desai commemoration volume*: Bharatiya Kala Prakashan.

Dixit K. (2012). The lake that was once Kathmandu.

Eaaswarkhanth M, Haque I, Ravesh Z, Romero IG, Meganathan PR, Dubey B, et al. (2010). Traces of sub-Saharan and Middle Eastern lineages in Indian Muslim populations. *Eur J Hum Genet, 18*(3), 354-363. doi:10.1038/ejhg.2009.168

Eltsov N, & Volodko N. (2009). mtPhyl-software tool for human mtDNA analysis and phylogeny reconstruction.

Enard W, Gehre S, Hammerschmidt K, Holter SM, Blass T, Somel M, et al. (2009). A humanized version of Foxp2 affects cortico-basal ganglia circuits in mice. *Cell, 137*(5), 961-971. doi:10.1016/j.cell.2009.03.041

Enard W, Przeworski M, Fisher SE, Lai CS, Wiebe V, Kitano T, et al. (2002). Molecular evolution of FOXP2, a gene involved in speech and language. *Nature, 418*(6900), 869-872. doi:10.1038/nature01025

Excoffier L, & Lischer HE. (2010). Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Mol Ecol Resour, 10*(3), 564-567. doi:10.1111/j.1755-0998.2010.02847.x

Fagundes NJ, Kanitz R, Eckert R, Valls AC, Bogo MR, Salzano FM, et al. (2008). Mitochondrial population genomics supports a single pre-Clovis origin with a coastal route for the peopling of the Americas. *Am J Hum Genet, 82*(3), 583-592. doi:10.1016/j.ajhg.2007.11.013

Fedorova SA, Reidla M, Metspalu E, Metspalu M, Rootsi S, Tambets K, et al. (2013). Autosomal and uniparental portraits of the native populations of Sakha (Yakutia): implications for the peopling of Northeast Eurasia. *BMC Evol Biol, 13*, 127. doi:10.1186/1471-2148-13-127

Fornarino S, Pala M, Battaglia V, Maranta R, Achilli A, Modiano G, et al. (2009). Mitochondrial and Y-chromosome diversity of the Tharus (Nepal): a reservoir of genetic variation. *BMC Evol Biol, 9*, 154. doi:10.1186/1471-2148-9-154

Forster P, Harding R, Torroni A, & Bandelt H-J. (1996). Origin and evolution of Native American mtDNA variation: a reappraisal. *American journal of human genetics, 59*(4), 935.

Fu Q, Posth C, Hajdinjak M, Petr M, Mallick S, Fernandes D, et al. (2016). The genetic history of Ice Age Europe. *Nature, 534*(7606), 200-205. doi:10.1038/nature17993

Fu Y-X. (1997). Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics, 147*(2), 915-925.

Fu YX. (1997). Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics, 147*(2), 915-925.

Fuller DQ. (2007). Non-human genetics, agricultural origins and historical linguistics in South Asia. In M. D. Petraglia & B. Allchin (Eds.), *The Evolution and History of Human Populations in South Asia*. Dordrecht: Springer Netherlands.

Furer-Haimendorf Cv. (1956). "Elements of Newar Social Structure". Royal Anthropological Institute of Great Britain and Ireland. *86*(2), 15-38.

Gayden T, Perez A, Persad PJ, Bukhari A, Chennakrishnaiah S, Simms T, et al. (2013). The Himalayas: barrier and conduit for gene flow. *Am J Phys Anthropol, 151*(2), 169-182. doi:10.1002/ajpa.22240

Gazi NN, Tamang R, Singh VK, Ferdous A, Pathak AK, Singh M, et al. (2013). Genetic structure of Tibeto-Burman populations of Bangladesh: evaluating the gene flow along the sides of Bay-of-Bengal. *PLoS One, 8*(10), e75064. doi:10.1371/journal.pone.0075064

Ge RL, Simonson TS, Cooksey RC, Tanna U, Qin G, Huff CD, et al. (2012). Metabolic insight into mechanisms of high-altitude adaptation in Tibetans. *Mol Genet Metab, 106*(2), 244-247. doi:10.1016/j.ymgme.2012.03.003

Gellner DN. (2009). Sociology : Language , caste , religion and territory : Newar identity ancient and modern. (1986), 102-148. doi:10.1017/S0003975600004549

Gellner DN, & Quigley D. (1995). *Contested hierarchies: A collaborative ethnography of caste among the Newars of the Kathmandu Valley, Nepal*: Clarendon Press Oxford.

Gentleman R, Ihaka R, & Bates D. (2009). The R project for statistical computing. *URL: http://www. r-project. org/254*.

Green RE, Krause J, Briggs AW, Maricic T, Stenzel U, Kircher M, et al. (2010). A draft sequence of the Neandertal genome. *Science, 328*(5979), 710-722. doi:10.1126/science.1188021

Grossman SR, Andersen KG, Shlyakhter I, Tabrizi S, Winnicki S, Yen A, et al. (2013). Identifying recent adaptations in large-scale genomic data. *Cell, 152*(4), 703-713. doi:10.1016/j.cell.2013.01.035

Grossman SR, Shlyakhter I, Karlsson EK, Byrne EH, Morales S, Frieden G, et al. (2010). A composite of multiple signals distinguishes causal variants in regions of positive selection. *Science, 327*(5967), 883-886. doi:10.1126/science.1183863

Grun R, Stringer C, McDermott F, Nathan R, Porat N, Robertson S, et al. (2005). U-series and ESR analyses of bones and teeth relating to the human burials from Skhul. *J Hum Evol, 49*(3), 316-334. doi:10.1016/j.jhevol.2005.04.006

Gurung HB. (1980). Vignettes of Nepal. *Sajha Prakashan*.

Hamblin MT, & Di Rienzo A. (2000). Detection of the signature of natural selection in humans: evidence from the Duffy blood group locus. *Am J Hum Genet, 66*(5), 1669-1679. doi:10.1086/302879

Hanaoka M, Droma Y, Basnyat B, Ito M, Kobayashi N, Katsuyama Y, et al. (2012). Genetic variants in EPAS1 contribute to adaptation to high-altitude hypoxia in Sherpas. *PLoS One, 7*(12), e50566. doi:10.1371/journal.pone.0050566

Holsinger KE, & Weir BS. (2009). Genetics in geographically structured populations: defining, estimating and interpreting F(ST). *Nature reviews. Genetics, 10*(9), 639-650. doi:10.1038/nrg2611

Hublin JJ. (2009). Out of Africa: modern human origins special feature: the origin of Neandertals. *Proc Natl Acad Sci U S A, 106*(38), 16022-16027. doi:10.1073/pnas.0904119106

Hublin JJ, Ben-Ncer A, Bailey SE, Freidline SE, Neubauer S, Skinner MM, et al. (2017). New fossils from Jebel Irhoud, Morocco and the pan-African origin of Homo sapiens. *Nature, 546*(7657), 289-292. doi:10.1038/nature22336

Irwin JA, Ikramov A, Saunier J, Bodner M, Amory S, Röck A, et al. (2010). The mtDNA composition of Uzbekistan: a microcosm of Central Asian patterns. *International Journal of Legal Medicine, 124*(3), 195-204.

Jarrige J-F JC. (2006). Premiers pasteurs et agriculteurs dans le sous-continent Indo-Pakistanais. *Comptes Rendus Palevol*(5), 463–472.

Jeong C, Ozga AT, Witonsky DB, Malmstrom H, Edlund H, Hofman CA, et al. (2016). Long-term genetic stability and a high-altitude East Asian origin for the peoples of the high valleys of the Himalayan arc. *Proc Natl Acad Sci U S A, 113*(27), 7485-7490. doi:10.1073/pnas.1520844113

Jiang C, Cui J, Liu F, Gao L, Luo Y, Li P, et al. (2014). Mitochondrial DNA 10609T promotes hypoxia-induced increase of intracellular ROS and is a risk factor of high altitude polycythemia. *PLoS One, 9*(1), e87775. doi:10.1371/journal.pone.0087775

Kamberov YG, Wang S, Tan J, Gerbault P, Wark A, Tan L, et al. (2013). Modeling recent human evolution in mice by expression of a selected EDAR variant. *Cell, 152*(4), 691-702. doi:10.1016/j.cell.2013.01.016

Kang L, Zheng HX, Chen F, Yan S, Liu K, Qin Z, et al. (2013). mtDNA lineage expansions in Sherpa population suggest adaptive evolution in Tibetan highlands. *Mol Biol Evol, 30*(12), 2579-2587. doi:10.1093/molbev/mst147

Kang L, Zheng HX, Zhang M, Yan S, Li L, Liu L, et al. (2016). MtDNA analysis reveals enriched pathogenic mutations in Tibetan highlanders. *Sci Rep, 6*, 31083. doi:10.1038/srep31083

Karmin M, Saag L, Vicente M, Wilson Sayres MA, Järve M, & Talas UG. (2015). A recent bottleneck of Y chromosome diversity coincides with a global change in culture. *Genome Res, 25*. doi:10.1101/gr.186684.114

Kingwell-Banham E, Petrie CA, & Fuller DQ. (2015). Early agriculture in South Asia. In C. A. Barker & C. Goucher (Eds.), *The Cambridge World History: Volume 2, A World with Agriculture, 12,000 BCE–500 CE*. Cambridge: Cambridge University Press.

Kirkpatrick C. (1811). *An Account of the Kingdom of Nepaul*: London: William Miller.

Kivisild T, Reidla M, Metspalu E, Rosa A, Brehm A, Pennarun E, et al. (2004). Ethiopian mitochondrial DNA heritage: tracking gene flow across and around the gate of tears. *Am J Hum Genet, 75*(5), 752-770. doi:10.1086/425161

Kloss-Brandstätter A, Pacher D, Schönherr S, Weissensteiner H, Binna R, Specht G, et al. (2011). HaploGrep: a fast and reliable algorithm for automatic classification of mitochondrial DNA haplogroups. *Human Mutation, 32*(1), 25-32.

Ko AM, Chen CY, Fu Q, Delfin F, Li M, Chiu HL, et al. (2014). Early Austronesians: into and out of Taiwan. *Am J Hum Genet, 94*(3), 426-436. doi:10.1016/j.ajhg.2014.02.003

Kovach W. (1999). MVSP–A multivariate statistical package for Windows, ver. 3.1. *Kovach Computing Services, Pentraeth, Wales, UK, 137*.

Kumar S, Padmanabham PB, Ravuri RR, Uttaravalli K, Koneru P, Mukherjee PA, et al. (2008). The earliest settlers' antiquity and evolutionary history of Indian populations: evidence from M2 mtDNA lineage. *BMC Evol Biol, 8*, 230. doi:10.1186/1471-2148-8-230

Kutanan W, Kampuansai J, Srikummool M, Kangwanpong D, Ghirotto S, Brunelli A, et al. (2017). Complete mitochondrial genomes of Thai and Lao populations indicate an ancient origin of Austroasiatic groups and demic diffusion in the spread of Tai-Kadai languages. *Hum Genet, 136*(1), 85-98. doi:10.1007/s00439-016-1742-y

Lamason RL, Mohideen MA, Mest JR, Wong AC, Norton HL, Aros MC, et al. (2005). SLC24A5, a putative cation exchanger, affects pigmentation in zebrafish and humans. *Science, 310*(5755), 1782-1786. doi:10.1126/science.1116238

Lane CS, Chorn BT, & Johnson TC. (2013). Ash from the Toba supereruption in Lake Malawi shows no volcanic winter in East Africa at 75 ka. *Proc Natl Acad Sci U S A, 110*(20), 8025-8029. doi:10.1073/pnas.1301474110

Levine NE. (1987). Caste, State, and Ethnic Boundaries in Nepal. *The Journal of Asian Studies, 46*(1), 71-71. doi:10.2307/2056667

Lewis TTaDRS. (1988). Contributions to the History of Nepal: Eastern Newar Diaspora Settlements' in Contributions to Nepalese Studies.

Li H, Cai X, Winograd-Cort ER, Wen B, Cheng X, Qin Z, et al. (2007). Mitochondrial DNA diversity and population differentiation in southern East Asia. *American journal of physical anthropology, 134*(4), 481-488.

Li YC, Wang HW, Tian JY, Liu LN, Yang LQ, Zhu CL, et al. (2015). Ancient inland human dispersals from Myanmar into interior East Asia since the Late Pleistocene. *Sci Rep, 5*, 9473. doi:10.1038/srep09473

Macaulay V, Hill C, Achilli A, Rengo C, Clarke D, Meehan W, et al. (2005). Single, rapid coastal settlement of Asia revealed by analysis of complete mitochondrial genomes. *Science, 308*(5724), 1034-1036. doi:10.1126/science.1109792

Majumder PP. (2010). The human genetic history of South Asia. *Curr Biol, 20*(4), R184-187. doi:10.1016/j.cub.2009.11.053

Malaspinas AS, Westaway MC, Muller C, Sousa VC, Lao O, Alves I, et al. (2016). A genomic history of Aboriginal Australia. *Nature, 538*(7624), 207-214. doi:10.1038/nature18299

Mallick S, Li H, Lipson M, Mathieson I, Gymrek M, Racimo F, et al. (2016). The Simons Genome Diversity Project: 300 genomes from 142 diverse populations. *Nature, 538*(7624), 201-206. doi:10.1038/nature18964

Mellars P, Gori KC, Carr M, Soares PA, & Richards MB. (2013). Genetic and archaeological perspectives on the initial modern human colonization of southern Asia. *Proc Natl Acad Sci U S A, 110*. doi:10.1073/pnas.1306043110

Metspalu M, Romero IG, Yunusbayev B, Chaubey G, Mallick CB, Hudjashov G, et al. (2011). Shared and unique components of human population structure and genome-wide signals of positive selection in South Asia. *Am J Hum Genet, 89*(6), 731-744. doi:10.1016/j.ajhg.2011.11.010

Miller LH, Mason SJ, Clyde DF, & McGinniss MH. (1976). The resistance factor to Plasmodium vivax in blacks. The Duffy-blood-group genotype, FyFy. *N Engl J Med, 295*(6), 302-304. doi:10.1056/NEJM197608052950602

Mishmar D, Ruiz-Pesini E, Golik P, Macaulay V, Clark AG, Hosseini S, et al. (2003). Natural selection shaped regional mtDNA variation in humans. *Proc Natl Acad Sci U S A, 100*(1), 171-176. doi:10.1073/pnas.0136972100

Mishra S, Chauhan N, & Singhvi AK. (2013). Continuity of microblade technology in the Indian Subcontinent since 45 ka: implications for the dispersal of modern humans. *PLoS One, 8*. doi:10.1371/journal.pone.0069280

Moorjani P, Thangaraj K, Patterson N, Lipson M, Loh P-R, Govindaraj P, et al. (2013). Genetic evidence for recent population mixture in India. *American journal of human genetics, 93*(3), 422-438. doi:10.1016/j.ajhg.2013.07.006

Negi N, Tamang R, Pande V, Sharma A, Shah A, Reddy AG, et al. (2016). The paternal ancestry of Uttarakhand does not imitate the classical caste system of India. *J Hum Genet, 61*(2), 167-172. doi:10.1038/jhg.2015.121

Nei M. (1978). Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics, 89*(3), 583-590.

Nielsen R, Akey JM, Jakobsson M, Pritchard JK, Tishkoff S, & Willerslev E. (2017). Tracing the peopling of the world through genomics. *Nature, 541*(7637), 302-310. doi:10.1038/nature21347

Nielsen R, Hellmann I, Hubisz M, Bustamante C, & Clark AG. (2007). Recent and ongoing selection in the human genome. *Nat Rev Genet, 8*(11), 857-868. doi:10.1038/nrg2187

Norton HL, Kittles RA, Parra E, McKeigue P, Mao X, Cheng K, et al. (2007). Genetic evidence for the convergent evolution of light skin in Europeans and East Asians. *Mol Biol Evol, 24*(3), 710-722. doi:10.1093/molbev/msl203

Oven M. (2015). PhyloTree Build 17: Growing the human mitochondrial DNA tree. *Forensic Sci Int Genet, 5*. doi:10.1016/j.fsigss.2015.09.155

Pagani L, Lawson DJ, Jagoda E, Morseburg A, Eriksson A, Mitt M, et al. (2016). Genomic analyses inform on migration events during the peopling of Eurasia. *Nature, 538*(7624), 238-242. doi:10.1038/nature19792

Palanichamy MG, Mitra B, Zhang C-L, Debnath M, Li G-M, & Wang H-W. (2015). West Eurasian mtDNA lineages in India: an insight into the spread of the Dravidian language and the origins of the caste system. *Hum Genet, 134*. doi:10.1007/s00439-015-1547-4

Pankratov V, Litvinov S, Kassian A, Shulhin D, Tchebotarev L, Yunusbayev B, et al. (2016). East Eurasian ancestry in the middle of Europe: genetic footprints of Steppe nomads in the genomes of Belarusian Lipka Tatars. *Sci Rep, 6*, 30197. doi:10.1038/srep30197

Patnaik R, & Chauhan P. (2009). India at the cross-roads of human evolution. *J Biosci, 34*(5), 729.

Perez-Morga D, Vanhollebeke B, Paturiaux-Hanocq F, Nolan DP, Lins L, Homble F, et al. (2005). Apolipoprotein L-I promotes trypanosome lysis by forming pores in lysosomal membranes. *Science, 309*(5733), 469-472. doi:10.1126/science.1114566

Petraglia MD. (2011). Archaeology: Trailblazers across Arabia. *Nature, 470*(7332), 50-51. doi:10.1038/470050a

Petrie CA. (2015). Case Study: Mehrgarh. In C. A. Barker & C. Goucher (Eds.), *The Cambridge World History: Volume 2, A World with Agriculture, 12,000 BCE–500 CE*. Cambridge: Cambridge University Press.

Poznik GD, Xue Y, Mendez FL, Willems TF, Massaia A, & Wilson Sayres MA. (2016). Punctuated bursts in human male demography inferred from 1,244 worldwide Y-chromosome sequences. *Nat Genet, 48*. doi:10.1038/ng.3559

Pradhan I, Manandhar KD and Shrestha TR (2014). Genetic affinity of Manandhar population of Kathmandu valley. M.Sc. thesis submitted in Central Department of Biotechnology, Tribhuvan University, Nepal.

Pranjal. (2015). *Travel and Tourism of Nepal - Pranjal - Google Books*: AuthorHouse.

Prufer K, Racimo F, Patterson N, Jay F, Sankararaman S, Sawyer S, et al. (2014). The complete genome sequence of a Neanderthal from the Altai Mountains. *Nature, 505*(7481), 43-49. doi:10.1038/nature12886

Prehistoric Colonization and Demographic History of Modern Humans on the Tibetan Plateau (John Wiley & Sons, Ltd 2014).

Qi X, Cui C, Peng Y, Zhang X, Yang Z, Zhong H, et al. (2013). Genetic evidence of paleolithic colonization and neolithic expansion of modern humans on the Tibetan plateau. *Molecular biology and evolution, 30*(8), 1761-1778. doi:10.1093/molbev/mst093

Qi X, Cui C, Peng Y, Zhang X, Yang Z, Zhong H, et al. (2013). Genetic evidence of paleolithic colonization and neolithic expansion of modern humans on the tibetan plateau. *Mol Biol Evol, 30*(8), 1761-1778. doi:10.1093/molbev/mst093

Quintana-Murci L, Chaix R, Wells RS, Behar DM, Sayar H, Scozzari R, et al. (2004). Where west meets east: the complex mtDNA landscape of the southwest and Central Asian corridor. *Am J Hum Genet, 74*(5), 827-845. doi:10.1086/383236

Raghavan M, Skoglund P, Graf KE, Metspalu M, Albrechtsen A, Moltke I, et al. (2014). Upper Palaeolithic Siberian genome reveals dual ancestry of Native Americans. *Nature, 505*(7481), 87-91. doi:10.1038/nature12736

Raghavan M, Steinrücken M, Harris K, Schiffels S, Rasmussen S, DeGiorgio M, et al. (2015). Genomic evidence for the Pleistocene and recent population history of Native Americans. *Science (New York, N.Y.), 349*(6250), aab3884-aab3884. doi:10.1126/science.aab3884

Rajkumar R, Banerjee J, Gunturi HB, Trivedi R, & Kashyap VK. (2005). Phylogeny and antiquity of M macrohaplogroup inferred from complete mt DNA sequence of Indian specific lineages. *BMC Evol Biol, 5*, 26. doi:10.1186/1471-2148-5-26

Rasmussen M, Guo X, Wang Y, Lohmueller KE, Rasmussen S, Albrechtsen A, et al. (2011). An Aboriginal Australian genome reveals separate human dispersals into Asia. *Science, 334*(6052), 94-98. doi:10.1126/science.1211177

Reddy BM, Langstieh BT, Kumar V, Nagaraja T, Reddy AN, Meka A, et al. (2007). Austro-Asiatic tribes of Northeast India provide hitherto missing genetic link between South and Southeast Asia. *PLoS One, 2*(11), e1141. doi:10.1371/journal.pone.0001141

Regmi DR. (1960). Ancient and Medieval Nepal. *vol I*, 5-20.

Regmi DR. (1965). *Medieval Nepal* (Vol. 1): Firma KL Mukhopadhyay.

Rosenberg NA, Pritchard JK, Weber JL, Cann HM, Kidd KK, Zhivotovsky LA, et al. (2002). Genetic structure of human populations. *Science, 298*(5602), 2381-2385. doi:10.1126/science.1078311

Sahakyan H, Hooshiar Kashani B, Tamang R, Kushniarevich A, Francis A, Costa MD, et al. (2017). Origin and spread of human mitochondrial DNA haplogroup U7. *Sci Rep, 7*, 46044. doi:10.1038/srep46044

Saillard J, Forster P, Lynnerup N, Bandelt HJ, & Norby S. (2000). mtDNA variation among Greenland Eskimos: the edge of the Beringian expansion. *Am J Hum Genet, 67*(3), 718-726. doi:10.1086/303038

Sankararaman S, Mallick S, Dannemann M, Prufer K, Kelso J, Paabo S, et al. (2014). The genomic landscape of Neanderthal ancestry in present-day humans. *Nature, 507*(7492), 354-357. doi:10.1038/nature12961

Sankararaman S, Patterson N, Li H, Paabo S, & Reich D. (2012). The date of interbreeding between Neandertals and modern humans. *PLoS Genet, 8*(10), e1002947. doi:10.1371/journal.pgen.1002947

Sarkar J, & Ghosh GC. (2003). *Populations of the SAARC countries : bio-cultural perspectives*: Sterling Publishers.

Schuenemann VJ, Peltzer A, Welte B, van Pelt WP, Molak M, Wang CC, et al. (2017). Ancient Egyptian mummy genomes suggest an increase of Sub-Saharan African ancestry in post-Roman periods. *Nat Commun, 8*, 15694. doi:10.1038/ncomms15694

Seguin-Orlando A, Korneliussen TS, Sikora M, Malaspinas AS, Manica A, Moltke I, et al. (2014). Paleogenomics. Genomic structure in Europeans dating back at least 36,200 years. *Science, 346*(6213), 1113-1118. doi:10.1126/science.aaa0114

Shakya MB. (2000). A Historical Analysis of Monasticism in Newar.

Silva M, Oliveira M, Vieira D, Brandao A, Rito T, Pereira JB, et al. (2017). A genetic chronology for the Indian Subcontinent points to heavily sex-biased dispersals. *BMC Evol Biol, 17*(1), 88. doi:10.1186/s12862-017-0936-9

Simonson TS, Yang Y, Huff CD, Yun H, Qin G, Witherspoon DJ, et al. (2010). Genetic evidence for high-altitude adaptation in Tibet. *Science, 329*(5987), 72-75. doi:10.1126/science.1189406

Singh KS. (1994). *People of India : Pondicherry*: Affiliated East-West Press.

Soares P, Ermini L, Thomson N, Mormina M, Rito T, Rohl A, et al. (2009). Correcting for purifying selection: an improved human mitochondrial molecular clock. *Am J Hum Genet, 84*(6), 740-759. doi:10.1016/j.ajhg.2009.05.001

Stewart JB, & Chinnery PF. (2015). The dynamics of mitochondrial DNA heteroplasmy: implications for human health and disease. *Nature reviews. Genetics, 16*(9), 530-542. doi:10.1038/nrg3966

Tajima F. (1989). Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics, 123*(3), 585-595.

Thangaraj K, Chaubey G, Kivisild T, Reddy AG, Singh VK, Rasalkar AA, et al. (2005). Reconstructing the origin of Andaman Islanders. *Science, 308*(5724), 996. doi:10.1126/science.1109987

Thangaraj K, Chaubey G, Singh VK, Vanniarajan A, Thanseem I, Reddy AG, et al. (2006). In situ origin of deep rooting lineages of mitochondrial Macrohaplogroup 'M' in India. *BMC Genomics, 7*, 151. doi:10.1186/1471-2164-7-151

ThermoScientific DreamTaq Green PCR Master Mix (2X). (2016).

Tishkoff SA, Reed FA, Friedlaender FR, Ehret C, Ranciaro A, Froment A, et al. (2009). The genetic structure and history of Africans and African Americans. *Science, 324*(5930), 1035-1044. doi:10.1126/science.1172257

Tuladhar PS. (2000). Nepal Bhasa Sahityaya Itihas: The History of Nepalbhasa

Literature. Kathmandu. *Nepal Bhasa Academy ISBN 99933-56- 00-X*, 14 & 306.

Udpa N, Ronen R, Zhou D, Liang J, Stobdan T, Appenzeller O, et al. (2014). Whole genome sequencing of Ethiopian highlanders reveals conserved hypoxia tolerance genes. *Genome biology, 15*(2), R36.

Underhill PA, Myres NM, Rootsi S, Metspalu M, Zhivotovsky LA, & King RJ. (2010). Separating the post-Glacial coancestry of European and Asian Y chromosomes within haplogroup R1a. *Eur J Hum Genet, 18*. doi:10.1038/ejhg.2009.194

van Oven M, & Kayser M. (2009). Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. *Hum Mutat, 30*(2), E386-394. doi:10.1002/humu.20921

Vernot B, & Akey JM. (2015). Complex history of admixture between modern humans and Neandertals. *Am J Hum Genet, 96*(3), 448-453. doi:10.1016/j.ajhg.2015.01.006

Vitti JJ, Grossman SR, & Sabeti PC. (2013). Detecting Natural Selection in Genomic Data. *Annual Review of Genetics, 47*(1), 97-120. doi:10.1146/annurev-genet-111212-133526

Wall JD, Yang MA, Jay F, Kim SK, Durand EY, Stevison LS, et al. (2013). Higher levels of neanderthal ancestry in East Asians than in Europeans. *Genetics, 194*(1), 199-209. doi:10.1534/genetics.112.148213

Wang HW, Li YC, Sun F, Zhao M, Mitra B, Chaudhuri TK, et al. (2012). Revisiting the role of the Himalayas in peopling Nepal: insights from mitochondrial genomes. *J Hum Genet, 57*(4), 228-234. doi:10.1038/jhg.2012.8

Watters D. (1975). Siberian shamanistic traditions among the Kham Magars of Nepal, Contributions to Nepalese Studies. Centre for Nepal and Asian Studies (CNAS), Tribhuvan University, Kathmandu.

Whelpton J. (2005). *A history of Nepal*: Cambridge University Press.

White TD, Asfaw B, DeGusta D, Gilbert H, Richards GD, Suwa G, et al. (2003). Pleistocene Homo sapiens from Middle Awash, Ethiopia. *Nature, 423*(6941), 742-747. doi:10.1038/nature01669

Wilson JF, Weiss DA, Richards M, Thomas MG, Bradman N, & Goldstein DB. (2001). Genetic evidence for different male and female roles during cultural transitions in the British Isles. *Proc Natl Acad Sci U S A, 98*. doi:10.1073/pnas.071036898

Yi X, Liang Y, Huerta-Sanchez E, Jin X, Cuo ZX, Pool JE, et al. (2010). Sequencing of 50 human exomes reveals adaptation to high altitude. *Science, 329*(5987), 75-78. doi:10.1126/science.1190371

Zhang X, Qi X, Yang Z, Serey B, Sovannary T, Bunnath L, et al. (2013). Analysis of mitochondrial genome diversity identifies new and ancient maternal lineages in Cambodian aborigines. *Nat Commun, 4*, 2599. doi:10.1038/ncomms3599