3-19-2012

Genetic Diversity in the Himalayan Populations of Nepal and Tibet

Tenzin Gayden
Florida International University, tgayd001@fiu.edu

DOI: 10.25148/etd.FI12042312

FLORIDA INTERNATIONAL UNIVERSITY

Miami, Florida

GENETIC DIVERSITY IN THE HIMALAYAN POPULATIONS OF NEPAL AND TIBET

A dissertation submitted in partial fulfillment of the

requirements for the degree of

DOCTOR OF PHILOSOPHY

in

BIOLOGY

by

Tenzin Gayden

2012

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

This dissertation, written by Tenzin Gayden, and entitled Genetic Diversity in the Himalayan Populations of Nepal and Tibet, having been approved in respect to style and intellectual content, is referred to you for judgment.

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

Timothy M. Collins

Bruce R. McCord

DeEtta K. Mills

George T. Duncan

Rene J. Herrera, Major Professor

Date of Defense: March 19, 2012

The dissertation of Tenzin Gayden is approved.

Dean Kenneth G. Furton
College of Arts and Sciences

Dean Lakshmi N. Reddi
University Graduate School

Florida International University, 2012

DEDICATION

Dedicated to my Grandmother, Mrs. Choekyi Dolma (1927-2012)

ACKNOWLEDGMENTS

I would like to express my sincere gratitude to my advisor, Dr. Rene J. Herrera, and the members of my committee, Dr. Timothy Collins, Dr. Bruce McCord, Dr. DeEtta Mills and Dr. George Duncan, for their constant support, patience and guidance throughout the course of my Ph.D. work. Dr. Maria Regueiro deserves my warmest thanks for her unconditional help, encouragement and advice. I want to take this opportunity to also thank all the members of Dr. Herrera's laboratory for their cooperation and support.

I gratefully acknowledge the financial support I received from the University Graduate School of Florida International University: Dissertation Evidence Aquisition (Spring and Summer 2010) and Dissertation Year (2011-2012) Fellowships, which helped me to finish my doctoral research in a timely fashion. My academic journey would not have been possible without the financial help and encouragement of Mr. Maurice Gallet and Mrs. Patricia Gallet since my high school. Finally, I am indebted to my family for their patience, understanding, love and unwavering support throughout my life.

The following chapters of the dissertation have already been published and are reprinted here with the kind permission of Nature Publishing Group (Chapter 2) and Springer Science and Business Media (Chapters 3 and 4) via the copyright clearance center.

Chapter 2. "*Genetic insights into the origins of Tibeto-Burman populations in the Himalayas*", Journal of Human Genetics 54:216-223 (2009) by Gayden T, Mirabal S, Cadenas AM, Lacau H, Simms TM, Morlote D, Chennakrishnaiah S and Herrera RJ.

Chapter 3. "*Y-STR diversity in the Himalayas*", International Journal of Legal Medicine 125:367-375 (2011) by Gayden T, Chennakrishnaiah S, La Salvia J, Jimenez S, Regueiro M, Maloney T, Persad PJ, Bukhari A, Perez A, Stojkovic O and Herrera RJ.

Chapter 4. "*Y-chromosomal microsatellite diversity in three culturally defined regions of the historical Tibet*", Forensic Science International: Genetics (2012) (In Press) by Gayden T, Bukhari A, Chennakrishnaiah S, Stojkovic O and Herrera RJ.

ABSTRACT OF THE DISSERTATION

GENETIC DIVERSITY IN THE HIMALAYAN POPULATIONS OF NEPAL AND TIBET

by

Tenzin Gayden

Florida International University, 2012

Miami, Florida

Professor Rene J. Herrera, Major Professor

The Himalayan Mountain range encompasses an unparalleled landscape featuring some of the planet's highest peaks, including Mount Everest. In the heart of this massive orographic barrier lies Nepal, sandwiched in the historically geostrategic position between the Tibetan plateau to the north and India in the south. Until recently, Nepalese and Tibetan populations remained poorly characterized genetically, partly because of their inaccessible geographical locations. In the present study, the genetic diversity of these two Himalayan populations is evaluated using different marker systems, including mitochondrial DNA (mtDNA) and Short Tandem Repeats (STRs) in the autosomes as well as on the Y-chromosome (Y-STR). While autosomal STRs are distributed throughout the genome and are biparentally inherited, the Y-chromosome and mtDNA are haploid markers and provide the paternal and maternal histories of the population, respectively.

Fifteen autosomal STR loci were typed in 341 unrelated individuals from three Nepalese populations (188), namely Tamang (45), Newar (66) and Kathmandu (77), and a general collection from Tibet (153). These samples were also sequenced for the mtDNA

control region and all of them were subsequently assigned to 75 different mtDNA haplogroups and sub-haplogroups by screening their diagnostic sites in the coding region using Restriction Fragment Length Polymorphism analysis and/or sequencing, thus achieving an unprecedented level of resolution. The results from the autosomal and mtDNA data suggest a Northeast Asian origin for the Himalayan populations, with significant genetic influence from the Indian subcontinent in Kathmandu and Newar, corroborating our previous Y-chromosome study. In contrast, Tibet displays a limited Indian component, suggesting that the Himalayan massif acted as a natural barrier for gene flow from the south. The presence of ancient Indian mtDNA lineages in Nepal implies that the region may have been inhabited by the earliest settlers who initially populated South Asia.

In addition, seventeen Y-STR loci were analyzed in 350 Tibetan males from three culturally defined regions of historical Tibet: Amdo (88), Kham (109) and U-Tsang (153). The results demonstrate that the 17 Y-STR loci studied are highly polymorphic in all the three Tibetan populations examined and hence are useful for forensic cases, paternity testing and population genetic studies.

TABLE OF CONTENTS

LIST OF FIGURES

FIGURE PAGE

Chapter II

Chapter III

Chapter IV

Chapter V

LIST OF TABLES

TABLE PAGE

Chapter II

Chapter III

Chapter IV

Chapter V

I. INTRODUCTION

The Himalayas are home to the world's tallest mountains, including Mount Everest. This mountain range extends from Pakistan in the west to Burma in the east, along the frontiers of northern India, Nepal, Tibet, and Bhutan, thereby presenting a formidable natural barrier between the Tibetan plateau and the Indian subcontinent. The region also constitutes an ethnic and linguistic contact zone as it marks a crossroads of three different religions (Islam, Hinduism and Buddhism) and two language families (Indo-European and Sino-Tibetan), with Sino-Tibetan partitioning into the Chinese and Tibeto-Burman (TB) subfamilies. Speakers of the latter represent the majority of the Himalayan inhabitants, while the Indo-Europeans are restricted to the south of the cordillera.

The Tibetan Plateau has remained relatively isolated throughout history as a result of its high altitude, cold climate and rugged topography. Tibet first appears in modern history during the seventh century AD when several clans were unified by Songtsen Gampo, the first emperor of Tibet. From the time of its unification in the seventh century to its disintegration in the ninth century after the assassination of Emperor Langdarma, the Tibetan empire extended as far east as the Qinghai and Sichuan provinces of China and as far west as Nepal, India, and the Hunza Valley in modern-day Pakistan (Richardson, 1962). Following this period, Tibet was subject to a series of foreign invasions, beginning in 1247 with Mongolia, followed by Nepal (late 18th century), India (mid-19th century), Britain (1904) and China (1959). Prior to the Chinese invasion in 1959, Tibet covered an area of about 2.5 million km^2, encompassing three main provinces: Amdo in the northeast, Kham in the southeast and U-Tsang in west and central

Tibet. In 1965, China created the Tibet Autonomous Region (TAR), reducing Tibet's area to 1.2 million km^2, and assimilating parts of Amdo and Kham into the adjacent Chinese provinces of Qinghai, Gansu, Sichuan and Yunnan (Richardson, 1962).

Nepal's physical and cultural landscape can be divided into three distinct regions (Whelpton, 1990), beginning with the mountainous north, which is located along the southern slopes of the Himalayas and whose people are culturally synonymous to the Buddhists of Tibet. 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. The Terai Arc landscape is an importan bioregion. Between these two extremes lie the intermediate hills and valleys where a majority of the Nepalese population resides, and whose cultural practices incorporate both Buddhist and Hindu teachings (Berreman, 1963). According to historical records, Nepal originally consisted only of Kathmandu Valley, located in the east central hills of present-day Nepal, whose indigenous inhabitants, the Newars, are of postulated Austro-Asiatic, Dravidian, Indo-Mongoloid and Aryan origin (Regmi, 1969). Today, Newars constitute only 5.5% of the total population and speak the Tibeto-Burman language of Newari or Nepal Bhasa, with the country's official language being the Indo-Aryan language of Nepali, which is closely related to Hindi, the official language of India. Presently, Nepal is the only constitutionally Hindu country in the world, with 80.6% of its population practicing this religion.

Our previous study on Y-chromosomal diversity in the Himalayan populations of Nepal and Tibet revealed high frequencies of East Asian specific haplogroup O3 derivatives, consistent with earlier observations among Tibeto-Burman groups (Su et al.,

2000; Cordaux et al., 2004; Sahoo et al. 2006; Sengupta et al., 2006), indicating a shared common ancestry among this language subfamily (Gayden et al., 2007). Furthermore, Nepalese populations, specifically those from Kathmandu and Newar, experienced significant genetic influences from the Indian subcontinent, whereas the Himalayas served as a barrier for gene flow from the south to the Tibetan plateau (Gayden et al., 2007). These findings are consistent with a recent mtDNA study (Thangaraj et al., 2008) which reported reciprocal maternal gene flow between India and Nepal.

Although several reports derived from autosomal (Ying et al., 2006, Yan et al., 2007; Ota et al., 2007; Kang et al., 2007) and Y-chromosomal (Zhang et al., 2006; Zhu et al., 2006; Kang et al., 2008) STR data from Tibet have been published previously, the present work is the first of its kind to perform comprehensive phylogenetic analyses in a statistically significant sample size. In addition, with the exception of the malaria-resistant Tharu people of the Terai region (Thangaraj et al., 2008; Fornarino et al., 2009), limited work has been done to characterize the maternal lineages present in other Nepalese groups, especially the indigenous Newar population. Therefore, the present study was conducted to provide a comprehensive view of the genetic diversity of the Himalayan populations of Nepal and Tibet by employing three different marker systems, namely Autosomal Short Tandem Repeat (Autosomal STR), Y-chromosomal microsatellite (Y-STR) and mitochondrial DNA (mtDNA).

In Chapter II, the genetic variation in three populations from Nepal (n=188), including Tamang (n=45), Newar (n=66) and Kathmandu (n=77) and a general population of Tibet (n=153) was examined using fifteen autosomal STR loci included in the AmpF*l*STR® Identifiler kit (Applied Biosystems, Foster City, CA). The data from

these four Himalayan groups were compared to geographically targeted worldwide populations as well as Tibeto-Burman speaking groups from Northeast India. The results suggest a Northeast Asian origin for the Himalayan populations with subsequent gene flow from South Asia into the Kathmandu valley and the Newar population, supporting a previous Y-chromosome study (Gayden et al., 2007). In contrast, Tamang and Tibet exhibit limited genetic contributions from South Asia, possibly because of the orographic obstacle presented by the Himalayan massif. The Tibeto-Burman groups from Northeast India are genetically distinct compared to their counterparts from the Himalayas probably resulting from prolonged isolation and/or founder effects (Gayden et al., 2009).

Chapter III focuses on the forensic and population genetic applications of seventeen Y-chromosomal short tandem repeat (Y-STR) loci in three Nepalese and one Tibetan populations. The latter displays the highest haplotype diversity (0.9990) followed by Kathmandu (0.9977), Newar (0.9570) and Tamang (0.9545). The overall haplotype diversity for the Himalayan populations at 17 Y-STR loci was 0.9973 and the corresponding values for the extended (11-loci) and minimal (9-loci) haplotypes were 0.9955 and 0.9942, respectively. No Y-STR profiles are shared across the four Himalayan collections at the 17-, 11- and 9-loci resolutions considered, indicating a lack of recent gene flow among them. A median-joining network of haplogroup O3a5c-M134 based on 15 Y-STR loci from the four Himalayan populations suggests either a male founder effect in Tamang, possibly from Tibet, or a recent bottleneck following their arrival south of the Himalayas from Tibet leading to their highly reduced Y-SNP and Y-STR diversity. The genetic uniqueness of the four Himalayan populations examined in

this study merits the creation of separate databases for individual identification, parentage analysis and population genetic studies (Gayden et al., 2011a).

In Chapter IV, seventeen Y-STR loci were analyzed in 350 Tibetan males to establish an anthropologically well-characterized forensic databases for three culturally defined regions of historical Tibet: Amdo (88), Kham (109) and U-Tsang (153). A total of 299 haplotypes were identified, 272 (90.9%) of which were unique. Only one Y-STR profile is shared across the three Tibetan groups and, incidentally, is also the most frequent haplotype (4.0%), represented by two, five and seven individuals from U-Tsang, Kham and Amdo, respectively. The overall haplotype diversity for the three Tibetan populations at 17 Y-STR loci was 0.9978 and the corresponding values for the extended (11-loci) and minimal (9-loci) haplotypes were 0.9935 and 0.9909, respectively. Both neighbor-joining and *Rst* pairwise analyses suggest a close genetic relationship between the Amdo and Kham populations, while U-Tsang is genetically distinct from the aforementioned groups (Gayden et al., 2011b).

The mtDNA genome variation in three Nepalese collections (Tamang, Newar and Kathmandu) and a general Tibetan population is presented in Chapter V. In this study, the mtDNA hypervariable regions I and II (HVRI and HVRII) of 344 Himalayan samples were sequenced in both the forward and reverse directions. The resulting sequences were aligned and compared to the revised Cambridge Reference Sequence (rCRS) (Andrews et al., 1999). The haplogroup affiliation of each mtDNA molecule inferred on the basis of the control region motif was subsequently confirmed by a hierarchical RFLP screening of diagnostic sites in the coding region according to the most updated mtDNA phylogeny (van Oven and Kayser 2009, Build 13). The results reveal a predominantly East Asian

specific component in Tibet and Tamang, while the populations from Newar and Kathmandu are both characterized by a combination of East and South Asian lineages. Also, in contrast to Tibet, all three Nepalese groups display significant West Eurasian haplogroups in their maternal gene pools. It is notable that, of the four Himalayan populations examined, Kathmandu shows the highest variation in control-region sequences as well as in haplogroup composition, which is likely attributed to this region's role as capital city of Nepal. Tamang and Newar, on the other hand, exhibit relatively reduced levels of haplotype and haplogroup diversity, possibly the result of genetic drift and/or smaller sample sizes. The presence of ancient Indian maternal lineages such as M2, R5 and U2 (coalescence ages >50 KYA), although limited, in Nepal suggests that this region may have been inhabited by the earliest settlers during the initial peopling of South Asia (Metspalu et al., 2004). This chapter is written according to the format required by the European Journal of Human Genetics.

REFERENCES

Andrews RM, Kubacka I, Chinnery PF, Lightowlers RN, Turnbull DM, Howell N (1999) Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. Nat Genet. 23:147.

Applied, B (2001) AmpFlSTR identifiler PCR amplification kit user's manual. (Foster City, CA).

Berreman, GD (1963) People and Cultures of the Himalayas. Asian Survey, 3:289-304.

Cordaux R, Weiss G, Saha N, and Stoneking M (2004) The Northeast Indian passageway: A barrier or corridor for human migrations? Mol Biol Evol. 21:1525-1533

Fornarino S, Pala M, Battaglia V, Maranta R, Achilli A, Modiano G, Torroni A, Semino O, Santachiara-Benerecetti SA (2009). Mitochondrial and Y-chromosome diversity of the Tharus (Nepal): A reservoir of genetic variation. BMC Evol Biol. 9:154-170.

Gayden T, Cadenas AM, Regueiro M, Singh NB, Zhivotovsky LA, Underhill PA, Cavalli-Sforza LL, Herrera RJ (2007). The Himalayas as a directional barrier to gene flow. Am J Hum Genet. 80:884-894.

Gayden T, Mirabal S, Cadenas AM, Lacau H, Simms TM, Morlote D, Chennakrishnaiah S and Herrera RJ (2009) Genetic insights into the origins of Tibeto-Burman population in the Himalayas. Journal of Human Genetics 54:216-223

Gayden T, Chennakrishnaiah S, La Salvia J, Jimenez S, Regueiro M, Maloney T, Persad PJ, Bukhari A, Perez A, Stojkovic O, Herrera RJ (2011). Y-STR Diversity in the Himalayas. Int J Legal Med. 125:367-375.

Gayden T, Bukhari A, Chennakrishnaiah S, Stojkovic O, Herrera RJ (2011) Y-chromosomal microsatellite diversity in three culturally defined regions of the historical Tibet. Forensic Sci Int Genet. (In Press)

Kang LL, Liu K, Ma Y (2007) Y chromosome STR haplotypes of Tibetan Living Tibet Lassa. Forensic Sci Int 172:79-83

Kang LL, Zhao J, Liu K, Li X (2007). Allele frequencies of 15 STR loci of Tibetan lived in Tibet Lassa. Forensic Sci Int 168:236-240

Metspalu M, Kivisild T, Metspalu E, Parik J, Hudjashov G, Kaldma K, Serk P, Karmin M, Behar DM, Gilbert MT, Endicott P, Mastana S, Papiha SS, Skorecki K, Torroni A, Villems R. 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.

Ota M, Droma Y, Basnyat B, Katsuyama Y, Asamura H, Sakai H, Fukuhsima H (2007) Allele frequencies for 15 STR loci in Tibetan populations from Nepal. Forensic Sci Int 169:234-238

Regmi, D. (1969). Ancient Nepal. Nepal: Firma K. L. Mukhopadhyay.

Richardson H (1962) A short history of Tibet, E.P. Dutton & Co Inc., New York.

Sahoo S, Singh A, Himabindu G, Banerjee J, Sitalaximi T, Gaikwad S, Trivedi R, Endicott P, Kivisild T, Metspalu M, Villems R, Kashyap VK (2006) A prehistory of Indian Y chromosomes: evaluating demic diffusion scenarios. Proc Natl Acad Sci U S A. 103:843-848.

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

Su B, Xiao C, Deka R, Seielstad T, Kangwanpong D, Xiao J, Lu D, Underhill P, Cavalli-Sforza L, Chakraborty R, Jin Li (2000) Y chromosome haplotypes reveal prehistorical migrations to the Himalayas. Hum Genet. 107:582-590.

Thangaraj K, Chaubey G, Kivisild T, Rani DS, Singh VK, Ismail T, Carvalho-Silva D, Metspalu M, Bhaskar LVKS, Reddy AG, Chandra S, Pande V, Prathap Naidu B, Adarsh N, Verma A, Jyothi IA, Mallick CB, Shrivastava N, Devasena R, Kumari B, Singh AK, Dwivedi SK, Singh S, Rao G, Gupta P, Sonvane V, Kumari K, Basha A, Bhargavi KR, Lalremruata A, Gupta AK, Kaur G, Reddy KK, Rao AP, Villems R, Tyler-Smith C, Singh L (2008) Maternal footprints of Southeast Asians in north India. Hum Hered. 66:1-9

Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation (2009) van Oven M and Kayser M Hum Mutat. 30:E386-394.

Whelpton J (1990) World Bibliographical Series: Nepal. Denver: Clio Press.

Whelpton J (2005). A History of Nepal. Cambridge, United Kingdom: Cambridge University Press.

Yan J, Shen C, Li Y, Yu X, Xiong X, Mu H, Huang Y, Deng Y, Yu J (2007). Genetic analysis of 15 STR loci on Chinese Tibetan in Qinghai Province. Forensic Sci Int 169: e3-6.

Ying B, Liu T, Chen G, Zhao Z, Fan H, Wen F (2006).Genetic profile of two Tibetan populations from China by analysis of 15 STR loci. Hum Biol. 78:757-770.

Zhang Q, Yan J, Tang H, Jiao Z, Liu Y (2006) Genetic polymorphisms of 17 Y-STRs haplotypes in Tibetan ethnic minority group of China. Leg Med 8:300-305.

Zhu B, Liu S, Ci D, Huang J, Wang Y, Chen L, Zhu J, Xu Y, Zhao Q, Li S, Liu Y (2006) Population genetics for Y-chromosomal STRs haplotypes of Chinese Tibetan ethnic minority group in Tibet. Forensic Sci Int 161:78-83.

II. GENETIC INSIGHTS INTO THE ORIGINS OF TIBETO-BURMAN POPULATIONS IN THE HIMALAYAS

A. INTRODUCTION

The Himalayan mountain range encompasses an unparalleled landscape featuring some of the planet's highest peaks providing a formidable barrier separating the Tibetan plateau from the Indian subcontinent. The region also constitutes an ethnic and linguistic contact zone as it marks a crossroads of three different religions (Islam, Hinduism and Buddhism) and two language families (Indo-European and Sino-Tibetan), with Sino-Tibetan partitioning into the Chinese and Tibeto-Burman (TB) subfamilies. Speakers of the latter represent the majority of the Himalayan inhabitants while the Indo-Europeans are restricted to the south of the cordillera.

The geo-political partitioning of TB speaking groups includes Tibet, Nepal, Northeast India, Bhutan, Burma and parts of Southeast Asia[2]. Two opposing views on the origin of TB populations have been proposed. Both models invoke demic diffusion for the dispersal of the language subfamily. Su et al.[3, 4] postulate the upper and middle Yellow River basin as the ancestral source of TB people. In contrast, van Driem[5, 6] argues for the Yangtse River of Sichuan as the homeland of TB speakers, with subsequent northward migrations to the fertile plains of the Yellow River valley where Neolithic civilizations flourished.

Our recent study on Y-chromosomal diversity in the Himalayan populations of Nepal and Tibet revealed high frequencies of East Asian specific haplogroup O3a5-M134, congruent with earlier reports on TB groups[4, 7, 8, 9], suggesting a common ancestry for this language subfamily[1]. This investigation also revealed elevated levels of South

Asian derived Y-haplogroup R1a1-M198 in the Nepalese populations of Kathmandu and Newar, indicating significant genetic influences from the Indian subcontinent, whereas the Himalayas served as a barrier to gene flow from the south into the Tibetan plateau[1]. The above findings are consistent with a recent mtDNA study[10] which reported reciprocal maternal gene flow between India and Nepal.

On the other hand, previous Y-chromosome studies characterizing the Tibetan people[4, 11, 12] argue for Central Asian genetic contributions to account for the presence of Asian specific YAP (Y *Alu* polymorphism) chromosomes (D-M174) in the plateau. Haplogroup D-M174 lineages are found at appreciable frequencies in the Andaman Islands of the Indian Ocean[13] but only minimally in all East Asian populations with the exception of Japan[1, 4, 11, 12, 14, 15]. Distinctive D-M174 subclades are observed in the Andaman Islands (D-M174*) and Japan (D2-M55) while Tibetans exhibit D1-M15 and D3-P47 lineages.

In addition to the demic diffusion scenarios portrayed by linguistic and genetic studies[4, 5, 6], Aldenderfer and Yinong[16] have proposed an acculturation mechanism to explain the spread of Neolithic civilization throughout northwestern Tibet. The authors deem it unlikely that agricultural migrants from relatively low elevations in the eastern plateau acclimated to the alpine and desert steppes of the Changtang (northwestern Tibet) considering the physiological stresses imposed by the high altitudes. As a result, they find it more plausible for cultural transfer rather than migration to be the mechanism leading to the nomadic pastoralism currently practiced in the region.

Northeast Indian populations have been extensively studied because of their strategic location as well as their socio-cultural and linguistic diversity[9, 17, 18, 19, 20, 21, 22].

The region is bordered by the Himalayas to the north and the Bay of Bengal in the south, and connects the Indian subcontinent to East Asia. In a recent autosomal microsatellite study, the genetic affinities present among TB groups from Northeast India[21] have been ascribed to geographic contiguity rather than linguistic affiliation. The same group suggested that, given its geographical proximity, Burma most likely provided the source population for the Manipur and Tripura territories, whereas those from the northern region (Arunachal Pradesh, Mizoram and Sikkim) have been populated primarily by Tibetans[21]. Unfortunately, the study did not incorporate both probable parental source populations to assess their phylogenetic relationships to the Northeast Indian groups. Previous studies on Bhutanese and Nepalese Y-STR profiles revealed isolation and drift among these Himalayan collections, with a pronounced effect of drift observed in Bhutan. Nevertheless, haplotype sharing indicates possible gene flow between Bhutan and Nepal or from a common source population[23, 24].

Autosomal STRs are selectively neutral, hypervariable markers that vary solely on the basis of mutation and drift[25]. They are distributed throughout the genome and biparentally inherited, as opposed to the uniparental Y-chromosome and mtDNA polymorphisms, providing a more holistic representation of the evolutionary and migratory events that have shaped modern populations. These attributes endow them with the high resolution necessary to assess phylogenetic relationships among closely related human groups[26, 27, 28, 29].

Although several studies related to the Tibetan population have been previously published[30, 31, 32, 33, 34, 35], the current project is the first of its kind to perform comprehensive phylogenetic analyses with a stastistically significant sample size. A set

of 15 autosomal STR loci was employed to characterize three populations from Nepal (Newar, Tamang and Kathmandu) as well as a general populace from Tibet. These collections were compared to previously published geographically and ethnically targeted groups from East Asia, South Asia and Southwest Asia in an attempt to establish genetic relationships among them.

The present study aims to provide information on the genetic origins of Tibetan and Nepalese populations. Our prior results suggest that these Himalayan populations are of Northeast Asian ancestry and that the Kathmandu and Newari collections have been significantly influenced by subsequent gene flow from South Asia. However, the genetic contribution of South Asia to the Tamang and Tibet is negligible, probably because of the physical barrier presented by the Himalayan massif. It is noteworthy that the three Nepalese populations (Kathmandu, Newar and Tamang) display strong genetic differentiation despite sharing close geographic confines, underscoring the need for studies of anthropologically well-defined ethnic groups. A second objective is to assess the phylogenetic relationships between our Himalayan populations and Northeast Indian TB peoples. Contrary to previously published Y-chromosome data[9], our work indicates that Northeast Indians are genetically distinct from the Himalayan and East Asian collections, a disparity that may be attributable to male founder effects.

B. MATERIALS AND METHODS

Sample collection and DNA isolation

Blood samples were collected with informed consent from 341 unrelated individuals from Tibet (153) and three populations from Nepal (188), namely Tamang

(45), Newar (66) and Kathmandu (77). The ancestry of every person was recorded for a minimum of two generations. With the exception of the people of Kathmandu, who speak an Indo-European language (Nepali), members of the remaining collections are TB speakers. The geographic locations of the populations are shown in Figure 1. Sample collections were performed in accordance with the ethical guidelines put forth by the institutions involved in this study. The DNA was extracted from the blood samples using the standard phenol-chloroform and ethanol precipitation method as described in Antunez de Mayolo et al.[36] and stored at -80°C.

DNA amplification and STR genotyping

The DNA samples were amplified at 15 autosomal STR loci in a multiplex reaction using the AmpF*l*STR® Identifiler kit[37]. The PCR amplification was performed in a GeneAmp PCR System 9600 thermocycler[37] following the cycling conditions described elsewhere[26, 27]. Amplicons were separated by multi-capillary electrophoresis in an ABI Prism 3100 Genetic Analyzer[37] and ABI Genescan 500 LIZ was utilized as an internal size standard.Genescan® 3.7 was employed to determine the fragment sizes and alleles were designated by comparison to an allelic ladder from the manufacturer using Genotyper® 3.7 NT software.

Statistical and phylogenetic analyses

The STR allelic frequencies were calculated utilizing the GENEPOP program in the web software v3.4[38]. The Arlequin software package version 2.000[39] was employed to estimate observed and expected heterozygosities (Ho and He, respectively) as well as to determine gene diversity indices (GDI). Several parameters of population genetics

interest, including power of discrimination (PD), matching probability (MP), polymorphic information content (PIC), power of exclusion (PE) and typical paternity index (TPI) were obtained with the PowerStats program version 1.2[40, 41, 42].

Allelic frequencies of the four collections in the current study as well as the reference populations (Table 1) were utilized to generate Neighbor Joining (NJ) dendrograms based on Nei's genetic distances for two different datasets involving 15 and 13 STR loci, respectively, using the PHYLIP v3.6 program[43]. The loci D2S1338 and D19S433 were excluded from the 13 loci analyses as the data for these loci were not reported for all reference populations. The robustness of the phylogenetic relationships established by the NJ tree was assessed using bootstrap analysis with 1000 replicates. Correspondence Analyses (CA) were also performed with the two sets of 15 and 13 STR loci with the NTSYSpc-2.02i software[44].

Pair-wise comparisons were conducted to ascertain genetic affinity between a given pair of populations with the Carmody program's G test[45] involving 100,000 simulations. The Bonferroni correction was applied to account for the possibility of Type I errors resulting from the multiplicity of the tests performed. The DISPAN program[46] was employed to compute inter-, intra- and total population variance (G_{st}, H_s and H_t, respectively) at the level of 13 STR loci. The populations analyzed were classified according to their geographical distributions as displayed in Table 1. In addition, DISPAN was utilized to calculate the average heterozygosity for each population.

Admixture estimations were generated by a Weighted Least Squares (WLS) method[25, 47] using the Statistical Package for the Social Sciences (SPSS) 14.0 software.

The WLS model calculates admixture proportions using the following equation:

$$p_{ih} = \sum_{j=1}^{J} p_{ij} \cdot \mu_j$$

where

p_{ih} is the frequency of the i^{th} allele in the hybrid population,
p_{ij} represents the frequency of the i^{th} allele in the j^{th} parental group (j = 1, ….., J),
μ_j is the proportionate contribution of the j^{th} parental gene pool to the hybrid population, and $\sum_j \mu_j = 1$ (where, $0 \leq \mu_j \leq 1$).

Parental groups were partitioned according to geographical location as follows: Northeast Asia (Japan, Korea and Shaanxi), Southeast Asia (Chao Shan, Thailand, Malaysia and Philippines), South Asia (Pakistan, Punjab and Bangladesh) and Southwest Asia (Afghanistan, Iran and Iraq). Their contributions to the five Himalayan collections (Tibet, Tamang, Newar, Kathmandu and Bhutan) were examined across the 15 loci.

C. RESULTS

Intra-population diversity

The allelic frequency distributions for Tibet, Tamang, Newar and Kathmandu are listed in Supplementary Tables 1 through 4. Combined Matching Probability (CMP), Combined Power of Exclusion (CPE), Combined Power of Discrimination (CPD) and Average Gene Diversity indices (GDI) are presented in Supplementary Table 5. Loci D7S820 and THO1 in Kathmandu and D13S317 in Tibet were found to depart from Hardy-Weinberg equilibrium (HWE) expectations, however, these deviations were rendered statistically insignificant after applying the Bonferroni correction ($\alpha = 0.05/15 = 0.0033$). The highest GDI in Tibet is found at the FGA locus (0.8792) whereas the maximum values for this parameter are observed at D18S51 (0.8597) and D2S1338

(0.8854 and 0.8823) in Tamang, Newar and Kathmandu, respectively. Intra-population variance (Hs) values are presented in Supplementary Table 6. The Southwest Asians possess the highest intra-population variance (0.7801) followed by the Himalayan (0.7771) and Southeast Asian (0.7767) collections, all of which are higher than the All Population group (0.7739), while the Northeast Asian assemblage exhibits the lowest value (0.7713).

Inter-population diversity: 15 STR loci analyses

Phylogenetic analyses based on allelic frequencies were performed to investigate the genetic relationships between the four Himalayan collections and other geographically targeted populations using the set of 15 STR loci. Four separate aggregates are evident in the CA plot (Fig. 2): both Northeast Asian and Himalayan clusters (except for Kathmandu and Newar) in the upper right quadrant, the Southeast Asian assemblage in the lower right portion and the South Asia/Southwest Asia cluster on the left side of the graph. Kathmandu plots intermediate between the South/Southwest Asian group and the Himalayan and Northeast Asian collections. Newar is an outlier near the perimeter of the upper left quadrant whereas Tibet and Tamang map close to each other within the Himalayan assemblage. In the NJ phylogram (Fig. 4), the Himalayan groups occupy an intermediate position between the East and South/Southwest Asian populations. Tibet, Tamang and Bhutan form a separate clade while Newar branches between Kathmandu and the general population from Nepal.

The G-test results at the 15 STR loci level are presented in Table 2. Only the Kathmandu and Bangladesh pair displayed non-significant genetic differences (G =

186.0376, $P = 0.1139$) from a total of 171 pair-wise population comparisons at $\alpha = 0.05$. However, after applying the Bonferroni correction ($\alpha = 0.05/171 = 0.0002924$) several comparisons yielded additional statistically insignificant co-relationships (Table 2). Pair-wise analyses using the13 STR loci dataset generated similar results to those using 15 loci with the exception of an additional statistically insignificant populations pair (Iran/Punjab) (data not shown).

Admixture proportions were generated for five Himalayan populations, namely Tibet, Tamang, Newar, Kathmandu and Bhutan using Northeast Asia, Southeast Asia, South Asia and Southwest Asia as parental populations (Table 3). When examining the genetic contributions from the parental groups, Northeast Asia arises as the major donor to Tibet (63.4%), Tamang (59.7%) and Newar (44.7%), and also contributes considerable proportions of the Bhutan (41.1%) and Kathmandu (22.3%) gene pools. Bhutan (44.3%), on the other hand is mainly influenced by Southeast Asia while relatively low inputs from the same region were observed in Tamang (29.0%), Tibet (21.1%) and Kathmandu (5.4%). In contrast, Kathmandu (63.3%) and Newar (41.5%) are primarily affected by the South Asian group while the latter impacts the other Himalayan populations only minimally (Tibet 2.7%) or not at all (Tamang, 0.0% and Bhutan, 0.0%). All of the Himalayan collections experienced similar levels of gene flow from Southwest Asia (9-15%).

Inter-population diversity: 13 CODIS Core STR loci analyses

In order to compare our Himalayan populations to TB groups from Northeast India and other Asian populations, allelic frequencies from the 13 CODIS STR loci

shared among 25 collections were employed to perform additional phylogenetic analyses. Three well delineated groupings are apparent in the CA plot (Fig. 3): the Northeast Indian assemblage (except for the Naga) in the lower right quadrant, the Southwest Asians with some South Asian populations in the lower left portion and the Himalayan/Northeast/Southeast Asian aggregate (except for Kathmandu and Newar) in the upper half of the graph. The NJ dendrogram (Fig. 5) supports the information provided by the CA plot with the exception of three Himalayan groups (Tibet, Tamang and Bhutan) which form a sister clade with the Northeast Indian populations. Interestingly, populations within this Himalayan and Northeast Indian cluster belong to the same TB language subfamily.

Inter- and total population variance (Gst and Ht, respectively) are reported in Supplementary Table 6. South Asia displays the highest Gst (0.0222) while also sharing the greatest Ht values (0.7894) with the All-population group. In contrast, Northeast Asia exhibits the lowest Gst and Ht diversities (0.0036 and 0.7740, respectively). The Himalayan populations possess intermediate Gst and Ht values (0.0120 and 0.7811, respectively).

D. DISCUSSION

In the present study, 15 autosomal STR loci were typed in three Nepalese populations (Tamang, Newar and Kathmandu) and in a general collection from Tibet to investigate their genetic ancestry and phylogenetic relationships to other TB communities and worldwide populations. The current work improves on an earlier study[48] by examining ethnic groups from Nepal that are anthropologically well-defined.

Furthermore, this report complements previous Y-chromosome data[1] thereby providing a comprehensive analysis of the genetic diversity in the Himalayas.

On average, the Himalayan populations (Newar, Kathmandu, Tamang and Tibet) possess 126 alleles whereas Northeast, Southeast and the South Asian populations examined in this study average 151, 140 and 137 allelic variants, respectively. The lower genetic diversity of the Himalayan collections is also reflected in their relatively reduced average heterozygosity values (0.7862) when compared to those of Northeast Asia (0.7890), Southeast Asia (0.7907) and South Asia (0.8009). Of the four, Tamang displays the lowest genetic variance reflected in both its average gene diversity index value (0.7632) and average heterozygosity value (0.7702), a characteristic echoed in its limited Y-chromosomal diversity[1].

The inter-population diversity (G_{st}) (Supplementary Table 6) among the Himalayan cluster (0.0120) is considerably higher than in Northeast (0.0036), Southwest (0.0053) and Southeast Asian (0.0068) populations, but is lower than among the South Asian collections (0.0222) and the All-populations group (0.0196). The concurrence of limited heterozygosity and elevated G_{st} value in the Himalayan populations could be results of multiple genetic sources, genetic drift and/or founder effect. Additionally, selective pressure and adaptation to high altitudes, possibly in combination with inbreeding and patrilocality, could also promote homozygosity and genetic differences among the populations. The relatively high G_{st} value observed among the South Asians is also reflected in the disperse partitioning of the Northeast Indian populations in the CA plot (Fig. 3) and may reflect known socio-cultural and genetic barriers. High degree of genetic differentiation resulting from various source populations may account for the

greater variance among Himalayan populations in comparison to the Northeast, Southeast and Southwest Asian assemblages.

It is noteworthy that all three Nepalese groups not only exhibit considerable genetic diversity among themselves but also with respect to the general population of Nepal (Figs. 2 and 3). Kathmandu plots closer to the South Asian groups while still maintaining genetic proximity with the Himalayan and Northeast Asian clusters (Figs. 2 and 3). Pairwise G-test comparisons between Kathmandu and Bangladesh (G = 186.0376, $P = 0.1139$), and Kathmandu and Pakistan (G = 213.5188, $P = 0.0014$) are statistically insignificant, after application of Bonferroni correction, indicating some degree of genetic homogeneity between these South Asian and Nepalese populations. Although Newar segregates from all other populations at the periphery of the upper left quadrant in the 15 loci CA graph (Fig. 2), its location is intermediate between the South Asia and other Himalayan collections in the NJ phylogram (Fig. 4). It is possible that the Newari population (an ethnic minority) has been subject to genetic isolation and drift generating this divergent phylogenetic profile. Conversely, genetic impacts from distinct source populations may have contributed to Newar's unique genetic partitioning. The latter possibility is also reflected in a previous study[19] derived from 11 genetic markers (blood groups, red cell enzyme and serum proteins) in which Newaris were found to cluster with two geographically distant Manipuri groups from Northeast India in the NJ tree rather than with Tamang and Gurung, both of Nepalese descent.

The admixture profiles (Table 3) reveal a substantial proportion of South Asian genes (63.3%) in Kathmandu as compared to Northeast Asia (22.3%), supporting the relationships illustrated by both the G test and CA plot. Newar, on the other hand,

experienced similar contributions from both regions (44.7 vs. 41.5%), possibly accounting for its equidistant positioning to these groups of populations in the NJ tree (Fig. 4). The South Asian influence in the Nepalese populations may be associated with the entry of Aryans from the Indian plains who introduced the Indo-European language (Nepali) in the Kathmandu valley. A recent mtDNA study[10] also reported reciprocal maternal gene flow between North India and Nepal. These results are consistent with the elevated levels of Y-haplogroup R in Newar (62.1%) and Kathmandu (46.7%), which are most likely derived from South Asia, particularly North India, given the geographic vicinity as well as the historical and socio-cultural affinities shared between these two neighboring regions[1]. In contrast, Tamang and Tibet exhibit minimal percentages of Y-haplogroup R (8.8 and 2.5%, respectively), indicating that the Himalayas served as a formidable orographic barrier to gene flow from the south[1]. Findings from the current investigation lend support to the aforementioned statement as admixture results reveal a null contribution from South Asia to both Tamang and Bhutan and only a minor genetic impact onto the Tibetan collection (2.7%). The absence of the South Asian signature in the gene pools of Tamang and Bhutan may be the result of geographic isolation and/or founder effects from another source population(s).

Close genetic ties have been reported between the Tamang and Tibet[1]. It is likely that Tamangs are descendants of Tibetans who migrated south and settled in the southern region of the Himalayan range[1]. The genetic affinity between Tamang and Tibet is also reflected in both CA plots (Figs. 2 and 3) and NJ dendrograms (Figs. 4 and 5). In addition, the Tibetan connection to the Tamang is evident in their shared cultural and religious practices. The partitioning of these two populations with Bhutan and their

proximity to the general collection from Nepal (Figs. 2, 3, 4 and 5) may be associated with Neolithic migrants carrying Y-haplogroup O3a5-M134, an East Asian specific marker, shared among TB populations[1, 3, 4, 9, 49]. The Himalayan populations, with the exception of Newar and Kathmandu, segregate close to the Northeast Asian cluster in agreement with the admixture analyses results (Table 3). Northeast Asia is the major contributor to both Tibet (63.4%) and Tamang (59.7%) while Newar (44.7%) and Bhutan (41.1%) received equivalent percentages, followed by Kathmandu (22.3%). These results corroborate studies indicating a shared common ancestry between Tibet and the Northeast Asian collections of Japan and Korea by a variety of marker systems, including classical[50, 51], autosomal[52], Y-chromosome[1, 12, 53, 54] and mtDNA studies[12, 53, 55, 56].

More than half of the Tibetan males possess the YAP polymorphic *Alu* insertion in their Y-chromosome, which is believed to have originated in Central Asia[1, 4, 11, 14], although its source remains highly debated[53, 57, 58]. In the present study, however, given the lack of representative Central Asian populations due to the paucity of the data available from the region, no clear connections were made between Tibet and its possible Central Asian genetic contributors. Afghanistan is the sole Central Asian collection included in the analyses and appears to make no contributions to any of the Himalayan groups except for a minor influence in Kathmandu (12.9%).

In order to evaluate the genetic relationships between the Himalayan collections and the neighboring TB speaking populations at the regional level, six Northeast Indian TB groups were included in the phylogenetic and statistical analyses performed using the 13 core CODIS STR loci. These Northeast Indian TB groups map distantly from both the Himalayan and East Asian populations in the CA graph (Fig. 3), inconsistent with

previous Y-chromosome and mtDNA studies which report a high degree of genetic homogeneity between Himalayan and Northeast Indian TB groups[3, 4, 9, 59]. The discrepancy observed between Y-chromosome and microsatellite polymorphisms in the Northeast Indian TB groups may be explained by a male founder effect from Northeast Asia and their subsequent genetic isolation for an extended period of time following their arrival[9]. Alternatively, the practice of patrilocal residence among these groups may have introduced the maternal genes from the local populations leading to the microsatellite diversity patterns in these Northeast Indian TB collections that are phylogenetically different from their Y-chromosomal profiles[9].

Altogether, our results suggest a Northeast Asian ancestry for the Himalayan populations with subsequent genetic admixture in Kathmandu and Newar populations from South Asia. However, South Asian influences in Tibet and Tamang are negligible most likely the result of the natural barrier presented by the Himalayas[1]. Tamang, Tibet and Bhutan display close genetic affiliations in all analyses possibly indicating a shared common ancestry. The biparental markers examined in the present study reveal unique genetic profiles for the Northeast Indian TB groups which are distinct from their Himalayan counterparts implying limited gene flow, geographic isolation and/or founder effects. The phylogeny also supports the genetic divergence between Northeast and Southeast Asian collections with a possible southern origin for the East Asian populations.

REFERENCES

1. Gayden, T., Cadenas, A. M., Regueiro, M., Singh, N. B., Zhivotovsky, L. A., Underhill, P. A. et al. The Himalayas as a directional barrier to gene flow. *Am. J. Hum. Genet.* **80**, 884-94 (2007).

2. Gordon, R. G. *Ethnologue: Languages of the World, 15th* edn (SIL International, Dallas, 2005).

3. Su, B., Xiao, J., Underhill, P., Deka, R., Zhang, W., Akey, J. et al. Y-Chromosome evidence for a northward migration of modern humans into Eastern Asia during the last Ice Age. *Am. J. Hum. Genet.* **65**, 1718-24 (1999).

4. Su, B., Xiao, C., Deka, R., Seielstad, M. T., Kangwanpong, D., Xiao, J.et al. Y chromosome haplotypes reveal prehistorical migrations to the Himalayas. *Hum. Genet.* **107**, 582-90 (2000).

5. van Driem, G. L. in *Tibeto-Burman phylogeny and prehistory: languages, material culture, and genes* (eds Bellwood, P. & Renfrew, C.) 233-249 (MacDonald Institute for Archaeological Research, University of Cambridge, United Kingdom, 2002).

6. van Driem, G. L. in *The peopling of East Asia: Putting together archaeology, linguistics and genetics* (eds Sagart L., Blench. R. & Sanches-Mazas, A.) 81-106 (Routledge Curzon, Taylor and Francis group, London, 2005).

7. Sengupta, S., Zhivotovsky, L. A., King, R., Mehdi, S. Q., Edmonds, C. A., Chow, C. E., et al. Polarity and temporality of high-resolution y-chromosome distributions in India identify both indigenous and exogenous expansions and reveal minor genetic influence of Central Asian pastoralists. *Am. J. Hum. Genet.* **78**, 202-21 (2006).

8. Sahoo, S., Singh, A., Himabindu, G., Banerjee, J., Sitalaximi, T., Gaikwad, S. et al. A prehistory of Indian Y chromosomes: evaluating demic diffusion scenarios. *Proc.Natl. Acad. Sci. U.S.A.* **103**, 843-848 (2006).

9. Cordaux, R., Weiss, G., Saha, N. & Stoneking, M. The northeast Indian passageway: a barrier or corridor for human migrations? *Mol. Biol. Evol.* **21**, 1525-33 (2004).

10. Thangaraj, K., Chaubey, G., Kivisild, T., Selvi Rani, D., Singh, V. K., Ismail, T. et al. Maternal footprints of Southeast Asians in North India. *Hum. Hered.* **66**, 1-9 (2008).

11. Qian, Y., Qian, B., Su, B., Yu, J., Ke, Y., Chu, Z., et al. Multiple origins of Tibetan Y chromosomes. *Hum. Genet.* **106**, 453-454 (2000).

12. Hammer, M. F., Karafet, T. M., Park, H., Omoto, K., Harihara, S., Stoneking, M. et al. Dual origins of the Japanese: common ground for hunter-gatherer and farmer Y chromosomes. *J. Hum. Genet.* **51**, 47-58 (2006).

13. Thangaraj, K., Singh, L., Reddy, A. G., Rao, V. R., Sehgal, S. C., Underhill, P. A., et al. Genetic affinities of the Andaman Islanders, a vanishing human population. *Curr. Biol.* **13**, 86-93 (2003).

14. Hammer, M. F., Spurdle, A. B., Karafet, T., Bonner, M. R., Wood, E. T., Novelletto, A., et al. The geographic distribution of human Y chromosome variation. *Genetics* **145**, 787-805 (1997).

15. Ding, Y. C., Wooding, S., Harpending, H. C., Chi, H. C., Li, H. P., Fu, Y. X. et al. Population structure and history in East Asia. *Proc. Natl. Acad. Sci. U.S.A.* **97**, 14003-14006 (2000).

16. Aldenderfer, M. & Yinong, Z. The prehistory of the Tibetan plateau to the seventh century A.D.: perspectives and research from China and the West since 1950. *J. World. Prehist.* **18**, 1-55 (2004).

17. Mastana, S. S., Murry, B., Sachdeva, M. P., Das, K., Young, D., Das, M. K. et al. Genetic variation of 13 STR loci in the four endogamous tribal populations of Eastern India. *Forensic. Sci. Int.* **169**, 266-273 (2007).

18. Maity, B., Nunga, S. C. & Kashyap, V. K.. Genetic polymorphism revealed by 13 tetrameric and 2 pentameric STR loci in four Mongoloid tribal population. *Forensic. Sci. Int.* **132**, 216-222 (2003).

19. Kumar, V., Basu, D. & Reddy, B. M. Genetic heterogeneity in northeastern India: reflection of Tribe-Caste continuum in the genetic structure. *Am. J. Hum. Biol.* **16**, 334-45 (2004).

20. Krithika, S., Trivedi, R., Kashyap, V. K. & Vasulu, T. S. Genetic diversity at 15 microsatellite loci among the Adi Pasi population of Adi tribal cluster in Arunachal Pradesh, India. *Leg. Med. (Tokyo)* **7**, 306-310 (2005).

21. Krithika, S., Trivedi, R., Kashyap, V. K., Bharati, P. & Vasulu, T. S. Antiquity, geographic contiguity and genetic affinity among Tibeto-Burman populations of India: a microsatellite study. *Ann. Hum. Biol.* **33**, 26-42 (2006).

22. Kashyap, V. K., Chattopadhyay, P., Dutta, R. & Vasulu, T. S. Genetic structure and affinity among eight ethnic populations of Eastern India: based on 22 polymorphic DNA loci. *Am. J. Hum. Biol.* **16**, 311-27 (2004).

23. Parkin, E. J., Kraayenbrink, T., van Driem, G. L., Tshering Of Gaselo, K., de Knijff, P. & Jobling, M. A. (2006). 26-Locus Y-STR typing in a Bhutanese population sample. *Forensic Sci Int* **161**, 1-7.

24. Parkin, E. J., Kraayenbrink, T., Opgenort, J. R., van Driem, G. L., Tuladhar, N. M., de Knijff, P. et al. Diversity of 26-locus Y-STR haplotypes in a Nepalese population sample: isolation and drift in the Himalayas. *Forensic. Sci. Int.* **166**, 176-181 (2007).

25. Perez-Miranda, A. M., Alfonso-Sanchez, M. A., Pena, J. A. & Herrera, R. J. Qatari DNA variation at a crossroad of human migrations. *Hum. Hered.* **61**, 67-79 (2006).

26. Shepard, E. M. & Herrera, R. J. Iranian STR variation at the fringes of biogeographical demarcation. *Forensic. Sci. Int.* **158**, 140-148 (2006).

27. Shepard, E. M. & Herrera, R. J. Genetic encapsulation among Near Eastern populations. *J. Hum. Genet.* **51**, 467-476 (2006).

28. Shepard, E. M., Chow, R. A., Suafo'a, E., Addison, D., Perez-Miranda, A. M., Garcia-Bertrand, R. L. et al. Autosomal STR variation in five Austronesian populations. *Hum. Biol.* **77**, 825-851 (2005).

29. Rowold, D. J. & Herrera, R. J. Inferring recent human phylogenies using forensic STR technology. *Forensic. Sci. Int.* **133**, 260-265 (2003).

30. Ying, B., Liu, T., Chen, G., Zhao, Z., Fan, H. & Wen, F. Genetic profile of two Tibetan populations from China by analysis of 15 STR loci. *Hum. Biol.* **78**, 757-770 (2006).

31. Yan, J., Shen, C., Li, Y., Yu, X., Xiong, X., Mu, H.et al. Genetic analysis of 15 STR loci on Chinese Tibetan in Qinghai Province. *Forensic. Sci. Int.* **169**, e3-6 (2007).

32. Ota, M., Droma, Y., Basnyat, B., Katsuyama, Y., Asamura, H., Sakai, H. et al. Allele frequencies for 15 STR loci in Tibetan populations from Nepal. *Forensic. Sci. Int.* **169**, 234-238 (2007)..

33. Kang, L. L., Zhao, J., Liu, K. & Li, X. Allele frequencies of 15 STR loci of Tibetan lived in Tibet Lassa. *Forensic. Sci. Int.l* **168**, 236-240 (2007).

34. Kang, L., Yuan, D., Yang, F., Liu, K. & Za, X. Genetic polymorphisms of 15 STR loci in two Tibetan populations from Tibet Changdu and Naqu, China. *Forensic. Sci. Int.* **169**, 239-43 (2007).

35. Kang, L. & Li, S. Allele frequencies of 15 STR loci of Luoba ethnic group in Tibet (Southwestern China). *Forensic. Sci. Int.* **155**, 219-221 (2005).

36. Antunez-de-Mayolo, G., Antunez-de-Mayolo, A., Antunez-de-Mayolo, P., Papiha, S. S., Hammer, M., Yunis, J. J. et al. Phylogenetics of worldwide human populations as determined by polymorphic Alu insertions. *Electrophoresis* **23**, 3346-3356 (2002).

37. Applied, B. *AmpFlSTR identifiler PCR amplification kit user's manual.* (Foster City, CA, 2001).

38. Raymon, M. & Rousset, F. Genepop (version 1.2): population genetics software for exact tests and ecumenicism. *J. Hered.* **86**, 248-249 (1995).

39. Schneider, S., Kueffer, J. M., Roessli, D. & Excoffier, L. *Arlequin v. 2.000: a software for population genetic analysis.* (Genetics and Biometry Laboratory, University of Geneva, Geneva, 2000).

40. Jones, D. A. Blood samples: probability of discrimination. *J. Forensic. Sci. Soc.* **12**, 355-359 (1972).

41. Brenner, C. H. & Morris, J. W. in *Proceedings for the international symposium on human identification.* (Promega Corporation, *Madison,*1990).

42. Tereba, A. (1999). in *Profiles in DNA*. (Promega Corporation, Madison, 1999).

43. Felsentein, J. *Phylogeny inference package (PHYLIP) Version 3.6a3.* (Department of Genetics, University of Washington, Seattle 2002).

44. Rohlf, F. *NTSYSpc.* (Exter Publishing, Setauket, NY, 2002).

45. Carmody, G. *G-test*. (Carleton University, Ottowa, 1990)

46. Ota, T. *DISPAN: genetic distance and phylogenetic analysis*. (Institute of Molecular Evolutionary Genetics, Pennsylvania State University, Pennsylvania, 1993).

47. Long, J. C., Williams, R. C., McAuley, J. E., Medis, R., Partel, R., Tregellas, W. M., e t al. Genetic variation in Arizona Mexican Americans: estimation and interpretation of admixture proportions. *Am. J. Phys. Anthropol.* **84**, 141-157 (1991).

48. Kraaijenbrink, T., van Driem, G. L., Opgenort, J. R., Tuladhar, N. M. & de Knijff, P. Allele frequency distribution for 21 autosomal STR loci in Nepal. *Forensic. Sci. Int.* **168**, 227-231 (2007).

49. Black, M. L., Wise, C. A., Wang, W. & Bittles, A. H. Combining genetics and population history in the study of ethnic diversity in the People's Republic of China. *Hum. Biol.* **78**, 277-293 (2006).

50. Cavalli-Sforza, L. L., Menozzi, P. & Piazza, A. *The history and geography of human genes*. (Princeton University Press, Princeton, New Jersey 1994).

51. Nei, M. & Roychoudhury, A. K. Evolutionary relationships of human populations on a global scale. *Mol. Biol. Evol.* **10**, 927-943 (1993).

52. Chu, J. Y., Huang, W., Kuang, S. Q., Wang, J. M., Xu, J. J., Chu, Z. T. Genetic relationship of populations in China. *Proc. Natl. Acad. Sci. U.S.A* **95**, 11763-11768 (1998).

53. Tajima, A., Hayami, M., Tokunaga, K., Juji, T., Matsuo, M., Marzuki, S. et al. Genetic origins of the Ainu inferred from combined DNA analyses of maternal and paternal lineages. *J. Hum. Genet.* **49**, 187-193 (2004).

54. Tajima, A., Pan, I. H., Fucharoen, G., Fucharoen, S., Matsuo, M., Tokunaga, K. et al. Three major lineages of Asian Y chromosomes: implications for the peopling of east and southeast Asia. *Hum. Genet.* **110**, 80-88 (2002).

55. Tanaka, M., Cabrera, V. M., Gonzalez, A. M., Larruga, J. M., Takeyasu, T., Fuku, N., et al. Mitochondrial genome variation in eastern Asia and the peopling of Japan. *Genome. Res.* **14**, 1832-1850 (2004).

56. Torroni, A., Miller, J. A., Moore, L. G., Zamudio, S., Zhuang, J., Droma, T. et al. Mitochondrial DNA analysis in Tibet: implications for the origin of the Tibetan population and its adaptation to high altitude. *Am. J. Phys. Anthropol.* **93**, 189-199 (1994).

57. Jin, L. & Su, B. Natives or immigrants: modern human origin in east Asia. *Nat. Rev. Genet.* **1**, 126-133 (2000).

58. Underhill, P. A. in *The Peopling of East Asia: Putting Together Archaeology, Linguistics, and Genetics* (eds. L. Sagart., R. Blench. & A. Sanchez-Mazas.) 301-319 (Routledge Curzon, Taylor and Francis group, London, 2004).

59. Wen, B., Xie, X., Gao, S., Li, H., Shi, H., Song, X. 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 (2004)

Appendix I

Figure 1. Geographic locations of the Himalayan and reference populations examined in this study.

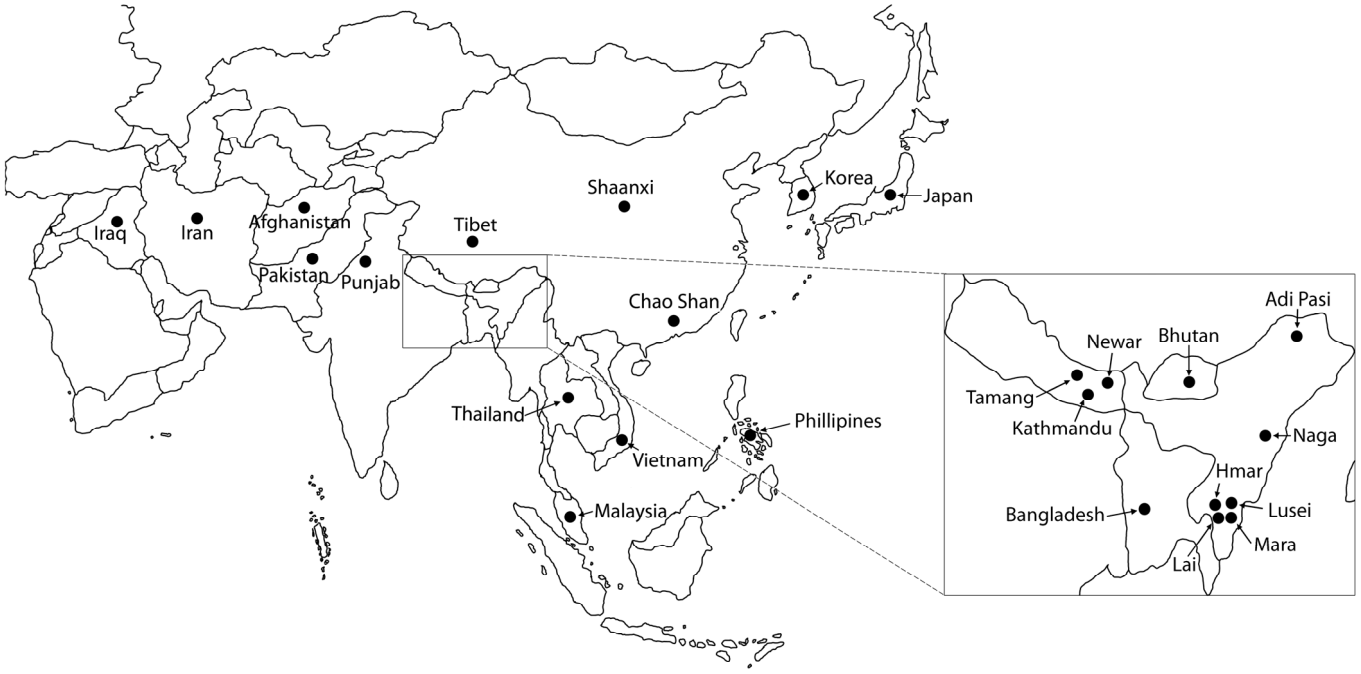

Figure 2. Correspondence Analysis (CA) employing 15 STR loci.

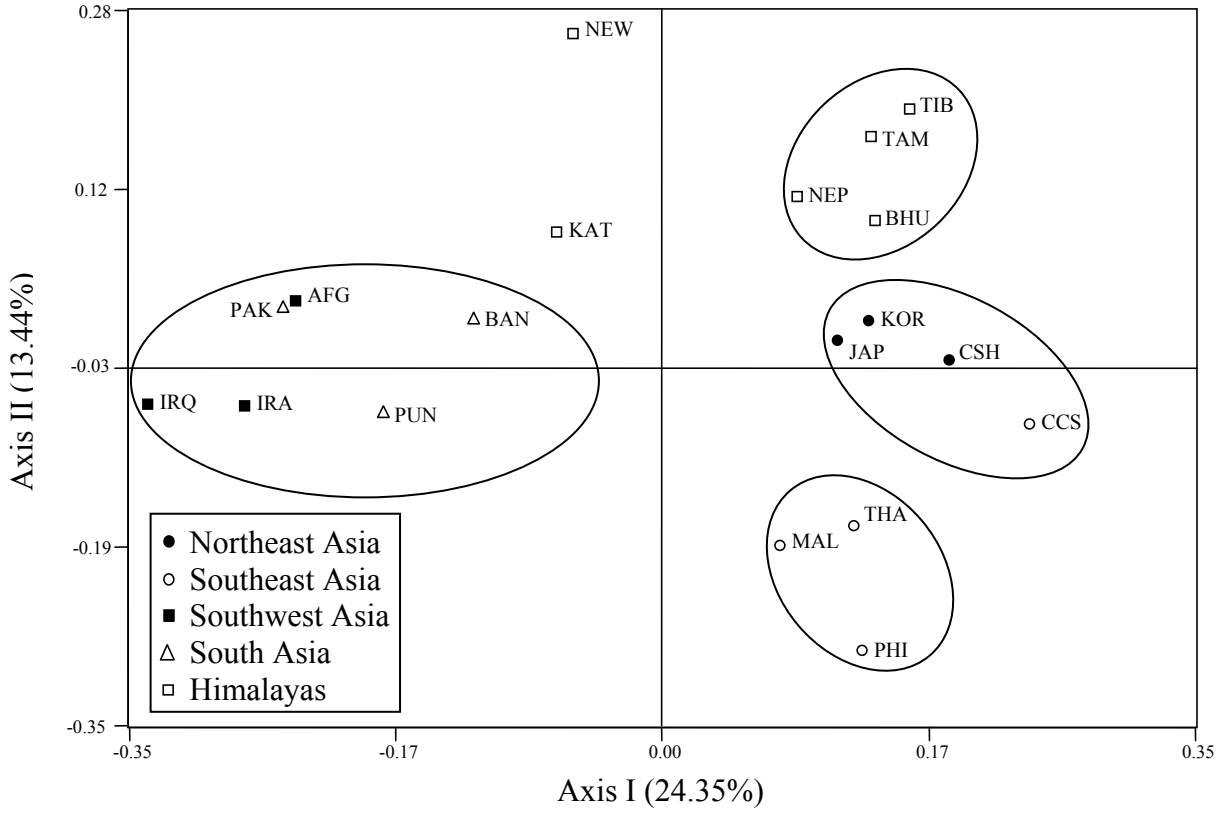

Figure 3. Correspondence Analysis (CA) using 13 STR loci.

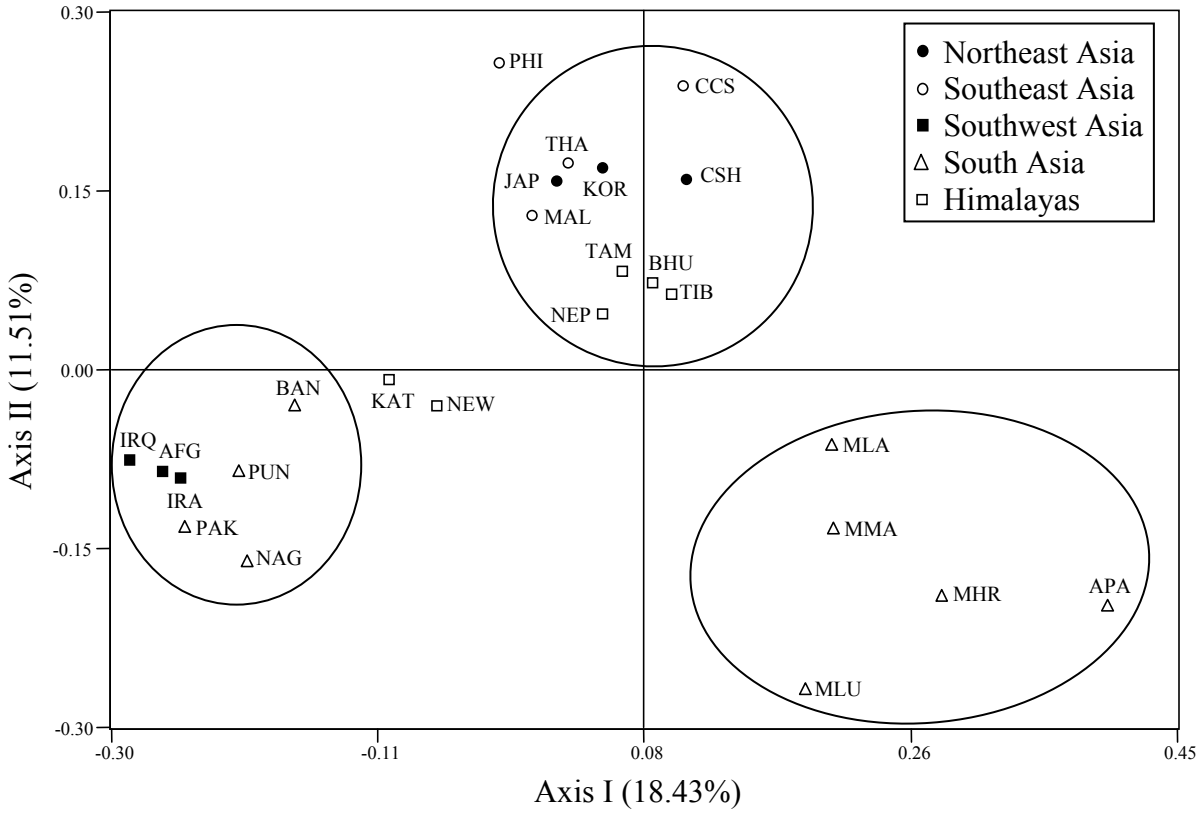

Figure 4. Neighbor-Joining (NJ) tree based on Nei's genetic distances generated using allele frequencies from 15 STR loci. The numbers at the nodes represent bootstrap values estimated from 1000 replications.

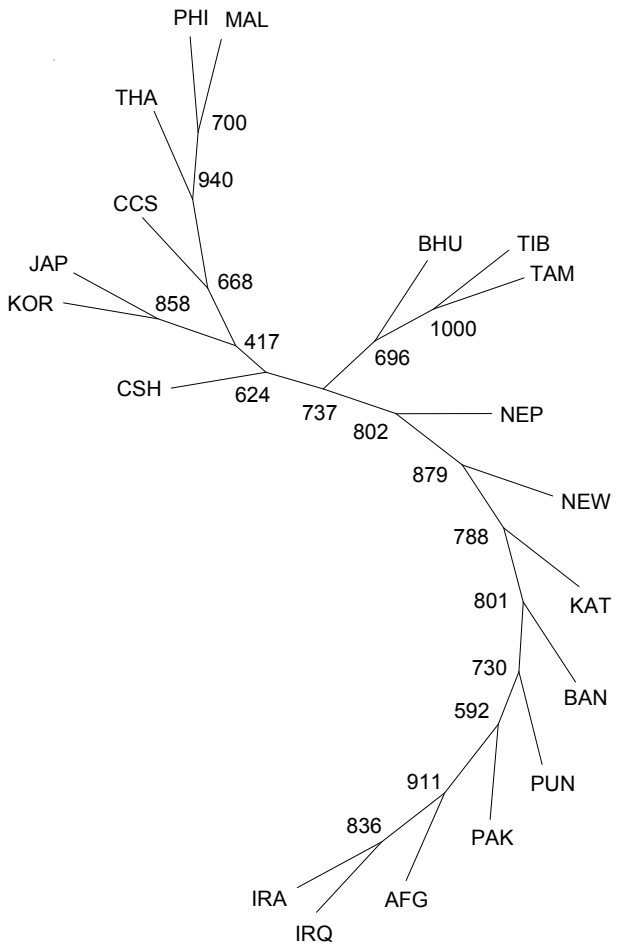

Figure 5. Neighbor-Joining (NJ) tree based on Nei's genetic distances generated using allele frequencies from 13 CODIS core STR loci. The numbers at the nodes represent bootstrap values estimated from 1000 replications.

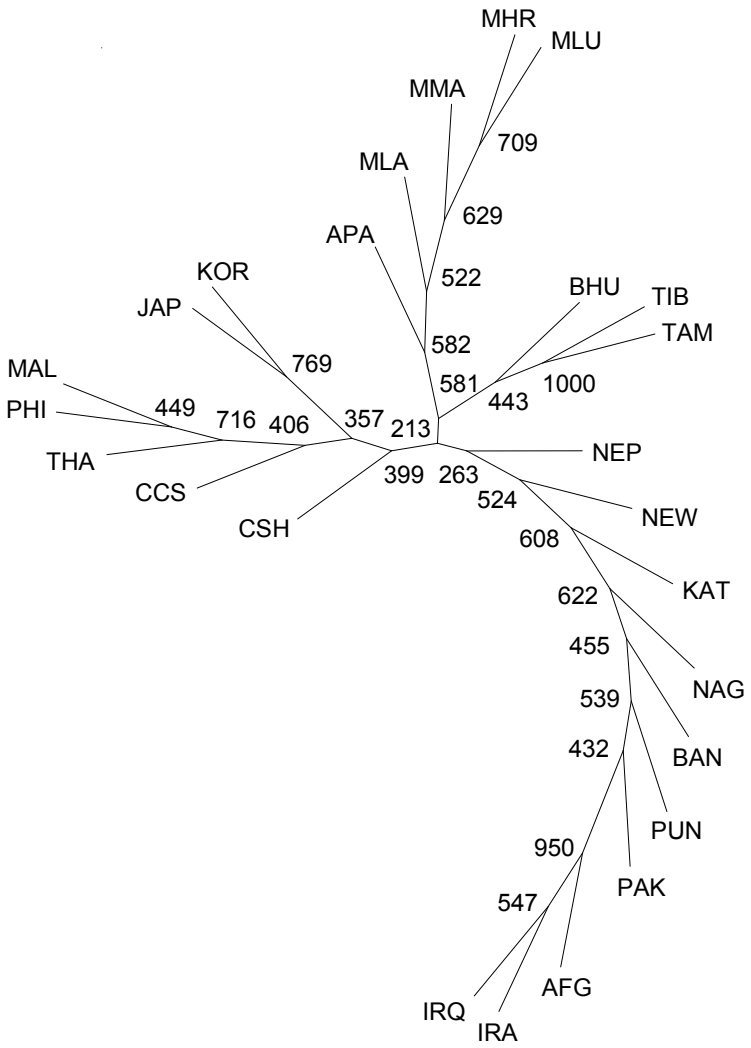

Table 1. Populations Analyzed

Population	Abbreviation	n	Language Family/Subfamily	Number of Loci	References
NORTHEAST ASIA					
Japan	JAP	525	Japanese	15	Hashiyada et al., 2003
Korea	KOR	231	Korean	15	Kim et al., 2003
Shaanxi Han (NW China)	CSH	203	Sino-Tibetan/Chinese	15	Wang et al., 2005
SOUTHEAST ASIA					
Chao Shan (South China)	CCS	144	Sino-Tibetan/Chinese	15	Hu et al., 2005
Malaysia	MAL	210	Austronesian/Malayo-polynesian	15	Seah et al., 2003
Philippines	PHI	106	Austronesian/Malayo-polynesian	15	De Ungria et al., 2005
Thailand	THA	210	Tai-Kadai/Kam-Tai	15	Rerkamnuaychoke et al. 2006
SOUTHWEST ASIA					
Afghanistan*	AFG	130	Indo-European/Indo-Iranian	15	Berti et al., 2005
Iran	IRA	150	Indo-European/Indo-Iranian	15	Shepard and Herrera, 2006b
Iraq	IRQ	206	Afro-Asiatic/Semitic	15	Barni et al., 2007
SOUTH ASIA					
Adi Pasi	APA	203	Sino-Tibetan/Tibeto-Burman	13	Krithika et al., 2005
Bangladesh	BAN	127	Indo-European/Indo-Iranian	15	Dobashi et al., 2005
Mizoram Hmar	MHR	80	Sino-Tibetan/Tibeto-Burman	13	Maity et al., 2003
Mizoram Mara	MMA	90	Sino-Tibetan/Tibeto-Burman	13	Maity et al., 2003
Mizoram Lai	MLA	92	Sino-Tibetan/Tibeto-Burman	13	Maity et al., 2003
Mizoram Lusei	MLU	92	Sino-Tibetan/Tibeto-Burman	13	Maity et al., 2003
Naga	NAG	105	Sino-Tibetan/Tibeto-Burman	13	Mastana et al., 2007
Pakistan	PAK	94	Indo-European/Indo-Iranian	15	Shepard and Herrera, 2006a
Punjab	PUN	86	Indo-European/Indo-Iranian	15	Shepard and Herrera, 2006a
HIMALAYAS					
Bhutan	BHU	936	Sino-Tibetan/Tibeto-Burman	15	Kraaijenbrink et al., 2006b
Nepal (General)	NEP	953	Sino-Tibetan/Tibeto-Burman	15	Kraaijenbrink et al., 2006a
Newar	NEW	66	Sino-Tibetan/Tibeto-Burman	15	present study
Kathmandu	KAT	77	Indo-European/Indo-Iranian	15	present study
Tamang	TAM	45	Sino-Tibetan/Tibeto-Burman	15	present study
Tibet	TIB	153	Sino-Tibetan/Tibeto-Burman	15	present study

Table 2. G-Test results for populations using 15 STR loci

		TIB	TAM	KAT	NEW	NEP	BHU	CCS	CSH	JAP	KOR	THA	MAL	PHI	PUN	PAK	BAN	AFG	IRA	IRQ
Tibet	TIB		265.5208	331.8308	331.4096	324.0563	294.1442	569.5753	381.6566	702.9995	479.7827	541.7274	600.4279	602.4645	502.2867	538.1968	498.6204	669.1839	717.9542	721.5934
Tamang	TAM	0.0000		267.0016	299.0512	301.2350	293.3618	394.2369	317.0794	358.2123	329.4326	334.2655	349.9590	378.8693	366.5260	395.1062	334.1797	401.8018	413.1159	442.2608
Katmandu	KAT	0.0000	0.0000		211.4277	219.1517	379.7130	489.3741	325.3644	421.1048	355.1138	358.4886	344.3952	382.4201	239.4038	213.5188	186.0736	280.6998	303.0143	350.7009
Newar	NEW	0.0000	0.0000	*0.0013*		276.1129	457.9844	527.0288	386.9984	458.1038	380.6284	461.1357	466.7141	464.8463	293.5210	299.9154	267.8649	385.1393	376.1184	380.2312
Nepal	NEP	0.0000	0.0000	*0.0091*	0.0000		462.2492	793.4504	403.5356	750.9483	546.0898	620.3414	667.3276	599.9863	461.8581	481.9894	422.7901	692.2495	807.3051	790.6835
Bhutan	BHU	0.0000	0.0000	0.0000	0.0000	0.0000		921.8695	540.6271	1014.0556	740.6976	695.8540	725.0259	618.6726	608.4052	708.0062	677.1872	869.2333	980.1330	911.5501
Chao Shan	CCS	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000		403.2074	728.9062	497.5356	403.3299	629.9239	456.1825	632.5628	760.9640	663.7009	976.2883	1019.8555	1012.0766
Shaanxi Han	CSH	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000		517.6439	343.3018	406.4190	476.1605	380.4575	469.6430	520.5606	476.5106	656.5607	746.2988	735.4686
Japan	JAP	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000		303.5310	692.6937	795.7937	511.8138	573.7626	743.9097	616.2498	886.4263	993.4788	921.0026
Korea	KOR	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000		461.4273	751.7308	454.0783	485.3125	621.9713	524.8886	691.8756	789.9065	770.7301
Thailand	THA	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000		224.9624	294.3215	404.7916	520.6833	424.8582	683.5734	636.2730	659.3774
Malaysia	MAL	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	*0.0033*		286.8904	376.7725	461.1105	392.4875	644.0248	545.6874	604.6087
Philippines	PHI	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000		399.6782	524.6376	463.3870	617.9642	552.4801	598.9198
Punjab	PUN	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000		213.7720	205.7492	321.4622	255.6687	309.8482
Pakistan	PAK	0.0000	0.0000	*0.0014*	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	*0.0040*		216.0772	303.8267	263.6936	396.3192
Bangladesh	BAN	0.0000	0.0000	**0.1139**	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	*0.0188*	*0.0060*		335.8770	327.9699	379.1754
Afghanistan	AFG	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000		316.7571	325.9964
Iran	IRA	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000		191.0005
Iraq	IRQ	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	*0.0346*	

The G statistics and P -values occupy the upper and lower levels of the diagonals, respectively

P values in bold and italics represent statistically insignificant differences before and after applying the bonferroni correction, respectively.

Table 3: Admixture analysis using regional groups

Parental Groups	Hybrid Populations				
	Tibet	Tamang	Newar	Kathmandu	Bhutan
Northeast Asia	0.634 ± 0.084	0.597 ± 0.108	0.447 ± 0.080	0.223 ± 0.058	0.411 ± 0.061
Southeast Asia	0.211 ± 0.085	0.290 ± 0.109	0.000 ± 0.081	0.054 ± 0.058	0.443 ± 0.061
South Asia	0.027 ± 0.087	0.000 ± 0.112	0.415 ± 0.083	0.633 ± 0.060	0.000 ± 0.063
Southwest Asia	0.127 ± 0.070	0.113 ± 0.090	0.138 ± 0.067	0.090 ± 0.048	0.146 ± 0.050

Supplementary Table 1. Tibet Allelic Frequencies (n=153)

Allele	D8S1179	D21S11	D7S820	CSF1PO	D3S1358	TH01	D13S317	D16S539	D2S1338	D19S433	vWA	TPOX	D18S51	D5S818	FGA
6						0.0882									
7				0.0065		0.2582								0.0065	
8	0.0033		0.1928			0.0719	0.1830	0.0229				0.6144			
9			0.0752	0.0621		0.5294	0.1144	0.2092				0.1634		0.0752	
9.3						0.0490									
10	0.1111		0.1765	0.2157		0.0033	0.1863	0.1536				0.0033		0.1340	
11	0.0261		0.2353	0.2582			0.2941	0.2941				0.2026	0.0033	0.4314	
12	0.1438		0.2876	0.3987			0.1569	0.2026		0.0523		0.0098	0.0163	0.2059	
12.2										0.0033					
13	0.2157		0.0294	0.0458			0.0523	0.1078		0.2745		0.0065	0.3235	0.1209	
13.2										0.0359					
14	0.1961		0.0033	0.0131	0.0490		0.0131	0.0098		0.3627	0.1144		0.1993	0.0229	
14.2										0.1046					
15	0.2124				0.3595					0.0621	0.0229		0.1307	0.0033	
15.2										0.0588					
16	0.0621				0.3399					0.0065	0.2320		0.0719		
16.2										0.0261					
17	0.0294				0.1765				0.0294	0.0065	0.2778		0.0621		
17.2										0.0065					
18					0.0686				0.1242		0.2157		0.0490		0.0261
19					0.0065				0.2092		0.1209		0.0654		0.0882
20									0.1340		0.0163		0.0392		0.0261
20.2															0.0098
21									0.0327				0.0163		0.0915
22									0.0294				0.0098		0.1242
22.2															0.0098
23									0.2549				0.0098		0.1569
23.2															0.0196
24									0.1275				0.0033		0.2255
24.2															0.0196
25									0.0458						0.1111
25.2															0.0131
26									0.0098						0.0359
26.2															0.0131
27		0.0196													0.0098
28		0.0425							0.0033						0.0163
28.2		0.0523													
29		0.2712													0.0033
29.2		0.0033													
30		0.2026													
30.2		0.0458													
31		0.1078													
31.2		0.0752													
32		0.0098													
32.2		0.1078													
33.2		0.0458													
34.2		0.0163													
Ho	0.8235	0.8693	0.8105	0.7255	0.7386	0.5882	0.7843	0.7778	0.9020	0.7190	0.7974	0.5556	0.8235	0.7843	0.8889
He	0.8460	0.8498	0.7897	0.7715	0.7564	0.6398	0.8327	0.8040	0.8394	0.8130	0.8070	0.7210	0.8232	0.7356	0.8792
HWE	0.4050	0.9773	0.4991	0.8012	0.6900	0.3986	0.0154	0.7377	0.6972	0.5249	0.5476	0.3629	0.3866	0.2735	0.4325
GDI	0.8342	0.8498	0.7897	0.7241	0.7193	0.6398	0.8074	0.7955	0.8394	0.7725	0.7966	0.5565	0.8232	0.7351	0.8792
MP	0.0536	0.0432	0.0833	0.1256	0.1290	0.1732	0.0735	0.0769	0.0600	0.0815	0.0763	0.2577	0.0565	0.1113	0.0325
PD	0.9464	0.9568	0.9167	0.8744	0.8710	0.8268	0.9265	0.9231	0.9400	0.9185	0.9237	0.7423	0.9435	0.8887	0.9675
PIC	0.8093	0.8310	0.7540	0.6762	0.6679	0.5895	0.7772	0.7617	0.8170	0.7402	0.7630	0.5012	0.8012	0.6983	0.8647
PE	0.6434	0.7332	0.6186	0.4688	0.4904	0.2770	0.5703	0.5585	0.7994	0.4582	0.5942	0.2409	0.6434	0.5703	0.7728
TPI	2.83	3.83	2.64	1.82	1.91	1.21	2.32	2.25	5.10	1.78	2.47	1.13	2.83	2.32	4.50

Ho: observed heterozygosity; He: expected heterozygosity; HWE: Hardy Weinberg Equilibrium p-values; GDI: Gene Diversity Index; MP: Matching Probability

PD:Power of Discrimination; PIC:Polymorphic Information Content; PE: Power of Exclusion; TPI: Typical Paternity Index

Supplementary Table 2: Tamang Allelic Frequencies (n=45)

Allele	D8S1179	D21S11	D7S820	CSF1PO	D3S1358	TH01	D13S317	D16S539	D2S1338	D19S433	vWA	TPOX	D18S51	D5S818	FGA
6						0.0882									
7				0.0065		0.2582								0.0065	
8	0.0033		0.1928			0.0719	0.1830	0.0229				0.6144			
9			0.0752	0.0621		0.5294	0.1144	0.2092				0.1634		0.0752	
9.3						0.0490									
10	0.1111		0.1765	0.2157		0.0033	0.1863	0.1536				0.0033		0.1340	
11	0.0261		0.2353	0.2582			0.2941	0.2941				0.2026	0.0033	0.4314	
12	0.1438		0.2876	0.3987			0.1569	0.2026		0.0523		0.0098	0.0163	0.2059	
12.2										0.0033					
13	0.2157		0.0294	0.0458			0.0523	0.1078		0.2745		0.0065	0.3235	0.1209	
13.2										0.0359					
14	0.1961		0.0033	0.0131	0.0490		0.0131	0.0098		0.3627	0.1144		0.1993	0.0229	
14.2										0.1046					
15	0.2124				0.3595					0.0621	0.0229		0.1307	0.0033	
15.2										0.0588					
16	0.0621				0.3399					0.0065	0.2320		0.0719		
16.2										0.0261					
17	0.0294				0.1765				0.0294	0.0065	0.2778		0.0621		
17.2										0.0065					
18					0.0686				0.1242		0.2157		0.0490		0.0261
19					0.0065				0.2092		0.1209		0.0654		0.0882
20									0.1340		0.0163		0.0392		0.0261
20.2															0.0098
21									0.0327				0.0163		0.0915
22									0.0294				0.0098		0.1242
22.2															0.0098
23									0.2549				0.0098		0.1569
23.2															0.0196
24									0.1275				0.0033		0.2255
24.2															0.0196
25									0.0458						0.1111
25.2															0.0131
26									0.0098						0.0359
26.2															0.0131
27		0.0196													0.0098
28		0.0425							0.0033						0.0163
28.2		0.0523													
29		0.2712													0.0033
29.2		0.0033													
30		0.2026													
30.2		0.0458													
31		0.1078													
31.2		0.0752													
32		0.0098													
32.2		0.1078													
33.2		0.0458													
34.2		0.0163													
Ho	0.8235	0.8693	0.8105	0.7255	0.7386	0.5882	0.7843	0.7778	0.9020	0.7190	0.7974	0.5556	0.8235	0.7843	0.8889
He	0.8460	0.8498	0.7897	0.7715	0.7564	0.6398	0.8327	0.8040	0.8394	0.8130	0.8070	0.7210	0.8232	0.7356	0.8792
HWE	0.4050	0.9773	0.4991	0.8012	0.6900	0.3986	0.0154	0.7377	0.6972	0.5249	0.5476	0.3629	0.3866	0.2735	0.4325
GDI	0.8342	0.8498	0.7897	0.7241	0.7193	0.6398	0.8074	0.7955	0.8394	0.7725	0.7966	0.5565	0.8232	0.7351	0.8792
MP	0.0536	0.0432	0.0833	0.1256	0.1290	0.1732	0.0735	0.0769	0.0600	0.0815	0.0763	0.2577	0.0565	0.1113	0.0325
PD	0.9464	0.9568	0.9167	0.8744	0.8710	0.8268	0.9265	0.9231	0.9400	0.9185	0.9237	0.7423	0.9435	0.8887	0.9675
PIC	0.8093	0.8310	0.7540	0.6762	0.6679	0.5895	0.7772	0.7617	0.8170	0.7402	0.7630	0.5012	0.8012	0.6983	0.8647
PE	0.6434	0.7332	0.6186	0.4688	0.4904	0.2770	0.5703	0.5585	0.7994	0.4582	0.5942	0.2409	0.6434	0.5703	0.7728
TPI	2.83	3.83	2.64	1.82	1.91	1.21	2.32	2.25	5.10	1.78	2.47	1.13	2.83	2.32	4.50

Ho: observed heterozygosity; He: expected heterozygosity; HWE: Hardy Weinberg Equilibrium p-values; GDI: Gene Diversity Index; MP: Matching Probability

PD:Power of Discrimination; PIC:Polymorphic Information Content; PE: Power of Exclusion; TPI: Typical Paternity Index

Supplementary Table 3: Newar Allelic Frequencies (n=66)

Allele	D8S1179	D21S11	D7S820	CSF1PO	D3S1358	TH01	D13S317	D16S539	D2S1338	D19S433	vWA	TPOX	D18S51	D5S818	FGA
6						0.1894									
7			0.0227	0.0152		0.1061	0.0076								
8			0.1970			0.1591	0.1515	0.0455				0.4015			
8.2			0.0152												
9			0.1061	0.0455		0.4318	0.0909	0.2727				0.1212	0.0076	0.1364	
9.3						0.1136									
10	0.1061		0.1742	0.1439			0.0909	0.0985				0.1136		0.1894	
11	0.0606		0.2727	0.2273			0.3333	0.2652				0.3409	0.0076	0.3106	
12	0.1818		0.1515	0.4167			0.2955	0.2121		0.0606		0.0152	0.0530	0.2273	
12.2										0.0076					
13	0.1591		0.0606	0.1212			0.0227	0.0985		0.4318		0.0076	0.2803	0.1288	
13.2										0.0455					
14	0.2121			0.0227	0.0076		0.0076	0.0076		0.2424	0.1818		0.2121	0.0076	
14.2										0.0152					
15	0.2348			0.0076	0.2424					0.0909	0.0303		0.0758		
15.2										0.0530					
16	0.0227				0.3561					0.0303	0.2424		0.1212		
16.2										0.0076					
17	0.0227				0.2879				0.0606	0.0076	0.2955		0.0379		
18					0.0833				0.0909	0.0076	0.1515		0.0227		0.0152
19					0.0227				0.1364		0.0909		0.1061		0.0682
20									0.0606		0.0076		0.0227		0.0682
20.2															0.0076
21									0.0606				0.0227		0.1136
22									0.0379				0.0076		0.1364
23									0.2045				0.0076		0.1818
24									0.1288						0.2879
25									0.1288				0.0152		0.0606
25.2															0.0076
26									0.0833						0.0455
27		0.0833							0.0076						0.0076
28		0.0455													
28.2		0.0379													
29		0.2424													
29.2		0.0152													
30		0.2576													
30.2		0.0303													
31		0.0758													
31.2		0.0758													
32.2		0.1136													
33.2		0.0227													
Ho	0.8182	0.8182	0.8182	0.7576	0.7273	0.7121	0.7879	0.6970	0.8333	0.7576	0.7576	0.6667	0.8333	0.7727	0.9091
He	0.8318	0.8490	0.8355	0.7421	0.7296	0.7354	0.7673	0.7948	0.8862	0.7922	0.8149	0.7000	0.8459	0.7867	0.8485
HWE	0.3870	0.2648	0.3001	0.5927	0.8587	0.8111	0.8139	0.4143	0.0599	0.8608	0.6784	0.6115	0.3300	0.3482	0.7627
GDI	0.8318	0.8448	0.8241	0.7421	0.7926	0.7337	0.7673	0.7948	0.8854	0.7422	0.7947	0.7000	0.8449	0.7867	0.8485
MP	0.0624	0.0569	0.0725	0.1079	0.1286	0.1157	0.0969	0.0771	0.0404	0.1024	0.0817	0.1570	0.0551	0.0937	0.0556
PD	0.9376	0.9431	0.9275	0.8921	0.8714	0.8843	0.9031	0.9229	0.9596	0.8976	0.9183	0.8430	0.9449	0.9063	0.9444
PIC	0.8020	0.8204	0.7933	0.7005	0.6744	0.6913	0.7252	0.7569	0.8668	0.7055	0.7568	0.6413	0.8210	0.7466	0.8255
PE	0.6332	0.6332	0.6332	0.5228	0.4717	0.4473	0.5767	0.4236	0.6623	0.5228	0.5228	0.3786	0.6623	0.5494	0.8140
TPI	2.75	2.75	2.75	2.06	1.83	1.74	2.36	1.65	3.00	2.06	2.06	1.50	3.00	2.20	5.50

Ho: observed heterozygosity; He: expected heterozygosity; HWE: Hardy Weinberg Equilibrium p-values; GDI: Gene Diversity Index; MP: Matching Probability

PD:Power of Discrimination; PIC:Polymorphic Information Content; PE: Power of Exclusion; TPI: Typical Paternity Index

Supplementary Table 4: Kathmandu Allelic Frequencies (n=77)

Allele	D8S1179	D21S11	D7S820	CSF1PO	D3S1358	TH01	D13S317	D16S539	D2S1338	D19S433	vWA	TPOX	D18S51	D5S818	FGA
6						0.2143									
7						0.1948						0.0065		0.0065	
8	0.0065		0.2468			0.1104	0.1234	0.0325				0.3961			
9	0.0065		0.0909	0.0649		0.3117	0.1623	0.1948				0.1623		0.0519	
9.3						0.1688									
10	0.1429		0.1818	0.1688			0.1039	0.0974				0.0649	0.0130	0.2013	
11	0.0325		0.2013	0.2013			0.3247	0.2987				0.3571	0.0130	0.2597	
12	0.1234		0.2468	0.4351			0.1948	0.2208		0.0455		0.0130	0.0260	0.3052	
13	0.1948		0.0325	0.1039			0.0714	0.1429		0.2857			0.1429	0.1688	
13.2										0.0519					
14	0.2597			0.0195	0.0195		0.0130			0.2532	0.1688		0.2792	0.0065	
14.2										0.0909					
15	0.1753			0.0065	0.2338		0.0065	0.0130		0.1169	0.0390		0.1688		
15.2										0.0844					
16	0.0519				0.4091					0.0519	0.2143		0.1169		
16.2										0.0130					
17	0.0065				0.2532				0.0779	0.0065	0.2662		0.0649		
18					0.0844				0.1623		0.2078		0.0390		0.0130
19									0.1558		0.0909		0.0779		0.1039
20									0.1169		0.0130		0.0325		0.0909
21									0.0584						0.0844
22									0.0649				0.0130		0.1234
23									0.1364				0.0130		0.2468
23.2															0.0065
24									0.1364						0.1429
24.2															0.0130
25									0.0909						0.0714
26															0.0779
27		0.0065													0.0260
27.2		0.0130													
28		0.1169													
28.2		0.0130													
29		0.2338													
30		0.2662													
30.2		0.0065													
31		0.0844													
31.2		0.1234													
32.2		0.0909													
33.2		0.0390													
34.2		0.0065													
Ho	0.7533	0.8312	0.9221	0.6883	0.7403	0.6494	0.7792	0.7922	0.76623	0.7922	0.7273	0.6753	0.8442	0.7792	0.8701
He	0.8338	0.8336	0.8005	0.8003	0.7110	0.7834	0.8381	0.7982	0.8836	0.8229	0.8068	0.6893	0.8566	0.7726	0.8748
HWE	0.1048	0.9206	0.0164	0.8380	0.2272	0.0412	0.9232	0.7274	0.3532	0.2790	0.3022	0.1515	0.1299	0.6810	0.9681
GDI	0.8927	0.8336	0.8005	0.7310	0.7110	0.7834	0.8042	0.7981	0.8823	0.8229	0.8068	0.6892	0.8508	0.7726	0.8707
MP	0.0632	0.0565	0.1017	0.1138	0.1580	0.0862	0.0717	0.0845	0.0329	0.0636	0.0703	0.1695	0.0555	0.1027	0.0363
PD	0.9368	0.9435	0.8983	0.8862	0.8420	0.9138	0.9283	0.9155	0.9671	0.9364	0.9297	0.8305	0.9445	0.8973	0.9637
PIC	0.8011	0.8073	0.7637	0.6897	0.6553	0.7431	0.7722	0.7624	0.8636	0.7953	0.7721	0.6270	0.8289	0.7293	0.8515
PE	0.5153	0.6581	0.8407	0.4104	0.4933	0.3543	0.5611	0.5847	0.5379	0.5847	0.4717	0.3911	0.6834	0.5611	0.7349
TPI	2.03	2.96	6.42	1.60	1.93	1.43	0.26	2.41	2.14	2.41	1.83	1.54	3.21	2.26	3.85

Ho: observed heterozygosity; He: expected heterozygosity; HWE: Hardy Weinberg Equilibrium p-values; GDI: Gene Diversity Index; MP: Matching Probability

PD:Power of Discrimination; PIC:Polymorphic Information Content; PE: Power of Exclusion; TPI: Typical Paternity Index

Supplementary Table 5: Parameters of population genetics interest

	Total # of alleles	CMP	CPD	CPE	Avg. He	Avg. GDI
Kathmandu	124	5.684×10^{16}	0.73724643	0.99999789	0.7740	0.8033
Newar	130	2.126×10^{16}	0.74659830	0.99999773	0.7778	0.7956
Tamang	113	3.872×10^{14}	0.84589854	0.99998260	0.7348	0.7632
Tibet	138	1.663×10^{16}	0.78531013	0.99999843	0.7725	0.7708

CMP: Combined Matching Probability, *CPE:* Combined Power of Exclusion, *CPD:* Combined Power of Discrimination, Avg. He: Average Heterozygosity, *Avg. GDI* : Average Gene Diversity Index

Supplementary Table 6: Intra- and Inter-population diversity

	Himalayan Populations			Northeast Asia			Southeast Asia			South Asia			Southwestern Asia			All populations		
Locus	G_{st}	H_s	H_t	G_{st}	H_s	H_t	G_{st}	H_s	H_t	G_{st}	H_s	H_t	G_{st}	H_s	H_t	G_{st}	H_s	H_t
D8S1179	0.0073	0.8230	0.8290	0.0040	0.8402	0.8436	0.0071	0.8482	0.8543	0.0128	0.8274	0.8381	0.0047	0.8348	0.8387	0.0118	0.8325	0.8425
D21S11	0.0085	0.8340	0.8412	0.0050	0.7942	0.7981	0.0050	0.8386	0.8428	0.0296	0.8367	0.8623	0.0028	0.8438	0.8462	0.0203	0.8320	0.8493
D7S820	0.0134	0.7930	0.8038	0.0042	0.7601	0.7633	0.0083	0.7622	0.7686	0.0182	0.7823	0.7969	0.0067	0.7903	0.7956	0.0182	0.7801	0.7946
CSF1PO	0.0059	0.7166	0.7208	0.0023	0.7177	0.7193	0.0047	0.7254	0.7288	0.0171	0.6890	0.7010	0.0051	0.7135	0.7171	0.0134	0.7082	0.7178
D3S1358	0.0098	0.7081	0.7151	0.0013	0.7139	0.7149	0.0121	0.7242	0.7331	0.0202	0.7067	0.7213	0.0068	0.7547	0.7598	0.0190	0.7172	0.7311
TH01	0.0213	0.7061	0.7215	0.0111	0.6778	0.6854	0.0088	0.7274	0.7339	0.0824	0.6360	0.6931	0.0060	0.7878	0.7926	0.0545	0.6900	0.7298
D13S317	0.0197	0.7894	0.8053	0.0026	0.8063	0.8084	0.0062	0.7949	0.7998	0.0187	0.8017	0.8170	0.0073	0.7762	0.7819	0.0209	0.7962	0.8132
D16S539	0.0071	0.7871	0.7927	0.0048	0.7706	0.7743	0.0081	0.7810	0.7873	0.0235	0.7891	0.8081	0.0050	0.7739	0.7778	0.0203	0.7827	0.7990
vWA	0.0025	0.7948	0.7968	0.0017	0.7964	0.7977	0.0068	0.7974	0.8028	0.0125	0.8024	0.8126	0.0065	0.7921	0.7973	0.0123	0.7977	0.8077
TPOX	0.0217	0.6289	0.6428	0.0018	0.6480	0.6492	0.0055	0.6036	0.6069	0.0125	0.6866	0.6953	0.0023	0.6410	0.6424	0.0177	0.6506	0.6624
D18S51	0.0117	0.8396	0.8495	0.0022	0.8587	0.8606	0.0078	0.8477	0.8544	0.0218	0.8381	0.8568	0.0035	0.8601	0.8631	0.0171	0.8462	0.8609
D5S818	0.0111	0.7576	0.7661	0.0044	0.7865	0.7900	0.0036	0.7846	0.7875	0.0154	0.7616	0.7736	0.0086	0.7217	0.7280	0.0195	0.7634	0.7786
FGA	0.0171	0.8542	0.8691	0.0021	0.8560	0.8578	0.0049	0.8615	0.8658	0.0113	0.8757	0.8857	0.0043	0.8522	0.8559	0.0130	0.8639	0.8753
All loci	0.0120	0.7771	0.7811	0.0036	0.7713	0.7740	0.0068	0.7767	0.7820	0.0222	0.7718	0.7894	0.0053	0.7801	0.7843	0.0196	0.7739	0.7894

III. Y-STR DIVERSITY IN THE HIMALAYAS

A. INTRODUCTION

Binary polymorphisms, including SNPs and insertions/deletions, located within the non-recombining region of the Y-chromosome (NRY) have been deemed useful for tracing ancient paternal lineages [1-3]. The Y-chromosome short tandem repeats (Y-STRs), on the other hand, mutate rapidly and are suitable for reconstructing the phylogeny of relatively recent demographic events as well as for forensic casework and paternity testing [2, 4-6]. Recently, Y-STR data have been increasingly used to investigate the evolution, migration and genetic diversity of modern human populations [2, 6-11]. The interest in Y-STRs as markers for ancestry and population studies have prompted the creation of a comprehensive worldwide Y-STR database called Y Chromosome Haplotype Reference Database (YHRD, www.yhrd.org) for calculating haplotype frequencies, matching probabilities and performing comparative population genetic analyses [12-15]. This global repository currently supports most frequently used haplotype formats, namely, 9-loci minimal (DYS19, DYS389I, DYS389II, DYS390, DYS391, DYS392 and DYS393, DYS385a/b), SWGDAM recommended 11-loci extended (Minimal + DYS438 and DYS439), 12-loci Powerplex (SWGDAM + DYS437) and 17-loci Yfiler haplotypes [15].

Our previous studies of Y-chromosomal biallelic [16] and autosomal STR polymorphisms [17] of Tibet and Nepal have revealed that these Himalayan groups arrived in the area during the Neolithic time from the Northeast Asia with subsequent gene flow from the Indian subcontinent into the Kathmandu valley and Newar population. The latter conclusion is congruent with recent mtDNA studies [18, 19], which

reported shared maternal lineage between Indian and Nepalese populations. In contrast, Tibet and the Tamang population display limited influence from the Indian subcontinent suggesting that the Himalayan massif has acted as a barrier for gene flow from the south into the Tibetan plateau [16, 17]. In addition to the Northeast Asian influence, the high frequency of the Asian-specific *Alu* insertion at the YAP (Y *Alu* polymorphism) locus in the Tibetan Y-chromosomes had previously led some researchers to argue for Central Asian contribution in the Tibetan gene pool [20-22].

Although several studies based on Y-STR data from Tibet have been previously published [23-26], the present study is the first of its kind to perform comprehensive phylogenetic analyses involving a considerable sample size. The knowledge of the phylogenetic relationships among populations is essential in order to assess whether or not populations should be considered as a separate entity as databases for forensic analysis and paternity testing. The present work also improves on an earlier Nepalese study [27] by examining populations that are anthropologically well-characterized. In addition, this report complements previous Y-chromosomal biallelic [16] and autosomal STR [17] data thus providing a comprehensive analysis of the genetic diversity in the Himalayas. In the present study, 17 Y-STR loci were typed in the three Nepalese populations of Tamang, Newar and Kathmandu as well as a general collection from Tibet to investigate their genetic ancestry and phylogenetic relationships to previously published geographically targeted groups from Northeast Asia, Southeast Asia, South Central Asia and Central Asia using 9-loci minimal haplotypes (DYS19, DYS389I, DYS389II, DYS390, DYS391, DYS392 and DYS393, DYS385a/b). Furthermore, the

haplotype data generated from the 17 Y-STR loci was utilized to ascertain the values of parameters of forensic interests.

B. MATERIALS AND METHODS

Sample collection and DNA isolation

Blood samples were collected with informed consent from 344 unrelated males from Tibet (156) and three populations from Nepal (188), namely Tamang (45), Newar (66) and Kathmandu (77). These samples have been previously typed for Y-SNP markers in one of our earlier studies [16] (Supplementary Table 1). Genealogical information of the donors was recorded for a minimum of two generations to ascertain their paternal ancestry. Sample collections were performed in strict compliance with the ethical guidelines put forth by the institutions involved in this study. The DNA was extracted by the standard phenol-chloroform method and ethanol precipitated as previously described [28] and stored at -80°C.

DNA amplification and STR genotyping

The DNA samples were amplified at 17 Y-STR loci in a multiplex reaction using the AmpF*l*STR® Yfiler kit [29]. PCR amplifications were performed as specified by the manufacturer using the recommended DNA amounts (0.5-1.25 ng) in an Eppendorf Master gradient cycler (Eppendorf AG, Germany). Amplicons were separated by multi-capillary electrophoresis in an ABI Prism 3130 Genetic Analyzer and the ABI Genescan 500 LIZ internal size standard was used as a basis for comparison. The software GeneMapper® v3.1 [29] was employed to determine fragment sizes, while alleles were

designated by comparison to an allelic ladder supplied by the manufacturer. The nomenclature of the Y-STR loci studied is as recommended by the DNA Commission of the International Society of Forensic Genetics for analysis of Y-STR systems [30] while the Y-SNP nomenclature followed is in accordance with the Y Chromosome Consortium [31, 32].

Statistical and phylogenetic analyses

Allelic frequencies were calculated by the gene counting method [6]. Gene and haplotype diversities were computed using the software package Arlequin v. 3.1 [33]. Chromosomes carrying null alleles or duplicated loci were excluded from the haplotype calculation at the 17 Y-STR loci level. Discrimination capacity (DC) and fraction of unique haplotypes (FUH) were estimated as percentage proportions of different and unique haplotypes, respectively, in a given population. All the statistical parameters were calculated for both minimal and extended haplotypes for the comparison purposes. The dw_{min} (minimum diversity within the population), mw_{min} (minimum matching probability within the population), mw_{max} (maximum matching probability within the population), and mb_{min} (minimum matching probability between two populations) were calculated as previously described [34, 35]. The ratio mw_{max}/mb_{min} gives an upper estimate of how much more probable it is to find a match within a population than between two populations and the ratio mw_{min}/mb_{min} provides a lower estimate for the same parameter.

A total of sixteen reference populations (Table 1) from the published literature were included for comparison across 9-loci minimal haplotypes (DYS19, DYS389I, DYS389II, DYS390, DYS391, DYS392, DYS393 and DYS385a/b), since the data for the

remaining loci typed in this study were not reported for all reference collections. Correspondence Analysis (CA) plots were generated using NTSYSpc 2.02i [44], while a Neighbor Joining (NJ) dendrogram was built utilizing the PHYLIP v3.6 program [45]. The statistical robustness of the phylogenetic relationships within the NJ tree was assessed using bootstrap analysis involving 1,000 replications. A median-joining network based on Y haplogroup O3a3c-M134 chromosomes in the Himalayan populations was constructed utilizing the 15 Y-STR loci, excluding the bilocal marker DYS385a/b, using the program NETWORK 4.2.0.0 (www.fluxus-engineering.com). For network calculations, microsatellite loci were weighted inversely to their variance such that higher weights were assigned to the least variable loci [22, 46]. The network generated was post-processed by employing the Maximum Parsimony (MP) parameter [47, 48] to obtain the simplest possible projection.

C. RESULTS AND DISCUSSION

Allelic frequencies and gene diversities for the 17 Y-STR loci analyzed in the Tamang, Newar, Kathmandu and Tibet collections are listed in Supplementary Tables 2 to 5. Tamang exhibits a relative (to the other 3 Himalayan populations) high degree of genetic homogeneity with 11 of the 17 loci displaying gene diversities lower than 0.5, whereas Kathmandu represents the other extreme with all loci registering values above 0.5 for the same parameter. As expected, Kathmandu also possesses the highest average gene diversity (0.6657) followed by Newar (0.6411), Tibet (0.6352) and Tamang (0.4195).

Null alleles were detected in two Newari and four Tibetan samples at locus DYS448 as previously described [6, 27], while five Kathmandu males were null at DYS458, consistent with the location of a deletion in the short arm of the Y chromosome encompassing the amelogenin locus (Supplementary Table 1) [16, 49]. The frequency of the amelogenin Y (AMGY) deletion in the Kathmandu populace (6.49%) is higher than levels previously reported in Nepal (1.2%) [27] and India (1.8%) [50] but lower than that of Sri Lankan males (8%) [51]. Duplication was observed in one individual from Newar at locus DYS458 (alleles 16, 18). All null alleles and duplications were confirmed by repeating the amplification process.

Table 2 shows the number of different and unique haplotypes, fraction of unique haplotypes (FUH), discrimination capacity (DC) and haplotype diversity (HD) for the three haplotype resolutions considered: Yfiler (17 loci), extended (11 loci) and minimal (9 loci) haplotypes. Overall, a total of 262 different haplotypes were identified, of which 228 (66.28%) were unique to a single individual (Table 2). Chromosomes carrying null alleles (n=11) or duplicated loci (n=1) were excluded from the haplotype calculation at the 17 Y-STR loci level. When analyzed at the 11-loci and 9-loci haplotype resolutions, the number of different haplotypes decreased to 236 and 214, respectively, among which 189 (54.94%) and 161 (46.80%) were distinct haplotypes, respectively (Table 2). The significant increase in the proportion of unique haplotypes using the Yfiler system (66.28%) compared to the minimal haplotype (46.80%) reflects the power of discrimination at 17 Y-STR loci.

The overall haplotype diversity for the Himalayan population at 17 Y-STR loci was 0.9973 while the corresponding values for the extended and minimal haplotypes

were 0.9955 and 0.9942, respectively (Table 2). The latter is slightly lower than values previously reported for East Asian (0.9996, n=700) [2] and European (0.9976, n=11,610) populations [14]. This relatively lower diversity may be attributed to the reduced heterogeneity observed in Tamang (0.9010) and Newar (0.9585), which in turn may be the result of bottlenecks and/or founder effects in the case of former and genetic drift in the latter. The genetic homogeneity in Tamang is also reflected in their reduced average gene diversity (0.4195), consistent with previous reports [16, 17]. In addition, the discrimination capacity in Tamang and Newar are considerably lower than that of Kathmandu and Tibet at all the three haplotype resolutions (Table 2).

Table 3 shows the Y-STR haplotype matching probabilities within and between the Himalayan populations. Similar to the limited haplotype and gene diversities values, Tamang (0.0667) shows the highest maximum match probability within the population followed by Newar (0.0582), Kathmandu (0.0162) and Tibet (0.0072) (Table 3). When compared among the four Himalayan collections, the maximum probability ($db_{max,}$ which is 1-db_{min}) of obtaining two different Y-haplotypes when sampling a pair of individuals from Tamang and Kathmandu, Tamang and Newar, Tamang and Tibet, Kathmandu and Newar, Kathmandu and Tibet, and Newar and Tibet are 99.82%, 100%, 100%, 99.98%, 100% and 100%, respectively. The high power of discrimination underscores the genetic uniqueness of the four Himalayan populations as well as the need for a well-defined ethnic and geographic sampling of populations for forensic applications.

There are no profiles shared across all four collections in any of the three data sets. Of the 34 shared haplotypes at the 17 Y-STR loci level (Table 2), 32 are common within a single population while two haplotypes are shared between two Nepalese groups

one between Newar and Kathmandu and another between Tamang and Kathmandu. However, when examined at the minimal 9-loci resolution, two haplotypes are shared across three Himalayan populations, with the exception of Newar. These two haplotypes differ from each other by a one-step mutation at the bilocal marker DYS385a/b (13, 18 vs. 13, 19 alleles) and one of them (DYS385a/b = 13, 19 alleles) is the most frequent minimal haplotype observed in our Himalayan populations at a frequency of 3.2% (14-12-28-23-10-14-12-13-19). The most common haplotype is represented by 6 Tibetans, 2 Tamangs and 2 individuals from Kathmandu. Comparison of the latter haplotype with the YHRD database (Release 32) returned 75 exact matches, the majority of which are from Bhutan (26), China (19; 13 of the 19 are from Eastern China) and Nepal (17), while the remaining matches are from East Asia, including South Korea (4), Japan (3) and Malaysia (2), with the exceptions of two Indian males, an admixed Hispanic American and a Tibetan from Lhasa. It is not surprising to find matching profiles from Nepalese samples in the YHRD database since three of the populations under study are from the same region. Matches with Bhutanese and Chinese samples, however, suggest recent gene flow and/or shared common patrimonies among these groups concordant with previous findings [16, 17, 20, 21, 43].

Phylogenetic relationships among the four Himalayan collections and other geographically targeted populations were assessed using CA (Figure 1) and NJ (Figure 2) analyses. Figure 1B represents the contribution of each of 94 alleles of 9 Y-STR loci to the partition of the populations. The Himalayans cluster loosely in the upper right quadrant of the plot (Figure 1A) and share the same clade in the tree (Figure 2), with the exception of Kathmandu which maps closer to the South Central Asian group in both the

analyses [16, 17]. Newar also seems to display some affinity to the South Central Asian assemblage along the X axis, whereas Tamang and Haryana are outliers from their respective groups (Figure 1A). The genetic similarities observed within the Himalayas based on their Y-STR loci (Figures 1A and 2) are reflected in the high frequencies of Y-haplogroup O3a3c-M134, common among Tibeto-Burman speakers [16, 21, 52]. The above inference is supported by the fact that the most frequent minimal haplotype (14-12-28-23-10-14-12-13-19), as well as its one-step mutation neighbor at DYS385 (DYS385a/b = 13, 18 allele), both belong to haplogroup O3a3c-M134 [16]. On the other hand, Kathmandu and Newar's affinity to the South Central Asian cluster in the CA and NJ tree (Figures 1A and 2, respectively) may be due to the presence of Indian Y-lineages (Haplogroups R, H and C5) in their gene pools [16].

The East Asian collections map toward the middle of the lower half of the graph while the Mongolians and Buryats segregate to the left of the chart (Figure 1A). There is no clear genetic partitioning between the northern and southern East Asian populations in both the CA and NJ tree (Figures 1A and 2) [2]. Overall, the NJ dendrogram mirrors the distributions of populations in the CA with the exceptions of Mongolia and Buryat which form a sister clade with the South Central Asian branch, and the general population of Nepal showing more affinities with Tamang than with Bhutan (Figure 2).

The lack of phylogenetic affinities exhibited by Tamang in relation to Tibet (Figures 1A and 2) are of interest given its proposed close genetic association with the latter in previous studies [16, 17]. Although both Tamang and Tibet share high frequencies of haplogroup O3a3c-M134 [16], their Y-STR profiles differ considerably. In order to gain further insight into the recent demographic history of these two groups, a

median-joining network based solely on Y-haplogroup O3a3c-M134 was constructed at the level of the 15 Y-STR loci utilizing our four Himalayan populations (Figure 3). It is notable that Tamang and Newar form distinct clusters because of their shared or closely related haplotypes, while Tibet and Kathmandu are highly divergent (Figure 3). This finding suggests either a male founder effect in Tamang, possibly from Tibet, or a recent bottleneck event as they settled south of the Himalayas from Tibet, leading to their highly reduced Y-SNP [18] and Y-STR diversity (Table 2). On the other hand, Newar's unique genetic profile may be due to isolation and/or drift [17].

In summary, our results confirmed previous Y-chromosomal and autosomal STR reports that Newar and Kathmandu experienced substantial gene flow from the India whereas Tamang and Tibet display no genetic influences from the subcontinent. A median-joining network of haplogroup O3a3c-M134 based on 15 Y-STR loci suggests recent bottleneck and/or founder effect in Tamang. A high value of combined haplotype diversity (0.9970) from the four Himalayan populations is indicative of genetic heterogeneity within the region. In addition, very high percentages (99.82-100%) of the maximum probability (db_{max}) of finding two different Y-haplotypes when sampling a pair of individuals between two different Himalayan populations underscores the genetic singularity of the four Himalayan collections reported. The uniqueness of our four Himalayan populations argues for independent databases for forensic analysis and paternity testing.

REFERENCES

1. Jobling MA, Tyler-Smith C (1993) The human Y chromosome: an evolutionary marker comes of age. Nat Rev Genet 8:598-612

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

3. Gayden T, Regueiro M, Martinez L, Cadenas AM, Herrera RJ (2008) Human Y-chromosome haplotyping by allele-specific polymerase chain reaction. Electrophoresis 29:2419-2423

4. Iida R and Kishi K (2005) Identification, characterization and forensic application of novel Y-STRs. Leg Med 7:255-258

5. Hanson EK, Ballantyne J (2007) An ultra-high discrimination Y chromosome short tandem repeat multiplex DNA typing system. PLoS ONE 8:e688

6. Budowle B, Ge J, Low J, Lai C, Yee WH, Law G, Tan WF, Chang YM, Perumal R, Keat PY, Mizuno N, Kasai K, Sekiguchi K, Chakraborty R (2009) The effects of Asian population substructure on Y STR forensic analyses. Leg Med 11:64-69

7. Alves C, Gusmão L, Barbosa J, Amorim A (2003) Evaluating the informative power of Y-STRs: a comparative study using European and new African haplotype data. Forensic Sci Int 34:126-133

8. Roewer L, Croucher PJ, Willuweit S, Lu TT, Kayser M, Lessig R, de Knijff P, Jobling MA, Tyler-Smith C, Krawczak M (2005) Signature of recent historical events in the European Y-chromosomal STR haplotype distribution. Hum Genet 116:279-291

9. Immel UD, Krawczak M, Udolph J, Richter A, Rodig H, Kleiber M, Klintschar M (2006) Y-chromosomal STR haplotype analysis reveals surname-associated strata in the East-German population. Eur J Hum Genet 14:577-582

10. Rebała K, Mikulich AI, Tsybovsky IS, Siváková D, Dzupinková Z, Szczerkowska-Dobosz A, Szczerkowska Z (2007) Y-STR variation among Slavs: evidence for the Slavic homeland in the middle Dnieper basin. J Hum Genet 52:406-414

11. Rodríguez V, Tomàs C, Sánchez JJ, Castro JA, Ramon MM, Barbaro A, Morling N, Picornell A (2009) Genetic sub-structure in western Mediterranean populations revealed by 12 Y chromosome STR loci. Int J Legal Med 123:137-141

12. Roewer L, Krawczak M, Willuweit S et al (2001) Online reference database of European Y-chromosomal short tandem repeat (STR) haplotypes. Forensic Sci Int 118:106-113

13. Kayser M, Brauer S, Willuweit S et al. (2002) Online Y-chromosomal short tandem repeat haplotype reference database (YHRD) for U.S. populations. J Forensic Sci 47:513-519

14. Lessig R, Willuweit S, Krawczak M, et al (2003) Asian online Y-STR Haplotype Reference Database. Leg Med 5:S160-163

15. Willuweit S, Roewer L, International Forensic Y Chromosome User Group (2007) Y chromosome haplotype reference database (YHRD): update. Forensic Sci Int Genet 1:83-87

16. Gayden T, Cadenas AM, Regueiro M, Singh NB, Zhivotovsky LA, Underhill PA, Cavalli-Sforza LL, Herrera RJ (2007) The Himalayas as a directional barrier to gene flow. Am J Hum Genet 80:884-894

17. Gayden T, Mirabal S, Cadenas AM, Lacau H, Simms TM, Morlote D, Chennakrishnaiah S, Herrera RJ (2009) Genetic insights into the origins of Tibeto-Burman populations in the Himalayas. J Hum Genet 54:216–223

18. Thangaraj K, Chaubey G, Kivisild T et al (2008) Maternal footprints of Southeast Asians in North India. Hum Hered 66:1-6

19. Fornarino S, Pala M, Battaglia V, Maranta R, Achilli A, Modiano G, Torroni A, Semino O, Santachiara-Benerecetti SA (2009) Mitochondrial and Y-chromosome diversity of the Tharus (Nepal): a reservoir of genetic variation. BMC Evol Biol 9:154

20. Qian Y, Qian B, Su B et al (2000) Multiple origins of Tibetan Y chromosomes. Hum Genet 106:453-454

21. Su B, Xiao C, Deka R et al (2000) Y chromosome haplotypes reveal prehistorical migrations to the Himalayas. Hum Genet 107:582-590

22. Hammer MF, Karafet TM, Park H, Omoto K, Harihara S, Stoneking M, Horai S (2006) Dual origins of the Japanese: common ground for hunter-gatherer and farmer Y chromosomes. J Hum Genet 51:47-58

23. Zhang Q, Yan J, Tang H, Jiao Z, Liu Y (2006) Genetic polymorphisms of 17 Y-STRs haplotypes in Tibetan ethnic minority group of China. Leg Med 8:300-305

24. Zhu B, Liu S, Ci D, et al (2006) Population genetics for Y-chromosomal STRs haplotypes of Chinese Tibetan ethnic minority group in Tibet. Forensic Sci Int 161:78-83

25. Kang LL, Liu K, Ma Y (2007) Y chromosome STR haplotypes of Tibetan Living Tibet Lassa. Forensic Sci Int 172:79-83

26. Zhu B, Wu Y, Shen C et al (2008) Genetic analysis of 17 Y-chromosomal STRs haplotypes of Chinese Tibetan ethnic group residing in Qinghai province of China. Forensic Sci Int 175:238-243

27. Parkin EJ, Kraaijenbrink T, Opgenort JRML, van Driem GL, Tuladhar NM, de Knijff P, Jobling MA (2007) Diversity of 26-locus Y-STR haplotypes in a Nepalese population sample: Isolation and drift in the Himalayas. Forensic Sci Int 166:176-181.

28. Antunez de Mayolo G, Antúñez de Mayolo A, Antúñez de Mayolo P (2002) Phylogenetics of Worldwide human populations as determined by polymorphic Alu insertions. Electrophoresis 23:3346-3356.

29. Applied Biosystems (2006) AmpF*l*STR Yfiler PCR amplification kit user's manual. Foster City, CA: Applied Biosystems.

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

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

32. Karafet TM, Mendez FL, Meilerman MB, Underhill PA, Zegura SL, Hammer MF (2008) New binary polymorphisms reshape and increase resolution of the human Y chromosomal haplogroup tree. Genome Res 18:830-838

33. Excoffier L, Laval LG, Schneider S (2005) Arlequin ver. 3.0: An integrated software package for population genetics data analysis. Evol Bioinform Online 1:47-50

34. Brinkmann C, Forster P, Schürenkamp M, Horst J, Rolf B, Brinkmann B (1999) Human Y-chromosomal STR haplotypes in a Kurdish population sample. Int J Legal Med 112:181-183

35. Pfeiffer H, Brinkmann B, Hühne J, Rolf B, Morris AA, Steighner R, Holland MM, Forster P (1999) Expanding the forensic German mitochondrial DNA control region database: genetic diversity as a function of sample size and microgeography. Int J Legal Med 112:291-298

36. Hashiyada M, Umetsu K, Yuasa I et al (2008) Population genetics of 17 Y-chromosomal STR loci in Japanese. Forensic Sci Int Genet 2:e69-70

37. Yan J, Tang H, Liu Y et al (2007) Genetic polymorphisms of 17 Y-STRs haplotypes in Chinese Han population residing in Shandong province of China. Leg Med 94:196-202

38. Chang YM, Perumal R, Keat PY, Kuehn DL (2007) Haplotype diversity of 16 Y-chromosomal STRs in three main ethnic populations (Malays, Chinese and Indians) in Malaysia. Forensic Sci Int 167:70-76

39. Huang TY, Hsu YT, Li JM, Chung JH, Shun CT (2008) Polymorphism of 17 Y-STR loci in Taiwan population. Forensic Sci Int 174:249-54

40. Dobashi Y, Kido A, Fujitani N, Susukida R, Hara M, Oya M (2005) Y-chromosome STR haplotypes in a Bangladeshi population. Leg Med 7:122-126

41. Nagy M, Henke L, Henke J et al (2007) Searching for the origin of Romanies: Slovakian Romani, Jats of Haryana and Jat Sikhs Y-STR data in comparison with different Romani populations. Forensic Sci Int 169:19-26

42. Woźniak M, Derenko M, Malyarchuk B et al (2006) Allelic and haplotypic frequencies at 11 Y-STR loci in Buryats from South-East Siberia. Forensic Sci Int 164:271-275

43. Parkin EJ, Kraayenbrink T, van Driem GL, Tshering Of Gaselô K, de Knijff P, Jobling MA (2006) 26-Locus Y-STR typing in a Bhutanese population sample. Forensic Sci Int. 161:1-7.

44. Rohlf F (2002) NTSYSpc. Exter Publishing, Setauket, NY.

45. Felsentein J (2002) Phylogeny inference package (PHYLIP), Version 3.6a3. Department of Genetics, University of Washington, Seattle. Distributed by author.

46. Kayser M, Brauer S, Weiss G et al (2003) Reduced Y-chromosome, but not mitochondrial DNA, diversity in human populations from West New Guinea. Am J Hum Genet 72:281-302

47. Polzin T and Daneschmand SV (2002) On Steiner trees and minimum spanning trees in hypergraphs. Oper Res Lett 31:12–20

48. Ferri G, Robino C, Alù M et al (2008) Molecular characterisation and population genetics of the DYS458 .2 allelic variant. Forensic Sci Int Genet: Genetics Supplement Series, Progress in Forensic Genetics 12 - Proceedings of the 22nd International ISFG Congress 1:203-205

49. Cadenas AM, Regueiro M, Gayden T, Singh N, Zhivotovsky LA, Underhill PA, Herrera RJ (2007) Male amelogenin dropouts: phylogenetic context, origins and implications. Forensic Sci Int 166:155-163

50. Thangaraj K, Reddy AG, Singh L (2002) Is the amelogenin gene reliable for gender identification in forensic casework and prenatal diagnosis? Int J Legal Med 116:121–123

51. Santos FR, Pandya A, Tyler-Smith C (1998) Reliability of DNA-based sex tests. Nat Genet 18:103

52. Wen B, Xie X, Gao S (2004) Analyses of genetic structure of Tibeto-Burman populations reveals sex-biased admixture in southern Tibeto-Burmans. Am J Hum Genet 74:856-865

Appendix II

Figure 1A. Correspondence Analysis based on allelic frequencies of 9 Y-STR loci from 20 populations.

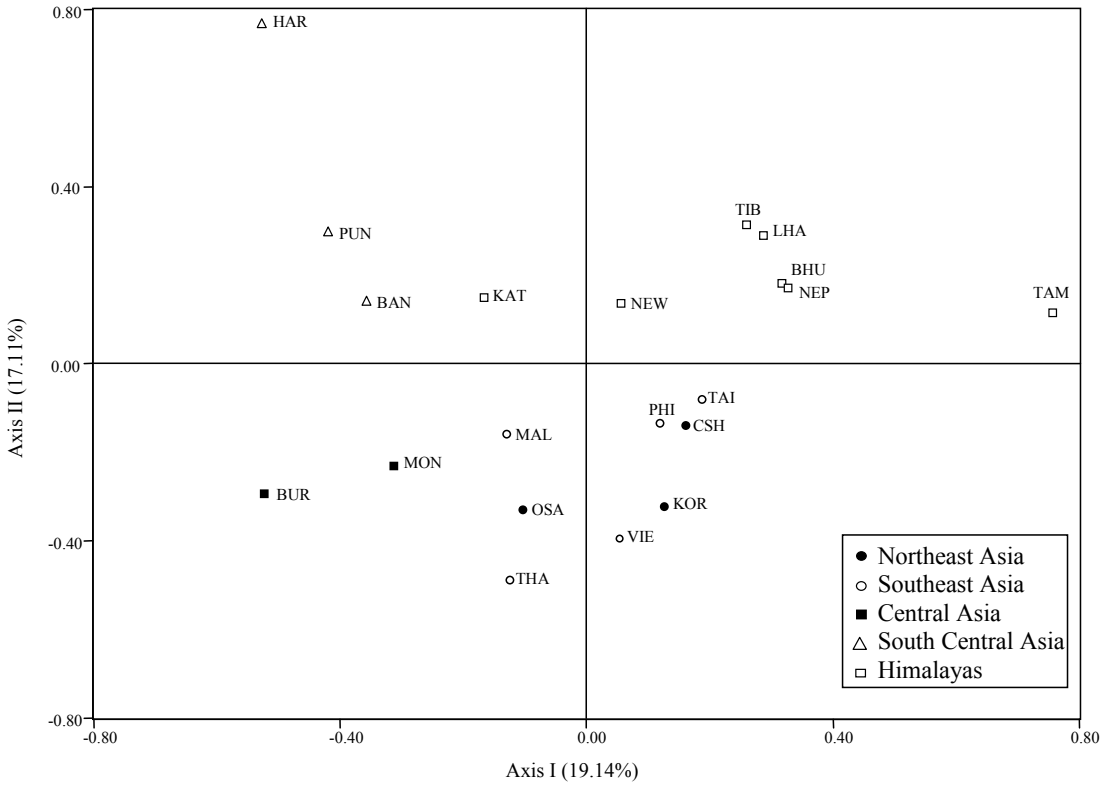

Figure 1B. Contribution of each of the ninety four alleles of the 9 Y-STR loci to the partition of populations in figure 1A.

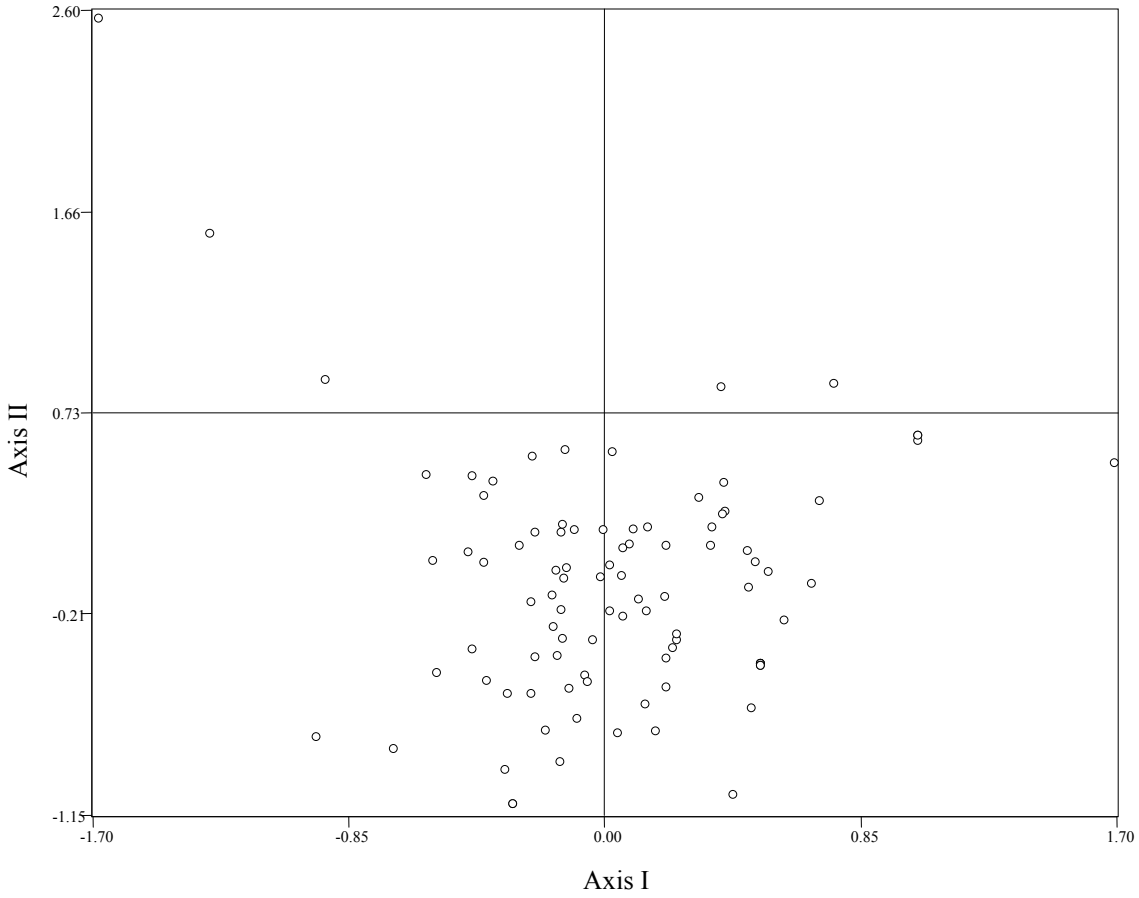

Figure 2. Neighbor-Joining (NJ) tree based on Nei's genetic distance. The numbers at the nodes represent bootstrap values estimated from 1000 replications.

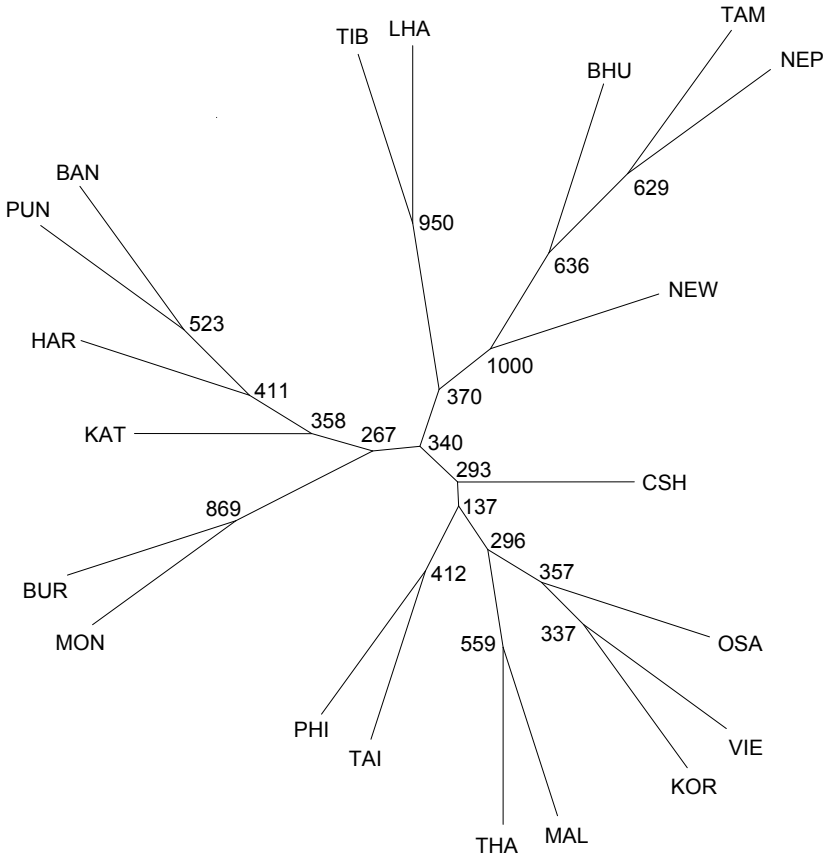

Figure 3: Median-joining network of the haplogroup O3a5-M134 lineage in the Himalayan populations. The area of the circles are proportional to the haplotype frequency and the smallest circle corresponds to one Y-chromosome

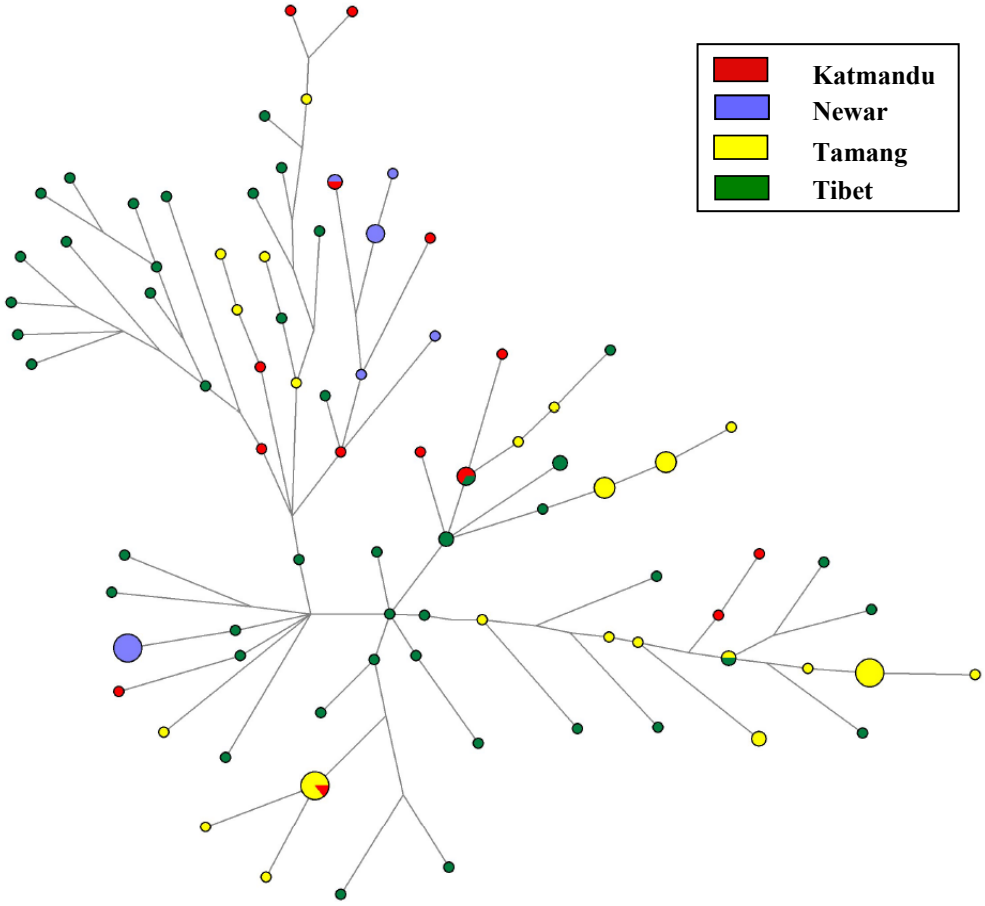

Table 1. Populations Analyzed

Population	Abbreviation	n	References
NORTHEAST ASIA			
Osaka	OSA	131	Hashiyada et al 2008 [36]
Korea	KOR	252	Kwak et al 2005 [2]
Shandong Han (NE China)	CSH	131	Yan et al. 2007 [37]
SOUTHEAST ASIA			
Malaysia	MAL	334	Chang et al. 2007 [38]
Philippines	PHI	76	Kwak et al 2005 [2]
Taiwan	TAI	200	Huang et al 2008 [39]
Thailand	THA	41	Kwak et al 2005 [2]
Vietnam	VIE	43	Kwak et al 2005 [2]
SOUTH CENTRAL ASIA			
Bangladesh	BAN	72	Dobashi et al 2005 [40]
Haryana	HAR	84	Nagy et al 2007 [41]
Punjab	PUN	80	Nagy et al 2007 [41]
CENTRAL ASIA			
Mongolia	MON	92	Kwak et al 2006 [2]
Buryat	BUR	215	Woźniak et al 2006 [42]
HIMALAYAS			
Bhutan	BHU	856	Parkin et al 2006 [43]
Nepal (General)	NEP	765	Parkin et al 2007 [27]
Newar	NEW	66	present study
Kathmandu	KAT	77	present study
Lhasa	LHA	112	Zhang et al 2006 [23]
Tamang	TAM	45	present study
Tibet	TIB	156	present study

Table 2: Parameters of forensic interest in Himalayan populations using the 9-loci, 11-loci and the Yfiler haplotypes

Haplotypes	Tamang	Newar	Kathmandu	Tibet	All Populations
9-loci Y-STR Haplotype					
Sample Size	45	66	77	156	344
Number of different Haplotypes	18	30	61	116	214
Number of Unique Haplotypes	10	18	49	95	161
Fraction of Unique Haplotypes	0.2222	0.2727	0.6364	0.6090	0.468
Discrimination Capacity	0.4000	0.4545	0.7922	0.7436	0.6221
Haplotype Diversity ± SD	0.9010 ± 0.0253	0.9585 ± 0.0097	0.9932 ± 0.0034	0.9940 ± 0.0018	0.9942 ± 0.0009
11-loci Y-STR Haplotype					
Sample Size	45	66	77	156	344
Number of Haplotypes	19	30	66	129	236
Unique Haplotypes	12	18	56	115	189
Fraction of Unique Haplotypes	0.2667	0.2727	0.7272	0.7372	0.5494
Discrimination Capacity	0.4222	0.4545	0.8571	0.8462	0.6860
Haplotype Diversity ± SD	0.9020 ± 0.0256	0.9585 ± 0.0097	0.9959 ± 0.0029	0.9967 ± 0.0013	0.9955 ± 0.0008
17-loci Y-STR Haplotype					
Sample Size[a]	45	63	72	152	332
Number of different Haplotypes	27	29	67	141	262
Number of Unique Haplotypes	21	17	63	130	228
Fraction of Unique Haplotypes	0.4667	0.2698	0.8750	0.8553	0.6867
Discrimination Capacity	0.6000	0.4603	0.9306	0.9276	0.7892
Haplotype Diversity ± SD	0.9545 ± 0.0167	0.9570 ± 0.0105	0.9977 ± 0.0029	0.9990 ± 0.0009	0.9970 ± 0.0007

[a]Excludes samples carrying null alleles and duplicated loci

Table 3. Y-STR haplotype matching probabilities within and between the Himalayan populations

Parameters	Tamang (Tam)	Newar (New)	Kathmandu (Kat)	Tibet (Tib)
N*	45	63	72	152
dw_{min}	0.9333	0.9418	0.9838	0.9926
mw_{max}	0.0667	0.0582	0.0162	0.0074
mw_{min}	45/990	56/1953	6/2556	9/11476
mb_{min}	Tam-New: 0.0000	New-Kat: 0.0002	Kat-Tib: 0.0000	Tib-Tam: 0.0000
	Tam-Kat: 0.0018	New-Tib: 0.0000	Kat-Tam: 0.0018	Tib-New: 0.0000
	Tam-Tib: 0.0000	New-Tam: 0.0000	Kat-New: 0.0002	Tib-Kat: 0.0000
mw_{max}/mb_{min}	Tam/New: 0.00	New/Kat: 291	Kat/Tib: 0.0000	Tib/Tam: 0.00
	Tam/Kat: 37.06	New/Tib: 0.00	Kat/Tam: 9.00	Tib/New: 0.00
	Tam/New: 0.00	New/Tam: 0.00	Kat/New: 81.00	Tib/Kat: 0.00
mw_{min}/mb_{min}	Tam/New: 0.00	New/Kat: 143.50	Kat/Tib: 0.00	Tib/Tam: 0.00
	Tam/Kat: 25.25	New/Tib: 0.00	Kat/Tam: 1.28	Tib/New: 0.00
	Tam/New: 0.00	New/Tam: 0.00	Kat/New: 11.50	Tib/Kat: 0.00

*Excludes samples carrying null alleles and duplicated loci

Supplementary Table 1: Y-STR data of the four Himalayan populations studied

Sample Info	YCC HG*	DYS456	DYS389I	DYS390	DYS389II	DYS458	DYS19	DYS385	DYS393	DYS391	DYS439	DYS635	DYS392	GATA	DYS437	DYS438	DYS448
NplT1	O3a3c*-M134	15	12	24	28	18	15	12-19	13	10	12	20	12	12	15	10	19
NplT2	O3a3c1-M117	16	12	24	28	16	14	13-19	12	10	12	20	15	11	15	11	20
NplT3	O3a3c1-M117	16	12	24	28	16	14	13-19	12	10	12	21	14	11	15	11	20
NplT4	O3a3c1-M117	16	12	24	28	16	14	13-19	12	10	12	20	15	11	15	11	20
NplT5	O3a3c1-M117	15	12	23	28	20	14	13-18	12	10	11	20	14	12	15	11	20
NplT6	O3a3c1-M117	15	12	23	29	19	14	12-19	12	10	12	20	14	11	15	11	20
NplT7	O3a3c1-M117	15	12	23	30	19	14	12-19	12	11	12	20	14	12	15	11	20
NplT8	O3a3c1-M117	15	12	23	29	19	14	12-19	12	10	12	20	14	11	15	11	20
NplT9	O3a3c1-M117	16	12	24	28	16	14	13-19	12	10	12	20	14	11	15	11	20
NplT10	O3a3c1-M117	15	12	24	28	17	14	13-18	12	10	12	20	13	12	15	11	20
NplT11	O3a3c1-M117	15	12	23	30	19	14	12-19	12	11	12	20	14	12	15	11	20
NplT12	O3a3c1-M117	15	12	24	28	17	14	13-18	12	10	12	20	13	12	15	11	20
NplT13	O3a3c1-M117	15	12	23	30	18	14	12-19	12	11	12	20	14	12	15	11	20
NplT14	O3a3c1-M117	15	12	23	30	18	14	12-19	12	11	12	20	14	12	15	11	20
NplT15	O3a3c1-M117	15	12	23	30	18	14	12-19	12	11	12	20	14	12	15	11	20
NplT16	R1a1*-M17	16	14	24	30	18	16	10-14	13	13	10	23	11	13	14	11	20
NplT17	O3a3c1-M117	15	12	23	29	18	14	12-19	12	10	12	20	14	11	15	11	20
NplT18	O3a3c1-M117	15	12	24	28	17	14	13-18	12	10	12	20	13	12	15	11	20
NplT19	O3a3c1-M117	18	12	23	29	19	14	14-20	12	10	13	20	14	12	15	11	20
NplT20	F*-M213	15	12	23	29	16	14	13-13	13	10	11	23	12	10	16	10	16
NplT21	O3a3c1-M117	15	12	23	29	19	14	12-19	12	10	12	20	14	11	15	11	20
NplT22	O3a3c1-M117	16	13	24	29	17	14	12-19	13	11	13	20	14	12	15	11	20
NplT23	R2-M124	15	13	23	29	18	14	13-13	14	10	11	25	10	12	16	11	19
NplT24	O3a3c1-M117	15	12	23	28	20	14	13-19	12	10	11	20	14	12	15	11	21
NplT25	O3a3c1-M117	15	12	24	28	17	14	13-18	12	10	12	20	13	12	15	11	20
NplT26	R2-M124	15	13	23	28	17	14	13-13	14	10	11	25	10	11	16	11	19
NplT28	O3a3c1-M117	14	12	24	28	17	14	13-18	12	10	12	20	13	12	15	11	20
NplT29	O3a3c1-M117	15	12	23	29	19	14	12-19	12	10	12	20	14	11	15	11	20
NplT30	O3a3c1-M117	15	12	23	28	17	14	13-19	12	10	12	20	14	11	15	11	20
NplT31	O3a3c1-M117	15	12	24	28	17	14	13-18	12	10	12	20	13	12	15	11	20
NplT32	O3a3c1-M117	15	12	24	28	16	14	13-18	12	10	12	20	13	12	15	11	20
NplT33	O3a3c1-M117	15	12	23	29	20	14	12-19	12	10	12	20	14	11	15	11	20
NplT34	O3a3c1-M117	15	12	23	30	19	14	12-19	12	11	12	20	14	12	15	11	20
NplT35	O3a3c1-M117	15	12	23	29	19	14	12-19	12	10	12	20	14	11	15	11	20
NplT36	O3a3c1-M117	15	12	24	28	17	14	14-18	12	10	12	20	13	12	15	11	20
NplT37	O3a3c1-M117	15	12	23	29	19	14	12-19	12	10	12	20	14	11	15	11	20
NplT38	O3a3c1-M117	15	12	23	30	20	14	12-19	12	11	12	20	14	12	15	11	20
NplT39	O3a3c1-M117	18	12	23	29	18	14	14-21	12	10	13	20	14	12	15	11	20
NplT40	O3a3c1-M117	15	12	23	29	19	14	12-19	12	10	12	20	14	11	15	11	20
NplT41	O3a3c1-M117	16	12	23	28	17	14	13-21	12	10	12	21	14	12	15	11	20
NplT42	O3a3c1-M117	15	12	23	28	17	14	14-19	12	10	12	20	13	12	15	11	20
NplT43	R1a1*-M17	16	14	24	30	17	16	10-14	13	13	10	23	11	13	14	11	20
NplT44	F*-M213	15	12	23	29	16	14	13-13	13	10	11	23	12	10	16	10	16

NplT45	O3a3c1-M117	15	12	23	30	18	14	12-19	12	11	12	20	14	12	15	11	20
NplT46	O3a3c1-M117	15	12	23	30	19	14	12-19	12	11	12	20	14	12	15	11	20
NplN1	R1a1*-M17	15	14	25	32	16	16	11-15	13	10	12	23	11	12	14	11	20
NplN2	J2a*-M410	14	14	23	30	18	14	15-18	12	10	11	23	11	11	14	9	21
NplN3	R2-M124	15	12	23	29	18	14	14-19	13	10	11	24	10	12	16	11	18
NplN4	R1b1*-P25	15	13	24	29	15	14	11-14	12	11	11	23	13	13	15	12	18
NplN5	R1a1*-M17	16	13	25	28	15	15	11-13	13	11	10	24	11	13	14	11	0
NplN6	J2a*-M410	14	14	23	30	19	14	15-18	12	10	11	23	11	11	14	9	21
NplN7	R2-M124	15	12	23	30	19	14	15-20	13	10	11	24	10	12	16	11	18
NplN8	R2-M124	15	12	23	28	18	14	14-19	13	10	11	24	10	12	16	11	18
NplN9	H1a*-M82	17	13	22	27	17	15	14-14	12	10	12	21	11	13	14	9	19
NplN10	J2a*-M410	14	15	23	31	18	14	15-18	12	10	11	23	11	11	14	9	21
NplN11	R1b1*-P25	15	13	24	29	15	14	11-14	12	11	11	23	13	13	15	12	18
NplN12	R1b1*-P25	15	13	24	29	15	14	11-14	12	11	11	23	13	13	15	12	18
NplN13	R2-M124	15	12	23	29	18	14	14-19	13	10	11	24	10	12	16	11	18
NplN14	C5*-M356	14	14	24	31	18	14	14-16	13	10	13	20	11	12	14	10	19
NplN15	R2-M124	15	12	23	29	18	14	14-19	13	10	11	24	10	12	16	11	18
NplN16	O3a3c1-M117	15	12	24	27	18	14	13-19	12	10	13	20	14	13	15	11	20
NplN17	H1a*-M82	15	14	22	30	18	15	14-16	12	10	12	20	11	13	14	9	20
NplN18	C5*-M356	14	14	24	31	18	14	14-16	13	10	13	20	11	12	14	10	19
NplN19	R1a1*-M17	15	14	25	32	16	16	11-14	13	9	12	23	11	12	14	11	20
NplN20	O3a3c1-M117	15	12	23	28	17	14	13-18	12	9	12	20	16	12	15	11	20
NplN21	R1a1*-M17	15	14	25	32	16	16	11-14	13	9	12	23	11	12	14	11	20
NplN22	R1a1*-M17	15	14	25	32	15	16	11-14	13	10	10	23	11	12	14	11	20
NplN23	R1a1*-M17	15	14	25	32	15	16	11-14	13	10	10	23	11	12	14	11	20
NplN24	R1a1*-M17	15	14	25	32	16	16	11-15	13	10	12	23	11	12	14	11	20
NplN25	R1a1*-M17	15	14	25	32	15	16	11-14	13	10	10	23	11	12	14	11	20
NplN26	O3a3c1-M117	15	12	24	27	18	14	13-19	12	10	13	20	14	13	15	11	20
NplN27	O3a3c1-M117	15	12	23	28	17	14	13-18	12	9	12	20	16	12	15	11	20
NplN28	R1a1*-M17	15	13	25	29	15	16	11-14	13	11	12	24	11	13	14	11	0
NplN30	R2-M124	12	14	23	29	18	17	13-17	15	10	10	24	10	12	16	12	19
NplN31	O3a3c1-M117	15	12	23	28	17	14	13-18	12	9	12	20	16	12	15	11	20
NplN32	O3a3c1-M117	15	12	23	28	17	14	13-18	12	9	12	20	16	12	15	11	20
NplN33	O3a3c1-M117	15	12	23	28	19	14	13-19	12	10	13	20	16	12	15	11	20
NplN34	H1a*-M82	17	12	22	29	18	17	15-16	12	10	11	20	11	11	14	9	19
NplN35	H1a*-M82	17	12	22	29	18	17	15-16	12	10	11	20	11	11	14	9	19
NplN36	O3a3c1-M117	15	12	24	27	16	14	13-18	12	10	13	20	14	13	15	11	20
NplN37	O3a3c1-M117	15	12	24	27	20	14	13-21	12	10	12	20	14	13	15	11	20
NplN38	R2-M124	15	13	23	29	18	14	12-18	14	10	10	24	10	12	16	11	19
NplN39	R2-M124	15	12	23	29	18	14	14-21	13	10	11	24	10	12	16	11	18
NplN40	O3a3c1-M117	15	12	24	27	19	14	13-19	12	10	13	20	14	12	15	11	20
NplN41	R2-M124	15	12	23	29	18	14	14-20	13	10	11	24	10	12	16	11	18
NplN42	R2-M124	15	12	23	29	18	14	14-20	13	10	11	24	10	12	16	11	18
NplN43	R2-M124	15	12	23	29	18	14	14-20	13	10	11	24	10	12	16	11	18
NplN44	R2-M124	15	12	23	29	18	14	14-19	13	10	11	24	10	12	16	11	18
NplN45	O3a3c1-M117	15	12	24	27	18	14	13-19	12	10	13	20	14	13	15	11	20
NplN46	R1a1*-M17	15	14	25	32	16	16	11-14	13	9	12	23	11	12	14	11	20
NplN47	R2-M124	15	12	23	29	18	14	14-20	13	10	11	24	10	12	16	11	18

NplN48	R1b1*-P25	15	13	24	29	15	14	11-14	12	11	11	23	13	13	15	12	18
NplN49	R1b1*-P25	15	13	24	29	15	14	11-14	12	11	11	23	13	13	15	12	18
NplN50	R1a1*-M17	15	14	24	33	16	17	11-14	13	10	11	23	11	13	14	11	20
NplN51	O3a3c1-M117	15	12	23	28	17	14	13-18	12	9	12	20	16	12	15	11	20
NplN52	R1a1*-M17	15	13	25	31	16	15	12-13	13	11	11	23	11	13	15	11	20
NplN53	O3a3c1-M117	15	12	23	28	17	14	13-18	12	9	12	20	16	12	15	11	20
NplN54	R1a1*-M17	15	14	25	32	16	16	11-14	13	9	12	23	11	12	14	11	20
NplN55	R1a1*-M17	15	14	25	32	16	16	11-15	13	10	12	23	11	12	14	11	20
NplN56	R2-M124	15	12	23	29	18	14	14-19	13	10	11	24	10	12	16	11	18
NplN57	R1a1*-M17	15	13	25	31	16	15	12-13	13	11	11	23	11	13	15	11	20
NplN58	J2a*-M410	14	14	23	30	18	14	15-18	12	10	11	23	11	11	14	9	21
NplN59	R2-M124	15	12	23	29	18	14	14-18	13	10	11	24	10	12	16	11	18
NplN60	R1a1*-M17	15	13	25	31	16	15	12-13	13	11	11	23	11	13	15	11	20
NplN61	R1b1*-P25	15	13	24	29	15	14	11-14	12	11	11	23	13	13	15	12	18
NplN62	J2b2*	12	13	24	29	16, 18	14	12-18	12	10	13	20	11	11	15	9	19
NplN63	O3a3c1-M117	15	12	23	28	17	14	13-18	12	9	12	20	16	12	15	11	20
NplN64	R1b1*-P25	15	13	24	29	15	14	11-14	12	11	11	23	13	13	15	12	18
NplN65	R2-M124	15	12	23	29	18	14	14-19	13	10	11	24	10	12	16	11	18
NplN66	R2-M124	12	14	23	29	16	16	13-18	15	10	10	24	10	13	16	12	19
NplN67	R1a1*-M17	15	13	25	31	16	16	11-15	13	11	10	23	11	13	14	11	20
NplK1	H1*-M52	15	14	22	30	17	15	14-15	12	10	10	23	11	12	14	11	18
NplK2	H1a*-M82	15	12	22	28	18	15	16-19	12	11	12	21	11	12	14	9	19
NplK3	H1*-M52	16	13	22	29	17	15	14-15	12	10	10	22	11	12	14	11	18
NplK4	R1a1*-M17	15	13	24	30	16	15	11-15	13	11	10	23	11	13	14	11	20
NplK5	O3a3c1-M117	15	11	23	26	19	14	13-19	12	10	12	19	15	12	14	11	20
NplK6	R1a1*-M17	15	13	25	30	15	16	11-14	13	10	10	23	11	13	14	11	20
NplK7	R1a1*-M17	15	12	25	28	16	16	11-14	13	11	10	25	11	13	14	11	20
NplK8	H1*-M52	15	14	22	30	17	15	14-15	12	10	10	23	11	12	14	11	18
NplK9	J2a*-M410	14	14	23	30	17	14	15-18	12	10	11	23	11	11	14	9	21
NplK10	R1a1*-M17	15	13	25	31	16	15	11-14	13	10	11	23	11	12	14	11	20
NplK11	C3*-M217	14	13	24	29	16	17	12-13	13	9	12	21	11	11	14	10	20
NplK12	C*-M216	15	13	24	30	17	16	11-14	13	10	10	23	11	11	14	11	19
NplK13	R2-M124	16	13	23	29	17	14	11-16	14	10	10	24	10	12	16	11	19
NplK14	R1a1*-M17	16	12	25	29	16	16	11-14	13	11	10	25	11	13	14	11	20
NplK15	R1a1*-M17	14	14	25	32	17	15	11-14	13	10	10	23	11	12	14	11	20
NplK16	J2b2*	13	12	23	28	0	15	12-16	12	10	12	21	11	11	15	9	19
NplK17	R1a1*-M17	16	13	25	31	16	16	11-14	13	12	10	23	11	13	14	11	20
NplK18	R*-M207	15	14	23	30	17	14	13-20	12	10	11	24	12	11	15	11	19
NplK19	R1a1*-M17	15	13	25	31	17	16	11-14	13	11	10	24	11	13	14	11	20
NplK20	O3a3c1-M117	15	12	23	28	18	14	14-19	12	10	13	20	14	12	15	11	19
NplK21	H1a*-M82	15	13	22	30	18	13	14-18	12	10	10	17	11	12	14	11	19
NplK22	C3*-M217	14	13	23	29	16	16	12-13	14	10	12	21	11	11	14	10	20
NplK23	R1a1*-M17	15	13	25	31	16	16	11-14	13	11	11	23	11	13	14	11	21
NplK24	R1a1*-M17	15	12	25	29	17	16	11-14	13	11	10	24	11	13	14	11	20
NplK25	R1a1*-M17	15	13	25	30	17	16	11-14	13	11	10	23	11	13	14	11	20
NplK26	R1a1*-M17	15	13	24	30	17	16	11-11	13	11	10	24	12	13	14	11	20
NplK27	R1a1*-M17	15	13	22	30	15	14	13-18	12	10	10	18	11	12	14	10	19
NplK28	J2b2*	13	12	23	28	0	15	12-16	12	10	13	21	11	11	15	9	19

NplK29	R1a1*-M17	17	13	25	30	14	16	11-14	13	10	11	23	11	13	14	11	20
NplK30	H*-M69	13	10	23	27	15	14	15-16	14	10	12	20	11	11	14	10	19
NplK31	R1a1*-M17	15	13	25	31	16	16	11-15	13	11	10	23	11	13	14	11	20
NplK32	R1a1*-M17	15	14	25	32	17	15	11-14	13	11	10	23	11	13	14	9	20
NplK33	H1*-M52	16	13	22	29	17	15	14-15	12	10	10	22	11	12	14	11	18
NplK34	R1a1*-M17	15	13	24	31	16	16	11-14	14	12	11	23	11	13	14	11	20
NplK35	O3a3c1-M117	15	12	24	28	18	15	13-16	12	9	12	20	14	11	15	11	19
NplK36	C*-M216	15	14	23	32	17	16	14-15	13	10	11	25	11	10	14	10	20
NplK37	R2-M124	12	14	23	29	18	15	13-18	14	11	11	24	10	12	16	12	19
NplK38	R2-M124	15	13	22	31	17	16	13-19	13	10	10	25	10	11	16	10	19
NplK39	O3a3c1-M117	15	12	23	28	19	14	13-19	12	10	11	20	14	12	15	11	20
NplK40	R1a1*-M17	15	13	25	31	14	16	11-14	13	11	10	23	11	13	14	11	20
NplK41	O3*-M122	15	12	24	27	18	15	15-16	13	10	13	23	13	12	15	9	18
NplK42	R1a1*-M17	16	13	25	31	16	15	11-14	13	10	10	23	11	12	14	11	20
NplK43	O3a3c*-M134	16	12	25	29	19	16	12-16	13	11	12	19	12	12	15	10	19
NplK44	R2-M124	15	13	22	31	17	15	13-19	13	10	10	25	10	11	16	10	19
NplK45	R2-M124	16	13	23	29	17	14	11-16	13	10	10	24	10	12	16	11	19
NplK46	C*-M216	15	14	23	32	17	16	14-15	13	10	12	24	11	10	14	10	20
NplK47	R1a1*-M17	15	13	25	31	16	15	11-14	13	10	10	23	11	12	14	11	20
NplK48	O3a3c1-M117	15	12	23	28	18	14	13-13	12	10	11	20	14	12	15	11	19
NplK49	C*-M216	15	14	23	32	17	16	14-15	13	11	12	25	11	10	14	10	20
NplK50	J2b2*	13	12	23	28	0	15	12-16	12	10	13	21	11	11	15	9	19
NplK51	O3a3c1-M117	15	12	23	28	19	14	13-18	12	10	13	20	14	12	15	11	20
NplK52	R1a1*-M17	15	14	25	32	16	15	11-13	13	10	10	24	11	13	14	11	20
NplK53	J2a*-M410	14	14	23	30	17	14	15-18	12	10	11	23	11	9	14	9	21
NplK54	R1a1*-M17	15	14	24	31	15	17	11-14	13	11	10	23	11	12	14	11	20
NplK55	N*-M231	15	14	25	32	15	14	11-12	13	11	11	24	14	13	14	11	18
NplK56	R1a1*-M17	16	13	25	30	16	16	12-14	13	11	10	23	13	11	14	11	20
NplK57	R2-M124	15	14	23	29	16	15	13-20	15	10	11	24	10	12	14	11	19
NplK58	O3a3c1-M117	18	12	23	28	18	14	14-20	12	10	13	20	14	12	15	11	20
NplK59	H1a*-M82	16	13	22	29	16	15	14-17	12	10	12	19	11	11	14	9	19
NplK60	O3a3c1-M117	15	12	24	28	17	14	13-18	12	10	12	20	13	12	15	11	20
NplK61	R1a1*-M17	15	12	24	30	16	15	11-13	13	11	10	23	11	13	14	10	20
NplK62	J2b2*	12	12	24	28	0	15	13-13	12	10	12	22	11	12	15	9	18
NplK63	R2-M124	15	13	22	31	17	16	13-19	13	10	10	25	10	11	16	10	19
NplK64	O3a3c1-M117	15	12	23	28	19	14	13-19	12	10	11	20	14	12	15	11	20
NplK65	O3a3c1-M117	15	12	24	28	17	15	13-16	12	9	12	20	14	11	15	11	19
NplK66	J2b2*	13	12	23	28	0	15	12-15	12	10	13	21	11	11	15	9	19
NplK67	O3a3c*-M134	16	12	24	29	17	15	12-16	13	11	11	19	12	12	15	10	19
NplK68	R2-M124	15	13	22	31	17	16	13-19	13	10	10	25	10	11	16	10	19
NplK69	N1c-TAT	14	12	24	27	20	15	11-12	13	10	11	22	14	12	14	10	20
NplK70	R1a1*-M17	16	13	24	30	16	16	12-14	13	11	10	23	11	13	14	11	20
NplK71	O3a3c1-M117	15	12	23	29	19	14	13-19	12	10	11	20	14	12	15	11	20
NplK72	R1a1*-M17	15	14	25	30	16	15	11-14	13	10	9	23	12	12	14	11	20
NplK73	R1a1*-M17	15	14	25	30	17	16	10-14	13	13	10	23	11	13	14	11	20
NplK74	O3a3c1-M117	16	12	25	27	17	14	13-19	12	10	13	20	14	12	14	11	20
NplK75	O3a3c1-M117	15	12	24	27	20	14	13-19	12	10	12	20	14	13	15	11	20
NplK76	C5*-M356	14	13	24	30	18	15	14-15	13	10	13	21	11	12	14	10	19

NplK77	J2a*-M410	14	13	23	29	17	14	15-19	12	10	11	24	11	11	14	9	21
Tib1	O3a3c1-M117	15	13	22	29	11	14	13-17	12	10	12	20	14	12	15	11	20
Tib2	D1-M15	15	13	24	29	18	14	14-19	13	10	10	21	11	11	14	10	16
Tib3	D1-M15	15	12	25	29	22	15	16-17	13	10	11	21	10	11	14	10	19
Tib4	D3*-P99	17	13	25	30	15	16	12-15	12	10	13	21	7	11	14	11	21
Tib5	N*-M231	15	14	23	31	15	15	12-12	13	11	11	22	14	13	14	11	19
Tib6	D1-M15	15	13	24	29	19	14	15-19	13	10	10	21	11	11	14	10	17
Tib7	O3a3c1-M117	15	12	23	28	18	15	13-16	12	10	14	21	14	12	16	11	19
Tib8	Q1a3*-M346	15	13	24	29	16	13	14-20	13	10	12	22	14	11	13	11	19
Tib9	H*-M69	15	13	22	29	17	16	14-19	14	10	11	18	11	10	14	10	19
Tib10	O3a3c1-M117	15	12	23	28	18	14	13-18	12	10	11	20	15	12	15	12	20
Tib11	O3a3c1-M117	15	12	23	28	18	14	13-18	12	10	11	20	15	12	15	12	20
Tib12	O3a3c1-M117	15	12	23	28	17	14	13-21	12	10	13	20	14	12	15	11	20
Tib13	O3a3c1-M117	15	12	23	29	18	15	13-17	12	10	12	21	14	12	16	11	20
Tib14	D1-M15	14	12	23	28	21	15	16-18	12	10	11	20	10	11	14	10	19
Tib15	C3*-M217	15	13	23	29	16	15	11-11	15	10	12	22	11	11	14	10	20
Tib16	O3a3c1-M117	15	12	23	28	19	15	13-17	12	10	12	20	14	12	16	11	20
Tib17	H*-M69	15	13	22	30	17	15	14-19	15	10	11	18	11	10	14	10	19
Tib18	D1-M15	15	12	25	29	20	15	16-18	13	10	11	21	10	11	14	10	19
Tib19	D1-M15	15	13	23	29	18	14	15-20	12	10	10	21	11	11	14	10	17
Tib20	Q1a1-M120	16	14	24	31	17	13	15-22	14	9	11	23	14	10	14	12	19
Tib21	D3a-P47	16	15	25	31	16	16	11-11	13	10	12	21	7	11	14	11	19
Tib22	D3a-P47	16	14	24	30	17	15	11-11	13	10	12	21	7	11	14	11	19
Tib23	N*-M231	15	14	24	29	15	14	10-12	13	10	12	24	14	13	14	11	19
Tib24	D1-M15	14	12	25	28	19	15	15-15	12	10	11	20	10	11	14	10	19
Tib25	D1-M15	16	13	24	29	17	14	14-20	13	9	10	21	11	11	14	10	17
Tib26	O3a3c1-M117	15	12	23	28	18	15	13-18	12	10	12	21	14	12	16	11	20
Tib27	D3*-P99	16	13	25	29	15	15	11-11	13	11	12	21	7	11	14	11	19
Tib28	Q1a3*-M346	15	13	24	29	16	13	14-20	13	10	12	22	14	11	13	11	19
Tib29	O3a3c1-M117	15	12	24	28	17	14	13-18	12	11	12	20	15	12	15	11	20
Tib30	D3*-P99	16	14	24	31	17	15	10-13	11	11	12	20	7	12	14	11	21
Tib31	O3a3c1-M117	15	12	23	28	17	14	13-17	12	10	12	20	15	12	15	11	20
Tib32	J2a*-M410	15	13	24	29	20	14	13-16	13	10	11	21	11	11	15	10	19
Tib33	Q1a1-M120	16	14	25	30	17	13	15-21	13	9	11	22	14	10	14	12	19
Tib34	D3a-P47	16	14	26	30	18	15	11-11	13	10	10	22	7	11	14	11	18
Tib35	D1-M15	14	14	24	30	17	14	14-19	13	10	10	21	11	11	14	10	17
Tib36	O3a3c1-M117	15	13	25	29	18	14	13-17	12	11	12	20	14	12	15	11	20
Tib37	O3a3c1-M117	15	12	23	29	20	14	13-20	12	10	11	20	14	12	15	12	21
Tib38	O3a*-M324	14	12	24	29	16	15	14-17	12	10	11	21	13	13	15	10	20
Tib39	D1-M15	15	13	24	29	19	15	14-18	12	10	10	21	11	11	14	10	17
Tib40	O3a3c1-M117	15	12	23	28	20	14	13-19	12	10	12	20	14	12	15	11	19
Tib41	O3a*-M324	14	12	25	27	20	15	12-20	12	10	12	23	13	11	14	10	19
Tib42	D3a-P47	15	13	25	29	16	15	11-11	13	10	11	21	7	11	14	11	0
Tib43	D3a-P47	16	14	26	30	18	15	11-11	13	10	10	22	7	11	14	11	18
Tib44	O3a3c1-M117	15	12	23	28	16	15	13-17	12	10	14	22	14	12	15	11	20
Tib45	O3a3c1-M117	15	12	23	28	20	14	13-19	12	10	13	20	14	12	15	11	20
Tib46	D3a-P47	15	14	25	30	17	15	11-11	13	10	13	22	7	11	14	11	19
Tib47	D3a-P47	16	14	25	30	17	15	11-11	13	10	13	22	7	11	14	11	19

Tib48	D3a-P47	15	15	26	31	16	15	11-11	13	10	12	21	7	10	14	11	19
Tib49	D1-M15	15	12	27	29	21	15	16-17	12	10	11	20	10	11	14	10	19
Tib50	D3a-P47	16	14	25	31	16	15	11-11	12	11	12	22	7	11	14	11	19
Tib51	D1-M15	16	12	27	30	21	15	17-18	12	10	11	20	10	11	14	10	19
Tib52	D3a-P47	16	14	25	30	18	16	11-11	13	10	12	21	7	11	14	11	19
Tib53	D3a-P47	15	15	26	31	16	15	11-11	13	10	12	21	7	10	14	11	19
Tib54	D1-M15	15	13	24	29	17	14	14-17	12	10	10	21	11	11	14	10	17
Tib55	N*-M231	16	14	24	30	15	14	11-12	13	11	11	22	14	12	14	11	19
Tib56	O3a3c1-M117	15	12	23	28	20	14	13-19	12	10	12	20	14	12	15	11	20
Tib57	O3a3c1-M117	15	12	23	27	19	15	13-18	12	10	12	21	14	12	15	11	20
Tib58	O3a3c1-M117	15	12	23	28	18	14	13-19	12	10	12	20	14	12	15	11	20
Tib59	O3a3c1-M117	15	12	24	28	16	14	13-19	12	10	12	21	14	12	15	11	20
Tib60	D1-M15	15	13	24	29	17	15	14-18	12	11	10	21	11	11	14	10	17
Tib61	D3a-P47	16	14	25	30	18	15	11-11	13	10	13	21	7	11	14	11	19
Tib62	D1-M15	15	13	24	29	17	14	14-19	13	10	10	21	11	11	14	10	17
Tib63	D3a-P47	16	14	26	31	16	15	11-11	13	10	12	21	7	10	14	11	19
Tib64	O3a3c1-M117	15	12	23	28	19	14	13-18	12	10	11	20	14	12	15	11	20
Tib65	D3a-P47	15	14	25	31	17	15	11-11	13	10	11	21	7	11	14	11	18
Tib66	D1-M15	15	14	25	30	18	14	14-18	12	10	10	21	11	11	14	10	17
Tib67	D3a-P47	16	14	23	30	17	15	11-11	13	11	12	21	7	11	14	11	18
Tib68	O3a3c1-M117	15	12	23	27	17	14	13-21	12	9	11	20	14	11	15	11	20
Tib69	N*-M231	15	14	22	29	17	14	12-12	13	11	12	21	14	11	14	10	19
Tib70	O3a*-M324	14	12	25	27	16	17	13-18	12	10	11	22	13	12	14	10	20
Tib71	D1-M15	15	13	24	29	17	15	14-18	12	10	10	21	11	11	14	10	17
Tib72	O3a3c1-M117	15	12	23	28	17	14	13-18	12	10	12	20	14	11	15	11	20
Tib73	D3a-P47	15	13	25	29	13	16	11-11	12	11	12	22	7	11	14	11	19
Tib74	D1-M15	15	12	27	29	22	15	17-17	12	10	12	20	10	11	14	10	19
Tib75	O3a3c1-M117	15	12	24	28	18	14	13-18	12	10	12	20	14	12	15	11	20
Tib76	D1-M15	15	13	23	29	17	14	13-18	13	10	10	21	11	11	14	10	17
Tib77	D1-M15	16	12	25	29	20	15	16-16	12	10	12	20	10	11	14	10	20
Tib78	O3a3c1-M117	15	13	23	29	18	14	14-18	13	10	13	20	14	12	15	11	20
Tib79	D1-M15	15	12	25	29	22	15	16-17	13	10	11	21	10	11	14	10	19
Tib80	D1-M15	14	13	24	29	21	15	16-16	12	10	11	20	10	11	14	11	19
Tib81	D1-M15	15	13	23	29	18	15	14-19	12	11	10	21	12	11	14	10	18
Tib82	D1-M15	15	12	27	29	22	15	16-17	12	10	11	20	10	11	14	10	19
Tib83	O3a3c1-M117	15	12	23	28	18	14	14-19	12	11	11	20	14	12	15	11	20
Tib84	R1a1*-M17	16	14	25	31	16	16	11-14	13	10	10	23	11	12	14	11	20
Tib85	O3a3c1-M117	16	13	23	29	19	15	13-16	12	10	13	21	14	12	15	11	20
Tib86	D3a-P47	17	13	25	29	17	15	11-11	13	10	12	21	7	11	14	11	19
Tib87	N*-M231	15	14	23	31	15	15	12-12	13	11	11	22	14	13	14	11	19
Tib88	R1a1*-M17	16	14	26	31	16	16	11-14	13	10	10	23	11	12	14	11	20
Tib89	C3*-M217	16	13	24	29	17	15	12-14	13	10	12	20	11	11	14	11	0
Tib90	O3a3c1-M117	15	12	25	28	20	14	12-17	12	10	12	20	14	12	15	12	20
Tib91	D1-M15	15	12	27	29	21	15	16-17	12	10	11	20	10	11	14	10	19
Tib92	O3a3c1-M117	15	12	23	28	18	14	13-18	12	10	11	20	14	12	15	11	20
Tib93	C3*-M217	15	14	23	31	17	15	11-20	14	10	11	21	11	10	14	10	21
Tib94	N*-M231	17	14	24	30	16	14	11-12	13	11	12	22	14	12	14	11	19
Tib95	D1-M15	15	13	24	30	18	16	14-18	12	10	10	21	11	11	14	10	17

Tib96	D3a-P47	16	14	25	30	17	15	11-14	13	10	11	21	7	11	14	11	19
Tib97	O3a3c1-M117	15	12	23	28	18	14	13-19	12	10	12	20	14	13	15	11	20
Tib98	O3a3c*-M134	15	12	23	27	16	15	11-17	13	10	12	19	12	13	15	11	19
Tib99	D1-M15	16	13	24	29	18	15	14-18	12	10	10	22	11	11	14	10	17
Tib100	D1-M15	15	13	23	29	18	14	15-20	12	10	10	21	11	11	14	10	17
Tib101	R1a1*-M17	15	13	25	32	16	16	11-14	13	10	10	23	11	13	14	11	20
Tib102	O3a3c1-M117	15	12	23	27	17	14	13-17	12	10	12	20	14	12	15	11	20
Tib103	D3a-P47	16	15	25	31	17	15	11-11	13	10	12	21	7	11	14	11	18
Tib104	O3a3c1-M117	15	12	23	29	17	14	13-19	12	10	12	21	14	11	14	11	20
Tib105	D3a-P47	16	13	25	29	17	15	11-14	13	10	11	21	7	11	14	11	19
Tib106	O3a3c1-M117	15	13	23	28	16	14	13-19	12	10	12	21	14	11	15	11	20
Tib107	H*-M69	15	13	22	29	17	15	14-19	15	10	11	18	11	10	14	10	19
Tib108	D1-M15	16	12	27	29	22	15	17-18	12	10	11	20	10	11	14	10	19
Tib109	D1-M15	15	13	24	29	19	14	15-19	13	10	10	21	11	11	14	10	17
Tib110	D3*-P99	15	13	23	30	18	17	12-15	12	10	12	21	7	11	14	11	21
Tib111	D1-M15	15	12	27	30	21	15	17-17	12	10	12	20	10	11	14	10	19
Tib112	D1-M15	15	13	24	29	18	14	15-19	13	10	10	21	11	11	14	10	18
Tib113	D1-M15	15	13	23	29	16	14	11-15	13	10	10	21	7	11	14	10	18
Tib114	O3a3c1-M117	16	12	23	28	18	14	13-20	13	10	11	20	14	12	15	11	20
Tib115	O3a3c1-M117	15	12	23	28	17	14	13-19	12	10	12	20	14	12	14	11	20
Tib116	O3a3c*-M134	16	13	24	29	17	14	13-19	12	10	12	21	14	12	15	11	20
Tib117	D3a-P47	15	13	24	29	15	15	11-14	13	10	12	22	7	11	14	11	20
Tib118	D1-M15	15	12	27	29	20	15	17-17	13	10	11	20	10	11	14	10	18
Tib119	D1-M15	15	13	25	30	17	14	14-19	12	10	10	23	11	11	14	10	17
Tib120	D1-M15	15	12	27	29	22	15	17-17	12	10	11	20	10	11	14	10	19
Tib121	D3a-P47	16	14	26	30	16	15	11-11	13	10	12	21	7	10	14	11	19
Tib122	O3a3c1-M117	15	13	23	29	17	15	12-16	13	10	13	19	12	12	15	10	19
Tib123	D1-M15	15	12	27	29	22	15	18-18	12	10	11	20	10	11	14	10	19
Tib124	C3*-M217	17	13	24	28	17	15	12-14	13	10	12	21	11	11	14	11	0
Tib125	R2-M124	15	13	23	29	19	14	13-17	14	10	10	25	10	9	16	11	19
Tib126	R2-M124	15	13	23	29	18	14	13-17	14	10	11	25	10	9	16	11	19
Tib127	O3a3c1-M117	16	12	23	28	17	14	13-20	12	11	12	21	14	12	15	11	20
Tib128	D1-M15	15	12	27	29	24	15	17-17	12	10	11	21	10	11	14	10	19
Tib129	D1-M15	15	13	25	29	18	15	14-19	12	10	10	21	11	11	14	10	17
Tib130	N*-M231	15	14	24	30	15	14	11-12	13	11	12	23	14	12	14	11	19
Tib131	Q1a3*-M346	15	13	24	29	17	13	14-19	13	10	13	22	14	11	13	11	19
Tib132	D3a-P47	16	14	26	30	16	15	11-11	13	10	12	21	7	10	14	11	19
Tib133	O3a3c1-M117	15	12	23	28	18	14	13-19	12	10	11	20	14	12	15	11	20
Tib134	D3a-P47	15	13	25	29	16	15	11-11	12	11	12	22	7	11	14	11	19
Tib135	D3a-P47	16	13	25	29	17	15	11-11	14	11	12	22	7	11	14	11	19
Tib136	D3a-P47	16	14	25	30	16	15	11-11	13	10	12	22	7	11	14	11	19
Tib137	O3a3c*-M134	15	12	23	28	19	14	12-15	13	11	12	20	12	12	15	11	19
Tib138	O3*-M122	15	14	22	29	18	15	11-14	14	10	11	21	13	12	15	11	18
Tib139	D1-M15	17	12	25	29	21	15	14-16	12	10	11	20	10	11	14	10	20
Tib140	D1-M15	14	12	24	28	19	15	16-16	12	10	11	20	10	11	14	10	19
Tib141	O3a3c1-M117	15	12	22	28	20	15	13-17	12	10	14	21	14	12	15	11	20
Tib142	D3a-P47	16	13	25	29	17	15	11-12	13	10	13	21	7	11	14	11	18
Tib143	O3a3c1-M117	15	12	22	28	19	15	14-17	12	10	12	21	15	12	16	11	20

Tib144	D3a-P47	16	14	25	31	17	15	11-11	13	10	12	22	7	11	14	11	18
Tib145	C3*-M217	16	13	24	29	17	15	12-14	14	10	12	20	11	11	14	11	0
Tib146	D3a-P47	15	13	25	29	16	15	11-11	13	10	12	22	7	11	14	11	19
Tib147	O3a3c1-M117	15	12	23	28	18	15	13-17	12	10	13	21	14	12	15	11	19
Tib148	O3a3c1-M117	15	12	22	29	17	14	14-19	12	10	12	20	14	12	16	11	20
Tib149	O3a3c1-M117	15	12	23	28	18	14	13-20	12	10	12	21	14	12	15	11	20
Tib150	O3a3c1-M117	15	12	23	28	18	14	18-19	13	10	13	20	14	12	15	11	20
Tib151	D1-M15	15	13	24	29	17	15	14-18	12	10	10	21	11	11	14	10	17
Tib152	D1-M15	14	12	27	29	21	15	17-17	12	10	11	20	10	11	14	10	19
Tib153	D3a-P47	15	14	25	30	16	16	11-11	13	10	13	22	7	11	14	11	19
Tib154	O3a3c1-M117	16	12	24	28	20	15	13-17	12	10	13	21	14	12	15	11	20
Tib155	O3a3c1-M117	15	12	23	28	12	14	14-18	12	10	13	21	14	12	15	11	19
Tib156	O3a3c1-M117	15	12	22	28	17	14	14-19	12	10	12	20	14	12	16	11	20

Null Allele
Duplicated Allele

Supplementary Table 2: Allelic Frequencies for the 17 Y-STR loci in the Tamang population (n=45)

Allele	DYS19	DYS389I	DYS389II	DYS390	DYS391	DYS392	DYS393	DYS437	DYS438	DYS439	DYS448	DYS456	DYS458	DYS635	GATA H4	Genotype	DYS385
10					0.7333	0.0444			0.0667	0.0444					0.0444	10-14	0.0444
11					0.2222	0.0444			0.9333	0.1333					0.3333	12-19	0.4444
12		0.8888				0.0667	0.8222			0.7556					0.5778	13-13	0.0888
13		0.0666			0.0444	0.2000	0.1333			0.0667					0.0444	13-18	0.1778
14	0.9333	0.0444				0.6000	0.0444	0.0444				0.0222				13-19	0.1333
15	0.0222		0.0222			0.0444		0.8667				0.7555				13-21	0.0222
16	0.0444		0.4889					0.0889			0.0444	0.1777	0.1556			14-18	0.0222
17			0.2889										0.2889			14-19	0.0222
18			0.2000									0.0444	0.2000			14-20	0.0222
19											0.0667		0.2667			14-21	0.0222
20											0.8667		0.0889	0.8222			
21											0.0222			0.0444			
22														0.0000			
23				0.6444										0.0889			
24				0.3556													
25														0.0444			
h	0.1293	0.2081	0.6515	0.4687	0.4202	0.6030	0.3111	0.2444	0.1273	0.4141	0.2475	0.4040	0.7909	0.3192	0.5636		0.7576

Supplementary Table 3: Allelic Frequencies for the 17 Y-STR loci in the Newar population (n=66)

Allele	DYS19	DYS389I	DYS389II	DYS390	DYS391	DYS392	DYS393	DYS437	DYS438	DYS439	DYS448	DYS456	DYS458	DYS635	GATA H4	Genotype	DYS385
9					0.1667				0.1364							11-13	0.0152
10					0.6364	0.2576			0.0303	0.1212						11-14	0.2424
11					0.1970	0.4242			0.6970	0.4697					0.1061	11-15	0.0606
12		0.4545					0.4545		0.1364	0.2727		0.0455			0.5606	12-13	0.0455
13		0.2424				0.1061	0.5000			0.1364					0.3333	12-18	0.0303
14	0.6515	0.2879	0.0152			0.0909	0.0152	0.3636				0.0909				13-17	0.0152
15	0.0909	0.0152	0.1364				0.0303	0.3788				0.8030	0.1818			13-18	0.1364
16	0.1970		0.3636			0.1212		0.2576				0.0152	0.2121			13-19	0.0758
17	0.0606		0.2424									0.0455	0.1212			13-21	0.0152
18			0.2273								0.3182		0.3939			14-14	0.0152
19			0.0152								0.1364		0.0606			14-16	0.0455
20											0.4545		0.0152	0.3030		14-18	0.0152
21											0.0606			0.0152		14-19	0.1061
22				0.0606												14-20	0.0606
23				0.4394										0.3939		14-21	0.0152
24				0.2576										0.2879		15-16	0.0303
25				0.2424												15-18	0.0606
Duplicated Locus																15-20	0.0152
16-18													0.0152				
Null											0.0303						
***h*[a]**	0.5329	0.6615	0.7497	0.6886	0.5366	0.7305	0.5506	0.6681	0.4838	0.6821	0.6592	0.3478	0.7519	0.6802	0.5720		0.9007

[a]Excludes samples carrying null alleles and duplicated loci

Supplementary Table 4: Allelic Frequencies for the 17 Y-STR loci in the Kathmandu population (n=77)

Allele	DYS19	DYS389I	DYS389II	DYS390	DYS391	DYS392	DYS393	DYS437	DYS438	DYS439	DYS448	DYS456	DYS458	DYS635	GATA H4	Genotype	DYS385
7																10-14	0.0130
9					0.0390				0.1558	0.0130					0.0130	11-11	0.0130
10		0.0130			0.6494	0.1039			0.2078	0.4416					0.0390	11-12	0.0260
11		0.0130			0.2727	0.6104			0.6234	0.2338					0.2468	11-13	0.0260
12		0.3377			0.0260	0.0649	0.3896		0.0130	0.1948		0.0260			0.4286	11-14	0.2468
13	0.0130	0.4156			0.0130	0.0390	0.5325			0.1169		0.0649			0.2727	11-15	0.0260
14	0.2597	0.2208				0.1688	0.0649	0.6494				0.1039	0.0260			11-16	0.0260
15	0.3766		0.0909			0.0130	0.0130	0.2597				0.6104	0.0649			12-13	0.0260
16	0.3247		0.4026					0.0909				0.1688	0.2468			12-14	0.0260
17	0.0260		0.2338									0.0130	0.3766	0.0130		12-15	0.0130
18			0.2727								0.0909	0.0130	0.1169	0.0130		12-16	0.0649
19											0.3377		0.0779	0.0519		13-13	0.0260
20											0.5195		0.0260	0.1688		13-16	0.0260
21											0.0519			0.1039		13-18	0.0519
22				0.1558										0.0519		13-19	0.1299
23				0.3247										0.3377		13-20	0.0260
24				0.2208										0.1558		14-15	0.1039
25				0.2987										0.1039		14-17	0.0130
Null													0.0649			14-18	0.0130
																14-19	0.0130
																14-20	0.0130
																15-16	0.0260
																15-18	0.0260
																15-19	0.0130
																16-19	0.0130
***h*[a]**	0.6934	0.6729	0.7098	0.7420	0.5082	0.5899	0.5677	0.5092	0.5509	0.7078	0.6131	0.5906	0.7496	0.8165	0.6883		0.9074

[a]Excludes samples carrying null alleles

Supplementary Table 5: Allelic Frequencies for the 17 Y-STR loci in the Tibet population (n=156)

Allele	DYS19	DYS389I	DYS389II	DYS390	DYS391	DYS392	DYS393	DYS437	DYS438	DYS439	DYS448	DYS456	DYS458	DYS635	Y GATA H4	Genotype	DYS385
7						0.2244										10-12	0.0064
9					0.0256										0.0128	10-13	0.0064
10					0.8462	0.1474			0.3397	0.1794					0.0705	11-11	0.1795
11					0.1282	0.2051	0.0064		0.6218	0.2885			0.0064		0.5449	11-12	0.0256
12		0.4167				0.0256	0.5192		0.0385	0.4103			0.0064		0.3269	11-14	0.0449
13	0.0321	0.3462				0.0256	0.4038	0.0192		0.1026			0.0064		0.0449	11-15	0.0064
14	0.3718	0.2115				0.3397	0.0513	0.6474		0.0192		0.0577				11-17	0.0064
15	0.5192	0.0256	0.0705			0.0321	0.0192	0.2756				0.6667	0.0513			11-20	0.0064
16	0.0641		0.6795					0.0577			0.0064	0.2436	0.1667			12-12	0.0192
17	0.0128		0.2308								0.1154	0.0321	0.2949			12-14	0.0192
18			0.0128								0.0769		0.2115	0.0192		12-15	0.0192
19			0.0064								0.4231		0.0769	0.0128		12-16	0.0064
20											0.3205		0.0769	0.3205		12-17	0.0064
21											0.0321		0.0513	0.4231		12-20	0.0064
22				0.0641									0.0449	0.1603		13-16	0.0192
23				0.3269										0.0449		13-17	0.0769
24				0.2372									0.0064	0.0064		13-18	0.0705
25				0.2436										0.0128		13-19	0.0705
26				0.0513												13-20	0.0256
27				0.0769												13-21	0.0128
																14-16	0.0064
																14-17	0.0192
																14-18	0.0577
																14-19	0.0833
																14-20	0.0192
																15-15	0.0064
																15-19	0.0192
																15-20	0.0128
																15-21	0.0064
																15-22	0.0064
																16-16	0.0192
																16-17	0.0321
																16-18	0.0128
																17-17	0.0385
																17-18	0.0128
																18-18	0.0064
																18-19	0.0064
Null											0.0256						
***h*[a]**	0.5907	0.6654	0.4830	0.7698	0.2687	0.7730	0.5679	0.5044	0.4997	0.7099	0.6864	0.4950	0.8226	0.6943	0.5929		0.9376

[a]Excludes samples carrying null alleles

IV. Y-CHROMOSOMAL MICROSATELLITE DIVERSITY IN THREE CULTURALLY DEFINED REGIONS OF HISTORIC TIBET

A. INTRODUCTION

The Tibetan Plateau in Central Asia has remained relatively isolated throughout history mostly because of its encapsulation on three sides by the highest mountain ranges in the world, including the KunLun and Tang La ranges to the north, the Karakoram and Ladakh Mountains to the west, and the Himalayas in the south. A break in the mountainous terrain to the east serves as a narrow migratory route into and out of the region. These unique geographical features of the plateau have played an important role in shaping the genetic landscapes of the Tibetan populations. Previous studies have revealed that the Himalayan Mountain range acts as a biased bidirectional barrier to gene flow, limiting genetic influence from India to the Tibetan highland [1, 2].

Prior to the Chinese invasion in 1959, Tibet covered approximately 2.5 million km^2, encompassing three main provinces: Amdo in the northeast, Kham in the southeast and U-Tsang in west and central Tibet. In 1965, China created the Tibet Autonomous Region (TAR), reducing Tibet's area to 1.2 million km^2, and assimilating parts of Amdo and Kham into the adjacent Chinese provinces of Qinghai, Gansu, Sichuan and Yunnan [3]. Although the official language of Tibet is Chinese, the majority of the population speaks the native Tibetan language, which belongs to the Tibeto-Burman subgroup of the Sino-Tibetan family. The people of Amdo, Kham, and U-Tsang each speak a different dialect of Tibetan. They are united, however, by their devout practice of Buddhism, introduced to the country in the seventh century C.E. during the rule of Songtsen Gampo [4].

Archaeological records indicate late Paleolithic inhabitation of the Tibetan plateau [3], while Y-chromosomal data [1, 5, 6] suggest that the peopling of the highland occurred during the Neolithic period. Recent articles on mtDNA genome diversity in Tibetan populations [7, 8] revealed evidence of successful late Paleolithic settlement on the plateau, thereby bridging the gap between the findings from genetic and archaeological studies.

The human Y-chromosome is a powerful molecular tool for forensic and population genetic studies [9]. In addition to Single Nucleotide Polymorphism (SNP) and insertion/deletion (indel) sites, the non-recombining region of the Y-chromosome (NRY) contains a number of short tandem repeat (STR) loci, which, besides their forensic applications, are now used to investigate the evolution, migration and genetic diversity of modern human populations [10-16]. In this chapter, I report the allelic frequencies for 17 Y- STR loci (DYS19, DYS389I, DYS389II, DYS390, DYS391, DYS392, DYS393, DYS385a/b, DYS437, DYS438, DYS439, DYS448, DYS456, DYS458, DYS635 and GATA H4) of three Tibetan populations. The aim of the current investigation is to evaluate the forensic and population genetic applications of the aforementioned 17 Y-chromosomal microsatellite loci in three different populations from historical Tibet, namely Amdo (N = 88), Kham (N = 109) and U-Tsang (N = 153). In addition, the data from this study were compared with previously published, geographically targeted reference populations from the Himalayas, South Central Asia, Southeast Asia, Central Asia and Northeast Asia using the 11-loci extended haplotypes (DYS19, DYS389I, DYS389II, DYS390, DYS391, DYS392, DYS393, DYS385a/b, DYS438, and DYS439) to assess genetic relationships among them.

B. MATERIALS AND METHODS

Sample collection and DNA isolation

Blood samples were collected with informed consent from 350 unrelated Tibetan males from three culturally defined regions, namely Amdo (N = 88), Kham (N = 109) and U-Tsang (N = 153). Genealogical history for at least two generations was recorded for each donor. Samples were collected in accordance with the ethical guidelines specified by the institutions involved in this study. The DNA was extracted using the standard phenol-chloroform method, ethanol-precipitated as described previously [17] and stored at -80° C.

Reference populations and previously reported Y-STR data

Allelic frequencies from a total of 23 previously published populations (Table 1) [10, 18-35] were chosen for comparison across the 11 Y-STR loci (DYS19, DYS389I, DYS389II, DYS390, DYS391, DYS392, DYS393, DYS385a/b, DYS438, and DYS439), given that the alleles for the remaining loci typed in this study (DYS437, DYS448, DYS456, DYS458, DYS635 and GATA H4) were not reported for all of the geographically-targeted reference populations. In addition, three individuals from Amdo, 20 from Kham and 133 from U-Tsang were incorporated in this study from previously reported data [10].

DNA amplification and STR genotyping

The DNA samples were amplified in a multiplex reaction at 17 Y-STR loci in an Eppendorf Master gradient cycler (Eppendorf AG, Hamburg, Germany) using the AmpF*l*STR® Yfiler kit (Applied Biosystems, Foster City, CA) [36] according to the

manufacturer's specifications. Amplicons were subsequently separated by multi-capillary electrophoresis on an ABI Prism 3130xl Genetic Analyzer using the ABI GeneScan 500 LIZ internal size standard as a basis for comparison. Fragment sizes were obtained using the software GeneMapper® v3.1 (Applied Biosystems, Foster City, CA) [36] and alleles were determined through comparison with an allelic ladder provided by the manufacturer (Applied Biosystems, Foster City, CA).

Quality Control

Our laboratory has participated in the Y-STR haplotype reference database (YHRD) [37] quality assurance exercise by typing the YHRD core loci as well as additional loci DYS437, DYS448, DYS456, DYS458, DYS635 and Y-GATA-H4 (Certificate dated: July 09, 2010). The accession numbers generated by the YHRD for the three Tibetan populations studied are YA003694 for Amdo, YA003695 for Kham and YA003696 for U-Tsang.

Statistical and phylogenetic analyses

Allelic frequencies were calculated for the three Tibetan populations [Amdo, Kham and U-Tsang] using the gene counting method [38]. Gene and haplotype diversities for each population were assessed using the software package Arlequin v3.1 [39]. Haplotype diversities were calculated at the minimal 9-, extended 11- and Yfiler 17-loci levels, excluding chromosomes carrying null alleles. Discrimination capacity (DC) and fraction of unique haplotypes (FUH) were estimated as the percent proportions of different and unique haplotypes, respectively, within a given population. Diversity parameters including dw_{min} (minimum diversity within the population), mw_{max}

(maximum matching probability within the population), mw_{min} (minimum matching probability within the population], mb_{min} [minimum matching probability between two populations) and db_{max} (maximum diversity between two populations) were calculated for the Tibetan populations as described previously [40, 41]. The ratio mw_{max}/mb_{min} gives an estimate for the upper limit of how many times more probable it is to find a match within a population rather than between two populations. The ratio mw_{min}/mb_{min} gives an estimate for the lower limit of the same parameter. The program Arlequin v3.1 [39] was utilized to determine the pairwise genetic distance (*Rst)* between a given pair of populations based on 11 Y-STR loci (DYS19, DYS389I, DYS389II, DYS390, DYS391, DYS392, DYS393, DYS385a/b, DYS438, and DYS439). The *Rst* values were ascertained at a significance level of 0.01 using 10,000 repetitions [42]. A Bonferroni adjustment ($\alpha = 0.01/325 = 0.000031$) was employed to compensate for potential type I errors. Samples carrying null alleles and microvariants were excluded from *Rst* calculations.

A correspondence analysis (CA) plot including all the 23 reference populations was generated using the program NTSYSpc 2.02i [43]. A neighbor joining (NJ) dendrogram was constructed utilizing the PHYLIP v3.6 program [44], with the statistical robustness of the phylogenetic relationships within the tree assessed using bootstrap analysis with 1,000 iterations. For all population comparisons, the repeat length of DYS389II was obtained by subtracting the size of the corresponding DYS389I allele.

C. RESULTS AND DISCUSSION

Tables 2 to 4 list the allelic frequencies and gene diversity values for the 17 Y-STR loci analyzed in the Tibetan populations from Amdo, Kham and U-Tsang, respectively. All three of the Tibetan collections exhibit a high degree of genetic heterogeneity, with at least 10 out of 17 loci possessing gene diversity values greater than 0.5. The U-Tsang population displays the highest average gene diversity (0.6415), followed by Amdo (0.6040), and Kham (0.5228). Microvariants were observed at loci DYS458 (18.2) and DYS448 (18.1) in one sample each from Amdo and Kham, respectively. Null alleles were detected at locus DYS448 in five individuals from U-Tsang and one sample from Kham. Both microvariants and null alleles were confirmed by repeating the amplification process.

Forensic parameters calculated from the minimal 9-, extended 11-, and Yfiler 17-loci haplotypes for the three Tibetan provinces, as well as the Tibetan collection as a whole, are provided in Table 5. A total of 299 haplotypes were observed at the 17-loci resolution for the entire Tibetan collection, 272 (90.9%) of which were unique (Supplementary Table 1). At the 11- and 9-loci resolutions, the total number of haplotypes decreased to 242 and 221, respectively, with only 199 (82.2%) and 117 (52.9%) unique haplotypes, respectively, demonstrating the superior discrimination power of the 17-loci profile. While all three of the populations analyzed possessed haplotype diversity values greater than 0.99 at the 17-loci resolution, at the 9- and 11-loci datasets, Amdo and Kham displayed haplotype diversity values less than 0.99. The lowest haplotype diversity was observed in Amdo at the minimal 9-loci resolution (0.9799). The latter value is relatively high compared to the corresponding values for

Nepalese populations, with the exception of Kathmandu (0.9932) [10], suggesting a high degree of genetic heterogeneity within the Tibetan populations.

Only one Yfiler profile (DYS19-15, DYS389I-14, DYS389II-30, DYS390-26, DYS391-10, DYS392-7, DYS393-13, DYS385a/b-11/11, DYS437-14, DYS438-11, DYS439-12, DYS448-19, DYS456-16, DYS458-16, DYS635-21 and GATA H4-10), present in 14 (4.0%) individuals, is shared among Amdo (7), Kham (5) and U-Tsang (2). This particular haplotype and its shorter versions are also the most common haplotype for the entire collection at all three resolutions considered herein. At the extended 11-loci resolution, a smaller version of the haplotype (DYS19-15, DYS389I-14, DYS389II-30, DYS390-26, DYS391-10, DYS392-7, DYS393-13, DYS385a/b-11/11, DYS438-11 and DYS439-12) is present in 17 (4.86%) Tibetan individuals: eight from Amdo, six from Kham and three from U-Tsang. A query of the extended haplotype against the YHRD database returned only 4 exact matches, not surprisingly from Qinghai (3) and Lhasa (1). The result attests to the genetic uniqueness of the Tibetan population and the limited gene flow from neighboring populations to the south and west.

Table 6 presents the matching probabilities within and among the three Tibetan populations. Consistent with the haplotype diversity results, Amdo (0.0183) displays the highest maximum probability of finding a match within the population (mw_{max}) followed by Kham (0.0132) and U-Tsang (0.0077). The maximum probabilities of obtaining dissimilar haplotypes (db_{max}) when sampling two individuals from Amdo and Kham, Amdo and U-Tsang, and U-Tsang and Kham are 0.9953, 0.9988 and 0.9992, respectively. These values are supported by the observed haplotype sharing between populations in the above mentioned pairs. The highest haplotype sharing occurs between

Amdo and Kham, which have five different Yfiler haplotypes in common. U-Tsang, on the other hand, is more unique, as it shares only one and two distinct Yfiler haplotypes with Amdo and Kham, respectively. The elevated db_{max} values for each population pair indicate genetic uniqueness not only within the Tibetan populations, but also among them. The high power of discrimination among the populations argues for sampling specific populations for population genetics or forensic studies.

As indicated by the ratio mw_{max}/mb_{min} (Table 6), it is more probable to find a match within Amdo, Kham or U-Tsang than between any two of these populations. Specifically, it is about 16 times more likely to find a match within Amdo or Kham than between either of these populations and U-Tsang. These particular ratios are also prominent at the lower estimate of the parameter (mw_{min}/mb_{min}; about 6 and 5 times more probable, respectively), with all other minimum probabilities close to 1 (equal likelihood of finding a match within and between two populations). These ratios indicate a greater degree of genetic heterogeneity and uniqueness in U-Tsang than in either Amdo or Kham. It is possible that U-Tsang's ancient monasteries, sacred religious and cultural sites, including the capital city Lhasa, may have attracted people from all over the Buddhist world, contributing to the genetic diversity in the region.

The phylogenetic relationships between the three Tibetan collections and 23 geographically targeted reference populations from the Himalayas, Southeast Asia, South Central Asia, Central Asia and Northeast Asia were assessed via CA (Fig. 1) and NJ (Fig. 2) analyses on the basis of their allele frequencies at 11 Y-STR loci. The three Tibetan provinces loosely cluster within the upper right quadrant of the CA plot, together with a Tibetan population from Qinghai and a collection from Lhasa. These five populations

also share the same clade on the NJ tree, with Amdo and Kham sharing the terminal node of the branch and U-Tsang diverging earlier, supporting the results of a recent study using autosomal STRs [45]. While Bhutan, Newar and Nepal form a sister clade with the Tibetan populations in the NJ tree, the CA plot indicates that these Himalayan populations possess stronger genetic ties with the East Asian populations. Moreover, Kathmandu shows genetic affinity towards the South Asian populations, as reported previously [10]. Overall, the phylogenetic relationships established in the NJ dendrogram reflect the distribution of the populations in the CA plot.

Pairwise genetic distance (*Rst)* results are presented in Table 7. Supporting the conclusions derived from the matching probabilities (Table 6), U-Tsang was found to be significantly different ($\alpha = 0.01$) from both Amdo and Kham (*Rst* = 0.0063 and 0.0065, *P<0.00001*), while no significant difference was detected between Amdo and Kham (*Rst* =0.0032, *P=0.0535*). Insignificant differences were also found between Lhasa and both the Amdo and U-Tsang collections. Subsequent to the application of the Bonferroni adjustment for type I errors ($\alpha = 0.01/325 = 0.000031$), 62 additional pairwise combinations yielded statistically insignificant genetic differences. The *P* values of pairs of populations that demonstrated insignificant genetic distances before and after the Bonferroni adjustment are indicated in bold and italics, respectively, in Table 7.

D. CONCLUSION

In summary, our results reveal a high degree of Y-chromosomal microsatellite diversity both within and among the three Tibetan provinces of Amdo, Kham and U-Tsang. The *Rst* results and NJ phylogeny suggest that the populations from Amdo and

Kham are more closely related to each other than either is to U-Tsang, possibly because of the cosmopolitan nature of the U-Tsang region and its importance as a worldwide religious center. Alternatively, the genetic affinity between Amdo and Kham populations may be attributed to similar genetic impact these two provinces experienced, given their relative geographic proximity to East Asia. The genetic differentiation observed among these populations may reflect their diverse cultural heritage as well as their ancient and complex genetic histories. The data presented in this study argues for the genetic uniqueness of these Tibetan regions and suggest independent consideration in forensic and population genetics studies.

This paper follows the ISFG guidelines [46] for publication of population data requested by the journal [47].

REFERENCES

1. T. Gayden, A.M. Cadenas, M. Regueiro, N.B. Singh, L.A. Zhivotovsky, P.A. Underhill, L. Cavalli-Sforza, R.J. Herrera, The Himalayas as a directional barrier to gene flow, Am. J. Hum. Genet. 80 (2007) 884–894.

2. T. Gayden, S. Mirabal, A.M. Cadenas, H. Lacau, T.M. Simms, D. Morlote, S. Chennakrishnaiah, R.J. Herrera, Genetic insights into the origins of Tibeto-Burman populations in the Himalayas, J. Hum. Genet. 54 (2009) 216–223.

3. M. Aldenderfer, Z. Yinong, The prehistory of the Tibetan plateau to the seventh century A.D.: Perspectives and research from China and the West since 1950, J. World. Prehistory. 18 (2004) 1–55.

4. H. Richardson, A short history of Tibet, E.P. Dutton & Co Inc., New York, 1962.

5. Y. Qian, B. Qian, B. Su, J. Yu, Y. Ke, Z. Chu, L. Shi, D. Lu, J. Chu, L. Jin, Multiple origins of Tibetan Y chromosomes, Hum. Genet. 106 (2000) 453–454.

6. B. Su, C. Xiao, R. Deka, M.T. Seielstad, D. Kangwanpong, J. Xiao, D. Lu, P. Underhill, L. Cavalli-Sforza, R. Chakraborty, L. Jin, Y chromosome haplotypes reveal prehistorical migrations to the Himalayas, Hum. Genet. 107 (2000) 582–590.

7. M. Zhao, Q.P. Kong, H.W. Wang, M.S. Peng, X.D. Xie, W.Z. Wang, Jiayang, J.G. Duan, M.C. Cai, S.N. Zhao, Cidanpingcuo, Y.Q. Tu, S.F. Wu, Y.G. Yao, H.J. Bandelt, Y.P. Zhang, Mitochondrial genome evidence reveals successful Late Paleolithic settlement on the Tibetan Plateau, Proc. Natl. Acad. Sci. USA. 106 (2009) 21230–21235.

8. Z. Qin, Y. Yang, L. Kang, S. Yan, K. Cho, X. Cai, Y. Lu, H. Zheng, D. Zhu, D. Fei, S. Li, L. Jin, H. Li; Genographic Consortium, A Mitochondrial revelation of early human migrations to the Tibetan plateau before and after the last glacial maximum, Am. J. Phys. Anthropol. 143 (2010) 555–569.

9. M.A. Jobling, C. Tyler-Smith, The human Y chromosome: an evolutionary marker comes of age, Nat. Rev. Genet. 8 (2003) 598–612.

10. T. Gayden, S. Chennakrishnaiah, J. La Salvia, S. Jimenez, M. Regueiro, T. Maloney, P.J. Persad, A. Bukhari, A. Perez, O. Stojkovic, R.J. Herrera, Y-STR diversity in the Himalayas, Int. J. Legal. Med. 125 (2011) 367–375.

11. U.D. Immel, M. Krawczak, J. Udolph, A. Richter, H. Rodig, M. Kleiber, M. Klintschar, Y-chromosomal STR haplotype analysis reveals surname-associated strata in the East-German population, Eur. J. Hum. Genet. 14 (2006) 577–582.

12. K.D. Kwak, H.J. Jin, D.J. Shin, J.M. Kim, L. Roewer, M. Krawczak, C. Tyler-Smith, W. Kim, Y-chromosomal STR haplotypes and their applications to forensic and population studies in east Asia, Int. J. Legal. Med. 119 (2005) 195–201.

13. K. Rebała, A.I. Mikulich, I.S. Tsybovsky, D. Siváková, Z. Dzupinková, A. Szczerkowska-Dobosz, Z. Szczerkowska, Y-STR variation among Slavs: evidence for the Slavic homeland in the middle Dnieper basin, J. Hum. Genet. 52 (2007) 406–414.

14. V. Rodríguez, C. Tomàs, J.J. Sánchez, J.A. Castro, M.M. Ramon, A. Barbaro, N. Morling, A. Picornell, Genetic sub-structure in western Mediterranean populations revealed by 12 Y chromosome STR loci, Int. J. Leg. Med. 123 (2009) 137–141.

15. L. Roewer, P.J. Croucher, S. Willuweit, T.T. Lu, M. Kayser, R. Lessig, P. de Knijff, M.A. Jobling, C. Tyler-Smith, M. Krawczak, Signature of recent historical events in the European Y-chromosomal STR haplotype distribution, Hum. Genet. 116 (2005) 279–291.

16. L. Roewer, S. Willuweit, C. Krüger, M. Nagy, S. Rychkov, I. Morozowa, O. Naumova, Y. Schneider, O. Zhukova, M. Stoneking, I. Nasidze, Analysis of Y chromosome STR haplotypes in the European part of Russia reveals high diversities but non-significant genetic distances between populations, Int. J. Legal. Med. 122 (2008) 219–223.

17. G. Antunez-de-Mayolo, A. Antunez-de-Mayolo, P. Antunez-de-Mayolo, S.S. Papiha, M. Hammer, J.J. Yunis, E.J. Yunis, C. Damodaran, M. Martinez de Pancorbo, J.L. Caeiro, V.P. Puzyrev, R.J. Herrera, Phylogenetics of worldwide human populations as determined by polymorphic *Alu* insertions, Electrophoresis. 23 (2002) 3346–3356.

18. E.J. Parkin, T. Kraayenbrink, G.L. van Driem, K. Tshering Of Gaselô, P. de Knijff, M.A. Jobling, 26-Locus Y-STR typing in a Bhutanese population sample, Forensic. Sci. Int. 161(2006) 1–7.

19. B. Zhu, Y. Wu, C. Shen, T. Yang, Y. Deng, X. Xun, Y. Tian, J. Yan, T. Li, Genetic analysis of 17 Y chromosomal STRs haplotypes of Chinese Tibetan ethnic group residing in Qinghai province of China, Forensic. Sci. Int. 175 (2008) 238–243.

20. Q. Zhang, J. Yan, H. Tang, Z. Jiao, Y. Liu, Genetic polymorphisms of 17 Y-STRs haplotypes in Tibetan ethnic minority group of China, Leg. Med. 8 (2006) 300–305.

21. E.J. Parkin, T. Kraayenbrink, J.R. Opgenort, G.L. van Driem, N.M. Tuladhar, P. de Knijff, M.A. Jobling, Diversity of 26-locus Y-STR haplotypes in a Nepalese population sample: isolation and drift in the Himalayas, Forensic. Sci. Int. 166 (2007) 176–181.

22. M. Woźniak, M. Derenko, B. Malyarchuk, I. Dambueva, T. Grzybowski, D. Miścicka-Sliwka, Allelic and haplotypic frequencies at 11 Y-STR loci in Buryats from South- East Siberia, Forensic. Sci. Int. 164 (2006) 271–275.

23. A. Fechner, D. Quinque, S. Rychkov, I. Morozowa, O. Naumova, Y. Schneider, S. Willuweit, O. Zhukova, L. Roewer, M. Stoneking, I. Nasidze, Boundaries and clines in the West Eurasian Y-chromosome landscape: insights from the European part of Russia, Am. J. Phys. Anthropol. 137 (2008) 41–47.

24. B. Zhu, X. Li, Z. Wang, H. Wu, Y. He, J. Zhao, Y. Liu. Y-STRs haplotypes of Chinese Mongol ethnic group using Y-PLEX™ 12, Forensic. Sci. Int. 153 (2005) 260–263.

25. H. Lacau, A. Bukhari, T. Gayden, J. La Salvia, M. Regueiro, O. Stojkovic, R.J. Herrera, Y-STR profiling in two Afghanistan populations, Leg. Med. 13 (2010) 103–108.

26. J. Yan, H. Tang, Y. Liu, Y. Jing, Z. Jiao, Q. Zhang, J. Gao, L. Shang, H. Guo, J. Yu, Genetic polymorphisms of 17 YSTRs haplotypes in Chinese Han population residing in Shandong province of China, Leg. Med. 94 (2007) 196–202.

27. H. Guo, J. Yan, Z. Jiao, H. Tang, Q. Zhang, L. Zhao, N. Hu, H. Li, Y. Liu, Genetic polymorphisms for 17 Y-chromosomal STRs haplotypes in Chinese Hui population, Leg. Med. 10 (2008) 163–169.

28. M. Hashiyada, T. Nagashima, Y. Itakura, J. Sakai, Y. Kanawaku, J. Kanetake, M. Nata, M. Funayama, 12 Y-chromosomal STR haplotypes in Japanese, Forensic. Sci. Int. 158 (2006) 204–212.

29. M.J. Park, H.Y. Lee, J. Yoo, U. Chung, S.Y. Lee, K. Shin, Forensic evaluation and haplotypes of 19 Y-chromosomal STR loci in Koreans, Forensic. Sci. Int. 152 (2005) 133–147.

30. S.P. Hu, Polymorphism of Y-chromosomal STR haplotypes in the Chaoshan Han Chinese in South China, Forensic. Sci. Int. 158 (2006) 80–85.

31. H. Zhang, L. Yun, Y. Li, J. Zhang, J. Wu, J. Yan, Y. Hou, Haplotype of 12 Y-STR loci of the PowerPlex Y-system in Sichuan Han ethnic group in west China, Forensic. Sci. Int. 175 (2008) 244–249.

32. Y.M. Chang, R. Perumal, P.Y. Keat, D.L. Kuehn, Haplotype diversity of 16 Y-chromosomal STRs in three main ethnic populations [Malays, Chinese and Indians] in Malaysia, Forensic. Sci. Int. 167 (2007) 70–76.

33. T.Y. Huang, Y.T. Hsu, J.M. Li, J.H. Chung, C.T. Shun, Polymorphism of 17 Y-STR loci in Taiwan population, Forensic. Sci. Int. 174 (2008) 249–254.

34. S. Alam, M. Eunus Ali, A. Ferdous, T. Hossain, M.M. Hasana, S. Akhteruzzaman. Haplotype diversity of 17 Y-chromosomal STR loci in the Bangladeshi population. Forensic. Sci. Int: Genet. 4 (2010) e59–e60.

35. M. Nagy, L. Henke, J. Henke, P.K. Chatthopadhyay, A. Völgyi, A. Zalán, O. Peterman, J. Bernasovská, H. Pamjav, Searching for the origin of Romanies: Slovakian Romani, Jats of Haryana and Jat Sikhs YSTR data in comparison with different Romani populations, Forensic. Sci. Int. 169 (2007) 19–26.

36. Applied Biosystems. AmpFlSTR Yfiler PCR amplification kit User's manual. Applied Biosystems, Foster City, California, USA, 2006.
37. S. Willuweit, L. Roewer, International Forensic Y Chromosome User Group, Y chromosome haplotype reference database (YHRD): update. Forensic. Sci. Int. Genet. 1 (2007) 83–87.

38. C.C. Li, First course in population genetics. Pacific Grove: Boxwood Press, 1976.

39. L. Excoffier, L.G. Laval, S. Schneider, Arlequin ver. 3.0: an integrated software package for population genetics data analysis, Evol. Bioinform. Online. 1 (2005) 47–50.

40. C. Brinkmann, P. Forster, M. Schürenkamp, J. Horst, B. Rolf, B. Brinkmann, Human Y-chromosomal STR haplotypes in a Kurdish population sample, Int. J. Leg. Med. 112 (1999) 181–183.

41. H. Pfeiffer, B. Brinkmann, J. Hühne, B. Rolf, A.A. Morris, R. Steighner, M.M. Holland, P. Forster, Expanding the forensic German mitochondrial DNA control region database: genetic diversity as a function of sample size and microgeography, Int. J. Leg. Med. 112 (1999) 291–298.

42. M. Kayser, S. Brauer, H. Schädlich, M. Prinz, M.A. Batzer, P.A. Zimmerman, B.A. Boatin, M. Stoneking, Y chromosome STR haplotypes and the genetic structure of U.S. populations of African, European, and Hispanic ancestry, Genome. Res. 13 (2003) 624–634.

43. F. Rohlf, NTSYSpc, Exter, Setauket, 2002.

44. J. Felsentein, Phylogeny inference package (PHYLIP), version 3.6a3, Department of Genetics, University of Washington, Seattle, 2002.

45. L. Kang, S. Li, S. Gupta, Y. Zhang, K. Liu, J. Zhao, L. Jin, H. Li, Genetic structures of the Tibetans and the Deng people in the Himalayas viewed from autosomal STRs, J. Hum. Genet. 55 (2010) 270-277.

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

47. A. Carracedo, J.M. Butler, L. Gusmao, W. Parson, L. Roewer and P.M. Schneider, Publication of population data for forensic purposes, Forensic Sci. Int. Genet. 4 (2010) 145–147.

Appendix III

Figure 1. Correspondence analysis based on allele frequencies of 11 Y-STR loci from 26 populations.

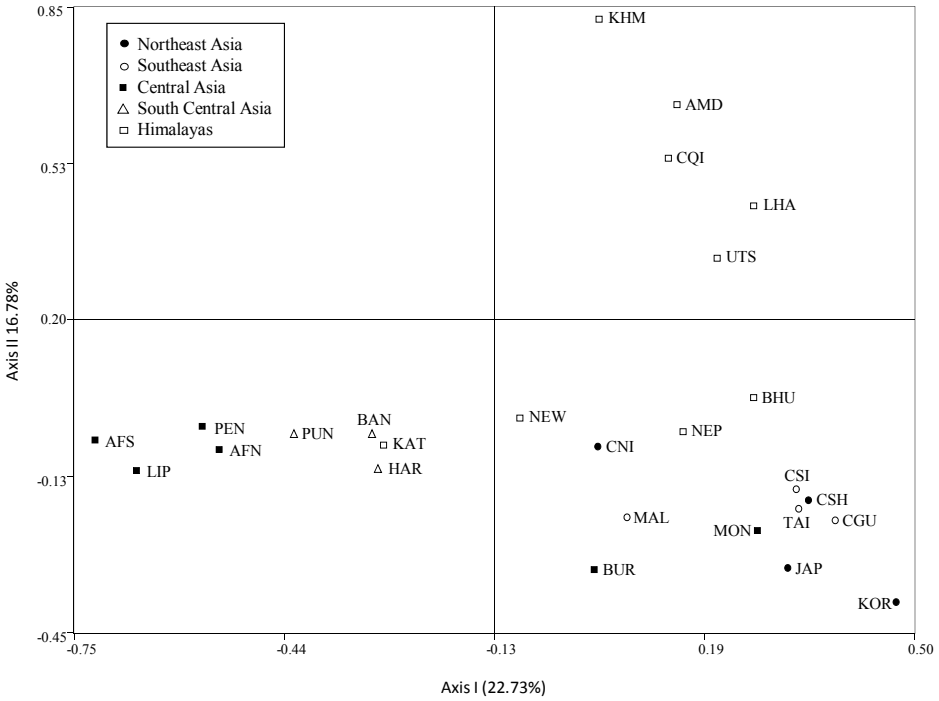

Figure 2. NJ tree based on Nei's genetic distance. Numbers at the nodes represent bootstrap values estimated from 1,000 iterations.

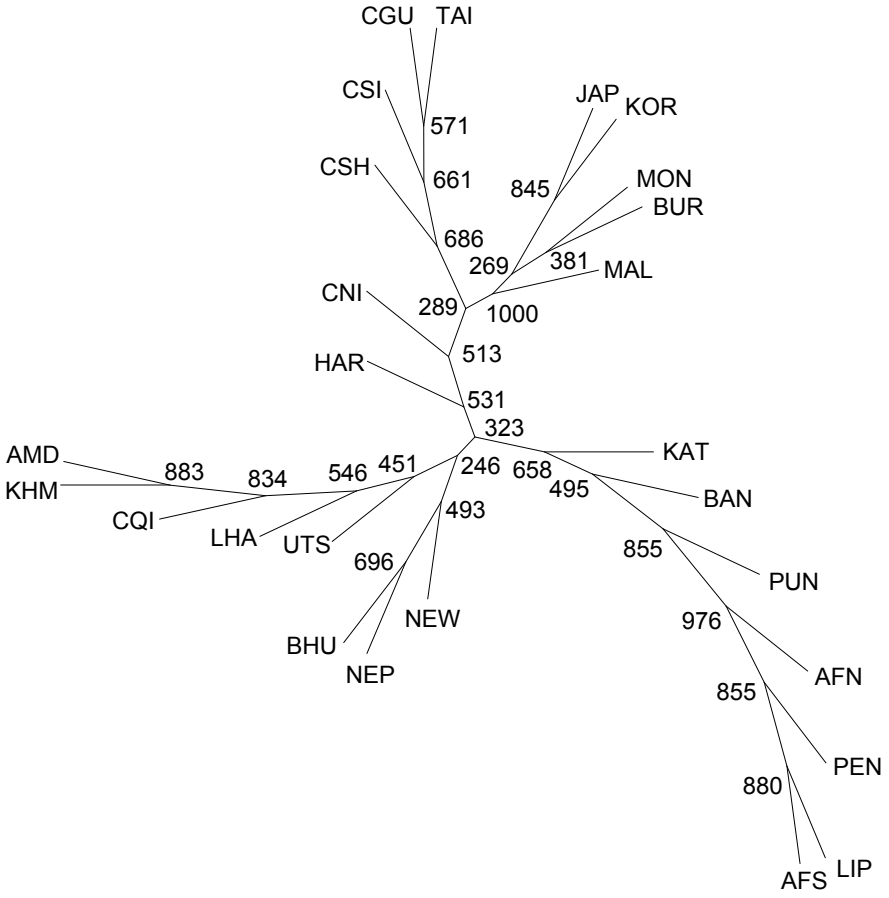

Table 1. Populations analyzed

Population	Abbreviation	n	References
HIMALAYAS			
Amdo (Tibet)	AMD	88	Present study
Kham (Tibet)	KHM	109	Present study
U-Tsang (Tibet)	UTS	153	Present study
Bhutan	BHU	856	Parkin et al. 2006 [24]
China Qinghai	CQI	167	Zhu et al. 2008 [25]
Kathmandu (Nepal)	KAT	77	Gayden et al. 2007 [10]
Lhasa	LHA	112	Zhang et al. 2006 [26]
Newar (Nepal)	NEW	66	Gayden et al. 2010 [10]
Nepal	NEP	769	Parkin et al. 2007 [27]
CENTRAL ASIA			
Buryat	BUR	215	Wozniak et al. 2006 [28]
Lipezkaja (Russia)	LIP	47	Fechner et al. 2008 [29]
Mongolia	MON	96	Zhu et al. 2005 [30]
North Afghanistan	AFN	43	Lacau et al. 2010 [31]
Pensenskaja (Russia)	PEN	81	Fechner et al. 2008 [29]
South Afghanistan	AFS	146	Lacau et al. 2010 [31]
NORTHEAST ASIA			
China Shangdong	CSH	131	Yan et al. 2007 [32]
China Ningxia	CNI	143	Guo et al. 2008 [33]
Japan	JAP	381	Hashiyada et al. 2006 [34]
Korea	KOR	301	Park et al. 2005 [35]
SOUTHEAST ASIA			
China Guangdong	CGU	120	Hu 2006 [36]
China Sichuan	CSI	237	Zhang et al. 2008 [37]
Malaysia	MAL	334	Chang et al. 2007 [38]
Taiwan	TAI	200	Huang et al. 2008 [39]
SOUTH CENTRAL ASIA			
Bangladesh	BAN	216	Alam et al. 2010 [40]
Haryana	HAR	84	Nagy et al. 2007 [41]
Punjab	PUN	80	Nagy et al. 2007 [41]

Table 2. Allelic frequencies of 17 Y-STR loci in Amdo (n=88)

allele	DYS19	DYS389I	DYS389II	DYS390	DYS391	DYS392	DYS393	DYS437	DYS438	DYS439	DYS448	DYS456	DYS458	DYS635	Y GATA H4	Genotype	DYS385a/b
7						0.4659										9,14	0.0114
9					0.0227										0.0114	11,11	0.3068
10					0.9205	0.0909			0.2045	0.0796					0.2160	11,13	0.0341
11					0.0568	0.1364			0.7614	0.3295			0.0114		0.4432	11,14	0.1364
12		0.3068					0.3523		0.0341	0.5114					0.2727	11,15	0.0227
13	0.0227	0.2614				0.0795	0.5455			0.0682		0.0227			0.0568	11,16	0.0227
14	0.1932	0.4318				0.1932	0.0909	0.7159				0.0682	0.0227			12,12	0.0227
15	0.6932					0.0227	0.0114	0.2386		0.0114		0.4091	0.0795			12,14	0.0227
16	0.0682					0.0114		0.0455				0.4773	0.4091			12,18	0.0114
17	0.0227										0.0114	0.0227	0.1705			12,19	0.0227
18											0.0114		0.1705			13,15	0.0114
18.2													0.0114			13,17	0.0568
19											0.6250		0.0341	0.0114		13,18	0.0114
20											0.3068		0.0341	0.2045		13,19	0.0795
21				0.0114							0.0455		0.0114	0.5341		13,20	0.0568
22				0.0341									0.0455	0.1250		14,17	0.0114
23				0.2159										0.1250		14,18	0.0114
24				0.1705												14,20	0.0114
25				0.3750												15,17	0.0341
26				0.1591												15,19	0.0114
27			0.0909	0.0341												15,22	0.0114
28			0.1477													16,17	0.0227
29			0.2159													16,18	0.0227
30			0.3864													16,19	0.0114
31			0.1477													17,17	0.0227
32			0.0114														
h	0.4820	0.6586	0.7607	0.7646	0.1507	0.7200	0.5765	0.4334	0.3817	0.6259	0.5188	0.6061	0.7717	0.6489	0.6870		0.8767

Table 3. Allelic Frequencies for 17 Y-STR loci in Kham (n= 109)

Allele	DYS19	DYS389I	DYS389II	DYS390	DYS391	DYS392	DYS393	DYS437	DYS438	DYS439	DYS448	DYS456	DYS458	DYS635	Y GATA H4	Genotype	DYS385a/b
7						0.6514										10,14	0.0092
9					0.0183											11,11	0.4037
10					0.8807	0.0826			0.2477	0.0550					0.1560	11,12	0.0367
11					0.0826	0.1101			0.7339	0.2844					0.7339	11,14	0.1927
12		0.1193			0.0183	0.0183	0.1651		0.0183	0.5688					0.0917	11,15	0.0367
13	0.0367	0.3670				0.0459	0.7431	0.0092		0.0826					0.0183	11,17	0.0183
14	0.0917	0.4954				0.0826	0.0734	0.9266		0.0092		0.0642				12,12	0.0183
15	0.6789	0.0183				0.0092	0.0183	0.0642				0.3853	0.0734			12,14	0.0275
16	0.1743											0.4771	0.4679			12,16	0.0183
17	0.0183										0.0459	0.0642	0.2844			12,18	0.0092
18											0.0642	0.0092	0.0642			13,13	0.0092
18.1											0.0092					13,14	0.0092
19											0.6881		0.0550	0.0183		13,17	0.0092
20											0.1560		0.0275	0.1193		13,18	0.0275
21											0.0183		0.0183	0.5963		13,19	0.0183
22				0.0275							0.0092		0.0092	0.1743		13,20	0.0092
23				0.1101										0.0917		14,16	0.0092
24				0.2385												14,17	0.0092
25				0.4862												14,18	0.0183
26			0.0092	0.1193												15,15	0.0183
27				0.0183												15,16	0.0183
28			0.0550								0.0092					15,17	0.0092
29			0.2936													15,18	0.0092
30			0.5505													15,20	0.0092
31			0.0550													15,22	0.0092
32			0.0275													16,16	0.0183
33			0.0092													17,17	0.0183
Null																	
h [a]	0.5032	0.6109	0.6094	0.6855	0.2188	0.5525	0.4186	0.1385	0.4033	0.5911	0.4907	0.6213	0.6928	0.5965	0.4322		0.7990

[a]Excludes samples carrying null alleles

Table 4. Allelic Frequencies for 17 Y-STR loci in U-Tsang (n= 153)

Allele	DYS19	DYS389I	DYS389II	DYS390	DYS391	DYS392	DYS393	DYS437	DYS438	DYS439	DYS448	DYS456	DYS458	DYS635	Y GATA H4	Genotype	DYS385a/b
7						0.6514										10,14	0.0092
9					0.0183											11,11	0.4037
10					0.8807	0.0826			0.2477	0.0550					0.1560	11,12	0.0367
11					0.0826	0.1101			0.7339	0.2844					0.7339	11,14	0.1927
12		0.1193			0.0183	0.0183	0.1651		0.0183	0.5688					0.0917	11,15	0.0367
13	0.0367	0.3670				0.0459	0.7431	0.0092		0.0826					0.0183	11,17	0.0183
14	0.0917	0.4954				0.0826	0.0734	0.9266		0.0092		0.0642				12,12	0.0183
15	0.6789	0.0183				0.0092	0.0183	0.0642				0.3853	0.0734			12,14	0.0275
16	0.1743											0.4771	0.4679			12,16	0.0183
17	0.0183										0.0459	0.0642	0.2844			12,18	0.0092
18											0.0642	0.0092	0.0642			13,13	0.0092
18.1											0.0092					13,14	0.0092
19											0.6881		0.0550	0.0183		13,17	0.0092
20											0.1560		0.0275	0.1193		13,18	0.0275
21											0.0183		0.0183	0.5963		13,19	0.0183
22				0.0275							0.0092		0.0092	0.1743		13,20	0.0092
23				0.1101										0.0917		14,16	0.0092
24				0.2385												14,17	0.0092
25				0.4862												14,18	0.0183
26			0.0092	0.1193												15,15	0.0183
27				0.0183												15,16	0.0183
28			0.0550								0.0092					15,17	0.0092
29			0.2936													15,18	0.0092
30			0.5505													15,20	0.0092
31			0.0550													15,22	0.0092
32			0.0275													16,16	0.0183
33			0.0092													17,17	0.0183
Null																	
h [a]	0.5032	0.6109	0.6094	0.6855	0.2188	0.5525	0.4186	0.1385	0.4033	0.5911	0.4907	0.6213	0.6928	0.5965	0.4322		0.7990

[a]Excludes samples carrying null alleles

Table 5. Forensic parameters of the Tibetan populations using the minimal, extended and the Yfiler haplotypes

Haplotypes	Amdo	Kham	U-Tsang	All Populations
Minimal 9 loci Y-STR Haplotype				
Sample Size	88	109	153	350
Number of Haplotypes	60	78	113	221
Unique Haplotypes	48	68	92	117
Fraction of Unique Haplotypes	54.55	62.39	60.13	33.43
Discrimination Capacity	68.18	71.56	73.86	63.14
Haplotype Diversity ± SD	0.9799 ± 0.0070	0.9862 ± 0.0044	0.9937 ± 0.0019	0.9909 ± 0.0017
Extended 11 loci Y-STR Haplotype				
Sample Size	88	109	153	350
Number of Haplotypes	64	78	125	242
Unique Haplotypes	53	68	108	199
Fraction of Unique Haplotypes	60.23	62.39	70.59	56.86
Discrimination Capacity	72.73	71.56	81.70	69.14
Haplotype Diversity ± SD	0.9835 ± 0.0064	0.9862 ± 0.0044	0.9962 ± 0.0015	0.9935 ± 0.0013
Y-filer 17 loci Y-STR Haplotype				
Sample Size[a]	88	108	148	344
Number of Haplotypes	76	95	138	299
Unique Haplotypes	69	88	128	272
Fraction of Unique Haplotypes	78.41	81.48	86.49	79.07
Discrimination Capacity	86.36	87.96	93.24	86.92
Haplotype Diversity ± SD	0.9929 ± 0.0045	0.9960 ± 0.0024	0.9991 ± 0.0009	0.9978 ± 0.0009

[a]Excludes samples carrying null alleles

Table 6. Y-STR haplotype matching probabilities within and between the Tibetan populations

Parameters	Amdo (Amd)	U-Tsang (Uts)	Kham (Khm)
N	88	148	108
dw_{min}(Haplotype diversity)	0.9817	0.9923	0.9868
mw_{max} (1-dw_{min})	0.0183	0.0077	0.0132
mw_{min}	0.0071	0.0009	0.0040
mb_{min}	Amd/Uts 0.0012 Amd/Khm 0.0047	Uts/Amd 0.0012 Uts/Khm 0.0008	Khm/Amd 0.0047 Khm/Uts 0.0008
db_{max} =(1-mb_{min})	Amd/Uts 0.9988 Amd/Khm 0.9953	Uts/Amd 0.9988 Uts/Khm 0.9992	Khm/Amd 0.9953 Khm/Uts 0.9992
mw_{min}/mb_{min}	Amd/Uts 6.1241 Amd/Khm 1.4897	Uts/Amd 0.7661 Uts/Khm 1.2245	Khm/Amd 0.8469 Khm/Uts 4.9758
mw_{max}/mb_{min}	Amd/Uts 15.9212 Amd/Khm 3.8727	Uts/Amd 6.3915 Uts/Khm 10.2162	Khm/Amd 2.8092 Khm/Uts 16.504

Table 7: *Rst* values (above diagonal) and associated p-values (below diagonal) between pairs of Tibetan and reference populations based on 10,000 repetitions at significance level 0.01

	AMD	UTS	KHM	NEW	KAT	LHA	NEP	BHU	CQI	TAI	MAL	CGU	CSI	BAN	HAR	PUN	BUR	AFN	AFS	LIP	PEN	JAP	KOR	CSH	CNI	MON
AMD	*	0.0063	0.0032	0.0319	0.0105	0.0027	0.0096	0.0117	0.0040	0.0087	0.0084	0.0095	0.0080	0.0087	0.0237	0.0143	0.0821	0.0191	0.0402	0.0099	0.0100	0.0099	0.0091	0.0084	0.0084	0.0092
UTS	0.0000	*	0.0065	0.0249	0.0041	0.0012	0.0025	0.0039	0.0016	0.0024	0.0020	0.0031	0.0019	0.0023	0.0170	0.0079	0.0732	0.0123	0.0330	0.0034	0.0036	0.0036	0.0028	0.0020	0.0021	0.0028
KHM	**0.0535**	0.0000	*	0.0303	0.0089	0.0035	0.0081	0.0103	0.0041	0.0074	0.0071	0.0081	0.0067	0.0074	0.0222	0.0131	0.0799	0.0177	0.0386	0.0085	0.0087	0.0086	0.0079	0.0070	0.0072	0.0079
NEW	0.0000	0.0000	0.0000	*	0.0239	0.0249	0.0233	0.0264	0.0236	0.0233	0.0229	0.0243	0.0230	0.0231	0.0383	0.0295	0.0999	0.0347	0.0561	0.0251	0.0250	0.0245	0.0236	0.0231	0.0236	0.0241
KAT	0.0000	0.0000	0.0000	0.0000	*	0.0033	0.0027	0.0054	0.0029	0.0026	0.0019	0.0034	0.0021	0.0022	0.0162	0.0058	0.0757	0.0115	0.0317	0.0024	0.0023	0.0039	0.0031	0.0023	0.0025	0.0031
LHA	**0.0355**	**0.0342**	*0.0054*	0.0000	*0.0003*	*	0.0026	0.0045	0.0008	0.0022	0.0019	0.0030	0.0015	0.0022	0.0169	0.0077	0.0740	0.0113	0.0310	0.0032	0.0034	0.0035	0.0028	0.0020	0.0023	0.0027
NEP	0.0000	0.0000	0.0000	0.0000	*0.0010*	*0.0001*	*	0.0043	0.0024	0.0020	0.0017	0.0027	0.0015	0.0018	0.0163	0.0072	0.0690	0.0117	0.0311	0.0029	0.0031	0.0033	0.0023	0.0016	0.0021	0.0025
BHU	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	*	0.0044	0.0041	0.0038	0.0047	0.0036	0.0041	0.0185	0.0095	0.0711	0.0138	0.0334	0.0050	0.0052	0.0054	0.0046	0.0039	0.0044	0.0046
CQI	*0.0019*	*0.0001*	0.0000	0.0000	*0.0002*	**0.0699**	0.0000	0.0000	*	0.0012	0.0009	0.0019	0.0009	0.0013	0.0158	0.0068	0.0716	0.0112	0.0318	0.0022	0.0025	0.0025	0.0017	0.0009	0.0013	0.0017
TAI	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	*	0.0005	0.0002	0.0003	0.0009	0.0154	0.0065	0.0708	0.0109	0.0313	0.0019	0.0021	0.0022	0.0014	0.0006	0.0010	0.0013
MAL	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	*0.0001*	*	0.0014	0.0004	0.0005	0.0150	0.0060	0.0691	0.0101	0.0297	0.0016	0.0017	0.0019	0.0012	0.0003	0.0008	0.0011
CGU	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	**0.2613**	0.0000	*	0.0009	0.0017	0.0163	0.0073	0.0731	0.0117	0.0325	0.0027	0.0029	0.0029	0.0021	0.0013	0.0017	0.0020
CSI	0.0000	0.0000	0.0000	0.0000	*0.0001*	*0.0001*	0.0000	0.0000	0.0000	**0.0223**	0.0000	*0.0038*	*	0.0008	0.0152	0.0063	0.0701	0.0107	0.0310	0.0017	0.0020	0.0020	0.0012	0.0004	0.0008	0.0012
BAN	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	*	0.0151	0.0057	0.0706	0.0098	0.0290	0.0015	0.0019	0.0022	0.0015	0.0006	0.0011	0.0014
HAR	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	*	0.0139	0.0900	0.0235	0.0445	0.0150	0.0151	0.0167	0.0159	0.0152	0.0151	0.0161
PUN	0.0000	0.0000	0.0000	0.0000	*0.0001*	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	*	0.0798	0.0107	0.0297	0.0032	0.0039	0.0077	0.0070	0.0062	0.0061	0.0070
BUR	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	*	0.0868	0.1064	0.0759	0.0745	0.0705	0.0703	0.0717	0.0721	0.0731
AFN	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	*0.0014*	0.0000	*	0.0053	0.0107	0.0072	0.0122	0.0114	0.0107	0.0103	0.0115
AFS	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	**0.1010**	*	0.0312	0.0274	0.0323	0.0317	0.0313	0.0309	0.0318
LIP	*0.0011*	*0.0038*	*0.0009*	0.0000	**0.0404**	*0.0053*	*0.0058*	*0.0028*	*0.0072*	*0.0025*	*0.0011*	*0.0058*	*0.0016*	**0.0162**	*0.0003*	**0.0721**	0.0000	*0.0005*	*0.0002*	*	0.0005	0.0032	0.0024	0.0014	0.0018	0.0024
PEN	0.0000	*0.0001*	0.0000	0.0000	**0.0159**	0.0000	*0.0002*	*0.0002*	*0.0001*	0.0000	0.0000	0.0000	0.0000	*0.0002*	0.0000	*0.0089*	0.0000	*0.0046*	0.0000	**0.3316**	*	0.0034	0.0027	0.0019	0.0020	0.0026
JAP	0.0000	0.0000	0.0000	0.0000	*0.0008*	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	*0.0001*	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	*0.0069*	*0.0008*	*	0.0021	0.0019	0.0025	0.0025
KOR	0.0000	0.0000	0.0000	0.0000	*0.0001*	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	*0.0065*	0.0000	0.0000	*	0.0011	0.0015	0.0017
CSH	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	*0.0001*	0.0000	*0.0002*	*0.0006*	*0.0093*	0.0000	*0.0069*	*0.0033*	0.0000	0.0000	0.0000	0.0000	0.0000	*0.0049*	0.0000	*0.0007*	*0.0010*	*	0.0009	0.0011
CNI	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	*0.0001*	0.0000	0.0000	*0.0001*	0.0000	*0.0005*	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	**0.0178**	*0.0012*	*0.0001*	*0.0001*	*0.0002*	*	0.0016
MON	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	*0.0002*	0.0000	0.0000	*0.0005*	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	*0.0071*	*0.0001*	*0.0005*	*0.0010*	*0.0001*	0.0000	*

P-values in bold and italics represent statistically insignificant differences before and after applying the Bonferroni adjustment, respectively.

Supplementary Table 1: Haplotype data for Amdo (88), Kham (109) and U-Tsang (153) populations studied

Haplotype	DYS456	DYS389	DYS390	DYS389II	DYS458	DYS19	DYS385a	DYS385b	DYS393	DYS391	DYS439	DYS635	DYS392	GATAH4	DYS437	DYS438	DYS448	Amdo	U-Tsang	Kham
1	15	13	22	29	11	14	13	17	12	10	12	20	14	12	15	11	20	1		
2	15	13	24	29	18	14	14	19	13	10	10	21	11	11	14	10	16		1	
3	15	12	25	29	22	15	16	17	13	10	11	21	10	11	14	10	19		2	
4	17	13	25	30	15	16	12	15	12	10	13	21	7	11	14	11	21		1	
5	15	14	23	31	15	15	12	12	13	11	11	22	14	13	14	11	19		2	
6	15	13	24	29	19	14	15	19	13	10	10	21	11	11	14	10	17		2	
7	15	12	23	28	18	15	13	16	12	10	14	21	14	12	16	11	19		1	
8	15	13	24	29	16	13	14	20	13	10	12	22	14	11	13	11	19		2	
9	15	13	22	29	17	16	14	19	14	10	11	18	11	10	14	10	19		1	
10	15	12	23	28	18	14	13	18	12	10	11	20	15	12	15	12	20		2	
11	15	12	23	28	17	14	13	21	12	10	13	20	14	12	15	11	20		1	
12	15	12	23	29	18	15	13	17	12	10	12	21	14	12	16	11	20		1	
13	14	12	23	28	21	15	16	18	12	10	11	20	10	11	14	10	19	1	1	
14	15	13	23	29	16	15	11	11	15	10	12	22	11	11	14	10	20			1
15	15	12	23	28	19	15	13	17	12	10	12	20	14	12	16	11	20		1	
16	15	13	22	30	17	15	14	19	15	10	11	18	11	10	14	10	19		1	
17	15	12	25	29	20	15	16	18	13	10	11	21	10	11	14	10	19		1	
18	15	13	23	29	18	14	15	20	12	10	10	21	11	11	14	10	17		2	
19	16	14	24	31	17	13	15	22	14	9	11	23	14	10	14	12	19	1		
20	16	15	25	31	16	16	11	11	13	10	12	21	7	11	14	11	19		1	
21	16	14	24	30	17	15	11	11	13	10	12	21	7	11	14	11	19		1	1
22	15	14	24	29	15	14	10	12	13	10	12	24	14	13	14	11	19		1	
23	14	12	25	28	19	15	15	15	12	10	11	20	10	11	14	10	19			2
24	16	13	24	29	17	14	14	20	13	9	10	21	11	11	14	10	17		1	
25	15	12	23	28	18	15	13	18	12	10	12	21	14	12	16	11	20		1	
26	16	13	25	29	15	15	11	11	13	11	12	21	7	11	14	11	19		1	
27	15	12	24	28	17	14	13	18	12	11	12	20	15	12	15	11	20		1	
28	16	14	24	31	17	15	10	13	11	11	12	20	7	12	14	11	21		1	
29	15	12	23	28	17	14	13	17	12	10	12	20	15	12	15	11	20		1	
30	15	13	24	29	20	14	13	16	13	10	11	21	11	11	15	10	19		1	
31	16	14	25	30	17	13	15	21	13	9	11	22	14	10	14	12	19		1	
32	16	14	26	30	18	15	11	11	13	10	10	22	7	11	14	11	18		2	
33	14	14	24	30	17	14	14	19	13	10	10	21	11	11	14	10	17		1	
34	15	13	25	29	18	14	13	17	12	11	12	20	14	12	15	11	20		1	
35	15	12	23	29	20	14	13	20	12	10	11	20	14	12	15	12	21		1	
36	14	12	24	29	16	15	14	17	12	10	11	21	13	13	15	10	20		1	
37	15	13	24	29	19	15	14	18	12	10	10	21	11	11	14	10	17		1	
38	15	12	23	28	20	14	13	19	12	10	12	20	14	12	15	11	19		1	
39	14	12	25	27	20	15	12	20	12	10	12	23	13	11	14	10	19		1	
40	15	12	23	28	16	15	13	17	12	10	14	22	14	12	15	11	20		1	
41	15	12	23	28	20	14	13	19	12	10	13	20	14	12	15	11	20		1	
42	15	14	25	30	17	15	11	11	13	10	13	22	7	11	14	11	19		1	
43	16	14	25	30	17	15	11	11	13	10	13	22	7	11	14	11	19		1	
44	15	15	26	31	16	15	11	11	13	10	12	21	7	10	14	11	19		1	1
45	15	12	27	29	21	15	16	17	12	10	11	20	10	11	14	10	19		2	
46	16	14	25	31	16	15	11	11	12	11	12	22	7	11	14	11	19		1	
47	16	12	27	30	21	15	17	18	12	10	11	20	10	11	14	10	19		1	
48	16	14	25	30	18	16	11	11	13	10	12	21	7	11	14	11	19		1	
49	15	13	24	29	17	14	14	17	12	10	10	21	11	11	14	10	17		1	

50	16	14	24	30	15	14	11	12	13	11	11	22	14	12	14	11	19		1	
51	15	12	23	28	20	14	13	19	12	10	12	20	14	12	15	11	20		1	
52	15	12	23	27	19	15	13	18	12	10	12	21	14	12	15	11	20		1	
53	15	12	23	28	18	14	13	19	12	10	12	20	14	12	15	11	20		1	
54	15	12	24	28	16	14	13	19	12	10	12	21	14	12	15	11	20		1	
55	15	13	24	29	17	15	14	18	12	11	10	21	11	11	14	10	17		1	
56	16	14	25	30	18	15	11	11	13	10	13	21	7	11	14	11	19		1	
57	15	13	24	29	17	14	14	19	13	10	10	21	11	11	14	10	17		1	
58	16	14	26	31	16	15	11	11	13	10	12	21	7	10	14	11	19		1	
59	15	12	23	28	19	14	13	18	12	10	11	20	14	12	15	11	20		1	
60	15	14	25	31	17	15	11	11	13	10	11	21	7	11	14	11	18		1	
61	15	14	25	30	18	14	14	18	12	10	10	21	11	11	14	10	17		1	
62	16	14	23	30	17	15	11	11	13	11	12	21	7	11	14	11	18		1	
63	15	12	23	27	17	14	13	21	12	9	11	20	14	11	15	11	20		1	
64	15	14	22	29	17	14	12	12	13	11	12	21	14	11	14	10	19			1
65	14	12	25	27	16	17	13	18	12	10	11	22	13	12	14	10	20		1	
66	15	13	24	29	17	15	14	18	12	10	10	21	11	11	14	10	17		2	
67	15	12	23	28	17	14	13	18	12	10	12	20	14	11	15	11	20		1	
68	15	13	25	29	13	16	11	11	12	11	12	22	7	11	14	11	19		1	
69	15	12	27	29	22	15	17	17	12	10	12	20	10	11	14	10	19		1	
70	15	12	24	28	18	14	13	18	12	10	12	20	14	12	15	11	20		1	
71	15	13	23	29	17	14	13	18	13	10	10	21	11	11	14	10	17		1	
72	16	12	25	29	20	15	16	16	12	10	12	20	10	11	14	10	20		1	
73	15	13	23	29	18	14	14	18	13	10	13	20	14	12	15	11	20		1	
74	14	13	24	29	21	15	16	16	12	10	11	20	10	11	14	11	19			1
75	15	13	23	29	18	15	14	19	12	11	10	21	12	11	14	10	18		1	
76	15	12	27	29	22	15	16	17	12	10	11	20	10	11	14	10	19		1	
77	15	12	23	28	18	14	14	19	12	11	11	20	14	12	15	11	20		1	
78	16	14	25	31	16	16	11	14	13	10	10	23	11	12	14	11	20		1	
79	16	13	23	29	19	15	13	16	12	10	13	21	14	12	15	11	20		1	
80	17	13	25	29	17	15	11	11	13	10	12	21	7	11	14	11	19			1
81	16	14	26	31	16	16	11	14	13	10	10	23	11	12	14	11	20			1
82	15	12	25	28	20	14	12	17	12	10	12	20	14	12	15	12	20		1	
83	15	12	23	28	18	14	13	18	12	10	11	20	14	12	15	11	20		1	
84	15	14	23	31	17	15	11	20	14	10	11	21	11	10	14	10	21		1	
85	17	14	24	30	16	14	11	12	13	11	12	22	14	12	14	11	19		1	
86	15	13	24	30	18	16	14	18	12	10	10	21	11	11	14	10	17			1
87	16	14	25	30	17	15	11	14	13	10	11	21	7	11	14	11	19	1		4
88	15	12	23	28	18	14	13	19	12	10	12	20	14	13	15	11	20		1	
89	15	12	23	27	16	15	11	17	13	10	12	19	12	13	15	11	19		1	
90	16	13	24	29	18	15	14	18	12	10	10	22	11	11	14	10	17			1
91	15	13	25	32	16	16	11	14	13	10	10	23	11	13	14	11	20			1
92	15	12	23	27	17	14	13	17	12	10	12	20	14	12	15	11	20		1	
93	16	15	25	31	17	15	11	11	13	10	12	21	7	11	14	11	18		1	
94	15	12	23	29	17	14	13	19	12	10	12	21	14	11	14	11	20		1	
95	16	13	25	29	17	15	11	14	13	10	11	21	7	11	14	11	19		1	
96	15	13	23	28	16	14	13	19	12	10	12	21	14	11	15	11	20		1	
97	15	13	22	29	17	15	14	19	15	10	11	18	11	10	14	10	19		1	
98	16	12	27	29	22	15	17	18	12	10	11	20	10	11	14	10	19		1	
99	15	13	23	30	18	17	12	15	12	10	12	21	7	11	14	11	21		1	
100	15	12	27	30	21	15	17	17	12	10	12	20	10	11	14	10	19		1	
101	15	13	24	29	18	14	15	19	13	10	10	21	11	11	14	10	18		1	
102	15	13	23	29	16	14	11	15	13	10	10	21	7	11	14	10	18		1	

103	16	12	23	28	18	14	13	20	13	10	11	20	14	12	15	11	20	1		
104	15	12	23	28	17	14	13	19	12	10	12	20	14	12	14	11	20		1	
105	16	13	24	29	17	14	13	19	12	10	12	21	14	12	15	11	20		1	
106	15	13	24	29	15	15	11	14	13	10	12	22	7	11	14	11	20			1
107	15	12	27	29	20	15	17	17	13	10	11	20	10	11	14	10	18			1
108	15	13	25	30	17	14	14	19	12	10	10	23	11	11	14	10	17		1	
109	15	12	27	29	22	15	17	17	12	10	11	20	10	11	14	10	19		1	
110	16	14	26	30	16	15	11	11	13	10	12	21	7	10	14	11	19	7	2	5
111	15	13	23	29	17	15	12	16	13	10	13	19	12	12	15	10	19		1	
112	15	12	27	29	22	15	18	18	12	10	11	20	10	11	14	10	19		1	
113	15	13	23	29	19	14	13	17	14	10	10	25	10	9	16	11	19		1	
114	15	13	23	29	18	14	13	17	14	10	11	25	10	9	16	11	19		1	
115	16	12	23	28	17	14	13	20	12	11	12	21	14	12	15	11	20		1	
116	15	12	27	29	24	15	17	17	12	10	11	21	10	11	14	10	19		1	
117	15	13	25	29	18	15	14	19	12	10	10	21	11	11	14	10	17		1	
118	15	14	24	30	15	14	11	12	13	11	12	23	14	12	14	11	19		1	
119	15	13	24	29	17	13	14	19	13	10	13	22	14	11	13	11	19		1	
120	15	12	23	28	18	14	13	19	12	10	11	20	14	12	15	11	20			1
121	15	13	25	29	16	15	11	11	12	11	12	22	7	11	14	11	19		1	
122	16	13	25	29	17	15	11	11	14	11	12	22	7	11	14	11	19			1
123	16	14	25	30	16	15	11	11	13	10	12	22	7	11	14	11	19	1		2
124	15	12	23	28	19	14	12	15	13	11	12	20	12	12	15	11	19		1	
125	15	14	22	29	18	15	11	14	14	10	11	21	13	12	15	11	18		1	
126	17	12	25	29	21	15	14	16	12	10	11	20	10	11	14	10	20			1
127	14	12	24	28	19	15	16	16	12	10	11	20	10	11	14	10	19			1
128	15	12	22	28	20	15	13	17	12	10	14	21	14	12	15	11	20		1	
129	16	13	25	29	17	15	11	12	13	10	13	21	7	11	14	11	18			1
130	15	12	22	28	19	15	14	17	12	10	12	21	15	12	16	11	20		1	
131	16	14	25	31	17	15	11	11	13	10	12	22	7	11	14	11	18		1	
132	15	13	25	29	16	15	11	11	13	10	12	22	7	11	14	11	19			1
133	15	12	23	28	18	15	13	17	12	10	13	21	14	12	15	11	19		1	
134	15	12	22	29	17	14	14	19	12	10	12	20	14	12	16	11	20		1	
135	15	12	23	28	18	14	13	20	12	10	12	21	14	12	15	11	20		1	
136	15	12	23	28	18	14	18	19	13	10	13	20	14	12	15	11	20		1	
137	14	12	27	29	21	15	17	17	12	10	11	20	10	11	14	10	19		1	
138	15	14	25	30	16	16	11	11	13	10	13	22	7	11	14	11	19		1	
139	16	12	24	28	20	15	13	17	12	10	13	21	14	12	15	11	20		1	
140	15	12	23	28	12	14	14	18	12	10	13	21	14	12	15	11	19		1	
141	15	12	22	28	17	14	14	19	12	10	12	20	14	12	16	11	20		1	
142	16	13	26	29	17	15	11	11	13	10	12	21	7	11	14	11	19	1		
143	15	12	23	29	17	15	13	18	12	10	12	20	14	12	15	11	20		1	
144	16	14	24	30	16	16	11	11	14	11	12	21	7	11	14	11	19		1	
145	15	13	25	31	15	15	11	14	13	11	10	23	11	13	14	11	21		1	
146	15	12	22	28	16	16	13	19	12	10	12	20	13	11	14	11	20		1	
147	15	12	27	29	21	15	17	17	12	10	11	20	10	11	14	10	19		1	
148	15	14	24	31	18	17	12	12	13	9	11	23	11	10	14	10	20	1		
149	16	14	25	29	16	15	11	11	13	10	12	21	7	10	14	11	19		1	
150	15	14	25	30	17	15	12	14	13	10	11	21	7	11	14	11	19			1
151	15	13	24	30	16	15	11	11	13	10	12	21	7	11	14	11	19			1
152	15	12	25	28	20	15	16	16	12	10	11	21	10	11	14	10	19		1	
153	15	14	25	30	17	15	11	11	12	10	12	22	7	11	14	11	19		1	
154	16	13	24	29	16	15	11	14	13	10	12	21	7	11	14	11	19			1
155	18	14	24	30	17	15	11	13	11	11	12	20	7	13	14	11	21		1	

156	15	14	25	30	17	16	11	11	13	10	12	21	7	11	14	11	19			1
157	15	12	23	27	17	16	12	17	13	10	12	19	12	13	15	11	19		1	
158	15	13	25	29	16	15	11	14	13	10	12	22	7	11	14	11	19		1	
159	16	12	25	29	19	15	15	16	13	10	12	20	10	11	14	10	20			1
160	15	13	24	29	16	14	14	18	12	11	10	22	11	11	14	10	17		1	
161	16	14	25	31	16	16	11	11	13	10	11	21	7	11	14	11	19			1
162	16	14	26	30	17	15	11	11	12	10	12	21	7	11	14	11	19			1
163	15	13	24	29	21	15	14	18	12	10	10	21	11	11	14	10	17		1	
164	15	13	24	29	16	13	15	17	13	10	14	22	14	11	13	11	19			1
165	16	14	26	30	16	15	11	11	13	12	12	22	7	11	14	11	19			1
166	15	13	24	30	16	15	11	14	13	10	12	21	7	11	14	11	19			2
167	15	13	24	30	16	15	12	14	13	10	12	21	7	11	14	11	19			1
168	17	14	25	30	17	15	11	11	13	10	12	21	7	11	14	11	19		1	
169	16	14	25	30	17	15	11	11	13	11	12	23	7	11	14	11	19			1
170	16	13	24	29	16	15	11	14	13	10	12	21	7	11	14	11	20			1
171	15	13	24	30	20	15	14	17	12	10	11	23	11	11	14	10	17			1
172	16	14	25	30	16	16	11	11	14	10	12	21	7	11	14	11	19			1
173	16	14	25	31	19	15	11	15	13	10	11	21	7	11	14	11	19			1
174	16	14	25	30	17	14	13	17	12	10	13	22	13	11	15	10	20			1
175	17	13	22	29	17	15	12	15	13	10	11	21	11	11	14	11	20		1	
176	16	12	23	27	16	14	13	22	12	10	13	20	14	12	15	11	20		1	
177	15	14	25	30	16	16	11	11	13	10	12	21	7	11	14	11	19			1
178	16	14	26	31	18	15	11	15	13	10	11	21	7	11	14	11	19	1		
179	16	14	25	30	16	15	11	11	13	10	12	21	7	11	14	11	19	2		
180	15	12	24	27	18	15	14	17	13	10	11	21	13	12	15	10	19	1		
181	16	14	25	30	16	15	11	11	13	10	11	21	7	10	14	11	19	1		
182	15	13	25	31	16	16	11	16	14	10	10	23	11	13	14	11	20	2		
183	15	12	27	29	22	15	17	17	12	10	11	21	10	11	14	10	19	2		
184	16	14	25	30	16	16	11	11	13	11	12	21	7	12	14	11	19			1
185	17	14	25	30	16	13	11	11	13	10	13	21	7	10	14	11	19			1
186	16	13	25	29	16	15	11	11	13	11	12	21	7	11	14	11	19			1
187	15	12	21	29	14	15	13	15	14	10	12	23	11	11	16	10	20	1		
188	16	14	25	31	16	15	11	11	13	10	11	21	7	10	14	11	19	1		
189	16	14	27	30	16	15	11	11	13	10	12	22	7	10	14	11	19			1
190	16	14	25	30	16	15	11	11	13	10	12	21	7	10	14	11	19	2		1
191	15	14	26	30	16	15	11	11	12	10	12	21	7	10	14	11	19			1
192	16	13	25	31	16	15	11	11	13	10	12	21	7	11	14	11	19	1		
193	15	13	23	30	16	15	11	14	13	10	12	21	7	11	14	11	19	1		1
194	16	14	25	30	17	15	11	14	13	10	12	21	7	11	14	11	19	1		
195	16	14	25	30	16	15	9	14	13	10	11	21	7	11	14	11	19	1		
196	16	14	25	30	15	15	11	11	14	10	12	21	7	10	14	11	19	1		
197	15	14	22	32	17	15	16	17	14	10	11	21	10	11	15	10	21	1		
198	15	12	23	28	18	14	13	19	12	10	13	20	14	12	15	11	20	2		
199	15	12	24	28	18	15	16	17	12	10	11	20	14	12	15	11	19	1		
200	16	13	25	29	16	15	11	14	13	10	12	21	7	11	14	11	19	1		
201	15	12	26	28	18	13	12	12	13	10	12	22	13	12	15	11	19	1		
202	16	14	25	30	17	15	11	14	13	10	11	22	7	11	14	11	19	2		
203	18	14	25	30	17	15	11	14	13	9	11	21	7	11	14	11	19			1
204	16	15	25	32	16	16	11	11	13	10	12	21	7	10	14	11	19			1
205	16	14	25	31	16	16	11	11	13	10	12	21	7	10	14	11	19	1		
206	16	13	22	29	17	14	11	14	13	10	11	21	15	12	14	10	19	1		
207	16	14	25	29	16	15	12	14	13	10	11	21	7	11	14	11	18			1
208	16	13	23	29	17	15	11	15	12	10	12	22	17	11	14	12	20			1

209	15	13	22	30	15	14	13	18	13	10	13	23	13	11	15	10	20			2
210	15	14	25	30	16	15	11	14	13	10	11	20	7	11	14	11	19			1
211	15	13	24	30	17	15	11	14	13	10	12	21	7	11	14	11	19			1
212	15	13	24	30	17	14	15	17	12	10	13	20	15	12	15	11	19	1		
213	16	14	26	30	16	15	11	11	13	10	12	22	7	10	14	11	19	1		
214	15	13	23	28	17	14	13	20	12	10	12	20	15	12	15	11	20		1	
215	16	14	24	30	17	14	11	11	13	10	12	21	7	11	14	11	19			1
216	16	14	25	30	18	15	11	14	13	10	11	21	7	11	14	11	19			1
217	15	12	26	29	24	16	17	16	12	10	11	20	10	11	14	10	19		1	
218	15	14	26	30	16	15	11	11	13	10	12	21	7	10	14	11	19		1	
219	13	12	25	29	17	16	15	19	12	10	11	23	13	11	15	10	20	1		
220	14	12	24	28	22	15	16	19	12	10	11	20	10	11	14	10	20	1		
221	16	14	24	30	16	15	13	20	14	10	13	21	11	11	14	10	21	1		1
222	15	13	25	31	15	15	11	13	13	10	10	23	11	13	14	11	20	1		
223	15	12	23	28	19	15	12	18	12	11	12	21	13	11	15	10	18	1		
224	16	12	23	29	20	14	13	19	12	10	12	21	14	13	15	11	19	1		
225	16	13	25	29	16	15	11	11	13	10	12	22	7	11	14	11	19	1		
226	16	13	25	29	16	15	12	12	13	10	12	21	7	11	14	11	19			1
227	17	13	25	29	17	17	11	14	13	10	12	21	7	10	14	11	19			1
228	15	13	25	29	17	15	11	11	13	10	12	23	7	11	14	11	20			1
229	16	13	25	29	16	15	11	11	13	10	12	21	7	11	14	11	18.1			1
230	14	12	25	28	20	15	15	16	12	10	12	20	10	11	14	10	19			1
231	15	13	24	30	15	15	11	14	13	10	10	23	11	12	14	11	20	1		
232	16	14	26	30	17	15	11	11	13	10	12	21	7	10	14	11	19			1
233	15	13	24	30	15	16	11	15	13	10	10	23	11	12	14	11	20	1		
234	15	12	24	28	16	14	13	18	12	10	15	21	14	13	15	11	20	1		
235	14	13	25	29	16	15	11	14	13	10	12	21	7	11	14	11	19	1		
236	14	12	23	28	19	15	16	18	12	10	11	20	10	11	14	10	19	1		
237	16	14	25	31	16	15	11	11	13	10	12	21	7	11	14	11	19	1		
238	14	14	25	30	16	15	11	11	13	10	12	22	7	11	14	11	19	1		
239	15	14	25	30	18	15	11	14	13	10	11	21	7	11	14	11	19	1		
240	16	13	25	29	16	16	11	14	13	10	11	23	7	9	14	12	19	1		
241	15	12	23	27	15	14	13	20	12	10	13	20	14	12	15	11	20	1		
242	16	13	23	29	15	15	12	19	15	10	12	22	11	11	14	10	21	1		
243	16	13	23	29	17	15	11	13	12	10	12	21	7	12	14	11	20	1		
244	15	13	23	29	18	15	13	19	12	10	13	21	14	12	15	11	19	1		
245	17	12	23	27	17	14	13	19	12	10	12	20	14	11	15	11	20	1		
246	15	12	24	27	18	14	13	19	12	10	11	19	14	12	16	11	20	1		
247	16	13	25	29	16	15	11	11	13	12	12	21	7	11	14	11	19			1
248	15	13	24	29	17	15	11	11	13	10	12	23	7	11	14	11	20			1
249	15	12	23	28	16	15	12	16	13	10	12	19	12	12	15	10	19			1
250	16	14	26	31	16	15	11	11	13	10	12	21	7	10	14	12	19	1		
251	15	14	23	30	18	14	11	13	14	11	10	20	16	12	14	11	19	1		
252	13	14	25	31	20	15	14	20	12	10	11	21	13	12	15	10	20	1		
253	15	12	27	29	22	15	15	17	12	10	11	20	10	11	14	10	19	1		
254	15	12	23	27	16	14	13	20	12	10	12	22	14	12	15	11	21	1		
255	16	14	25	30	16	16	11	11	13	10	12	21	7	11	14	11	19			3
256	15	14	24	30	15	14	11	12	13	11	11	23	14	12	14	11	19			1
257	15	14	26	30	16	15	11	11	13	11	12	21	7	10	14	11	19	1		
258	17	13	25	29	17	14	15	20	14	11	13	22	13	11	14	10	18			1
259	15	12	23	26	17	13	13	19	12	10	12	20	14	12	15	11	20			1
260	15	14	24	32	16	13	15	22	14	9	11	22	14	10	14	12	20			1
261	16	14	25	30	17	15	11	15	13	10	11	21	7	11	14	11	19			1

262	15	13	24	28	15	15	12	19	12	10	12	22	13	12	14	11	20	1		
263	15	14	23	29	17	14	13	13	13	11	12	22	14	12	14	10	19			1
264	15	13	24	30	18	16	13	18	14	11	13	21	13	11	14	10	18			1
265	15	13	24	30	19	15	15	18	12	10	11	21	11	11	14	10	17			1
266	16	13	25	29	16	16	11	11	13	10	12	21	7	11	14	11	19			1
267	15	13	25	29	16	16	11	14	13	10	11	21	7	11	14	11	19			1
268	15	13	23	29	18	16	11	17	14	10	11	21	11	11	14	10	21			1
269	16	12	23	27	17	14	13	20	12	10	12	21	14	12	15	11	20	1		
270	16	14	25	31	16	15	11	11	13	10	12	21	7	10	14	11	19			1
271	15	14	25	30	18	15	11	14	13	10	11	21	7	11	14	11	20			1
272	16	12	23	29	15	15	13	14	13	10	11	20	14	11	14	10	18			1
273	17	14	25	30	16	15	11	11	13	10	12	21	7	11	14	11	19	1		
274	15	12	23	27	18	15	13	17	12	10	12	21	13	12	16	11	20	1		
275	16	13	24	29	18	15	14	18	12	10	10	20	11	11	14	10	17	1		
276	15	12	25	29	20	15	15	17	12	10	11	20	10	11	14	10	19	1		
277	16	14	25	30	16	15	12	14	13	10	11	22	7	11	14	11	19	1		
278	16	13	25	31	17	15	11	11	13	10	12	20	7	11	14	11	19	1		
279	15	12	24	27	19	15	13	17	12	10	12	21	14	12	16	11	20	1		
280	15	14	24	30	16	15	11	14	13	11	12	23	11	12	14	11	20	1		
281	14	12	23	31	17	15	12	16	13	10	11	19	12	11	15	10	19			1
282	16	14	25	30	14	17	12	14	13	10	11	21	7	11	14	11	19	1		
283	15	14	25	30	17	15	11	11	13	10	12	22	7	11	14	11	19			1
284	16	14	24	33	15	15	11	15	13	10	10	23	11	13	14	11	19			1
285	15	13	24	30	17	15	12	18	12	10	10	21	11	12	14	10	17			1
286	15	14	25	30	17	15	10	14	13	10	11	21	7	11	14	11	19			1
287	17	14	23	30	15	14	11	12	13	10	12	22	14	12	14	11	19			1
288	15	12	23	28	17	14	13	17	12	10	12	20	14	12	15	11	20	1		
289	16	14	26	30	16	15	11	11	14	10	13	22	7	10	14	11	19			1
290	16	14	25	30	16	15	11	12	13	10	12	21	7	10	14	11	19			1
291	17	13	25	29	16	15	11	14	13	10	12	21	7	11	14	11	19			1
292	16	14	25	30	16	16	11	11	13	10	12	21	7	11	14	11	18			1
293	15	12	26	29	22	15	17	17	12	10	11	20	10	11	14	10	19			1
294	16	14	24	30	16	16	11	11	13	10	12	21	7	11	14	11	19			1
295	15	13	23	30	15	17	11	17	15	10	11	22	11	12	14	10	22			1
296	14	13	24	30	17	15	11	14	13	10	12	21	7	11	14	11	19			1
297	15	13	26	29	16	15	11	11	13	10	12	21	7	10	14	11	19	1		
298	14	13	23	30	18.2	14	13	19	12	11	11	21	11	11	14	11	20	1		
299	15	12	25	28	18	14	13	17	12	10	12	20	14	11	15	11	20	1		
300*	15	13	25	29	16	15	11	11	13	10	11	21	7	11	14	11	0		1	
301*	16	13	24	29	17	15	12	14	13	10	12	20	11	11	14	11	0		1	
302*	17	13	24	28	17	15	12	14	13	10	12	21	11	11	14	11	0		1	
303*	16	13	24	29	17	15	12	14	14	10	12	20	11	11	14	11	0		1	
304*	17	13	24	29	17	15	12	14	13	10	12	21	11	11	14	11	0		1	
305*	15	14	25	30	16	16	11	11	13	10	12	21	7	11	14	11	0			1

*Haplotypes 300-305 carry null allele at DYS448 and they were excluded for haplotype diversity and matching probabilities calculations.

V. MITOCHRONDRIAL GENOME VARIATION IN NEPALESE AND TIBETAN POPULATIONS

A. INTRODUCTION

While most of Asia was populated, the Tibetan plateau remained an uninhabitable, hyper arid, cold desert due to its high altitude of around 4000 meters above sea level (masl). Approximately 25-50 thousand years ago (kya), temperatures began to rise and glaciers retreated, making migration onto the plateau from the north more feasible.[1] In fact, recent archaeological evidence points toward this period of glacial retreat as the time when the peopling of the plateau first occurred. This notion is further supported by radiocarbon dating of recovered ostracod samples along with artifacts found in the Qiadam basin in the northeastern region of Tibet, which suggest that the initial inhabitation of the plateau occurred as early as 30 kya.[2] Brantingham and colleagues,[3] however, argue that this date is premature, and that occupation of the plateau did not take place until 23 to 25 kya. In either case, a pre-Last Glacial Maximum (LGM) inhabitation of Tibet is consistent with recent genetic studies.[4,5]

Tibet first appeared in historical records during the seventh century C.E. when several clans were unified by Songtsen Gampo, the first emperor of the country. Under his rule, Tibet acquired formidable military power and, as a result, extended its empire as far east as the Qinghai and Sichuan provinces of China and as far west as Nepal, India, and the Hunza Valley in modern-day Pakistan.[6] During this time, Tibet maintained a stronghold over the Silk Road trading route, which allowed the exchange of goods, ideas and culture between Tibet and the neighboring kingdoms of Nepal, India, China and Persia. Specifically, the introduction of Buddhism from northern India in the seventh

century C.E. had the greatest influence on Tibetan society, with Buddhist imprints detected in Tibetan art, music, literature as well as present-day cultural festivals, including Losar (Tibetan New Year). However, following the assassination of emperor Lang Dharma in the ninth century, Tibet was subject to a series of foreign invasions, beginning with Mongolia in 1247 C.E., whose empire at that time included all of China and stretched as far west as Europe. Although this event resulted in Tibet becoming a protectorate of the Mongolian Empire, by the 18th century the Mongol Empire disintegrated and the Manchurian Empire from northeast Asia took control of China while recognizing the independence of the demilitarized state of Tibet. Over the next two centuries, Tibet was invaded on different occasions by Nepal (late 18th century), India (mid-19th century) and Great Britain (1904), but in 1911, while China was weakened by a civil war under the Qing Dynasty, Tibet formally declared its independence from China.[6] This action, however, only served as a temporary deterrent, since China invaded Tibet again in 1959.

The Himalayan kingdom of Nepal encompasses a total area of 147,181 km^2 and its physical and cultural landscape can be divided into three distinct regions.[7] The mountainous north is characterized by the southern slopes of the Himalayas and its people are culturally linked to the Buddhists of Tibet. 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.[8]

Historically, Nepal was comprised of only the Kathmandu Valley, a region located in the east central hills of present-day Nepal. The Valley witnessed several different waves of migrations because of its rich soil, favorable climate, malaria free zone and trade location.[9] Therefore, it is not surprising that the indigenous inhabitants, the Newars, are postulated to be a mixture of Austro-Asiatic, Dravidian, Indo-Mongoloid and Aryan origins.[10] The earliest rulers of the Kathmandu Valley were the Tibeto-Burman Kiratas, who reigned until the 4th century C.E. The inscriptions of Kirati words, including the names of places, rivers, canals and tax offices in the ancient Nepalese epigraphy attest to their early settlement in the Valley.[11] The Kiratas 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. By the 13th century C.E., the Malla Dynasty came to power, which ultimately resulted in the development of trade, art and architecture throughout the valley. The Newaris excelled at the arts, dominating most forms of artistry and, as a result, began expanding beyond the Kathmandu Valley. Today, this group constitutes 5.5% of Nepal's population, with roughly 40% of them living outside of the Kathmandu Valley.[9] They speak the Tibeto-Burman language of Newari or Nepal Bhasa, which has been heavily influenced by Sanskrit, an Indo-Aryan language. Presently, Nepal is the only constitutionally Hindu country in the world, with 80.6% of its population practicing this religion.

Our previous study[12] revealed a high frequency of the Asian-specific Y *Alu* insertion (D-M174) and the East Asian haplogroup (O3a5c-M134) on the Tibetan plateau. The mitochondrial DNA (mtDNA) data indicates that Tibetans share a common maternal lineage with northern Mongol populations,[13] while Qin and colleagues[5] suggests

a northern East Asian ancestry for the Tibetan populations, an observation that is consistent with our autosomal STR data.[14] Previous reports[4,5] have also indicated the presence of both pre- and post-LGM lineages in the Tibetan maternal gene pool.

Although extensive genetic studies have been performed on the malaria-resistant Tharu people of the Terai region of Nepal,[15,16] limited work has been done to characterize the maternal lineages present in other Nepalese groups, especially the indigenous Newar population of the Kathmandu Valley. Therefore, the present study was undertaken to address this gap in the genetic profiling of Himalayan populations by examining the mtDNA of 344 unrelated individuals from Tibet (156) and Nepal (188) [Tamang (45), Newar (66) and Kathmandu (77)]. In addition, we evaluated the impact of the Himalayas on the current maternal genetic landscape of populations residing on either side of the mountain range. This report also complements our previous Y-chromosomal[12] and autosomal STR[14] data from the same four populations, thereby providing a comprehensive analysis of the genetic diversity present in the Himalayas.

B. MATERIALS AND METHODS

Sample collection and DNA isolation

Blood samples were collected with informed consent from 344 unrelated individuals belonging to three Nepalese populations, namely Tamang ($N = 45$), Newar ($N = 66$) and Kathmandu ($N = 77$), and a general collection from Tibet ($N = 156$). The genealogical history of each donor was recorded for a minimum of two generations to establish regional ancestry. All ethical guidelines were followed as stipulated by the Institutional Review Board at Florida International University. Genomic DNA was

isolated from peripheral blood lymphocytes using standard phenol-chloroform and ethanol precipitation methods as described by Antunez de Mayolo and colleagues.[17]

Sequencing and RFLP typing

The mtDNA hypervariable regions I and II (HVRI and HVRII) were amplified separately using the L15996/H16401 and L29/H408 primer pairs, respectively.[18,19] The resulting amplicons were sequenced in both directions on an ABI 3130xl genetic analyzer (Applied Biosystem) using the abovementioned PCR primers and the BigDye terminator v1.1 Sequencing kit (Applied Biosystem). The sequences generated were aligned and compared to the revised Cambridge Reference Sequence (rCRS)[20,21] with the BioEdit Sequence Alignment Editor.[22]

The haplogroup affiliation of each sample was inferred based on the HVRI [nucleotide positions (nps) 16024-16395] and HVRII [nucleotide positions (nps) 50-400] motifs and was subsequently confirmed by a hierarchical RFLP screening of diagnostic sites in the coding region according to the most updated mtDNA phylogeny,[23] (Build 13) and previously published reports[4,5,16,24,25] (Supplementary Tables 1 and 2). However, those haplogroups whose diagnostic point mutation variants were not naturally amenable to RFLP analysis were either sequenced or genotyped by primer-mediated RFLP[26] (Supplementary Table 2). Additionally, haplogroup B, which is defined by a 9-bp deletion (nps 8281–8289), was scored based on length polymorphism assessed by electrophoretic separation.

Data analyses

Diversity indices (haplotype diversity, nucleotide diversity, number of different haplotypes and mean number of pairwise differences) and neutrality tests (Tajima's *D* and Fu's F_s) based on HVRI and HVRII sequence data were calculated with the Arlequin package version 3.11.[27] Haplogroup-specific median-joining (MJ) networks were constructed with the Network 4.1 program (http://www.fluxus-engineering.com) using the HVRI sequences. All polymorphic nucleotide positions were assigned weights following the recommendations provided by Roostalu and colleagues.[28] In addition, the coalescence times of haplogroups A4, A11, D4, F1, M9a1, M9a1a1c1b and U2 were computed using rho statistics[29] calibrated with two different mutation rates, including i) the most widely used mutation rate of one transitional step between nps 16090-16365 equal to 20,180 years[29] and ii) the revised rate of one mutation every 18,845 years for the region between nps 16090-16365.[30] The standard deviation for the rho estimates was obtained according to Saillard et al.,[31] and the length variation in the polycytosine tract between nps 16180-16193 was excluded from the analysis.

To compare mtDNA variation among the four Himalayan groups and the 33 geographically targeted reference populations (Table 1),[5,16,25,32–44] a correspondence analysis (CA) based on mtDNA haplogroup and sub-haplogroup frequency data was performed with the NTSYSpc- 2.02i software.[45] Analysis of molecular variance (AMOVA) was computed with the Arlequin 3.11 program[27] to evaluate genetic structure among populations when grouped according to their geographical locations and linguistic affiliations (Table 1).

C. RESULTS

A total of 75 different haplogroups and sub-haplogroups were observed in the four Himalayan populations studied (Table 2). Among the four groups, Kathmandu shows the highest degree of heterogeneity, with 44 lineages, followed by Tibet (39), Newar (26) and Tamang (22), a pattern consistent with Y-chromosomal data.[12] Based on their phylogeographic origin,[4,5,16,24,46,47] these haplogroups can be assigned to three main sources, namely East Asia, South Central Asia and West Eurasia (i.e, territory including West Asia and Europe). Supplementary Table 1 provides the mtDNA control region sequence variation and haplogroup designations of all 344 samples analyzed.

While the majority of mtDNA diversity in Tibet (96.1%) and Tamang (66.7%) is represented by East Asian-specific haplogroups, including those from Northeast [*i.e.,* A, D, G and M8 (M8a, C and Z)] and Southeast (*i.e.,* F, M9 and M13) Asia, Kathmandu, Newar and, to a lesser extent, Tamang, display a considerable proportion of South Central Asian lineages (49.3%, 36.4% and 17.8%, respectively) in their maternal gene pools. These South Central Asian-specific markers include M2, M3, M4"67, M5, M30, M33, M34, M35, M38, M43, R2, R5, R6, R30, U2 and U4, whereas the West Eurasian mtDNA component in Kathmandu (9.1%), Tamang (15.6%) and Newar (18.2%) is characterized by a combination of haplogroups HV, H, J, N1e, T and U7. The near absence of South Central Asian lineages in Tibet (1.9%) corroborates our previous Y-chromosomal study[12] which suggests that the Himalayan mountain range acts as a geographic barrier to gene flow from the Indian subcontinent.

Phylogeography of lineages within superhaplogroup M

More than half of the Himalayan dataset belongs to macrohaplogroup M, with frequencies ranging from 67.3% in Tibet to 56.1% in Newar. Within this major haplogroup, the M9a1 (27.6%) sub-clade, particularly one of its terminal branches, M9a1a1c1b (21.8%), was found to predominate in the Tibetan mtDNA pool,[25] while haplogroup D lineages were the most frequent in the Tamang (26.5%) and Kathmandu (11.7%) populations. Specifically, sub-haplogroup D4, which is prevalent throughout Central Asia[39] and Northeast Asia, including Japan[32] and Korea,[33] represents the majority of haplogroup D samples in Tamang (15.5%) and Kathmandu (9.1%), as well as Tibet (8.3%). Newar, on the other hand, exhibits equivalent levels (12.1%) of haplogroups M5 and Z; however, sub-haplogroup M9a1 and its derivatives (M9a1a1c1b, M9a1a2 and M9a1b1), which occur at relatively high frequencies in Tamang (15.5%) and Tibet (27.6%), are completely absent from this collection.

Haplogroup M8 encompasses sub-clades M8a, C and Z, the latter of which has been previously reported at high frequencies in the Koryak and Itelmen populations from the Kamchatka peninsula in the Russian Far East.[48] Haplogroup Z is also present in several Siberian populations, including the Altaians, Evenks, Dolgans and Kents,[49] and the Volga-Ural region of Russia and the Sami from Finland and Sweden.[50] While sub-haplogroup Z1 predominates in the matrilineages of the above-mentioned populations, the majority of Z samples in Newar (9.1%) and Kathmandu (3.9%) belong to the Z3 lineage. In Tibet, however, this haplogroup is observed at relatively low levels (1.3%), while its sister clade, haplogroup C, is detected at a much higher frequency (8.3%). In Nepal, C lineages are either absent (Newar) or observed only at low levels (Tamang,

2.2%; Kathmandu, 3.9%). Additionally, haplogroup M8a, in contrast to other M8 sub-clades, is restricted to the collection from Tibet (1.3%).

Similar to haplogroup Z, G lineages are found at their greatest frequencies in the Kamchatka peninsula but they are also common among populations originating in Japan and Korea. This haplogroup accounts for an average of 20% of the mtDNA gene pool of the Tharus from Nepal[16] and reaches frequencies greater than 15% in the Tibetan populations of Nagqu and Garze.[5] In the present study, both the G2 (5.1%) and G3 (3.2%) sub-clades are observed in Tibet, whereas Tamang (11.11%) and Newar (4.5%) display mtDNAs that belong only to the former sub-haplogroup and are further defined by a transition at np 16193 (Figure 2). Interestingly, Kathmandu, despite being the most diverse group of the four Himalayan populations, does not contain any samples derived for haplogroup G.

Our previous report examining the Y-chromosomal diversity of the Himalayas[12] revealed the presence of several Indian-derived Y-chromosomes, namely R-M207, H-M69 and C5-M356 in both Kathmandu and Newar. Therefore, it is not surprising that in the current investigation, several M lineages typical of the Indian subcontinent were uncovered in the abovementioned Nepalese collections. For instance, haplogroup M5, although present at low levels in Kathmandu (3.9%), is one of the most common lineages in Newar (12.1%). Haplogroups M30 and M33 are also detected at frequencies greater than 4.5% in both populations, while M3 and M35 are present at much higher proportions in Kathmandu (both at 5.2%) relative to Newar (both at 1.5%). It is notable that, in contrast to the Y-chromosome data,[12] Tamang's maternal gene pool exhibits several Indian lineages, albeit limited, including M4"67 (4.4%), M31a2 (2.2%), M33 (2.2%) and

M43 (2.2%). Of particular interest is M31a2, the sister clade of the Andamanese-specific lineage M31a1,[51] which has been previously observed by Endicott and collaborators[52] in the Lodha, Lamdbadi and Chenchu tribal groups of India.

Tibetans also exhibit M descendants that are largely restricted to the plateau. Haplogroup M62, for example, was first documented in Northeast India[47] and since then has been reported in several populations throughout Tibet.[4,5] In this study, we observed haplogroup M62 in six Tibetans, four of whom belong to the M62b branch. Likewise, haplogroup M13 (4.5%) is restricted to the Tibetan collection, although it was found in a single sample from Kathmandu. The other Tibetan-specific haplogroups detected include M11, M61 and M70, with frequencies of 0.6%, 0.6% and 1.3%, respectively

Phylogeography of lineages within superhaplogroup N

Haplogroup R and its derivatives represent the majority of the lineages branching from the basal N trunk, accounting for 23.1%, 28.9%, 33.8% and 36.4% of the matrilineal diversity in Tibet, Tamang, Kathmandu and Newar, respectively. Within the R branch, a contrasting distribution pattern of clades F and U are observed in the Tibetan and Nepalese populations. Haplogroup F is characterized by sub-haplogroup F1 in all four Himalayan groups but its frequency decreases southward from Tibet (18.6%) to Nepal [Tamang (6.7%), Newar (10.6%) and Kathmandu (7.8%)]. It is largely represented by F1d in Newar (7.6%), Tibet (5.8%), Kathmandu (3.9%) and Tamang (2.2%) followed by the F1c (1.5%, 2.6%, 1.3% and 2.2%, respectively) and F1b (0.0%, 1.9%, 1.3% and 2.2%, respectively) lineages. The remaining F1 samples are unresolved, harboring a basal HVRI motif of 16183C-16189-16304, which likely reflects the initial migrants of this

lineage onto the plateau. Interestingly, a single mtDNA molecule belonging to F2 is present in the Newari population.

In contrast to haplogroup F1, U lineages, which are prevalent in the Indian subcontinent and West Eurasia,[46] are detected at appreciable levels in the Nepalese collections of Tamang (22.2%), Newar (19.7%) and Kathmandu (13.0%) but are almost absent in the Tibetan plateau (1.2%). It is noteworthy that haplogroup U sub-clades exhibit differential distributions among the three Nepalese collections. In Tamang (15.6%) and Newar (15.2%), a high proportion of the West Eurasian U sub-clade (U7) is observed compared to the South Central Asian component, U2 (6.7% and 4.5%, respectively), which is found at a higher frequency in Kathmandu (10.4%).[46]

The branch N*, a sister clade of haplogroup R, is almost exclusively represented by haplogroup A in our samples, with the exception of two N1e individuals (2.6%) from Kathmandu. Haplogroup A occurs at a frequency of 7.6% in the Himalayan collections, consistent with 5%-10% detected in East Asia.[39] Qin and colleagues[5] report the presence of haplogroup A at an average frequency of 8.6% in the plateau; however, it has not been observed in Nepal[16] until the current investigation. Unlike its derivative A2, which is common among Native American and northeast Siberian populations,[53,54] sub-clade A4 is found in Central and Northeast Asia.[39] This sub-haplogroup is present in all four studied populations, with the highest frequency in Newar (7.6%), followed by Tibet (5.8%), Kathmandu (5.2%) and Tamang (2.2%), while the paralogous branch, A11, is restricted to Tibet (3.8%) and Tamang (2.2%).[4,5]

Neutrality tests

Tables 3 and 4 display the intra-population diversity indices and the results of Tajima's *D* and Fu's F_s neutrality tests based on the HVRI (nps 16024-16395) and HVRII (nps 50-400) sequences of the control region. For the entire Himalayan collection, a total of 195 different haplotypes were defined by 125 variable sites in the HVRI dataset, and these numbers reduced to 140 and 64, respectively, in the case of HVRII. All four studied populations demonstrate a high degree of genetic heterogeneity, with the lowest haplotype diversity (HD) observed in Tamang (0.9576 for HVRI/0.9374 for HVRII) and the highest in Kathmandu (0.9949/0.9884). The level of diversity in the latter population may be partly inflated given the cosmopolitan nature of the region and this elevated value likely contributed to the greater HD in the combined Nepalese dataset (0.9936/0.9833) compared to the Tibetan population (0.9878/0.9700).

Fu's F_s values were statistically significant in all four Himalayan populations for both the HVRI and HVRII sequences, indicating historical population expansion (Tables 3 and 4). Tajima's *D* test also yielded significant negative values for each population, with the exception of Tamang and Newar for the HVRII data, where the rejection of the hypothesis of population growth may be the result of small sample sizes and/or genetic drift.[12,14] This notion is further supported by the fact that Newar exhibits the lowest nucleotide diversity (0.0161) and mean number of pairwise difference (6.0056) levels for the HVRI data, while Tamang displays the same pattern (0.0096 and 3.4061, respectively) for the HVRII region (Tables 3 and 4). Alternatively, the mutation rate differences of the HVRI and HVRII regions may account for the discrepancy in the Tajima's *D* test results between the two datasets, as previously reported.[5,39]

Time estimates

The coalescence times of major mtDNA lineages in the four Himalayan populations are presented in Table 5. Given that the time estimates generated using both mutation rates are marginally different, the coalescent ages based on Soares et al.[30] will be used throughout the narrative for comparative purposes, since recent studies[4,5] have employed the same calibration rate. In addition, unless otherwise stated, the haplogroup ages for Nepal are averaged across the three populations due to the limited sample size (less than five individuals) and/or haplotypes (less than three) within a given haplogroup for any individual Nepalese group. As such, the dates provided in this study should be viewed with caution since time estimates may be affected by several factors, including mutation rates, methodological approaches and demographic histories, such as bottleneck, effective population size and multiple founders.[5,30,55,56]

Of all the haplogroups examined, U2 lineages (which include the U2a and U2b sub-clades) were the most ancient, with average coalescence time estimates of 41.1 ± 10 kya and 40.4 ± 11.7 kya in Nepal and Kathmandu, respectively. The age estimates generated for sub-haplogroups D4 and F1 in Nepal and Tibet were found to range between 26-32 kya, suggesting the presence of these lineages in the Himalayas since the Upper Paleolithic period. On the other hand, haplogroups M5 (11.8 ± 7.8 kya) and U7 (11.3 ± 6.5 kya) yield early Holocene arrival times in Newar, most likely from the Indian sub-continent where age estimates of 36 ± 10 kya and 41.4 ± 15.8 kya, respectively, have been reported.[46,47] In the Tibetan collection, the M9a1 sub-clade and its derivative, M9a1a1c1b, date back to 12.6 ± 4.1 kya and 11.6 ± 4.1 kya, respectively, implying post-LGM expansions on the plateau. A similar time estimate was also obtained for Tibetan

A11 lineages (11.3 ± 5.3 kya), while the ages of haplogroups M13 and M62, which did not exhibit a star-like phylogeny, could not be determined for this population.

Population structure

Figure 1 displays a CA plot based on the mtDNA haplogroup frequencies of 37 populations (Table 1), including the four Himalayan groups examined in this study. Although the first two dimensions of the CA graph account for only a quarter of the genetic variance, three distinct groupings are evident: a Himalayan cluster occupying the top half of the plot, a South Central/Southwest Asian assemblage situated in the lower left quadrant and an East/Central Asian (except for Tajikistan) aggregate in the bottom right section. The Himalayan populations segregate away from the other two clusters along the *X*-axis, forming a loosely associated cluster in the upper right quadrant of the plot, with the exception of Newar and Kathmandu, which partition intermediate between the other Himalayan groups and the South Central/Southwest Asian assemblage. The Eastern Tharu population also displays a considerable degree of genetic separation from the other Himalayan collections, while the two Central Tharu collections exhibit stronger genetic affinities with the Tibetan populations compared to Tamang. The Tibetan populations form a tight cluster, except for Garze and Yushu, which partition away from this group, reflecting their distant geographical locations on the plateau.[5]

The results of the AMOVA using the four Himalayan populations and the 33 reference collections are presented in Table 6. As expected, the highest fraction of genetic diversity is observed within the population (93%) for both the geographic and linguistic clusters. When the populations are divided according to language sub-family

the proportion of among-group variance (3.43%) is slightly higher than among-populations-within-groups (3.17%). A similar pattern is observed when the populations are grouped based on geography, where the among-groups variance component accounts for 3.75% of the variation compared to 3.11% attributable to among-populations-within-groups (Table 6).

D. DISCUSSION

Gene flow from East Asia

Previous studies employing classical,[57,58] autosomal STR,[14,59] Y-chromosome[12,60] and mtDNA[4,5,13] polymorphisms have all suggested a shared common ancestry between the Tibetan population and those originating in Northeast Asia. Additionally, a recent report on the Tharus from Nepal revealed Northeast Asian haplogroups in their maternal gene pools.[16] Given these findings, it is not surprising that several mtDNA lineages typical of Northeast Asians, including haplogroups A, D, G and M8[42,61] were detected in the current investigation in the Himalayan populations. These haplogroups collectively account for more than 25% of the matrilineal diversity in the Himalayas, with the highest frequency observed in Tamang (44.2%), followed by Tibet (37.5%), Newar (31.7%) and Kathmandu (26%).

Of the Northeast Asian lineages present in the Himalayan maternal gene pools, the sub-clades A4, D4 and G2 are the most prevalent. These three sub-haplogroups also occur predominantly throughout Central Asia, while the remaining A, D and G sub-lineages show decreasing frequency gradients from east to west.[62] Interestingly, G2a, a sub-clade defined by the HVRI motif 16223, 16227, 16278 and 16362, with restricted

distribution in Central Asia, is observed exclusively in Tibet (Figure 2).[5,61] In addition, the M8 clade in Tibet is largely characterized by haplogroup C, which is the second most frequent East Asian lineage in Central Asia.[61] On the other hand, the high frequency of haplogroup Z in Newar may be the result of a founder effect and/or genetic drift, as reflected by the limited number of haplotypes (2) representing this lineage. The presence of haplogroups from East and Central Asia may suggest that the Himalayan region served as a conduit for gene flow between the east and west, possibly along the Silk Road.[40,63]

Southeast Asian haplogroups are relatively less prevalent in the Nepalese gene pool compared to the Tibetans. While F1 lineages are present across all four studied populations, M9a1 derivatives occur at high frequencies only in Tibet and Tamang. The presence of the latter sub-haplogroup in two Central Tharu populations from Nepal[16] may account for their affinity with the Tamang and Tibetan collections in the CA plot (Figure 1). Haplogroup B is observed at low frequencies in Newar (1.5%), Kathmandu (1.3%) and Tibet (1.3%), whereas M7 and E, which are common among Southeast Asian populations, are not detected in the Himalayan dataset.

Peng et al.[25] proposed South China and/or Southeast Asia as the source of post-LGM expansion of basal haplogroup M9a'b lineages throughout East Asia. The M9a1 sub-clade, in particular, has its origin in South China, where it most likely differentiated (around 17-21 kya) prior to its expansion westward into the Tibetan plateau and, later on, into northeast India and the southern Himalayan regions.[25] This observation is consistent with the average time estimate obtained for the Himalayan M9a1 samples in the present study (14.7 ± 4.1 kya) as well as with the distributions of this haplogroup in the Tibetan and Tamang populations, both of which exhibit primarily M9a1a1c1b lineages (21.8%

and 8.9%, respectively). The star-like network (Figure 3) and coalescent age of M9a1a1c1b (11.6 ± 4.1 kya) suggest that this lineage expanded onto the plateau soon after its separation from the parent haplogroup (M9a1). It is also interesting to note that all four Tamang mtDNA samples belonging to M9a1a1c1b were found to possess the same control-region haplotype, indicating a founder effect, possibly from Tibet, where higher sequence diversity within this lineage is observed (Supplementary Table 1).

The haplogroup F is found at its greatest frequency and diversity in Southeast Asia.[36,64] Although its derivative, F1, is present across East Asia, it shows a decreasing cline of frequencies from south to north.[65] The sub-haplogroup F1 has also been previously reported in Northeast India,[66] Nepal[16] and Tibet,[4,5] paralleling the distribution of Y-haplogroup O3a3c-M134, which has been associated with the spread of the Tibeto-Burman language.[12] This observation is consistent with the AMOVA results from the present study, which displays significant population structure along the linguistic affiliation (Table 6), although the difference in variance components between among-groups (3.4%) and among-populations-within-groups (3.17%) is not substantial. The presence of F1 in the Nepalese collection also suggests that Nepal may have served as a bridge for gene flow from Southeast Asia to India.[15] This notion is in agreement with the findings of Peng et al.[25] who describe the spread of F1c from South China to Northeast India and the southern portion of the Himalayas, mirroring the expansion of the M9a1b and M9a1a2 lineages. Alternatively, given the diversity of F1 sub-lineages and the presence of basal haplogroup in Tibet, it is possible that F1 derivatives in Nepal may be attributed to gene flow from the Tibetan plateau. In addition, haplogroups A, D, G, M8 (including M8a, C and Z) and M9a1 display a relatively higher level of diversity in Tibet

as compared to the Nepalese populations, lending further support to this southward trans-Himalayan migratory route.[12]

South Central Asian genetic signatures

Given their geographical proximity and strong historical ties, it is not surprising that India and Nepal share close linguistic, religious and cultural affinities. In the current investigation, we also observe genetic ties between the Nepalese and Indian collections in the CA plot (Figure 1), which shows Kathmandu and Newar partitioning away from the other Himalayan groups in the direction of the Kashmir and Punjab populations from India. This finding is likely the result of a high proportion of Indian-specific lineages in the maternal gene pools of Kathmandu (49.3%) and Newar (36.4%). In addition, significant frequencies (67.5%) of Indian-specific haplogroups have been reported in the Eastern Tharu population from Nepal,[16] which may explain this group's position proximal to Newar and Kathmandu along the Y-axis in Figure 1. In contrast to the Y-chromosome data of Tamang,[12] which is mostly composed of the East Asian derivative O3a5c-M134 (86.6%), this collection's mtDNA profile indicates a considerable level (17.8%) of gene flow from the Indian subcontinent. The high peaks, rough terrain and colder climates towards the north, however, may have discouraged human migration into the plateau, since the Tibetan population shows minimal maternal influence (1.9%) from India.[12,14]

The Nepalese mtDNA pool consists of several ancient Indian lineages emerging directly from the roots of macrohaplogroups M (M2, M5, M34, M35 and M43) and N (R5, R6, U2a and U2b), suggesting a deep common ancestry between these two

neighboring groups of populations. In particular, haplogroups M2 and R5 are reported as some of the oldest lineages in the Indian subcontinent, with coalescence ages >50 kya.[46,47] This finding is consistent with the deep coalescence ages obtained for the haplogroup U2 in Nepal (41.1 ± 10.0 kya) and Kathmandu (40.4 ± 11.7 kya), supporting an early settlement of modern humans in the region. On the other hand, haplogroup M5, which has been previously observed in the general Andhra Pradesh population (South India) and in Hindus from New Delhi,[16,46] exhibits a relatively younger age in Newar (11.8 ± 7.8 kya) than in India (36 ± 10 kya). It is interesting to note that this time estimate overlaps with the age of the pan-Indian Y-haplogroup R2-M124 in Newar (10.5 ± 3.1 kya), although Kathmandu exhibits a relatively older date (17.2 ± 4.6 kya) (data not shown). In addition, both of these populations are characterized by Y-haplogroup H1-M82, which is, for the most part, restricted to southern India, while Tamang displays two F*-M213 samples, also common in the Indian peninsula.[12] The prevalence of ancient maternal lineages in combination with the abovementioned south Indian Y-chromosomes in our Nepalese populations may represent genetic signatures of an early human dispersal from Africa to the Indian subcontinent along the southern coastal route.

Genetic input from West Eurasia

Genetic influences from West Eurasia are represented predominantly by haplogroup U7 in Newar and Tamang, whereas Kathmandu shows low levels of other West Eurasian lineages, namely H, HV, J and T. Haplogroup U7 is detected at high frequencies in populations throughout Iran, Pakistan, northwestern India and the Arabian Peninsula.[67] In India, U7 lineages are highest in Gujarat (12%) and Punjab (9%) but are

at frequencies ≤ 2% in the rest of the country.[46] Notably, our U7 samples possessed neither the Indian (16207-16309-16318T) nor the West Eurasian (16126-16309-16318T) founder HVRI motifs, whose coalescence times of 20-30 kya[46] are considerably older than those of Newar (11.3 ± 6.5 kya). While all U7 samples in the Himalayas carry the basal haplotype (16309-16318T), with the exception of one sample from Kathmandu (16309-16318C), all but one individual from Tamang harbor an additional mutation at np 16298. The majority of Newar U7 samples are characterized by a transition at np 16223, two of which also possess another mutation at np 16289 (Figure 4 and Supplementary Table 1). Given their restricted distribution in the region, these Tamang (16298-16309-16318T) and Newar (16223-16289-16309-16318T) specific U7 lineages suggest an in situ differentiation in Nepal and possibly represent new branches within this haplogroup. On the other hand, the homogeneity of the U7 control-region mutations in Tamang and Newar, as reflected by their distinct clusters in the network projection (Figure 4), may be the result of genetic drift and/or a founder effect. This notion is further supported by relatively low levels of haplotype and haplogroup diversity (Tables 2, 3 and 4) observed in these two populations. Additionally, the high incidence of U7 in Tamang (15.6%) contrasts with the near absence of Indian and West Eurasian Y-haplogroups.

Comparison between paternal and maternal lineages in the Himalayas

Y-chromosomal data indicates that the majority of Tibetan males (>50%) are characterized by haplogroup D-M174,[12] a lineage with restricted but distinctive branches in the Andaman Islands (D-M174*), Tibet (D1-M15 and D3-P47) and Japan (D2-M55).[60,68] Yet, in the present study, the Tibetan mtDNA pool did not display any lineages

equivalent to D-M174 in frequency and/or distribution. According to Zhao and colleagues,[4] clades M9a and M13 mirror the Y-haplogroup D-M174 profile, given the presence of both of these lineages in Japan. However, due to the limited resolution of the current Y-chromosome phylogeny, the same authors suggest that it may be more appropriate to equate D-M174 with macrohaplogroup M. Unfortunately, comparison of these uniparental genomes at their basal haplogroups is misleading since, unlike the sporadic occurrence of D-M174 in East Asia (with the exception of Tibet and Japan), mtDNA M lineages show widespread distributions throughout Central, East and South Central Asia, with the highest frequency in India.[47] Moreover, the M lineages prevalent in the Indian subcontinent do not overlap with those observed on the Tibetan plateau. Alternatively, as Shi and collaborators[69] suggested, the initial genetic signature of Y-haplogroup D-M174 in East Asia may have been erased by a combination of subsequent Neolithic expansion and the LGM in the region, while the enrichment of this haplogroup in Tibet and Japan may be the result of isolation and/or genetic drift.

The other half of the Tibetan paternal gene pool is represented predominantly by the East Asian-specific haplogroup O3a5c-M134, a marker common among Tibeto-Burman speakers. The distribution of this Y-haplogroup parallels that of F1 matrilineages, which are also observed in populations throughout East Asia. Like the Y-chromosome data, the Tibetan mtDNA pool exhibits limited Indian mtDNA haplogroups, further corroborating the notion that the Himalayan Mountain range acted as a natural barrier for gene flow from the Indian subcontinent.[12]

The majority of Tamang's Y-chromosomes is composed of Y-haplogroup O3a5c-M134 (86.6%), suggesting significant gene flow from East Asia.[12,70] In contrast, the

mtDNA pool of Tamang consists of moderate frequencies of South Central Asian (17.6%) and West Eurasian (15.6%) haplogroups. The inconsistencies between these two haploid marker systems may be attributed to sex-biased genetic flow resulting from male-driven migration from Tibet to Tamang, followed by genetic admixture with the local females. Alternatively, the small sample size of the Tamang and/or a founder effect may have contributed to the differences observed in Y-chromosome and mtDNA profiles.

On the other hand, the maternal gene pools of Newar and Kathmandu, like their paternal components, exhibit a substantial proportion of Indian-specific lineages, signaling a direct genetic connection between the two regions.[12] While the presence of Y-haplogroup R-M207 in Newar (62.1%) and Kathmandu (46.7%) may be associated with the recent entry of Aryans from the northern Indian plains, who likely introduced the Indo-European language (Nepali) in Nepal, the presence, albeit limited, of ancient Indian maternal lineages, such as M2, R5 and U2 in Nepal suggests that this region may have been inhabited by the earliest settlers during the initial peopling of South Central Asia.[46]

REFERENCES

1 Aldenderfer M, Yinong Z: The prehistory of the Tibetan Plateau to the seventh century AD: perspectives and research from China and the West since 1950. *J World Prehist* 2004; **18:** 1–55.

2 Huang W: The prehistoric human occupation of the Qinghai-Xizang plateau. Gottinger Geographische Abhandlungen, 1994; **95:** 201–219.

3 Brantingham P, Haizhou M, Olsen J, Xing G, Madsen D, Rhode D: Speculation on the timing and nature of Late Pleistocene hunter-gatherer colonization of the Tibetan Plateau. *Chin Sci Bull* 2003; **48:** 1510–1516.

4 Zhao M, Kong QP, Wang HW *et al:* Mitochondrial genome evidence reveals successful Late Paleolithic settlement on the Tibetan Plateau. *Proc Natl Acad Sci U S A* 2009; **106:** 21230–21235.

5 Qin Z, Yang Y, Kang L *et al:* A mitochondrial revelation of early human migrations to the Tibetan Plateau before and after the last glacial maximum. *Am J Phys Anthropol* 2010; **143:** 555–569.

6 Richardson H: A short history of Tibet, New York: E.P. Dutton & Co Inc, 1962.

7 Whelpton J: World Bibliographical Series: Nepal. Denver, Clio Press, 1990.

8 Berreman GD. People and Cultures of the Himalayas. *Asian Survey* 1963; **3:** 289–304.

9 Lewis TT, Shakya DR: Contributions to the history of Nepal: eastern Newar diaspora settlements. *Contributions to Nepalese studies* 1988; **15:** 25-65.

10 Regmi DR: Ancient Nepal. (3d ed.) Calcutta: Mukhopadhyay KL, 1969

11 Malla KP: Linguistic Archaeology of the Nepal Valley. A Preliminary Report. *Kailash* 1981; **8:** 5–23.

12 Gayden T, Cadenas AM, Regueiro M *et al:* The Himalayas as a directional barrier to gene flow. *Am J Hum Genet* 2007; **80:** 884–894.

13 Torroni A, Miller JA, Moore LG, Zamudio S, Zhuang J, Droma T, Wallace DC: Mitochondrial DNA analysis in Tibet: implications for the origin of the Tibetan population and its adaptation to high altitude. *Am J Phys Anthropol* 1994; **93:** 189–199.

14 Gayden T, Mirabal S, Cadenas AM *et al:* Genetic insights into the origins of Tibeto-Burman populations in the Himalayas. *J Hum Genet* 2009; **54:** 216–223.

15 Thangaraj K, Chaubey G, Kivisild T *et al:* Maternal footprints of Southeast Asians in North India. *Hum Hered* 2008; **66:** 1–9.

16 Fornarino S, Pala M, Battaglia V *et al:* Mitochondrial and Y-chromosome diversity of the Tharus (Nepal): a reservoir of genetic variation. *BMC Evol Biol* 2009; **9:** 154.

17 Antunez-de-Mayolo G, Antunez-de-Mayolo A, Antunez-de-Mayolo P *et al:* Phylogenetics of worldwide human populations as determined by polymorphic Alu insertions. *Electrophoresis* 2002; **23:** 3346–3356

18 Vigilant L, Pennington R, Harpending H, Kocher TD, Wilson AC: Mitochondrial DNA sequences in single hairs from a southern African population. *Proc Natl Acad Sci U S A* 1989; **86:** 9350-9354.

19 Stoneking M, Hedgecock D, Higuchi RG, Vigilant L, Erlich HA: Population variation of human mtDNA control region sequences detected by enzymatic amplification and sequence-specific oligonucleotide probes. *Am J Hum Genet* 1991; **48:** 370-382.

20 Anderson S, Bankier AT, Barrell BG, *et al:* Sequence and organization of the human mitochondrial genome. *Nature* 1981; **290:** 457–465.

21 Andrews RM, Kubacka I, Chinnery PF, Lightowlers RN, Turnbull DM, Howell N: Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. *Nat Genet* 1999; **23:** 147.

22 Hall TA: BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucl Acids Symp Ser* 1999; **41:** 95–98.

23 van Oven M, Kayser M: Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. *Hum Mutat* 2009; **30:** E386–394.

24 Kong QP, Bandelt HJ, Sun C *et al:* Updating the East Asian mtDNA phylogeny: a prerequisite for the identification of pathogenic mutations. *Hum Mol Genet* 2006; **15:** 2076–2086.

25 Peng MS, Palanichamy MG, Yao YG *et al:* Inland post-glacial dispersal in East Asia revealed by mitochondrial haplogroup M9a'b. *BMC Biol* 2011; **9:** 2.

26 Gayden T, Regueiro M, Martinez L, Cadenas AM, Herrera RJ: Human Y-chromosome haplotyping by allele-specific polymerase chain reaction. *Electrophoresis* 2008; **29:** 2419–2423.

27 Excoffier L, Laval LG, Schneider S: Arlequin (version 3.0): an integrated software package for population genetics data analysis. *Evol Bioinform Online* 2007; **1:** 47–50.

28 Roostalu U, Kutuev I, Loogväli, EL *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* 2007; **24:** 436–448.

29 Forster P, Harding R, Torroni A, Bandelt HJ: Origin and evolution of Native American mtDNA variation: a reappraisal. *Am J Hum Genet* 1996; **59:** 935–945.

30 Soares P, Ermini L, Thomson N *et al.* Correcting for purifying selection: an improved human mitochondrial molecular clock. *Am J Hum Genet* 2009; **84:** 740–759.

31 Saillard J, Forster P, Lynnerup N, Bandelt HJ, Norby S: mtDNA variation among Greenland Eskimos: the edge of the Beringian expansion. *Am J Hum Genet* 2000; **67:** 718–726.

32 Nohira C, Maruyama S, Minaguchi K: Phylogenetic classification of Japanese mtDNA assisted by complete mitochondrial DNA sequences. *Int J Legal Med* 2010; **124:** 7–12.

33 Lee HY, Yoo JE, Park MJ, Chung U, Shin KJ: Mitochondrial DNA control region sequences in Koreans: identification of useful variable sites and phylogenetic analysis for mtDNA data quality control. *Int J Legal Med* 2006; **120:** 5–14.

34 Zhang YJ, Xu QS, Zheng ZJ, Lin HY, Lee JB: Haplotype diversity in mitochondrial DNA hypervariable region I, II and III in northeast China Han. *Forensic Sci Int* 2005; **149:** 267–269.

35 Wen B, Li H, Lu D *et al:* Genetic evidence supports demic diffusion of Han culture. *Nature* 2004; **431:** 302–305.

36 Hill C, Soares P, Mormina M *et al.* A mitochondrial stratigraphy for island southeast Asia. *Am J Hum Genet* 2007; **80:** 29–43.

37 Li H, Cai X, Winograd-Cort ER, *et al.* Mitochondrial DNA diversity and population differentiation in southern East Asia. *Am J Phys Anthropol* 2007; **134:** 481–488.

38 Tabbada KA, Trejaut J, Loo JH *et al:* Philippine mitochondrial DNA diversity: a populated viaduct between Taiwan and Indonesia? *Mol Biol Evol* 2010; **27:** 21–31.

39 Derenko M, Malyarchuk B, Grzybowski T *et al:* Phylogeographic analysis of mitochondrial DNA in northern Asian populations. *Am J Hum Genet* 2007; **81:** 1025–1041.

40 Cheng B, Tang W, He L: Genetic imprint of the Mongol: signal from phylogeographic analysis of mitochondrial DNA. *J Hum Genet* 2008; **53:** 905–913.

41 Kolman CJ, Sambuughin N, Bermingham E: Mitochondrial DNA analysis of Mongolian populations and implications for the origin of New World founders. *Genetics* 1996; **142:** 1321–1334.

42 Kivisild T, Tolk HV, Parik J, Wang Y, Papiha SS, Bandelt HJ, Villems R: The emerging limbs and twigs of the East Asian mtDNA tree. *Mol Biol Evol* 2002; **19:** 1737–1751.

43 Cordaux R, Saha N, Bentley GR, Aunger R, Sirajuddin SM, Stoneking M: Mitochondrial DNA analysis reveals diverse histories of tribal populations from India. *Eur J Hum Genet* 2003; **11:** 253–264.

44 Rakha A, Shin KJ, Yoon JA *et al:* Forensic and genetic characterization of mtDNA from Pathans of Pakistan. *Int J Legal Med* 2011; **125:** 841–848.

45 Rohlf FJ: NTSYS pc: Numerical Taxonomy System, Version 2.1. Exeter Publishing, Setauket, NY, 2002.

46 Metspalu M, Kivisild T, Metspalu E *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* 2004; **5:** 26.

47 Chandrasekar A, Kumar S, Sreenath J *et al.* Updating phylogeny of mitochondrial DNA macrohaplogroup M in India: dispersal of modern human in South Asian corridor. *PLoS ONE* 2009; **4:** e7447.

48 Schurr TG, Sukernik RI, Starikovskaya YB, Wallace DC. Mitochondrial DNA variation in Koryaks and Itel'men: population replacement in the Okhotsk Sea-Bering Sea region during the Neolithic. *Am J Phys Anthropol* 1999; **108:** 1–39.

49 Tambets K, Rootsi S, Kivisild T *et al.* The western and eastern roots of the Saami--the story of genetic "outliers" told by mitochondrial DNA and Y chromosomes. *Am J Hum Genet* 2004; **74:** 661–682.

50 Ingman M, Gyllensten U: A recent genetic link between Sami and the Volga-Ural region of Russia. *Eur J Hum Genet* 2006; **15:** 115–120.

51 Wang HW, Mitra B, Chaudhuri TK, Palanichamy MG, Kong QP, Zhang YP: Mitochondrial DNA evidence supports northeast Indian origin of the aboriginal Andamanese in the Late Paleolithic. *J Genet Genomics* 2011; **38:** 117–122.

52 Endicott P, Metspalu M, Stringer C, Macaulay V, Cooper A, Sanchez JJ: Multiplexed SNP typing of ancient DNA clarifies the origin of Andaman mtDNA haplogroups amongst South Asian tribal populations. *PLoS ONE* 2006; **1,** e81.

53 Achilli A, Perego UA, Bravi CM *et al:* The phylogeny of the four pan-American MtDNA haplogroups: implications for evolutionary and disease studies. *PLoS ONE* 2008; **3:** e1764.

54 Kumar S, Bellis C, Zlojutro M, Melton PE, Blangero J, Curran JE: Large scale mitochondrial sequencing in Mexican Americans suggests a reappraisal of Native American origins. *BMC Evol Biol* 2011; **11:** 293.

55 Endicott P, Ho SY, Metspalu M, Stringer C: Evaluating the mitochondrial timescale of human evolution. *Trends Ecol Evol* 2009; **24:** 515-521.

56 Henn BM, Gignoux CR, Feldman MW, Mountain JL: Characterizing the time dependency of human mitochondrial DNA mutation rate estimates. *Mol Biol Evol* 2009; **26:** 217–230.

57 Nei M, Roychoudhury AK: Evolutionary relationships of human populations on a global scale. *Mol Biol Evol* 1993; **10:** 927–943.

58 Cavalli-Sforza LL, Menozzi P, Piazza A: The history and geography of human genes. Princeton University Press, Princeton, NJ, 1994.

59 Kang L, Li S, Gupta S, Zhang Y, Liu K, Zhao J, Jin L, Li H: Genetic structures of the Tibetans and the Deng people in the Himalayas viewed from autosomal STRs. *J Hum Genet* 2010; **55:** 270–277.

60 Hammer MF, Karafet TM, Park H, Omoto K, Harihara S, Stoneking M, Horai S. Dual origins of the Japanese: common ground for hunter-gatherer and farmer Y chromosomes. *J Hum Genet* 2006; **51:** 47–58.

61 Comas D, Plaza S, Wells RS, Yuldaseva N, Lao O, Calafell F, Bertranpetit J: Admixture, migrations, and dispersals in Central Asia: evidence from maternal DNA lineages. *Eur J Hum Genet* 2004; **12:** 495–504.

62 Metspalu M, Kivisild T, Bandelt HJ, Richards M, Villem R: The Pioneer Settlement of Modern Humans in Asia; in Bandelt HJ, Macaulay V, Richards M (eds): Human Mitochondrial DNA and the Evolution of Homo sapiens. Berlin, 2006, pp. 181-199

63 Comas D, Calafell F, Mateu E *et al:* Trading genes along the silk road: mtDNA sequences and the origin of central Asian populations. *Am J Hum Genet* 1998; **63:** 1824–1838.

64 Ballinger SW, Schurr TG, Torroni A *et al*: Southeast Asian mitochondrial DNA analysis reveals genetic continuity of ancient mongoloid migrations. *Genetics* 1992; **130:** 139–152.

65 Yao YG, Kong QP, Bandelt HJ, Kivisild T, Zhang YP: Phylogeographic differentiation of mitochondrial DNA in Han Chinese. *Am J Hum Genet* 2002; **70:** 635–651.

66 Reddy BM, Langstieh BT, Kumar V *et al:* Austro-Asiatic tribes of Northeast India provide hitherto missing genetic link between South and Southeast Asia. *PLoS ONE* 2007; **2:** e1141.

67 Quintana-Murci L, Chaix R, Wells RS: Where west meets east: the complex mtDNA landscape of the southwest and Central Asian corridor. *Am J Hum Genet* 2004; **74:** 827–845.

68 Thangaraj K, Singh L, Reddy AG: Genetic affinities of the Andaman Islanders, a vanishing human population. *Curr Biol* 2003; **13:** 86-93.

69 Shi H, Zhong H, Peng Y *et al:* Y chromosome evidence of earliest modern human settlement in East Asia and multiple origins of Tibetan and Japanese populations. *BMC Biol* 2008; **6:** 45.

70 Gayden T, Chennakrishnaiah S, La Salvia J *et al:* Y-STR diversity in the Himalayas. *Int J Legal Med* 2011; **125:** 367-375.

Appendix IV

Figure 1. Correspondence Analysis (CA) based on haplogroup frequencies from 37 populations. Population abbreviations correspond to those found in Table 1.

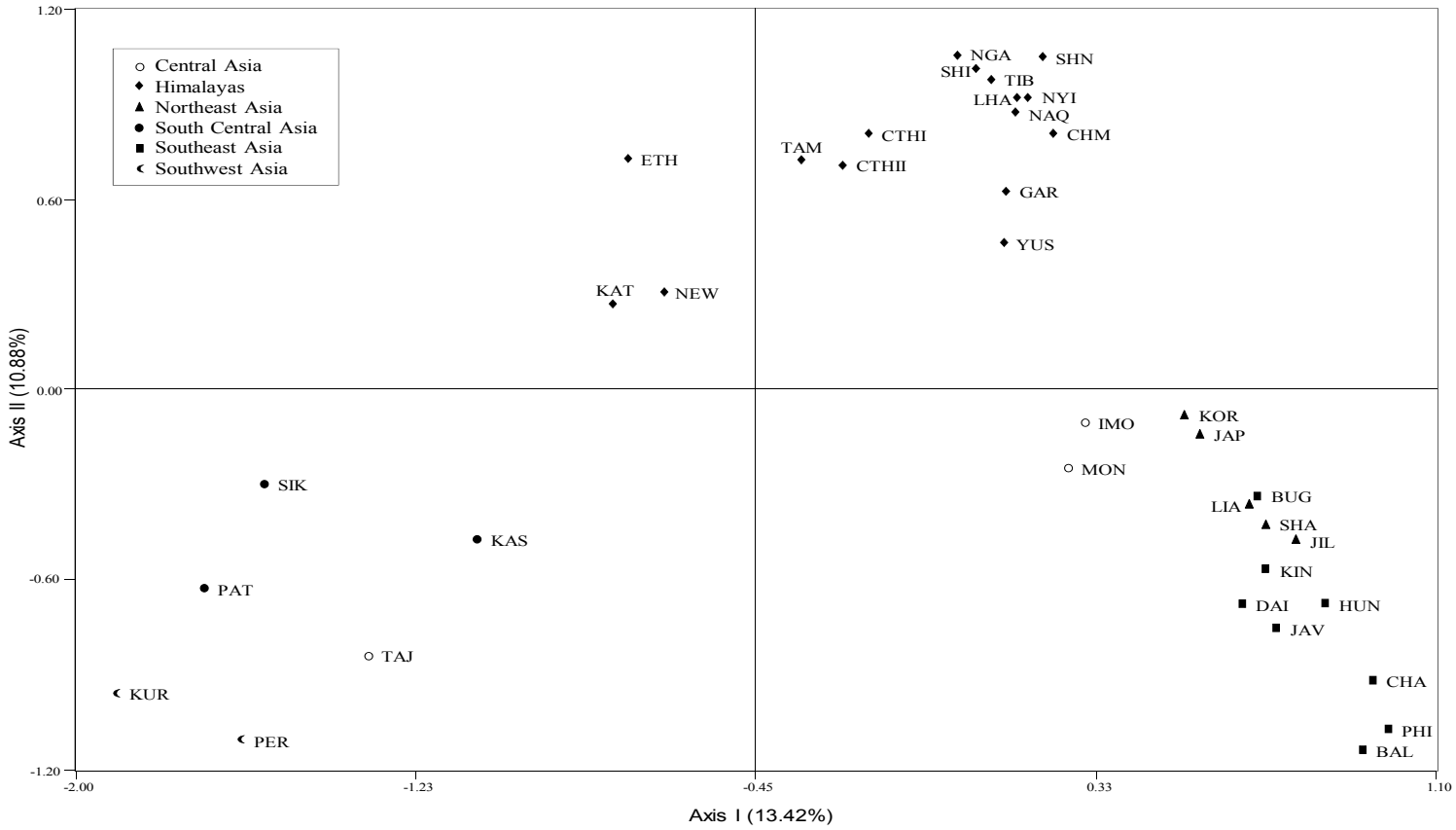

Figure 2. Median-joining network based on the HVRI data within haplogroup G2. Circle areas are proportional to the haplotype frequency, and the smallest circle corresponds to one individual. Mutation positions relative to rCRS are shown along the branches. HVRI data used in this analysis were reported by Qin et al. (2010) (Monba) and Fornarino et al. (2009) (Tharu).

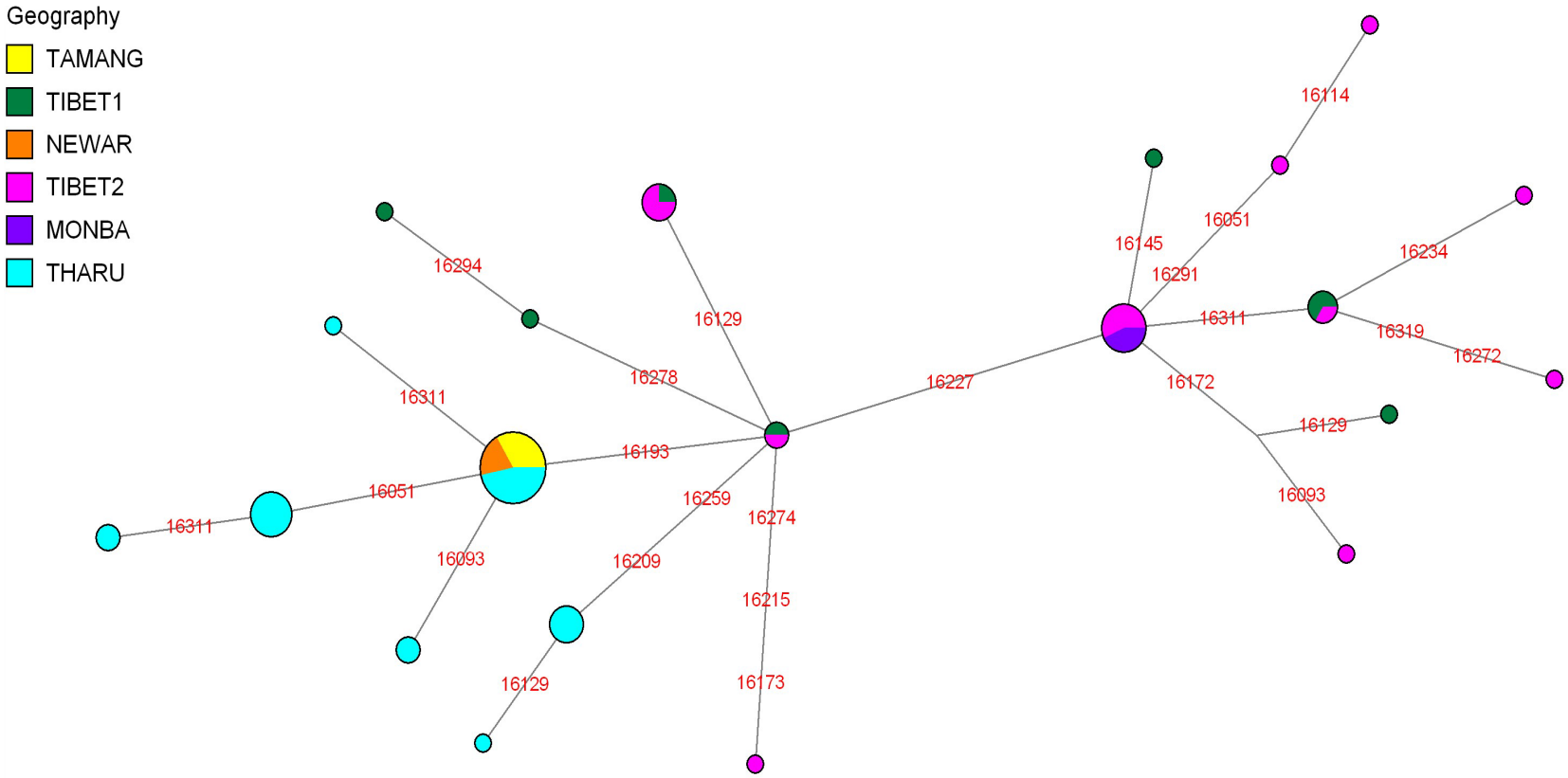

Figure 3. Phylogenetic network of haplogroup M9a1 lineages in the Himalayan and Northeast Indian populations. Circle areas are proportional to the haplotype frequency, and the smallest circle corresponds to one individual. HVRI mutation positions relative to rCRS are shown along the branches. HVRI data used in this analysis were reported by Qin et al., 2010 (Tibet2 and Monba), Fornarino et al. (2009) (Tharu) and Peng et al. (2011) [Northeast India (NEINDIA)].

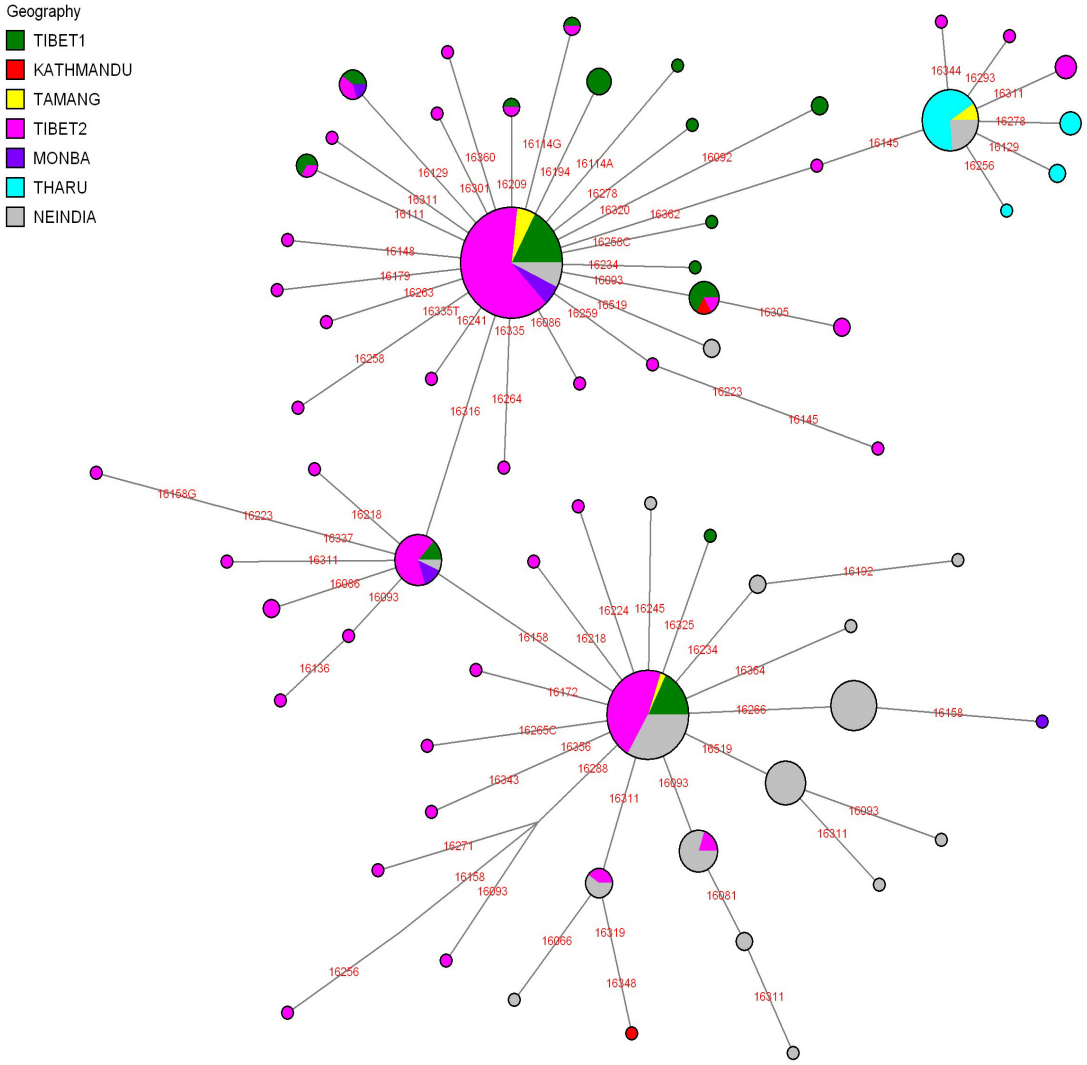

Figure 4. Network of the U7 lineage based on HVRI data. Circle areas are proportional to the haplotype frequency, and the smallest circle corresponds to one individual. Mutation positions relative to rCRS are shown along the branches. HVRI data used in this analysis were reported by Qin et al. (2010) (Tibet2), Fornarino et al. (2009) (Andhra Paradesh, New Delhi and Tharu) and Metspalu et al. (2004) (Gujarat, North Iran, South Iran and Pakistan).

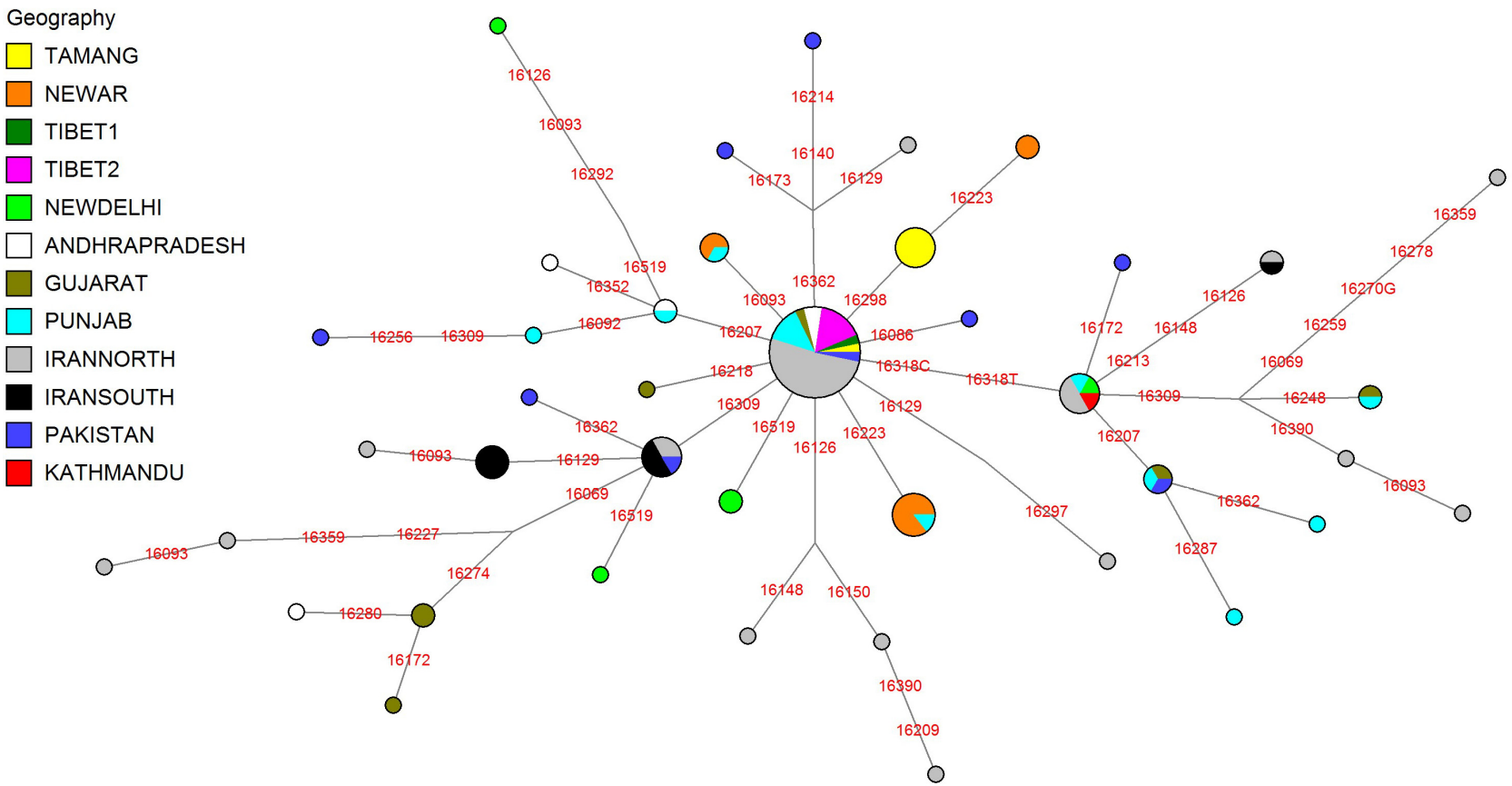

Table 1. Populations Analyzed

Population	Abbreviation	n	Language family/sub-family	References
NORTHEAST ASIA				
Japan	JAP	100	Japanese	Nohira et al. 2010
Korea	KOR	593	Korean	Lee et al. 2006
Jilin (China)	JIL	51	Sino-Tibetan/Chinese	Zhang et al. 2005
Shanghai (China)	SHA	56	Sino-Tibetan/Chinese	Wen et al. 2004
Liaoning (China)	LIA	51	Sino-Tibetan/Chinese	Wen et al. 2004
SOUTHEAST ASIA				
Bali	BAL	82	Austronesian/Malayo-polynesian	Hill et al. 2007
Java	JAV	46	Austronesian/Malayo-polynesian	Hill et al. 2007
Hunan (China)	HUN	16	Sino-Tibetan/Chinese	Wen et al. 2004
Bugan (China)	BUG	32	Austro-Asiatic/Mon-Khmer	Li et al. 2007
Dai-lu (China)	DAI	30	Tai-Kadai/Kam-Tai	Li et al. 2007
Cham (Vietnam)	CHA	168	Austronesian/Malayo-polynesian	Peng et al. 2010
Kinh (Vietnam)	KIN	139	Austro-Asiatic/Mon-Khmer	Peng et al. 2010
Philippines	PHI	423	Austronesian/Malayo-polynesian	Tabbada et al. 2010
SOUTHWEST ASIA				
Persian	PER	82	Indo-European/Indo-Iranian	Derenko et al. 2007
Kurds	KUR	25	Indo-European/Indo-Iranian	Derenko et al. 2007
CENTRAL ASIA				
Inner Mongolia	IMO	155	Altaic/Mongolic	Cheng et al. 2008
Mongolia	MON	103	Altaic/Mongolic	Kolman et al. 1996
Tajikistan	TAJ	44	Indo-European/Indo-Iranian	Derenko et al. 2007
SOUTH CENTRAL ASIA				
Kashmir	KAS	19	Indo-European/Indo-Iranian	Kivisild et al. 2002
Sikh	SIK	40	Indo-European/Indo-Iranian	Cordaux et al. 2003
Pathan (Pakistan)	PAT	230	Indo-European/Indo-Iranian	Rakha et al. 2011
HIMALAYAS				
Central Tharu I (Nepal)	CTHI	57	Indo-European/Indo-Iranian	Fornarino et al. 2009
Central Tharu I (Nepal)	CTHII	76	Indo-European/Indo-Iranian	Fornarino et al. 2009
East Tharu (Nepal)	ETH	40	Indo-European/Indo-Iranian	Fornarino et al. 2009
Newar	NEW	66	Sino-Tibetan/Tibeto-Burman	present study
Kathmandu	KAT	77	Indo-European/Indo-Iranian	present study
Tamang	TAM	45	Sino-Tibetan/Tibeto-Burman	present study
Tibet	TIB	153	Sino-Tibetan/Tibeto-Burman	present study
Chamdo	CHM	61	Sino-Tibetan/Tibeto-Burman	Qin et al. 2010
Garze	GAR	55	Sino-Tibetan/Tibeto-Burman	Qin et al. 2010
Lhasa	LHA	59	Sino-Tibetan/Tibeto-Burman	Qin et al. 2010
Nyingchi	NYI	53	Sino-Tibetan/Tibeto-Burman	Qin et al. 2010
Ngari	NGA	46	Sino-Tibetan/Tibeto-Burman	Qin et al. 2010
Nagqu	NAQ	58	Sino-Tibetan/Tibeto-Burman	Qin et al. 2010
Shannan	SHN	56	Sino-Tibetan/Tibeto-Burman	Qin et al. 2010
Shigatse	SHI	59	Sino-Tibetan/Tibeto-Burman	Qin et al. 2010
Yushu	YUS	44	Sino-Tibetan/Tibeto-Burman	Qin et al. 2010

Table 2. MtDNA Haplogroup frequencies of the four Himalayan populations studied.

Haplogroup	Tamang (45)	Newar (66)	Kathmandu (77)	Tibet (156)
A4	0.022	0.076	0.052	0.058
A11	0.022	0.000	0.000	0.038
B4	0.000	0.000	0.000	0.013
B5	0.000	0.015	0.013	0.000
C4	0.022	0.000	0.039	0.032
C4a1	0.000	0.000	0.000	0.032
C5	0.000	0.000	0.000	0.019
D4	0.111	0.030	0.013	0.013
D4e1a	0.044	0.000	0.000	0.000
D4g2a	0.000	0.000	0.013	0.013
D4i	0.000	0.000	0.052	0.006
D4j1a	0.000	0.000	0.013	0.006
D4j3	0.000	0.000	0.000	0.026
D4q	0.000	0.000	0.000	0.019
D5	0.022	0.000	0.000	0.000
D5a2	0.000	0.030	0.013	0.000
D5a2a1	0.022	0.000	0.013	0.006
D5a3	0.066	0.015	0.000	0.006
F1	0.000	0.015	0.013	0.077
F1a1	0.000	0.000	0.000	0.006
F1b	0.022	0.000	0.013	0.019
F1d	0.022	0.076	0.039	0.058
F1c1a	0.022	0.015	0.013	0.026
F2b	0.000	0.015	0.000	0.000
G2	0.111	0.045	0.000	0.026
G2a	0.000	0.000	0.000	0.026
G3	0.000	0.000	0.000	0.019
G3a1	0.000	0.000	0.000	0.013
HV	0.000	0.030	0.013	0.013
H	0.000	0.000	0.013	0.000
J	0.000	0.000	0.013	0.000
M2a1	0.000	0.015	0.000	0.000
M2a3a	0.000	0.015	0.000	0.000
M3	0.000	0.015	0.052	0.000
M4'67	0.044	0.000	0.000	0.006
M4	0.000	0.015	0.000	0.000
M5	0.000	0.121	0.039	0.000
M8a	0.000	0.000	0.000	0.013
M9a1a1c1b	0.089	0.000	0.013	0.218

M9a1a2	0.044	0.000	0.000	0.000
M9a1b1	0.022	0.000	0.013	0.058
M10a1	0.000	0.000	0.013	0.000
M10a2	0.000	0.000	0.013	0.000
M11a2	0.000	0.000	0.000	0.006
M13a	0.000	0.000	0.013	0.000
M13a1b	0.000	0.000	0.000	0.032
M13a2	0.000	0.000	0.000	0.013
M30	0.000	0.015	0.039	0.000
M30c1	0.000	0.030	0.000	0.000
M30e	0.000	0.000	0.013	0.000
M31a2	0.022	0.000	0.000	0.000
M33	0.000	0.000	0.039	0.000
M33a1a	0.000	0.061	0.013	0.000
M33b1	0.022	0.000	0.013	0.000
M34	0.000	0.015	0.013	0.000
M35	0.000	0.015	0.052	0.000
M38	0.000	0.000	0.013	0.000
M43	0.022	0.000	0.026	0.000
M61	0.000	0.000	0.000	0.006
M62a	0.000	0.000	0.000	0.006
M62b	0.000	0.000	0.000	0.026
M70	0.000	0.000	0.000	0.013
N1e	0.000	0.000	0.026	0.000
R2	0.000	0.000	0.000	0.006
R5	0.000	0.000	0.013	0.000
R6	0.000	0.000	0.039	0.000
R30	0.000	0.000	0.013	0.000
T	0.000	0.000	0.013	0.000
U2a	0.000	0.000	0.026	0.000
U2b	0.044	0.015	0.052	0.006
U2b2	0.022	0.030	0.026	0.000
U4	0.000	0.000	0.013	0.000
U7	0.156	0.152	0.013	0.006
Z	0.000	0.030	0.013	0.000
Z3	0.000	0.091	0.039	0.013

Table 3: Diversity indices and tests of selective neutrality based on HVRI (np 16024-16395) of mtDNA.

Population	n	k	UH	PS	HD ± SE	ND ± SE	MPD ± SE	Tajima's D	Tajima's D (P)	Fu's Fs	Fu's Fs (P)
Tamang	45	26	19	50	0.9576 ± 0.0146	0.0178 ± 0.0095	6.6081 ± 3.1796	-1.4782	0.0474	-9.9686	0.0016
Newar	66	37	23	56	0.9744 ± 0.0074	0.0161 ± 0.0086	6.0056 ± 2.8990	-1.6263	0.0271	-22.0162	0.0000
Kathmandu	77	66	58	79	0.9949 ± 0.0034	0.0179 ± 0.0095	6.6904 ± 3.1906	-1.9535	0.0054	-25.06791	0.0000
Nepal	188	110	74	94	0.9936 ± 0.0015	0.0175 ± 0.0092	6.5142 ± 3.0944	-1.8350	0.0052	-24.78318	0.0001
Tibet	156	96	65	88	0.9878 ± 0.0033	0.0186 ± 0.0097	6.9583 ± 3.2885	-1.7248	0.0137	-24.77519	0.0002
All Populations	344	195	130	125	0.9956 ± 0.0008	0.0181 ± 0.0095	6.8060 ± 3.2136	-1.9385	0.0031	-24.43928	0.0003

n = Sample Size
k = Number of different Haplotypes
UH = Number of Unique Haplotypes
PS = Number of polymorphic sites
HD = Haplotype Diversity
ND = Nucleotide Diversity
MPD = Mean number of pairwise differences
Tajima's D (P) = Probability of D statistics
Fu's Fs (P) = Probability of Fs statistics

Table 4: Diversity indices and tests of selective neutrality based on HVRII (np 50-400) of mtDNA.

Population	n	k	UH	PS	HD ± SE	ND ± SE	MPD ± SE	Tajima's D	Tajima's D (P)	Fu's Fs	Fu's Fs (P)
Tamang	45	21	12	21	0.9374 ± 0.0184	0.0096 ± 0.0057	3.4061 ± 1.7756	-1.2798	0.0872	-10.8707	0.0004
Newar	66	42	29	23	0.9772 ± 0.0083	0.0108 ± 0.0061	3.8378 ± 1.9543	-1.1786	0.1116	-25.9936	0.0000
Kathmandu	77	55	41	36	0.9884 ± 0.0045	0.0111 ± 0.0062	3.9238 ± 1.9883	-1.7231	0.0165	-25.9535	0.0000
Nepal	188	91	51	46	0.9833 ± 0.0031	0.0108 ± 0.0060	3.8087 ± 1.9258	-1.7884	0.0087	-25.7464	0.0000
Tibet	156	70	41	43	0.9700 ± 0.0062	0.0100 ± 0.0057	3.5613 ± 1.8199	-1.9829	0.0022	-25.9826	0.0000
All Populations	344	140	81	64	0.9800 ± 0.0030	0.0105 ± 0.0059	3.7223 ± 1.8845	-2.0703	0.0003	-25.4838	0.0001

n = Sample Size
k = Number of different Haplotypes
UH = Number of Unique Haplotypes
PS = Number of polymorphic sites
HD = Haplotype Diversity
ND = Nucleotide Diversity
MPD = Mean number of pairwise differences
Tajima's D (P) = Probability of D statistics
Fu's Fs (P) = Probability of Fs statistics

Table 5. Time estimates of major haplogroups in the Himalayan populations using two different mutation rates.

Populations	(Sub)haplogroup	*n*	Forster Rate (Forster et al. 1996)[29]		Soares Rate (Soares et al. 2009)[30]	
			ρ ± σ	T (ky)	ρ ± σ	T (ky)
Tibet	A4	9	0.89 ± 0.38	17.94 ± 7.77	0.89 ± 0.38	16.75 ± 7.25
	A11	6	0.60 ± 0.28	12.11 ± 5.71	0.60 ± 0.28	11.31 ± 5.33
	D4	13	1.67 ± 0.69	33.63 ± 14.00	1.67 ± 0.69	31.41 ± 13.08
	F1	29	1.45 ± 0.33	29.35 ± 6.68	1.45 ± 0.33	27.41 ± 6.23
	M9a1	43	0.67 ± 0.22	13.45 ± 4.41	0.67 ± 0.22	12.56 ± 4.12
	M9a1a1c1b	34	0.62 ± 0.22	12.46 ± 4.40	0.62 ± 0.22	11.64 ± 4.11
Newar	M5	8	0.62 ± 0.41	12.61 ± 8.37	0.62 ± 0.41	11.78 ± 7.81
	U7	10	0.60 ± 0.35	12.11 ± 7.00	0.60 ± 0.35	11.31 ± 6.53
Kathmandu	F1	6	1.40 ± 0.49	28.25 ± 9.89	1.40 ± 0.49	26.38 ± 9.23
	U2	8	2.14 ± 0.62	43.24 ± 12.57	2.14 ± 0.62	40.38 ± 11.73
Tamang	M9a1	7	0.86 ± 0.45	17.30 ± 9.12	0.86 ± 0.45	16.15 ± 8.51
Nepal	A4	10	1.57 ± 0.61	31.71 ± 12.23	1.57 ± 0.61	29.61 ± 11.42
	D4	16	1.53 ± 0.55	30.94 ± 11.17	1.53 ± 0.55	28.90 ± 10.44
	F1	16	1.70 ± 0.59	34.31 ± 11.94	1.70 ± 0.59	32.04 ± 11.49
	U2	14	2.18 ± 0.53	44.03 ± 10.70	2.18 ± 0.53	41.11 ± 9.99
Himalayas	A4	19	1.19 ± 0.37	23.96 ± 7.46	1.19 ± 0.37	22.38 ± 6.97
	D4	29	1.65 ± 0.50	33.30 ± 10.09	1.65 ± 0.50	31.09 ± 9.42
	F1	45	1.54 ± 0.33	31.05 ± 6.68	1.54 ± 0.33	28.99 ± 6.23
	M9a1*	52	0.78 ± 0.22	15.75 ± 4.40	0.78 ± 0.22	14.71 ± 4.11
	M9a1a1c1b*	39	0.56 ± 0.20	11.38 ± 4.14	0.56 ± 0.20	10.63 ± 3.87

*Absent in Newar

Table 6. Analysis of Molecular Variance (AMOVA)

Classification	Among Groups			Among Populations Within Groups			Within Population		
	Variance %	Φ_{CT}	P value	Variance %	Φ_{SC}	P value	Variance %	Φ_{ST}	P value
Geography (6 groups)[a]	3.75	0.037	p< 0.0001	3.11	0.032	p< 0.0001	93.14	0.067	p< 0.0001
Language (9 groups)[b]	3.43	0.034	p< 0.0001	3.17	0.033	p< 0.0001	93.41	0.066	p< 0.0001

[a] Geographical affiliation (6 groups): Northeast Asia, Southeast Asia, Southwest Asia, Central Asia, South Central Asia and Himalayas (see Table 1).
[b] Linguistic affiliation (9 groups): Japanese, Korean, Chinese, Malayo-polynesian, Mon-Khmer, Kam-Tai, Mongolic, Indo-Iranian and Tibeto-Burman (see Table 1).

Supplementary Table 1. MtDNA control and coding regions information of the four Himalayan populations examined.

Samples	Population	Haplogroup	HVRI (1600 to 16395)	HVRII (1 to 400)
Tam1	Tamang	M43	16223 16311 16362	73 152 263 309.1C 315.1C
Tam2	Tamang	U2b	16051	73 146 152 263 315.1C 373
Tam3	Tamang	D5	16095G 16182C 16183C 16189 16223 16356 16362	73 150 152 263 315.1C
Tam4	Tamang	U2b2	16051 16209 16239 16244 16352 16353	73 146 152 234 263 315.1C
Tam5	Tamang	A4	16223 16290 16295 16319 16362	73 114 146 152 235 263 315.1C
Tam6	Tamang	M4"67	16145 16188 16189 16192 16223 16293 16304	73 263 309.1C 315.1C
Tam7	Tamang	A11	16093 16223 16290 16293C 16319	73 152 235 263 309.1C 315.1C
Tam8	Tamang	M33b1	16077 16223 16259 16324 16362	73 263 309.1C 315.1C
Tam9	Tamang	M9a1b1	16158 16223 16234 16362	73 150 152 153 263 315.1C
Tam10	Tamang	G2	16051 16193 16223 16278 16362	73 260 263 315.1C
Tam11	Tamang	D5a3	16111 16181C 16182C 16183C 16189 16223 16360 16362	73 150 263 309.2C 315.1C
Tam12	Tamang	F1c1a	16111 16129 16266 16304	73 143 152 249d 263 309.1C 315.1C
Tam13	Tamang	U7	16298 16309 16318T	73 151 152 263 315.1C
Tam14	Tamang	D4e1a	16223 16311 16362	73 94 214 263 315.1C
Tam15	Tamang	D4	16223 16243 16362	73 146 263 309.2C 315.1C
Tam16	Tamang	U7	16298 16309 16318T	73 151 152 263 315.1C
Tam17	Tamang	D5a3	16093 16111 16181C 16182C 16183C 16189 16223 16360 16362	73 150 263 309.2C 315.1C
Tam18	Tamang	M9a1a1c1b	16223 16234 16316 16362	73 263 309.1C 315.1C
Tam19	Tamang	M31a2	16017 16093 16126 16145 16223 16325	73 195 263 315.1C
Tam20	Tamang	C4	16223 16298 16327	73 200 204 207 249d 263 309.1C 315.1C
Tam21	Tamang	G2	16051 16193 16223 16278 16362	73 260 263 315.1C
Tam22	Tamang	D4	16223 16243 16362	73 146 263 309.1C 315.1C
Tam23	Tamang	U2b	16051 16114	73 146 152 263 315.1C 373
Tam24	Tamang	M4"67	16145 16188 16189 16192 16223 16293 16304	73 263 309.1C 315.1C
Tam25	Tamang	M9a1a1c1b	16223 16234 16316 16362	74 263 309.1C 315.1C
Tam26	Tamang	M9a1a1c1b	16223 16234 16316 16362	75 263 309.1C 315.1C
Tam28	Tamang	D5a3	16111 16182C 16183C 16189 16218 16223 16311 16360 16362	73 150 263 309.2C 315.1C
Tam29	Tamang	D4e1a	16223 16311 16362	73 94 214 263 315.1C
Tam30	Tamang	U7	16298 16309 16318T	73 151 152 263 315.1C
Tam31	Tamang	G2	16051 16193 16223 16278 16362	73 260 263 315.1C

Tam32	Tamang	M9a1a2	16145 16223 16234 16316	73 153 263 309.1C 315.1C
Tam33	Tamang	D4	16223 16243 16362	73 146 263 309.1C 315.1C
Tam34	Tamang	G2	16051 16193 16223 16278 16362	73 260 263 315.1C
Tam35	Tamang	U7	16298 16309 16318T	73 151 152 263 315.1C
Tam36	Tamang	U7	16298 16309 16318T	73 151 152 263 315.1C
Tam37	Tamang	U7	16309 16318T	73 151 152 263 315.1C
Tam38	Tamang	F1d	16189 16284 16304	73 146 263 309.1C 315.1C
Tam39	Tamang	D5a2a1	16092 16164 16167 16182C 16183C 16188 16189 16223 16266 16362	73 150 195 263 309.2C 315.1C
Tam40	Tamang	D4	16223 16325 16362	73 263 309.1C 315.1C
Tam41	Tamang	M9a1a1c1b	16223 16234 16316 16362	73 263 309.2C 315.1C
Tam42	Tamang	U7	16298 16309 16318T	73 151 152 263 315.1C
Tam43	Tamang	G2	16051 16193 16223 16278 16362	73 260 263 315.1C
Tam44	Tamang	D4	16223 16243 16362	73 146 263 309.2C 315.1C
Tam45	Tamang	M9a1a2	16145 16223 16234 16316	73 153 263 309.1C 315.1C
Tam46	Tamang	F1b	16182C 16183C 16189 16232A 16249 16304	73 152 249d 263 309.2C 315.1C
New1	Newar	M5	16129 16223 16311	73 263 309.1C 315.1C
New2	Newar	A4	16223 16290 16311 16319 16362	73 152 234 235 263 309.2C 315.1C
New3	Newar	U7	16223 16289 16309 16318T	73 151 152 263 309.1C 315.1C
New4	Newar	D4b2b	16223 16362	73 194 263 309.1C 315.1C
New5	Newar	F1d	16129 16189 16263 16284 16304	73 146 249d 263 309.1C 315.1C
New6	Newar	U7	16223 16309 16318T	73 151 152 263 309.2C 315.1C
New7	Newar	U2b2	16051 16209 16239 16352 16353	73 146 152 195 234 263 309.1C 315.1C
New8	Newar	M33a1a	16223 16294	73 146 195 207 263 309.2C 315.1C
New9	Newar	D5a3	16111 16181C 16182C 16183C 16189 16223 16360 16362	73 150 263 309.2C 315.1C
New10	Newar	F1d	16189 16284 16304	73 146 249d 263 309.1C 315.1C
New11	Newar	Z3a	16150 16185 16223 16260 16298	73 152 204 207 249d 263 309.2C 315.1C
New12	Newar	D4	16223 16311 16362	73 263 309.2C 315.1C
New13	Newar	U7	16223 16309 16318T	73 151 152 263 315.1C
New14	Newar	F1	16183C 16189 16304	73 249d 263 309.2C 315.1C
New15	Newar	A4	16223 16274 16290 16295 16319 16362	73 152 235 263 309.2C 315.1C
New16	Newar	M33a1a	16223 16294	73 146 195 207 263 309.1C 315.1C
New17	Newar	M5	16129 16223 16249	73 146 263 315.1C
New18	Newar	M2a1	16223 16270 16319 16352	73 195 204 207 263 315.1C

New19	Newar	M35	16223	73 189 199 207 263 315.1C
New20	Newar	F1d	16189 16284 16304	73 146 249d 263 268 309.2C 315.1C
New21	Newar	Z3a	16185 16223 16260 16298	73 152 204 207 249d 263 315.1C
New22	Newar	M5	16048 16129 16223 16263	73 263 315.1C
New23	Newar	Z	16185 16223 16260 16298	73 152 249d 263 315.1C
New24	Newar	U2b2	16051 16209 16239 16352 16353	73 146 152 234 263 309.1C 315.1C
New25	Newar	M5	16048 16129 16223 16263	73 263 315.1C
New26	Newar	M5	16129 16223	73 263 315.1C
New27	Newar	Z3a	16185 16223 16260 16298	73 152 207 249d 263 309.1C 315.1C
New28	Newar	B5a1c	16129 16140 16182C 16183C 16189 16261 16266A	73 152 210 263 309.2C 315.1C
New30	Newar	U7	16093 16309 16318T	73 152 263 309.1C 315.1C
New31	Newar	G2	16051 16193 16223 16278 16362	73 260 263 309.1C 315.1C
New32	Newar	M5	16129 16223	73 263 315.1C
New33	Newar	M5	16048 16129 16223 16263	73 263 315.1C
New34	Newar	HV	16247	263 309.2C 315.1C
New35	Newar	HV	16247	263 309.1C 315.1C
New36	Newar	M33a1a	16223 16294	73 146 195 207 263 309.2C 315.1C
New37	Newar	Z3a	16150 16185 16223 16260 16298	73 152 207 249d 263 309.1C 315.1C
New38	Newar	G2	16051 16193 16223 16278 16362	73 260 263 315.1C
New39	Newar	U7	16223 16309 16318T	73 151 152 263 309.1C 315.1C
New40	Newar	F2b	16092A 16183C 16189 16288 16291 16304	73 152 249d 263 309.2C 315.1C
New41	Newar	F1d	16183C 16189 16284 16304	73 146 249d 263 309.2C 315.1C
New42	Newar	A4	16223 16274 16290 16295 16319 16362	73 152 235 263 309.2C 315.1C
New43	Newar	A4	16223 16274 16290 16295 16319 16362	73 152 235 263 309.2C 315.1C
New44	Newar	M30c1	16166d 16223	73 146 152 195A 263 315.1C
New45	Newar	M2a3a	16140 16223 16227 16265C 16274 16319	73 146 263 309.2C 315.1C
New46	Newar	Z	16185 16223 16260 16298	73 152 249d 263 315.1C
New47	Newar	F1d	16189 16284 16304	73 146 249d 263 309.2C 315.1C
New48	Newar	Z3a	16150 16185 16223 16260 16298	73 151 152 195 207 249d 263 309.2C 315.1C
New49	Newar	A4	16223 16274 16290 16295 16319 16362	73 152 235 263 309.2C 315.1C
New50	Newar	U7	16223 16289 16309 16318T	73 151 152 263 309.1C 315.1C
New51	Newar	Z3a	16185 16223 16260 16298	73 152 207 249d 263 309.1C 315.1C
New52	Newar	M4	16145 16176 16223 16261 16311	73 263 309.1C 315.1C

New53	Newar	U7	16093 16309 16318T	73 152 263 309.1C 315.1C
New54	Newar	F1c1a	16111 16129 16304 16318	73 152 234 249d 263 315.1C
New55	Newar	D5a2	16172 16182C 16183C 16189 16223 16294 16362	73 150 263 315.1C
New56	Newar	G2	16048 16193 16223 16278 16362	73 93 260 263 315.1C
New57	Newar	M30	16166d 16223	73 146 152 195A 263 315.1C
New58	Newar	M34a	16086 16223 16249 16359	73 263 315.1C
New59	Newar	U2b	16051 16179 16234 16256	73 93 95C 146 198 263 309.1C 315.1C
New60	Newar	M3	16126 16129 16223	73 200 204 263 309.1C 315.1C
New61	Newar	M33a1a	16223 16294	73 195 207 263 309.1C 315.1C
New62	Newar	U7	16223 16309 16318T	73 151 152 263 309.2C 315.1C
New63	Newar	M5	16129 16223	73 263 315.1C
New64	Newar	U7	16223 16309 16318T	73 151 152 263 309.1C 315.1C
New65	Newar	U7	16223 16309 16318T	73 151 152 263 309.1C 315.1C
New66	Newar	D5a2	16172 16182C 16183C 16189 16223 16294 16362	73 150 263 315.1C
New67	Newar	M30	16177 16223	73 195A 263 309.1C 315.1C
Kat1	Kathmandu	C4	16223 16298 16311 16327 16357	73 152 249d 263 310
Kat2	Kathmandu	U2a	16051 16206C	73 194 263 315.1C
Kat3	Kathmandu	U2b	16051 16168	73 146 200 263 309.1C 315.1C
Kat4	Kathmandu	T	16126 16172 16294 16296 16325	73 152 195 263 309.2C 315.1C
Kat5	Kathmandu	A4	16092 16223 16290 16319 16362	73 152 235 263 309.1C 315.1C
Kat6	Kathmandu	M35	16223 16304	73 199 263 309.1C 315.1C
Kat7	Kathmandu	U2b	16051 16126 16179 16234 16247	73 146 152 263 309.2C 315.1C
Kat8	Kathmandu	C4	16223 16298 16311 16327 16357	73 152 249d 263 310
Kat9	Kathmandu	M30	16179d 16223	73 150 195A 263 315.1C
Kat10	Kathmandu	M5	16048 16129 16223 16227	73 263 309.1C 315.1C
Kat11	Kathmandu	M3	16102 16126 16223 16301 16344	73 263 297 315.1C
Kat12	Kathmandu	HV	16356 16360	146 263 309.2C 315.1C
Kat13	Kathmandu	A4	16223 16290 16319 16362	73 152 200 235 263 315.1C
Kat14	Kathmandu	N1e'1	16129 16223 16309	73 143 199 204 250 263 309.1C 315.1C
Kat15	Kathmandu	M3	16102 16126 16223 16301 16344	73 263 297 315.1C
Kat16	Kathmandu	J	16069 16126	73 263 295 315.1C
Kat17	Kathmandu	M34	16189 16223 16243	73 243 263 309.2C 315.1C
Kat18	Kathmandu	M35	16223 16304	73 199 263 309.1C 315.1C

Kat19	Kathmandu	D4j1a	16086 16223 16362	73 263 315.1C
Kat20	Kathmandu	Z3a	16185 16223 16260 16298	73 152 207 249d 263 309.1C 315.1C
Kat21	Kathmandu	M35	16223 16304	73 199 263 309.2C 315.1C
Kat22	Kathmandu	R6	16093 16179 16227 16245 16266 16278 16293 16362	73 195 246 263 315.1C
Kat23	Kathmandu	M30	16111 16172 16223	73 195A 263 309.1C 315.1C
Kat24	Kathmandu	U2a	16051 16206C 16311	73 194 263 315.1C
Kat25	Kathmandu	H	16271	263 309.1C 315.1C
Kat26	Kathmandu	D4i	16093 16114 16223 16294 16318 16362	73 195 263 315.1C
Kat27	Kathmandu	M35	16223 16304	73 199 263 309.2C 315.1C
Kat28	Kathmandu	M9a1b1	16158 16223 16234 16311 16319 16348 16362	73 150 152 153 263 315.1C
Kat29	Kathmandu	M33a1a	16223 16294	73 195 207 263 309.2C 315.1C
Kat30	Kathmandu	M9a1a1c1b	16093 16223 16234 16316 16362	73 217 263 309.1C 315.1C
Kat31	Kathmandu	U2b	16051 16126 16179 16234 16247	73 146 152 263 309.1C 315.1C
Kat32	Kathmandu	D4i	16114 16223 16294 16318 16362	73 195 263 315.1C
Kat33	Kathmandu	U4	16356	73 195 249d 263 315.1C
Kat34	Kathmandu	M5	16048 16129 16223	73 263 309.1C 315.1C
Kat35	Kathmandu	U2b	16051	73 146 152 263 315.1C 373
Kat36	Kathmandu	F1c1a	16111 16129 16266 16304	73 152 249d 263 309.2C 315.1C
Kat37	Kathmandu	M10a1	16129 16193 16223 16311 16357	73 146 152 185 204 263 309.1C 315.1C
Kat38	Kathmandu	D4g2a	16213 16223 16274 16362	73 263 298 309.1C 315.1C
Kat39	Kathmandu	Z3a	16150 16185 16223 16260 16298	73 152 195 207 249d 263 309.1C 315.1C
Kat40	Kathmandu	M33b1	16077 16223 16259 16324 16362	73 263 309.1C 315.1C
Kat41	Kathmandu	Z	16185 16223 16260 16298	73 152 249d 263 315.1C
Kat42	Kathmandu	R5	16093 16304 16309 16325	73 152 263 309.1C 315.1C
Kat43	Kathmandu	M38	16181 16223 16311 16319	73 246 263 315.1C 316
Kat44	Kathmandu	M33	16051 16223 16249 16250	73 194 263 315.1C
Kat45	Kathmandu	F1d	16093 16189 16284 16304	73 146 249d 263 309.1C 315.1C
Kat46	Kathmandu	R30b1	16129 16183C 16189 16193.1C 16298 16299	73 152 263 299d 309.1C 315.1C 373
Kat47	Kathmandu	U2b2	16051 16209 16239 16352 16353	73 146 152 234 256 263 315.1C
Kat48	Kathmandu	M33	16129 16223 16240 16362	73 195 263 309.1C 315.1C
Kat49	Kathmandu	M3	16102 16126 16223 16301 16344	73 263 297 315.1C
Kat50	Kathmandu	R6	16093 16179 16227 16245 16266 16278 16362	73 195 246 263 315.1C
Kat51	Kathmandu	M10a2	16066 16223 16293 16311	73 263 315.1C

Kat52	Kathmandu	C4	16223 16298 16327 16357	73 249d 263 309.2C 315.1C
Kat53	Kathmandu	U7	16309 16318C	73 151 152 263 315.1C
Kat54	Kathmandu	M5	16129 16223	73 263 309.1C 315.1C
Kat55	Kathmandu	M3	16172 16223 16259 16324 16362	73 263 309.2C 315.1C
Kat56	Kathmandu	B5a1c	16129 16140 16182C 16183C 16189 16261 16266A	73 152 210 263 309.1C 315.1C
Kat57	Kathmandu	M43	16092 16223	73 263 315.1C
Kat58	Kathmandu	F1d	16189 16284 16304	73 146 249d 263 309.1C 315.1C
Kat59	Kathmandu	F1	16183C 16189 16193.1C 16304	73 249d 263 309.1C 315.1C
Kat60	Kathmandu	U2b2	16051 16209 16239 16244 16352 16353	73 146 152 234 263 315.1C
Kat61	Kathmandu	M33	16223 16271	146 152 263 309.1C 315.1C
Kat62	Kathmandu	D5a2	16172 16182C 16183C 16189 16223 16311 16362	73 150 263 309.2C 315.1C
Kat63	Kathmandu	M30	16179d 16223T	73 150 195A 263 315.1C
Kat64	Kathmandu	D4i	16093 16114 16223 16294 16318 16362	73 195 263 309.1C 315.1C
Kat65	Kathmandu	D4	16223 16243 16362	73 146 263 309.2C 315.1C
Kat66	Kathmandu	N1e'1	16223 16309	73 143 199 204 250 263 309.1C 315.1C
Kat67	Kathmandu	M43	16092 16223	73 263 315.1C
Kat68	Kathmandu	D5a2a1	16092 16164 16167 16182C 16183C 16189 16223 16266 16362	73 150 194 195 263 310 314d 315d
Kat69	Kathmandu	F1d	16183C 16189 16304 16311	73 143 146 154 249d 263 315.1C
Kat70	Kathmandu	A4	16125 16223 16290 16311 16319 16362	73 152 235 263 309.2C 315.1C
Kat71	Kathmandu	D4i	16114 16223 16294 16311 16318 16362	73 195 263 315.1C
Kat72	Kathmandu	M30e	16223 16234	73 152 195G 263 309.2C 315.1C
Kat73	Kathmandu	F1b	16182C 16183C 16189 16232A 16249 16304	73 152 249d 263 309.2C 315.1C
Kat74	Kathmandu	Z3a	16150 16185 16223 16260 16298	73 152 207 249d 263 315.1C
Kat75	Kathmandu	M13a	16145 16148 16188 16189 16223 16311	73 152 263 315.1C
Kat76	Kathmandu	R6	16093 16179 16227 16245 16266 16278 16362	73 195 246 263 309.1C 315.1C
Kat77	Kathmandu	A4	16223 16290 16311 16319 16362	72 73 152 235 263 315.1C
TIB1	Tibetan	G3	16186 16223 16274 16362	73 195 309.2C 315.1C
TIB2	Tibetan	M9a1a1c1b	16194 16223 16234 16316 16362	73 263 309.1C 315.1C
TIB3	Tibetan	M9a1b1	16158 16223 16234 16362	73 150 152 153 263 309.1C 315.1C
TIB4	Tibetan	M13a1b	16145 16148 16188 16189 16223 16381	73 152 263 309.1C 315.1C
TIB5	Tibetan	M9a1a1c1b	16194 16223 16234 16316 16362	73 263 309.1C 315.1C
TIB6	Tibetan	F1c1a	16111 16129 16266 16304	73 152 249d 263 315.1C
TIB7	Tibetan	M9a1a1c1b	16209 16223 16234 16316 16362	73 263 309.1C 315.1C

TIB8	Tibetan	A4	16223 16248 16290 16319 16362	73 150 152 182 235 263 309.1C 315.1C
TIB9	Tibetan	G2a	16223 16227 16278 16311 16362	73 165 263 309.1C 315.1C
TIB10	Tibetan	F1	16147 16168 16183C 16189 16193.1C 16304	73 249d 263 309.1C 315.1C
TIB11	Tibetan	M9a1b1	16158 16223 16234 16325 16362	73 150 152 153 155 263 315.1C
TIB12	Tibetan	U7	16309 16318T	73 152 153 263 309.1C 315.1C
TIB13	Tibetan	M9a1b1	16158 16223 16234 16362	73 150 152 153 183 263 315.1C
TIB14	Tibetan	A4	16092 16223 16290 16319 16362	73 152 235 263 309.2C 315.1CC
TIB15	Tibetan	M13a1b	16145 16148 16188 16189 16223 16381	73 152 263 309.1C 315.1C
TIB16	Tibetan	F1	16183C 16189 16304	73 249d 263 309.1C 315.1C
TIB17	Tibetan	F1	16183C 16189 16304	73 249d 263 309.2C 315.1C
TIB18	Tibetan	HV	rCRS	150 263 309.2C 315.1C
TIB19	Tibetan	G2a	16223 16227 16278 16311 16362	73 263 309.1C 315.1C
TIB20	Tibetan	M9a1b1	16158 16223 16234 16362	73 150 152 153 263 315.1C
TIB21	Tibetan	M9a1a1c1b	16223 16234 16316 16362	73 263 309.1C 315.1C
TIB22	Tibetan	M62b	16169 16223 16260	73 150 203 204 263 309.1C 315.1C
TIB23	Tibetan	F1	16147 16183C 16189 16193.1C 16304	73 249d 263 309.2C 315.1C
TIB24	Tibetan	D4g2a	16172 16223 16232A 16274 16362	73 263 298 315.1C
TIB25	Tibetan	G2a	16129 16172 16189 16223 16227 16278 16362	73 195 263 309.2C 315.1C
TIB26	Tibetan	M62b	16223 16260 16274 16295 16318 16320	73 150 203 204 263 309.1C 315.1C
TIB27	Tibetan	M9a1a1c1b	16223 16234 16316 16362	73 263 309.2C 315.1C
TIB28	Tibetan	A4	16223 16248 16290 16319 16362	73 150 152 182 235 263 309.1C 315.1C
TIB29	Tibetan	M9a1a1c1b	16114A 16223 16234 16316 16362	73 204 263 309.1C 315.1C
TIB30	Tibetan	M62b	16223 16260 16274 16295 16318 16320	73 150 203 204 263 309.1C 315.1C
TIB31	Tibetan	A4	16223 16290 16319 16362	73 146 152 235 263 315.1C
TIB32	Tibetan	M9a1b1	16158 16223 16234 16362	73 150 152 153 263 315.1C
TIB33	Tibetan	C4a1	16093 16129 16223 16224 16298 16327	73 150 249d 263 315.1C
TIB34	Tibetan	C4	16298 16327	73 143 152 249d 263 315.1C
TIB35	Tibetan	C4	16183C 16189 16223 16298 16311 16357	73 188 189 249d 263 310 315d
TIB36	Tibetan	F1	16182C 16183C 16189 16304	73 249d 263 309.2C 315.1C
TIB37	Tibetan	F1c1a	16111 16129 16189 16266 16304 16362	73 152 249d 263 309.1C 315.1CC
TIB38	Tibetan	C4	16183C 16189 16223 16298 16311 16357	73 188 189 249d 263 310 315d
TIB39	Tibetan	M9a1a1c1b	16129 16223 16234 16316 16362	73 146 263 297 309.1C 315.1C
TIB40	Tibetan	B4	16092 16182C 16183C 16189 16217 16261	73 152 200 204 263 309.2C 315.1C

TIB41	Tibetan	F1d	16048 16182C 16183C 16189 16210 16304 16309	73 249d 263 309.2C 315.1C
TIB42	Tibetan	G3a1	16148 16153 16215 16223 16274	73 143 150 263 309.2C 315.1C
TIB43	Tibetan	G3a1	16215 16223 16274	73 150 182 263 315.1C
TIB44	Tibetan	M9a1a1c1b	16223 16234 16316 16362	73 263 315.1C
TIB45	Tibetan	M9a1a1c1b	16223 16234 16316 16362	73 263 309.1C 315.1C
TIB46	Tibetan	C4a1	16093 16129 16223 16298 16311 16327	64 73 249d 263 315.1C
TIB47	Tibetan	F1d	16145 16189 16284 16304	73 146 249d 263 309.2C 315.1C
TIB48	Tibetan	C5	16086 16174 16223 16288 16298 16327	73 249d 263 309.1C 315.1C
TIB49	Tibetan	M9a1b1	16158 16223 16234 16362	73 150 152 153 263 315.1C
TIB50	Tibetan	M9a1b1	16158 16223 16234 16362	73 150 152 153 263 315.1C
TIB51	Tibetan	M9a1a1c1b	16223 16234 16316 16362	73 263 315.1C
TIB52	Tibetan	M9a1b1	16158 16166d 16223 16234 16362	73 150 152 153 263 309.1C 315.1C
TIB53	Tibetan	G2a	16145 16223 16227 16278 16362	73 263 309.1C 315.1C
TIB54	Tibetan	A11	16093 16223 16290 16293C 16302 16319	73 152 235 263 309.1C 315.1C
TIB55	Tibetan	D4	16223 16256 16311 16362	73 200 263 309.1C 315.1C
TIB56	Tibetan	M9a1a1c1b	16129 16223 16234 16316 16362	73 263 309.2C 315.1C
TIB57	Tibetan	D4	16223 16256 16311 16362	73 200 263 309.1C 315.1C
TIB58	Tibetan	C5	16086 16172 16174 16223 16288 16298 16327	73 249d 263 309.2C 315.1C
TIB59	Tibetan	M9a1a1c1b	16223 16234 16316 16362	73 263 315.1C
TIB60	Tibetan	A4	16125 16223 16290 16311 16319 16362	73 152 235 263 309.2C 315.1C
TIB61	Tibetan	F1	16168 16183C 16189 16193.1C 16304	73 249d 263 309.2C 315.1C
TIB62	Tibetan	C5	16086 16172 16174 16223 16288 16298 16327	73 249d 263 309.2C 315.1C
TIB63	Tibetan	M9a1a1c1b	16092 16223 16234 16316 16320 16362	73 257 263 309.1C 315.1C
TIB64	Tibetan	D4q	16223 16256 16311 16362	73 200 263 309.1C 315.1C
TIB65	Tibetan	M9a1a1c1b	16223 16234 16362	73 150 152 153 194 263 309.1C 315.1C
TIB66	Tibetan	M9a1a1c1b	16092 16223 16234 16316 16320 16362	73 257 263 309.1C 315.1C
TIB67	Tibetan	F1d	16048 16182C 16183C 16189 16210 16304 16309	73 249d 263 309.2C 315.1C
TIB68	Tibetan	B4	16182C 16183C 16189 16217 16299	73 189 193 263 309.1C 315.1C
TIB69	Tibetan	M62b	16169 16183C 16189 16193.1C 16223 16260 16295	73 143 150 203 204 263 309.1C 315.1C
TIB70	Tibetan	F1	16168 16183C 16189 16193.1C 16304	73 249d 263 309.2C 315.1C
TIB71	Tibetan	A11	16093 16223 16290 16293C 16319	73 152 235 263 315.1C
TIB72	Tibetan	M61	16223 16270 16362 16381	73 152 263 309.1C 315.1C
TIB73	Tibetan	M13a1b	16145 16148 16188 16189 16223 16381	73 152 263 315.1C

TIB74	Tibetan	M9a1a1c1b	16111 16223 16234 16316 16362	73 263 315.1C
TIB75	Tibetan	M13a2	16145 16168 16223 16257 16311	73 152 263 315.1C
TIB76	Tibetan	F1d	16145 16189 16193.2C 16255 16284 16304	73 146 234 249d 263 309.2C 315.1C
TIB77	Tibetan	G2	16223 16362	73 263 309.2C 315.1C
TIB78	Tibetan	A4	16223 16290 16319 16362	73 146 152 235 263 315.1C
TIB79	Tibetan	M9a1a1c1b	16223 16234 16316 16362	73 263 309.2C 315.1C
TIB80	Tibetan	D4j3	16184 16223 16311 16356 16362	73 263 309.1C 315.1C 338
TIB81	Tibetan	M9a1a1c1b	16114G 16223 16234 16316 16362	73 263 309.1C 315.1C
TIB82	Tibetan	F1	16183C 16189 16193.1C 16304	73 150 249d 263 315.1C
TIB83	Tibetan	D4g2a	16223 16274 16362	73 196 263 298 309.1C 315.1C
TIB84	Tibetan	F1	16182C 16183C 16189 16304	73 249d 263 309.1C 315.1C
TIB85	Tibetan	M9a1a1c1b	16223 16234 16258C 16262+C 16316 16362	73 263 309.1C 315.1C
TIB86	Tibetan	M9a1a1c1b	16194 16223 16234 16316 16362	73 263 309.2C 315.1C
TIB87	Tibetan	C4a1	16093 16298 16311 16327	73 152 249d 263 315.1C
TIB88	Tibetan	U2b	16051 16168 16172 16311	73 146 195 263 309.1C 315.1C
TIB89	Tibetan	M70	16214 16223 16297 16342 16381	73 236 263 315.1C
TIB90	Tibetan	F1a1	16129 16162 16172 16220 16304	73 152 249d 263 315.1C
TIB91	Tibetan	M8a	16223 16298 16319	73 263 309.1C 315.1C
TIB92	Tibetan	D4	16223 16362	73 263 309.1C 315.1C
TIB93	Tibetan	M9a1a1c1b	16223 16234 16316 16362	73 263 315.1C
TIB94	Tibetan	M9a1b1	16158 16223 16234 16362	73 150 152 153 183 263 315.1C
TIB95	Tibetan	M9a1a1c1b	16223 16234 16316 16362	73 263 309.1C 315.1C
TIB96	Tibetan	F1d	16183C 16189 16193.1C 16304	73 146 249d 263 309.1C 315.1C
TIB97	Tibetan	M9a1a1c1b	16093 16223 16234 16316 16362	73 263 315.1C
TIB98	Tibetan	M11a2	16173 16223	73 146 198 200 215 263 309.1C 315.1C 318 326
TIB99	Tibetan	C4	16298 16327	73 143 152 249d 263 315.1C
TIB100	Tibetan	D4i	16114 16223 16294 16362	73 195 263 315.1C
TIB101	Tibetan	F1d	16145 16189 16284 16304	73 146 249d 263 309.1C 315.1C
TIB102	Tibetan	HV	rCRS	150 263 309.2C 315.1C
TIB103	Tibetan	M13a2	16145 16168 16188 16223 16257 16311	73 152 263 315.1C 385
TIB104	Tibetan	M62a	16147 16189 16193.2C 16223 16295	73 146 150 263 310 315d
TIB105	Tibetan	A11	16223 16290 16293C 16319	73 152 235 263 309.1C 315.1C
TIB106	Tibetan	Z	16185 16223 16260 16278	73 152 249d 263 309.2C 315.1C

TIB107	Tibetan	M70	16214 16223 16297 16342 16381	73 236 263 315.1C
TIB108	Tibetan	C4a1	16093 16298 16327 16053G 16183C 16189 16193.1C 16223 16298 16311 16320 16327	73 152 195 249d 263 315.1C
TIB109	Tibetan	C4	16357	73 194 249d 263 310 315d
TIB110	Tibetan	M13a1b	16145 16148 16188 16189 16223 16381	73 151 152 263 315.1C
TIB111	Tibetan	M9a1a1c1b	16223 16234 16316 16362	73 263 309.1C 315.1C
TIB112	Tibetan	D5a3	16111 16182C 16183C 16189 16223 16360 16362	73 150 195 263 309.2C 315.1C
TIB113	Tibetan	D4j3	16184 16223 16311 16362	73 263 309.1C 315.1C 338
TIB114	Tibetan	D4	16223 16362	73 194 263 315.1C
TIB115	Tibetan	A4	16092 16223 16290 16319 16362	73 152 235 263 315.1C
TIB116	Tibetan	M9a1a1c1b	16194 16223 16234 16316 16362	73 263 309.1C 315.1C
TIB117	Tibetan	D4j3	16184 16223 16311 16362	73 263 309.2C 315.1C 338
TIB118	Tibetan	M9a1a1c1b	16223 16234 16316 16362	73 263 309.2C 315.1C
TIB119	Tibetan	M4"67	16092 16223 16289	73 263 315.1C
TIB120	Tibetan	D4j3	16184 16223 16311 16362	73 263 309.1C 315.1C 338
TIB121	Tibetan	F1d	16145 16154 16183C 16189 16193.1C 16284 16304	73 146 207 249d 263 315.1C
TIB122	Tibetan	Z	16185 16223 16260 16298 16319	73 152 249d 263 309.1C 315.1C
TIB123	Tibetan	F1	16183C 16189 16193.1C 16304	73 195 249d 263 309.2C 315.1C
TIB124	Tibetan	D4j1a	16086 16223 16311	73 263 315.1C
TIB125	Tibetan	G3	16223 16274 16286 16362	73 195 198 263 315.1C
TIB126	Tibetan	M9a1a1c1b	16223 16316 16362	73 263 309.1C 315.1C
TIB127	Tibetan	F1d	16183C 16189 16193.1C 16304	73 146 249d 263 309.1C 315.1C
TIB128	Tibetan	A11	16093 16223 16290 16293C 16319	73 152 235 263 315.1C
TIB129	Tibetan	G2	16223 16278 16362	73 263 309.2C
TIB130	Tibetan	A4	16223 16290 16319 16362	73 152 235 309.1C 315.1C
TIB131	Tibetan	F1c1a	16111 16129 16266 16295 16304	73 152 249d 263 309.1C 315.1C
TIB132	Tibetan	C4a1	16093 16129 16223 16298 16327	73 249d 263 309.1C 315.1C
TIB133	Tibetan	A4	16223 16230 16234 16290 16319 16362	73 152 235 263 309.1C 315.1C
TIB134	Tibetan	G2	16223 16294 16362	73 183 263 309.2C 315.1C
TIB135	Tibetan	M8a	16223 16298 16311 16319	73 263 309.1C 315.1C
TIB136	Tibetan	F1b	16182C 16183C 16189 16232A 16249 16295 16304 16311	73 249d 263 315.1C
TIB137	Tibetan	F1b	16183C 16189 16193.1C 16232A 16249 16304 16311	73 146 249d 263 309.1C 315.1C
TIB138	Tibetan	G2	16129 16223 16278 16362	73 263 309.1C 315.1C
TIB139	Tibetan	M9a1a1c1b	16093 16223 16234 16316 16362	73 263 309.1C 315.1C

TIB140	Tibetan	M13a1b	16140 16145 16148 16188 16189 16223 16381	73 152 263 315.1C
TIB141	Tibetan	D5a2a1	16164 16172 16182C 16183C 16189 16223 16266	73 150 263 309.2C 315.1C
TIB142	Tibetan	F1b	16183C 16189 16193.1C 16232A 16249 16304 16311	73 146 249d 263 309.2C 315.1C
TIB143	Tibetan	M9a1a1c1b	16223 16234 16362	73 263 309.2C 315.1C
TIB144	Tibetan	M9a1a1c1b	16093 16223 16234 16316 16362	73 263 309.2C 315.1C
TIB145	Tibetan	F1	16182C 16183C 16189 16304	73 195 249d 263 309.2C 315.1C
TIB146	Tibetan	F1d	16093 16145 16183C 16189 16193.1C 16255 16304	73 146 152 249d 263 309.1C 315.1C
TIB147	Tibetan	R2	16071 16111 16172 16189 16311	73 150 152 263 315.1C
TIB148	Tibetan	G3	16223 16274 16362	73 195 263 315.1C
TIB149	Tibetan	F1c1a	16111 16129 16266 16304	73 152 249d 263 315.1C
TIB150	Tibetan	M9a1a1c1b	16111 16223 16234 16316 16362	73 263 315.1C
TIB151	Tibetan	A11	16093 16223 16290 16293C 16319	73 152 235 263 315.1C
TIB152	Tibetan	M9a1a1c1b	16223 16234 16316 16362	73 263 309.2C 315.1C
TIB153	Tibetan	M9a1a1c1b	16223 16234 16278 16316 16362	73 263 315.1C
TIB154	Tibetan	F1	16183C 16189 16304	73 249d 263 309.2C 315.1C
TIB155	Tibetan	M9a1a1c1b	16093 16223 16234 16316 16362	73 263 309.1C 315.1C
TIB156	Tibetan	A11	16223 16234 16290 16293C 16319	73 152 235 263 309.2C 315.1C

Supplementary Table 2. MtDNA Haplogroups typing by RFLP, Primer-mediated RFLP and sequencing

Haplogroup	Nucleotide Position tested*	Typing Method
A	663	HaeIII (+)
B4'5	8281–8289d	size detection
C	13263	AluI (+)
C4	6026	Sau96I (-)
C5	595.1	BseRI (-)
D	5178A	AluI (-)
D4	3010	BccI (-)
D4j	11696	DdeI (+)**
D4e1a	14470	AccI (+)
D5	5301	BtscI (-)**
D5a	12026	HincII (+)
F1	12406	HincII (-)
F1a	4086	HphI (-)
F1b	12633	Sau96I (-)
F1c	10454	Sequencing
F1d	15402	MboII (+)
F2	13708	BstNI (-)
G	4833	HaeII (+)
G2	5601	Sequencing
HV	14766	MseI (+)
H	7028	AluI (+)
JT	15452A	HpyAv (-)
M	10400	AluI (+)
M2	447G	NIaIII (-)**
M2a	12810	BccI (+)
M3	482	BccI (+)
M4"67	12007	Bsp1286I (-)
M4	6620	Tsp45I (-)
M5	1888	Cac8I (-)
M8	4715	Hpy188III (-)
M8a	8684	MnlI (-)
M9	4491	BccI (+)
M9a1	1041	NIaIII (+)
M9a1a1c1b	7697	BfaI (-)
M9a1a2	7256	BtscI (-)
M10	8793	DdeI (-)**
M11	6531	MseI (+)
M13	6023	AluI (-)

M30	15431	HgaI (-)
M31	15440	HaeIII (+)
M33	2361	TspRI (-)
M34	3010	BccI (-)
M35	12561	AluI (-)
M38	15487	AvaII (+)
M43	11696	DdeI (-)**
M61	5582C	Sequencing
M62	15520	HaeIII (+)
M70	8867	Tsp45I (+)
N	10873	MnlI (+)
N1e'1	4529T	BtsI (+)
R	12705	MboII (-)
R2'JT	4216	NIaIII (+)
R5	8594	MboI (-)
R6	12285	BtsCI (+)
R30	8584	Sequencing
U	12308	HinfI (+)**
Z	6752	BfaI (-)
Z3	10208	HpyAv (+)

VI. CONCLUSION

In the present work, the genetic diversity of a general population of Tibet and three Nepalese collections, namely Tamang, Newar and Kathmandu, were examined using three different marker systems, including autosomal STR, Y-STR and mtDNA. The present study complements our previous Y-chromosome data from these four abovementioned populations (Gayden et al., 2007), thereby providing a comprehensive view of genetic variation in people of the Himalayas. The results from the autosomal STR data suggest a Northeast Asian ancestry for the Himalayan populations with subsequent genetic admixture in Kathmandu and Newar populations from South Asia. Given the geographical proximity of India and Nepal and their strong historical ties, a significant Indian contribution to Newar and Kathmandu gene pools is plausible. However, South Asian influences in Tibet and Tamang are negligible, most likely due to the natural barrier presented by the Himalayas. Interestingly, unlike the Y-chromosomal data of Tamang, which is mostly composed of the East Asian Y-haplogroup O3a5c-M134 (Gayden et al., 2007), its mtDNA pool consists of an appreciable level of Indian lineages, suggesting a male-mediated migration of the ancestors of the present day Tamang population from the Tibetan plateau to the southern slope of the Himalayas in Nepal, where they admixed with the local females. In addition, the Nepalese gene pool harbors ancient Indian mtDNA haplogroups such as M2, R5 and U2, whose spread may be associated with the initial peopling of South Asia (Metspalu et al., 2004). In contrast, Y-haplogroup R in Newar and Kathmandu may be a recent arrival in the valley, probably with the expansion of the Indo-European language. Finally, the Y-STR profiles generated

will be useful in establishing national databases for both the Nepalese and Tibetan populations, for individual identification, paternity testing and population genetic studies.

REFERENCES

Metspalu M, Kivisild T, Metspalu E, Parik J, Hudjashov G, Kaldma K, Serk P, Karmin M, Behar DM, Gilbert MT, Endicott P, Mastana S, Papiha SS, Skorecki K, Torroni A, Villems R. 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.

Gayden T, Cadenas AM, Regueiro M, Singh NB, Zhivotovsky LA, Underhill PA, Cavalli-Sforza LL, Herrera RJ (2007). The Himalayas as a directional barrier to gene flow. Am J Hum Genet. 80:884-894.

VITA

Tenzin Gayden
Miami, Florida 33172
tgayd001@fiu.edu

Education

- Ph.D. Candidate Biology, Florida International University, Miami, FL, 2012
- M.S. Biology, Florida International University, Miami, FL, 2006
- M.Sc. Biotechnology, University of Mysore, Mysore, India, 2003
- B.Sc.Ed. Regional Institute of Education, Mysore, India, 1998

Experience

- Visiting Researcher, Dr. Carlos Bustamante Lab, Stanford University (September 2011)
- Teaching Assistant, Florida International University, Miami, Fl, USA, 2006-2009
- Senior Science Teacher, Central School for Tibetans Manali, HP, India, 1998-2001

Awards

- Dalai Lama Trust Graduate Scholarship, 2011-2012
- Dissertation Year Fellowship, UGS, Florida International University, Miami, FL, 2011
- Dissertation Evidence Acquisition Fellowship, UGS, Florida International University, Miami, FL, 2010 (Summer and Fall semesters)
- Teaching Assistantship, Florida International University, Miami, FL, 2006-2009
- Tibetan Scholarship Program, Funded by the US Department of State through The Tibet Fund, NY, USA, 2004-2006
- Silver Medalist, M.Sc. Biotechnology, University of Mysore, India 2003

Projects

- Ph.D. Dissertation: Genetic diversity in the Himalayan populations of Nepal and Tibet, 2012, Florida International University, Miami, FL. Advisor: Dr. Rene J. Herrera
- M.S. Thesis: Y-chromosome polymorphisms in the Himalayas, 2006, Florida International University, Miami, FL. Advisor: Dr. Rene J. Herrera

Peer-Reviewed Publications (A total of 15 publications)

1. Gayden T, Perez A, Persad PJ, Bukhari A *et al*. (2012) Mitochondrial genome variation in the Himalayan populations of Nepal and Tibet. (Submitted to *Eur J Hum Genet*)

2. Lacau H, Gayden T, Chennakrishnaiah S, Bukhari A *et al*. (2012) Afghanistan from a Y-Chromosome perspective. *Eur J Hum Genet* (*Accepted*)

3. Gayden T, Bukhari A, Chennakrishnaiah S, Stojkovic O *et al*. (2011b) Y-chromosomal microsatellite diversity in three culturally defined regions of the historical Tibet. *Forensic Sci Int Genet* (*In Press*)

4. Morlote DM, Gayden T, Arvind P, Babu A *et al*. (2011) The Soliga, an isolated tribe from Southern India: genetic diversity and phylogenetic affinities. *J Hum Genet* 56:258-269
5. Lacau H, Bukhari A, Gayden T, La Salvia J, *et al*. (2011) Y-STR profiling in two Afghanistan populations. *Leg Med* 13:103-108

6. Gayden T, Chennakrishnaiah S, La Salvia J, Jimenez S *et al*. (2011a) Y-STR diversity in the Himalayas. *Int J Legal Med* 125:367-375

7. Mirabal S, Varlhenb T, Gayden T, Regueiro M *et al*. (2010) Human Y-chromosome short tandem repeats: A tale of acculturation and migrations as mechanisms for the diffusion of agriculture in the Balkan Peninsula. *Am J Phys Anthropol* 142:380-390

8. Gayden T, Mirabal S, Cadenas AM, Lacau H *et al*. (2009) Genetic insights into the origins of Tibeto-Burman populations in the Himalayas. *J Hum Genet* 54:216-223

9. Gayden T, Regueiro M, Martinez L, Cadenas AM *et al*. (2008). Human Y-chromosome haplotying by Allelic-Specific PCR. *Electrophoresis* 29:2419-2423

10. Gayden T, Cadenas AM, Regueiro M, Singh NB *et al*. (2007). The Himalayas as a directional barrier to gene flow. *Am J of Hum Genet* 80:884-994

11. Cadenas AM, Regueiro M, Gayden T, Singh N *et al*. (2007). Male amelogenin dropouts: an analysis based on Y-chromosome biallelic polymorphisms and Y-STRs. *Forensic Sci Int* 166:155-163

12. Regueiro M, Cadenas AM, Gayden T, Underhill PA *et al*. (2006). Iran: Tricontinental nexus for Y-driven migration. *Hum Hered* 61:132-143

Presentations (A total of 10 presentations)

1. Gayden T, Chennakrishnaiah S, Regueiro M, Underhill PA *et al*. Peopling of the Himalayas. Stanford University, California, September 26th, 2011.

2. Gayden T, Chennakrishnaiah S, Regueiro M, Underhill PA *et al*. Phylogeography of the human Y-chromosome in three culturally defined regions of historical Tibet. Evolution 2011, The University of Oklahoma, Oklahoma, June 17-21st, 2011.

3. Gayden T, Perez A, Chennakrishnaiah S, and Herrera RJ. Tracing the human maternal lineages in the Himalayas. The 87th Florida Annual Meeting and Exposition (FAME), Tampa, May 12-15th, 2011.

4. Gayden T, Cadenas AM, Regueiro M, Singh NB, Zhivotovsky LA, Underhill PA, Cavalli-Sforza LL and Herrera RJ. Genetic histories of Tibetan and Nepalese populations. Graduate Students' Association (GSA) Scholarly Forum, FIU, April 22nd, 2010. (First position in Biology and Marine Sciences category)

5. Gayden T, Cadenas AM, Regueiro M, Singh NB, Zhivotovsky LA, Underhill PA, Cavalli-Sforza LL and Herrera RJ. Genetic diversity in the Himalayas. The 12th Annual Biology Research Symposium, FIU, February 6th, 2010. (Best oral presentation award)