

The matrilineal ancestry of Nepali populations

Rajdip Basnet (✉ rajdipbasnet@gmail.com)

Tribhuvan University Central Department of Biotechnology <https://orcid.org/0000-0003-1981-8243>

Niraj Rai

Birbal Sahni Institute of Palaeosciences: Birbal Sahni Institute of Palaeobotany

Rakesh Tamang

University of Calcutta

Nagendra Awasthi

Deakin University Faculty of Health

Isha Pradhan

Tribhuvan University Central Department of Biotechnology

Pawan Parajuli

AAHL Library: CSIRO Australian Animal Health Laboratory

Deepak Kashyap

Centre for Cellular and Molecular Biology CSIR

Alla Govardhan Reddy

Centre for Cellular and Molecular Biology CSIR

Gyaneshwer Chaubey

Banaras Hindu University

Krishna Das Manandhar

Tribhuvan University Central Department of Biotechnology

Tilak Ram Shrestha

Tribhuvan University Central Department of Biotechnology

Kumarasamy Thangaraj

Centre for Cellular and Molecular Biology CSIR

Research Article

Keywords: Population and Evolutionary Genetics, Human Evolution, East Asian, Tibeto-Burman, Nepal, Tibet

Posted Date: June 14th, 2022

DOI: <https://doi.org/10.21203/rs.3.rs-1728898/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

The Tibetan plateau and high mountain ranges of Nepal are one of the challenging geographical regions inhabited by modern humans. While much of the ethnographic and population-based genetic studies were carried out to investigate the Tibetan and Sherpa highlanders, little is known about the demographic processes that enabled the colonization of the hilly areas of Nepal. Thus, the present study aimed to investigate the past demographic events that shaped the extant Nepalese genetic diversity using mitochondrial DNA (mtDNA) variations from ethnic Nepalese groups. We have analyzed mtDNA sequences from 999 Nepalese samples and compared the data with 38,622 published mtDNA sequences from the rest of the world. Our analysis revealed that the genomic landscapes of prehistoric Himalayan settlers of Nepal were similar to the low altitude extant Nepalese (LAN), especially to that of Newar and Magar population group; but differ from contemporary high-altitude Sherpas. LAN might have derived their East Eurasian ancestry mainly from low altitude Tibeto-Burmans, who might have migrated from East Asia and assimilated across the Eastern Himalayas extended from Eastern Nepal to North-East of India, Bhutan, the Tibet and Northern Myanmar. We also identified a clear genetic sub-structure across different ethnic groups of Nepal based on mtDNA haplogroups and ectodysplasin-A receptor (EDAR) gene. Our comprehensive high-resolution mtDNA-based genetic study of Tibeto-Burman communities reconstructs the maternal origins of prehistoric Himalayan populations and sheds light on migration events that have brought most of the East Eurasian ancestry to the present-day Nepalese population.

Background

Nepal is geographically located along the southern slopes of the high Himalayan Mountain ranges. Politically, it is bordered by Tibet autonomous region in the north and India in the south, east, and west (Dhital 2015; Whelpton 2005). Between the Himalayas and the Terai plains, the hilly area stretches from west to east with an elevation rising up to 3000 m above sea level (masl). In many places, the mid-hill ranges are fragmented by the rivers flowing north to south. This rugged topography has been acting as a natural barrier, only permeable to small-scale human movement facilitating the existence of several isolated indigenous communities in Nepal. Pahadi (comprising mainly Brahmin, Chhetri and Kami), the most populous ethnic group in this region, speaks Indo-European (IE) language (Whelpton 2005). The other prominent ethnic populations such as Newar, Magar, Gurung, Tamang, Rai, and Limbu speak Tibeto-Burman languages (Dhital 2015; Whelpton 2005) and are scattered throughout the country. The Sherpas inhabit the high-altitude regions of Nepal, whereas the Tamang are considered to be native to the mid-hilly areas of the adjoining districts of Kathmandu valley. Since ancient times, this valley was a melting pot for different reasons, like fertile soil, a favorable climate, a malaria free zone, and a primal trade center that linked India and Tibet. Newars are considered as the original inhabitant of the Kathmandu valley and its surrounding districts². According to some historians, the Sino-Tibetan ethnic group (Kirats/ Kirati) arrived from the north are the ancestors of the contemporary Newar (Malla 1981; Whelpton 2005). Kirats are considered to be the indigenous aboriginals of the Eastern Himalayan region, extending eastward from Nepal to India, and are primarily divided into two ethnic groups viz. Rai and

Limbu (Malla 1981). Newar makes up to 5.0% of the Nepal population (2011 census, Nepal). The Magar people are the third-largest indigenous ethnic group in Nepal, representing 7.1% of its total population (2011 census, Nepal). They speak a language belonging to the Tibeto-Burman family and are believed to have settled in Nepal like the Kirats (Whelpton 2005).

Archaeological records suggest that the Mesolithic settlers in the Eastern Terai region had unique features that show some striking similarities to those found in Southeast Asians, especially to the Hoabinhian culture of Vietnam (Corvinus 1996). Neolithic tools found in several parts of the country, including Kathmandu valley, dates back to the 2nd millennium BC show affinities with the Assamese of Northeast India (Banerjee and Sharma 1969; Sharma 1988). The Himalayas, though considered being a natural barrier to gene flow, the presence of Tibeto-Burmans in Nepal could plausibly indicate dispersals, by circumventing the Himalayas, from Tibet directly or via Northeastern India. Among the Nepalese Tibeto-Burmans, Sherpa's migration from Tibet directly across the Himalayas to the high-altitude Nepalese valleys of Khumbu is well established (Bhandari et al. 2015; Zhang et al. 2017). However, little is known about the migration route, admixture and/or isolation events that brought most of the East Eurasian mtDNA lineages detected in LAN Tibeto-Burman groups. Currently available ancient DNA and genetic data studies from the Himalayan region have suggested contrasting hypotheses of Sherpa/Tibetan or low-altitude East Asian origin for this populations (Cole et al. 2017; Fornarino et al. 2009; Gayden et al. 2013; Gneccchi-Ruscione et al. 2017; Jeong et al. 2016; Wang et al. 2012), though this was most probably due to limited sample size and low-resolution mtDNA sequences. In the present study, we focused on dissecting the maternal ancestry of the Nepalese populations by performing the high-resolution mtDNA phylogeographic study to understand the demographic history of the Nepalese population. Additionally, we have also examined the distribution of the derived variant of the EDAR gene (rs3827760) to find out the East Asian ancestry in extant Nepalese populations.

Results And Discussion

Table 1 | **The background information of the 15 Nepalese groups.** Among them, nine groups were newly investigated, whereas six groups were included from previous studies. In total, 999 Nepalese samples were included in this study of which 461 samples were newly generated. All the Nepalese samples belonging to the Newar Ethnic group were included in a single group (Newar).

Table 1

The background information of the 15 Nepalese groups. Among them, nine groups were newly investigated, whereas six groups were included from previous studies. In total, 999 Nepalese samples were included in this study of which 461 samples were newly generated. All the Nepalese samples belonging to the Newar Ethnic group were included in a single group (Newar).

SN	Population	Ethnic group	Language	Location	Size (N)	Reference
1	Maharjan	Newar	Tibeto-Burman	Kirtipur, Kathmandu, Nepal	100	Present Study
2	Shrestha	Newar	Tibeto-Burman	Kathmandu; Bhaktapur; Kavre, Nepal	36	Present Study
3	Newa_mix	Newar	Tibeto-Burman	Kathmandu; Bhaktapur; Kavre, Nepal	50	Present Study
4	Manandhar	Newar	Tibeto-Burman	Kathmandu, Nepal	93	Present Study
5	Bajracharya	Newar	Tibeto-Burman	Kathmandu, Nepal	20	Present Study
6	Shakya	Newar	Tibeto-Burman	Kathmandu, Nepal	19	Present Study
7	Udaya	Newar	Tibeto-Burman	Kathmandu, Nepal	58	Present Study
8	Brahmin	Brahmin	Indo-European	Tanahun, Nepal	48	Present Study
9	Magar	Magar	Tibeto-Burman	Tanahun, Nepal	37	Present Study
10	Kathmandu	Various (admixed)	Various (admixed)	Kathmandu, Nepal	77	Gayden et al. 2013
11	Nepali-mix /Nepali-other	Various (admixed)	Various (admixed)	Kathmandu and Eastern Nepal	245	Wang et al. 2012
12	Tamang	Tamang	Tibeto-Burman	Nepal	45	Wang et al. 2012
13	Tharu CI	Tharu	Indo-European	Chitwan (Central Terai), Nepal	55	Fornarino et al. 2009
14	Tharu CII	Tharu	Indo-European	Chitwan (Central Terai), Nepal	76	Fornarino et al. 2009
15	Tharu E	Tharu	Indo-European	Morang (Eastern Terai), Nepal	40	Fornarino et al. 2009
Total					999	

Figure 1 | Map of Nepal and the approximate geographic locations of the Nepalese populations. Numbers refer to the populations (1. Newar; 2. Nepali-mix; 3. Tamang; 4. Tharu; 5. Magar; 6. Sherpa; and 7. Brahmin). More details regarding the populations are displayed in Table 1.

mtDNA lineages in Nepalese populations

A total of 999 individuals were classified into known mtDNA haplogroups (Table 1 and Additional file 2: Table S1 and S2) that were previously identified in East Eurasian, South Asian, and West Eurasian populations (Chandrasekar et al. 2009; Chaubey et al. 2008; Li et al. 2019b; Palanichamy et al. 2015; Qi et al. 2013; Silva et al. 2017; Thangaraj et al. 2005; Thangaraj et al. 2006; van Oven and Kayser 2009). Among the 376 Newar mtDNAs, the majority (50.8%) could be assigned unambiguously into East Eurasian haplogroups such as Z, F, D, and A, whereas 35.9% were assigned to haplogroups of South Asian origin and 12.8% mtDNA belonged to West Eurasian lineages. Similarly, in Magar 54.1% were assigned to East Eurasian haplogroups such as D, M9 and C, whereas 37.8% and 5.41% fit to South Asian and West Eurasian haplogroups, respectively. Likewise, in Brahmin, the bulk (60.4%) could be allotted to South Asian haplogroups, followed by 25% and 14.6% belonging to East Asian and West Eurasian haplogroups. Among the Nepalese populations, we observed the highest frequency of East Eurasian ancestry in Sherpa (94.3%) followed by Tharu CI (68.4%), Tamang (65.9%) and so on, as shown in the Table 2.

Table 2
Frequencies of the East Eurasian lineages observed in Nepalese populations.

Population	East Eurasian Ancestry (%)	Major East Eurasian Haplogroup frequencies							
		Z	F	D	B5	C	G	M9	A
Sherpa	94.3	2.7	2.5	5.8	0.0	20.9	0.7	24.2	27.1
Tharu CI	68.4	1.8	7.3	5.3	8.8	0.0	26.3	19.3	0.0
Tamang	65.9	0.0	6.8	26.1	0.0	2.3	11.1	14	4.5
Tharu CII	64.5	3.9	5.3	11.8	2.6	5.3	25	11.8	0.0
Magar	54.1	5.4	0.0	24.3	2.7	8.1	0.0	13.5	0.0
Newar	50.8	18.7	15.5	7.7	0.4	0.2	2.3	0.7	4.9
Kathmandu	41.6	5.2	7.8	11.7	1.3	3.9	0.0	2.6	5.2
Nepali-other	36.6	4.9	6.5	9.3	0.4	0.8	4.9	1.6	6.1
Tharu Eastern	35	0.0	5.0	10.0	2.5	2.5	12.5	0.0	0.0
Brahmin	25	0.0	6.3	4.2	0.0	2.1	8.3	0.0	2.1
Hindu Terai	8.3	0.0	0.0	4.2	0.0	0.0	0.0	0.0	0.0

Genetic relationship of Nepalese populations and their surrounding populations

We calculated diversity indices and tests of selective neutrality of Nepalese populations based on the HVS-I sequences (Additional file 2: Table S5). We detect relatively lower nucleotide diversity and mean pair-wise differences among the Newar sub-caste, whereas the highest diversity was observed among the Kathmandu (Gayden et al. 2007) and other Nepalese (mix population from Kathmandu and Eastern Nepal) (Wang et al. 2012). The pair-wise F_{st} distance (Additional file 1: Figure S2) shows a relatively higher genetic distance between the Newar and Tamang (0.031), followed by Newar and Nepali-other (0.029), Newar and Brahmin (0.025), and Newar and Magar (0.018). The principal component analysis (PCA) among the Nepalese populations (Additional file 1: Figure S3), accounting for 43.27% of the total variation, reflected a broad overview of population structure across the Southern Himalaya. All the Newar sub castes clustered tightly and were placed closer to the mixed populations from Kathmandu and Eastern Nepal. Brahmin positioned itself near the eastern Tharu (TH-E), Kathmandu and Nepali-other. Magar (MGR), Tamang (TMG), and Tharu from Chitwan (TH-CI and TH-CII) cluster closely in the PCA plot. Sherpa, who inhabit the highest mountainous regions close to the Tibetan plateau, remained distant from LAN, suggesting a low level of recent gene flow from surrounding Nepalese populations.

Further to examine the genetic relationship of Nepalese with other Asian populations, we performed PCA (Figure. 2), based on mtDNA haplogroup frequencies of Nepalese and Asian populations (Additional file 2: Table S4). PC1 fairly separated the Tai-Kadai and Han Chinese from the other groups. PC2 was successful in delineating the South Asians, including the Nepalese from the rest of the group. Tamang, Magar and Tharu (CI & CII) group of Nepal are placed closer to the high-altitude Himalayan cluster formed by Sherpa and Tibetans, showing high-altitude barrier to gene flow is rather more permeable from North to South direction across the Himalayan arc. A Southern cluster consisted mostly of IE and Dravidian-speaking populations of India and Nepal. Newar, Nepali-mix (Nepali-other), Tharu (Eastern), Tibeto-Burman groups (Myanmar, Bangladesh and northeast India) and IE groups (Kshatriya and Shah of North India) appear as potentially admixed populations forming a central Himalayan cluster in the PCA plot.

Figure 2 | Principal component analysis among the Nepalese and 127 other Asian populations. Non-Nepalese populations were mostly grouped by language families. The first (PC1) and the second (PC2) Principal components explain 24.37% and 21.4% of the genetic variance, respectively.

This differentiation across Nepalese subgroups was further tested by the analysis of the derived variant rs3827760 of the EDAR gene (Figure. 3). Scans of positive selection on genome-wide surveys from the global human population have identified the derived coding variant of the EDAR gene as an interesting candidate for the positive selection in East Asians (Fujimoto et al. 2008; Kamberov et al. 2013). This has resulted in a very high frequency of non-synonymous SNP rs3827760 (370A; 1540C) in East Asian and the Native American origin but virtually absent among the Indians (Dravidian and Indo-European), European and African populations (Chaubey et al. 2011; Fujimoto et al. 2008). Interestingly, this derived allele was present in both the Tibeto-Burman (Magar and Newar) and IE (Brahmin) populations of Nepal. The highest 1540C allele frequency was observed in Magar (71%), followed by Newar (30%) and Brahmin (20%). The frequency observed in Magar was similar to those reported in HAAPs (Himalayan and adjoining populations) (Tamang et al. 2018) and Indian Tibeto-Burman populations (Chaubey et al. 2011). However, despite having nearly equal proportions of East Eurasian maternal components, Newar showed almost half of the derived *EDAR* allele frequency than that of Magar, reflecting their distinct population history. The presence of a substantial 1540c allele frequency in Brahmin implies admixture with the local Tibet-Burman populations.

Figure 3 | EDAR 1540C allele frequency distribution. (A) Geographic distribution of the EDAR 1540C allele frequency worldwide. The Map was generated using the Kriging linear model of Surfer 16.0.3 (Golden Software, LLC). (B) Geographic distribution of the EDAR 1540C allele frequency in different groups of Asia. The frequency is shown in proportion to the bubble size.

Phylogeographic patterns of the dominant East Eurasian haplogroups

To delve into the migratory and admixture events that might have occurred between the aboriginal Nepalese and other East Eurasians, we compared the dominant East Eurasian lineages in Nepalese

(Table 2) and East Eurasians and analyzed haplotype networks tree and parsimonious phylogenetic tree of major East Eurasian haplogroups. Further, the frequency counter maps of shared haplogroups were constructed.

Table 3
Estimated ages of the basal East Eurasian lineages in the Nepalese population.

Complete Genome Substitutions				Complete Genome Substitutions			
Soares Rate				Soares Rate			
(Soares et.al.2009)				(Soares et.al.2009)			
Haplogroup	n	$\rho \pm SE$	TMRCA (Kya)	Haplogroup	n	$\rho \pm SE$	TMRCA (Kya)
A + 152 + 16362	53	8.9 ± 0.2	24.5 (22.9, 26.0)	F1d1	29	2.9 ± 0.2	7.6 (6.4, 8.8)
A17	41	5.3 ± 0.1	13.9 (13.4, 14.9)	F1d1a*+15204	14	2.8 ± 0.2	7.3 (6.0, 8.6)
A17a*	8	3.2 ± 0.3	8.4 (6.8, 10)	F1d1a1*	14	1.8 ± 0.2	4.6 (3.4, 5.9)
A17a1*	2	1.5 ± 0.5	3.8 (1.3, 6.4)	F1d + 5460 (F1d2*)	7	3.5 ± 0.1	9.2 (8.28, 10.2)
A27*	11	3.2 ± 0.2	8.4 (7.3, 9.4)	F1c1a2a*	19	3.6 ± 0.2	9.5 (8.4, 10.5)
Z3	97	9.2 ± 0.1	25.4 (24.4, 26.3)	F1c1a2a2*	5	2 ± 0.07	5.2 (4.8, 5.5)
Z3a	59	6.9 ± 0.2	18.7 (17.6, 19.8)	F1c1a2a1	9	1 ± 0.1	2.5 (2, 3)
Z3a1	51	6.8 ± 0.2	18.4 (16.8, 20.0)	G2a1d2	14	5.4 ± 0.3	14.4 (12.5, 16.1)
Z3a1a	45	4.8 ± 0.2	12.8 (11.7, 13.9)	G2a1d2b*	7	5.1 ± 0.5	13.6 (10.9, 16.3)
Z3a1a1*	43	3.2 ± 0.1	8.4 (7.9, 8.9)	G2a1d2b1*	6	4.8 ± 0.4	12.8 (10.6, 15)

TMRCA (Time to most recent common ancestor).

The symbol * indicates new basal lineages identified in this study.

		Complete Genome Substitutions				Complete Genome Substitutions	
Z3a2	7	4.1 ± 0.5	10.8 (8.2, 13.5)	G2a1d2a	5	1.8 ± 0.5	4.6 (2.1, 7.2)
Z3a2a	5	1 ± 0.3	2.5 (1.0, 4.1)	G2b2a1*	3	3.3 ± 0.1	8.7 (8.1, 9.2)
Z3b	9	2.1 ± 0.3	5.4 (3.9, 7.0)	G3a3*	4	2.2 ± 0.4	5.7 (3.6, 7.8)
Z1	22	7.9 ± 0.4	21.6 (19.3, 23.8)	D4i	41	5 ± 0.1	13.3 (12.8, 13.9)
Z1a	21	4 ± 0.2	10.6 (9.5, 11.6)	D4i2	21	3.1 ± 0.3	8.1 (6.5, 9.7)
Z1a1	15	3.3 ± 0.3	8.7 (7.1, 10.3)	D4i2a*	21	1.1 ± 0.1	2.8 (2.3, 3.3)
Z1a1a	13	1.5 ± 0.1	3.8 (3.3, 4.4)	D4i3	5	3 ± 0.4	7.8 (5.7, 10)
18	0.8 ± 0.07	2.0 (1.7, 2.4)		D4i6*	4	5.5 ± 0.8	14.7 (10.4, 19.2)
Z7	7	2.1 ± 0.2	5.4 (4.1, 6.8)	D4i6a*	3	2.7 ± 0.6	7 (3.9, 10.2)
Z7 + 7471	7	2.1 ± 0.2	5.4 (4.0, 6.8)	D4h2b*	5	1.6 ± 0.1	4.1 (3.6, 4.6)
Z8*	2	1.5 ± 0.6	3.8 (0.8, 7.0)	D5a2b	11	4.7 ± 0.3	12.5 (11, 14.1)
F1d	59	3.6 ± 0.1	9.5 (8.7, 10.2)	D5a2a1a	7	6.6 ± 0.5	17.8 (15, 20.6)
TMRCA (Time to most recent common ancestor).							
The symbol * indicates new basal lineages identified in this study.							

The phylogeography of haplogroup Z

Figure 4 | Diversity and distribution pattern of haplogroup Z. (A) Phylogenetic tree of sub-clade Z3 based on complete mtDNA sequences. Symbol * represents the new haplogroup defined in this study. (B) Network of sub-haplogroup Z3 based on HVR I sequence, color-coded by geographic origin. (C) Phylogenetic tree of haplogroup Z7 and Z8 based on complete mtDNA sequences. (D) Spatial frequency distribution map of haplogroup Z. Populations (Additional file 2: Table S4) having a frequency of Z > 0 were included in the analysis and were marked by the blue dots on the contour map.

Haplogroup Z is one of the infrequent mtDNA haplogroups with sporadic distribution in Russia, East Asia, Mainland Southeast Asia (MSEA), Central Asia, Eastern and Northern Europe, and South Asia (Figure. 4) (Chandrasekar et al. 2009; Fedorova et al. 2013). Among the members (Z1, Z2, Z3, Z4 and Z7) of haplogroup Z, Nepalese populations were characterized by rare clades Z3a1a and Z7, of which Z3a1a was the most frequent sub-clade in Newar (Additional file 1: Figure S4). Lineage Z1 dominated most of the representatives of haplogroup Z in North Asia and Europe (Fedorova et al. 2013). In this region, we estimated the coalescent age of ~ 21.6 Kilo-years ago (Kya) for clade Z1 (Table 3), similar to the earlier estimate of ~ 20.4 Kya (Fedorova et al. 2013). Another clade Z3, found in East Asia, North Asia and MSEA, is the oldest member of haplogroup Z with an estimated age of ~ 25.4 Kya. Haplogroup Z has its highest clade diversity in China, with an overall prevalence of 2.8% (620/ 21668) and is partially more frequent in Northern Han Chinese (3.1%). Likewise, the overall rate of haplogroup Z in Tibetans was 1.92% (of 6109 samples) (Qi et al. 2013). In another recent study of Tibetans, with comparatively less sample, haplogroup Z was characterized by clades Z3a (5/682), Z3a1a (1/682), Z4a (1/682) and Z4 (1/682) (Qi et al. 2013). Overall, haplogroup Z reaches the highest incidence in Tibeto-Burman speaking populations from Thailand (Lisu_2, 20%) (Oota et al. 2001), Nepal (Newar, 18.6%) and Myanmar (Chin2, 12%) (Li et al. 2015), in which clade Z3a1a dominated most of the representatives of this haplogroup.

After constructing a phylogenetic tree (Fig. 4A and Additional file 1: Figure S5) based on 161 complete mtDNA sequences, we identified the ancestral nodes of Z3a were distributed mainly in Northern and Eastern Han Chinese, and thus most likely originated in these regions. This claim was further bolstered by the presence of Z3a1 lineage in Yakut of Sakha Republic (Additional file 1: Figure S6). Terminal clade Z3a1a1 detected in Tibet, Myanmar, Nepal, India, Thai-Laos and Vietnam trace their ancestral roots to China with a coalescent age of ~ 8.4 Kya (Table 3). Basal variant 9452A with a coalescent age of ~ 4.4 Kya were only shared between the members of Z3a1a1 from North-East India, Thailand and Vietnam but not with the Nepalese and Tibetans. Z3a1a1 lineage of Tibet, Nepal, and India does not share any further basal variants, and lineage expansion as denoted by a star-like pattern was confined to their respective geographic locations (Figure. 4B). Sub haplogroup Z7 (Figure. 4C) observed, so far, only in Tibet, Northeast India and Nepal coalesce at ~ 5.4 Kya (Table 3). Newly defined clade Z8 (Figure. 4C), so far, detected only in Newar ethnic group of Nepal with a coalescent age ~ 3.8 Kya (Table 3).

The phylogeography of haplogroup F

Figure 5 | Diversity and distribution pattern of haplogroup F1. (A) Network of haplogroup F1 (F1d, F1d1, F1c1a2a, F1g, and F1g1) complete mtDNA sequences, color-coded by geographic origin. (B) network of sub-haplogroup F1d1, color-coded by geographic origin. (C, D) Spatial frequency distribution map of sub-

haplogroup F1d and F1c in Asia. Out of 308 Asian populations (Additional File 2: Table S4), only the populations having a frequency of F1d > 0 were included in the analysis and were marked by the blue dots in the Map.

Haplogroup F, mainly clade F1a, is one of the most common haplogroups throughout Asia, with high frequency and haplotype diversities reported in MSEA and China, an indication of a dispersal center for this haplogroup (Duong et al. 2018; Kutanan et al. 2017; Li et al. 2015; Li et al. 2019b; Summerer et al. 2014; Zhang et al. 2013). Though the overall incidence of haplogroup F in Han Chinese is 16.5%, its sub-clades F1d (0.5%, 111/21668), F1c (0.85%) and F1g (0.8%) was scarce (< 1%) and were relatively more in Northern Han Chinese (Li et al. 2019b). In the present study, Nepalese populations were characterized by rare clades of F1 (F1c, F1d1, and F1g) and F2b, of which F1d1 was the second most frequent sub-clade in Newar (Fig. 5A and Additional file 1: Figure S4). Tibetan mtDNA data also revealed similar frequency patterns, in which F1g (4.5%) (Kang et al. 2016) were rather more frequent than F1d (2.6%, 18/671) (Li et al. 2019a) and F1c (2.16% of 6109 samples) (Qi et al. 2013). However, the Lhobas, one of the most isolated Himalayan tribes of southeastern Tibet, harbors a high frequency of both clade F1d (11%, 10/91) and F1c (7.2%, 7/96) (Kang et al. 2016). Likewise, in MSEA, among the Burmese of Myanmar and Mon people of west Thailand, the frequency of F1d reaches up to 7% and 8% respectively (Kutanan et al. 2017; Summerer et al. 2014). Overall, haplogroup F1d reaches the greatest proportion in Newar (11.97%) of Nepal and Kshatriya (16%) (Negi et al. 2016) of North India.

When we constructed a parsimonious phylogenetic tree (Additional file 1: Figure S6) based on 121 complete mtDNA sequences, we identified the ancestral node of clade F1d was mainly disseminated in Han Chinese with an estimated coalescent age of ~ 9.5 Kya (Table 3). Clade F1c, another sub-lineage of haplogroup F, attains its highest frequency in the Tibeto-Burman speaking group from Northeast India (Apatani, 16.8% and Nishi, 9.1%) and Myanmar (Chin 3: 15.4%, Naga 1: 6.3% and Rakhine 1: 8.3%). The estimated coalescent age of the sub-clade F1c1a2a and F1d1 was ~ 9.5 Kya and 7.8 Kya, respectively (Table 3). After the divergence from the ancestral Chinese F1d1 sequence, Tibeto-Burman groups (of Myanmar, Tibet, and Nepal) and Austroasiatic (AA) groups (of Thailand) shared a common variant 16284G (Figure. 5B) implying a common Tibeto-Burman ancestry. Intriguingly, these groups do not share any further basal variants and lineage expansion as denoted by star-like shape was confined to their respective geographic regions (Figure. 5B). Nepalese-specific branch (F1d1a1) coalesces at ~ 4.4 Kya, suggesting their ancient origin and lineage expansion in Nepal most probably via southeastern Tibet (Figure. 5C). Likewise, after bifurcating from ancestral Chinese/Vietnamese sequence, four basal variants (3970T, 5300T, 5663T, and 12892C) defining clade F1g1a remained restricted within Nepalese populations (Additional file 1: Figure S8). Analysis of other clades, F1c and F2b, also resulted in a similar mode of distribution, further verifying the notion of early entry into Nepal (Figure. 5D and Additional file 1: Figure S7 and S9).

The phylogeography of haplogroup A

Figure 6 | Maximum Parsimonious tree of haplogroup A17 and newly defined haplogroup A27. Complete phylogenetic tree based on 52 complete mtDNA sequences is given in the Additional File 1: Figure S10

Haplogroup A is one of the most prevalent haplogroups in northern and eastern Asia. It is relatively higher in Tibetans (14.63%) (Qi et al. 2013) than in Northern (8.53%) and Southern Han Chinese (6.54%) (Li et al. 2019b). Members of haplogroup A in Tibet are dominated by sub-clades A4 (7.8%) and A11 (5.5%) whereas the other sub-clades are absent or present in shallow frequency in Tibetans (< 2%) (Kang et al. 2016; Qi et al. 2013). Haplogroup A, dominated by its sub-lineage A15c1, was the most common maternal lineage in Sherpa of Nepal, with a frequency of 27.5% (Bhandari et al. 2015). This clade was not observed in the other Nepalese populations studied so far. In the examined LAN, lineage A was mirrored by its clades, A27, A14 and A17, of which A27 was the most abundant clade in Newar (3.99%). Newly defined clade A27 only discerned so far in Newar and Nepali-mix coalesce at ~ 8.4 Kya (Fig. 6 and Table 3) suggesting their ancient origin and potentially *in-situ* differentiation in Nepal. The other sub-clades, A14 and A17, were seen with shallow occurrence in Brahmin, Newar, Kathmandu and Nepali-mix. A17a lineage belonging to Nepalese, Chinese and Vietnamese shares the most recent common ancestor dating back to ~ 8.4 Kya (Fig. 5 and Table 3). Its sub-clade A17a1, branched off directly from the nodes occupied by the Chinese/Vietnamese lineages, is peculiar to Nepal with a coalesces time ~ 3.8 Kya.

The phylogeography of haplogroup D

Haplogroup D is found in Northeast Asia, and Central Asia; and is the second most dominant haplogroup in Tibetans with an average frequency of 16.5% in the Plateau (Comas et al. 2004; Kang et al. 2013; Kutanan et al. 2017; Li et al. 2015; Li et al. 2019b). Detected with noticeable frequency among the Nepalese, it also represents the most dominant maternal lineage in Tamang (26.1%) and Magar (24.3%). Lineage D4 and its representative D4i, D4j and D4e were the major sub-haplogroup in Nepalese, whereas D5 lineage was dominated by D5a2 and its sub-clades D5a2b and D5a2a1 (Additional file 1: Figure S4). We detect distinct sub clades of D4i distributed among the Nepalese populations. Magar and Nepali-mix shared the most recent common ancestor with Chinese, and Taiwanese D4i3 with a coalescent time of ~ 7.8 Kya (Table 3 and Additional file 1: Figure S11). Whereas, D4i detected in Newar, newly named D4i6, shared its ancestral root with the Turkic ethnic group (Uyghur) of northwest China (near to Central Asia) and coalesces at ~ 14.7 Kya (Table 3). Its sub-clade D4i6a so far only sighted in the Nepalese populations coalesce at ~ 7 Kya. Interestingly, newly defined clade D4h2b detected in Thailand and Nepal share a most recent common ancestor and coalesce at ~ 4.1 Kya. Three Nepalese samples shared a most recent common ancestor with the Chinese/Vietnamese and Tibetan D5a2b lineage (Additional file 1: Figure S12).

The phylogeography of haplogroup M9a

M9a is the most common haplogroup in Tibet, with an overall occurrence of 22.4% (Qi et al. 2013). In the Nepalese populations, it is prevalent mainly in Sherpa (27.4%), Tharu-CI (19.6%), Tamang (15.5%), Magar (13.5%), and Tharu-CII (Additional file 1: Figure S4). Among the four sub-haplogroups (M9a1a, M9a1a2, M9a1b1 and M9a1a1c1b1a) of M9a, M9a1a1c1b1a is predominant (58% of the total M9a individual) in both Tibetans (Peng et al. 2011) and Sherpa (Bhandari et al. 2015) but virtually absent in other Nepalese populations. Previous study estimated a shallow divergence between Tibetans and Sherpa, indicating recent gene flow from Tibetans to Sherpa (Bhandari et al. 2015). Likewise, M9a1a2 is prevalent in Tharu

and Magar, whereas Tamang encompasses M9a1a, M9a1a2 and M9a1b sub-clades. The shared prevalence of M9a1a2 between Nepalese populations (Tharu, Magar, Tamang and Newar) and the estimated age of 5.6 Kya (Table 3) show their rather ancient origin and most plausibly a *de novo* differentiation in Nepal (Wang et al. 2012).

The phylogeography of haplogroup G

Haplogroup G is distributed at a significant frequency across East Asia, Central Asia, and MSEA (Asari et al. 2007; Kutanan et al. 2017; Li et al. 2015; Li et al. 2019b; Qi et al. 2013). Among Nepalese, haplogroup G reaches the highest frequency in the Tharu (Th-CII, 28%, Th-CI, 26.7% and Th-E, 9.7%) ethnic group and also detected in considerable frequency in Tamang (11%) and Brahmin (8.3%). Complete mtDNA analysis shows clustering of Nepalese Brahmin G3 lineage with those identified in China, Tibet and Kashmir. Moreover, Brahmin and Kashmiri G3a3a lineage show divergence from ancestral Tibetans/Chinese G3 lineage ~ 5.7 Kya (Table 3 and Additional file 1: Figure S13). Similar to G3a, we identified several new lineages (G2a1d2b, G2b2a1 and G2c) of G2 characterized by population expansion within the Nepalese groups (Additional file 1: Figure S13).

Nepal holds a crucial and significant location in the geology of the Himalaya, serving as a bridge between the Western and Eastern Himalaya (Dhital 2015). It stretches from low-lying foothills in the south to the Great Himalayas that extend to the Tibetan plateau in the north, featuring several of the world's highest peaks (Dhital 2015). At a genetic level, Himalayan arc was also suggested to have played a pivotal role in shaping human population history and genetic diversity in Asia, acting as a natural genetic barrier that helped to keep large cultural and genetic distinctions observable between South Asian and East Asian populations (Bhandari et al. 2015; Gayden et al. 2007; Gayden et al. 2013; Wang et al. 2012). Tibetan desert high plateau and the high-altitude transverse valleys (> 3,000 masl) of Nepal were among the challenging places inhabited by prehistoric humans. According to previous genetic studies, most of the Tibetan genetic components can be attributed mainly to the Neolithic immigrations initiated from northern China ~ 7 Kya (Qi et al. 2013; Zhao 2009). Archaeological data from the Tibetan plateau suggest an initial extensive human settlement in the northeastern plateau (< 2500 masl) ~ 5.2 Kya, long before their further expansion to high-altitude plateau areas (> 2500 masl) ~ 3.6 Kya (Chen et al. 2015). Among the Nepalese Tibeto-Burmans, Sherpa migration from Eastern to Central Tibet and later, approximately in the 16th century, to the Solukhumbu region of Nepal is well-documented (Oppitz ; Whelpton 2005). These historical records correlate adequately with both uniparental and genome-wide perspective (Bhandari et al. 2015; Jeong et al. 2016; Zhang et al. 2017). A recent study showed that two clades, M9a1a1c1b1a and A11a1a, played significant roles in shaping the genetic landscapes of contemporary Tibetans (Li et al. 2019a). However, our findings show that these lineages were limited to Tibetans and Sherpas, and were essentially non-existent in the LAN. Except for few LAN lineages that show connections to northeastern Tibet e.g., D4i3, G2a1d2b, G3a1—most of them show settlements into Eastern Himalayas that extend from Eastern Nepal across Northeast India, Bhutan, Southeastern Tibet to Yunan in China and Northern Myanmar.

Hence, phylogeographic analysis of the East Asian haplogroups detected in LAN shows different characteristics than high altitude Sherpa/Tibetans: **(1)** LAN share common Tibeto-Burman precursor with few ethnic groups of Southeastern Tibet including Lhasa and are prevalent in Myanmar, Thailand and Northeast India; **(2)** majority of the LAN maternal components can be traced back to Neolithic immigrations from China ~ 8 Kya, which is in a good agreement with previous studies (Qi et al. 2013; Zhao 2009); **(3)** Significantly, after the divergence from ancestral Chinese sequence, the majority of LAN mtDNA lineage do not share any additional variants with Tibetans, Northeast Indians and Burmese.

This genetic differentiation between Sherpa/ Tibetans and LAN is further supported by the mtDNA profile of eight ancient dental samples derived from the high-altitude Annapurna Conservation Area (ACA) of Nepal (Mustang District), situated near the Tibetan plateau (Jeong et al. 2016). These samples, spanning 3,150 to 1,250 years before present, yielded sub-clades (Z3a1a, F1c1a, F1d1, D4j1b, and M9a) that were mostly present in the contemporary LAN and were virtually absent in high altitude Sherpa (except M9a, whose terminal haplogroups was uncharacterized). Therefore, it is conceivable that precursors of the earliest inhabitant of the Southern slopes of Himalaya (e.g., those having occupied the mustang region of Nepal) were genetically closer to the contemporary low-altitude Nepalese and plausibly crossed the high-altitude Tibetan/Nepalese passes. Our results rejected the previous claim of affiliation of ACA samples with the high-altitude contemporary Sherpas (Jeong et al. 2016). This shed new light on the migratory roadmap and admixture events that might have occurred on the southern slopes of the Himalayas. Furthermore, phylogeographic analysis of the majority of the East Eurasian lineages detected in LAN displayed *de novo* differentiation in contemporary Nepalese, albeit a few shows a recent genetic admixture with northeast Indians. Among the *de novo* differentiated lineages, most were > 3.8 Kya (Table 3) suggesting a subsequent entry into Nepal after the extensive human settlements on the plateau began.

Conclusion

The present study is the first extensive high-resolution mtDNA based population genetic analysis of Nepalese populations, which provides new insights into migration and admixture events that occurred on the populations speaking Tibeto-Burman languages. Populations living in the South of Himalaya (mainly Newar and Magar) reveals distinct population history than contemporary high-altitude Tibetans/Sherpas. In fact, they were similar to ACA samples (prehistoric Himalayan settlers) and their ancestry can be traced back to Neolithic immigration from East Asia around ~ 8 Kya. Finally, we argued that unlike Sherpas/ Tibetans, who remained confined in the high-altitude region, LAN primarily derived their East Asian ancestry from low altitude dwelling proto Tibeto-Burmans who seems to have diffused from East Asia across the eastern Himalaya. The ancient genetic makeup being gradually reshaped by several admixture events along the migratory route from China to Nepal, the bearers of these lineages might have entered Nepal directly across the Himalayas, most probably *via* Southeastern Tibet around 3.8-6 Kya. Overall, the result of this study provides a step forward in dissecting the complexity of the East Asian mtDNA landscape of the Himalayan populations residing in the South of the Himalaya and point to the need for further study using other genetic markers.

Methods

Sampling

About 5.0 ml of blood samples were collected from 461 unrelated male individuals belonging to 9 aboriginal Nepalese populations, representing the three ethnic groups, namely Newar, Magar, and Brahmin (Table 1 and Fig. 1). The individuals were healthy and unrelated, as recognized by the personal interrogation. Newar samples were collected from Kathmandu, Lalitpur, Bhaktapur and Kavrepalanchok district, and Magar & Brahmin from Tanahun district. This research is conducted under strict guidelines and regulations based on the experimental protocol on human subjects approved by the Review Board of Nepal Health Research Council, Kathmandu, Nepal and Institutional Ethical Committee of CSIR-Centre for Cellular and Molecular Biology, Hyderabad, India.

Mitochondrial DNA and EDAR 1540T/C Sequencing

Following the previously described phenol-chloroform method (THANGARAJ et al. 2002), the genomic DNA was extracted from the collected blood samples. DNA quantification was done using NanoDrop® ND-1000 (ThermoFisher Scientific, USA). Polymerase chain reaction (PCR) amplification and DNA sequencing of partial/ complete mtDNA segment(s) (Rieder et al. 1998) were performed in dedicated clean rooms with negative control and blank samples. Initially, we have sequenced the hypervariable region I & II (HVR I & II), followed by sequencing of selected coding region of mtDNA for accurate haplogroup assignment. Further, the whole mtDNA sequencing of the samples representing the dominant East Eurasian haplogroups was performed. In total, we obtained 139 complete mtDNA sequences from Nepalese individuals. The adaptive non-synonymous allele rs3827760, also known as *EDAR* 1540T/C single nucleotide polymorphism (SNP), was genotyped by PCR and direct sequencing using forward (5'-3') TCTGCACACAAGGACTCCAC and reverse (5'-3') GCGTTCTAGGTGTCGTTTGC primers to find East Asian genetic affinity. Altogether, this SNP was assayed in 461 Nepalese samples.

Data analyses

This study generated 461 mtDNA sequence data. Additionally, 538 Nepalese mtDNA sequences were collected from previous literature (Table 1). In total, haplogroups were assigned/reassigned to 999 Nepalese individuals using PhyloTree_{mt} Build 17 (van Oven and Kayser 2009) and Haplogrep2 (Kloss-Brandstätter et al. 2011). To designate the new basal haplogroup in this study, we followed the nomenclature system suggested by van Oven (van Oven and Kayser 2009). Molecular diversity indices, as well as the pair-wise genetic distance between the Nepalese populations, were calculated using Arlequin 3.5 (Excoffier and Lischer 2010) with HVS-I sequences. For the comparative analyses, we retrieved mtDNA data of a total of 39,293 individuals from 309 global populations, published elsewhere (Additional file 2: Table S4 and Additional file 1: Figure S1). To understand the genetic relationship of Nepalese with other populations, we performed PCA using a method developed by Richards et al. (Richards et al. 2002) in the MVSP (Multivariate Statistical Package for Windows, Ver. 3.22) software (Kovach 1999). The median-joining tree for the dominant haplogroups was constructed manually and

confirmed by Network10.1 (www.fluxus-engineering.com/sharenet.htm) program. The most parsimonious phylogenetic tree of various haplogroups was reconstructed manually based on the complete mtDNA sequences and verified by mtPhyl software (Eltsov and Volodko 2009). The isofrequency maps for dominant haplogroups were constructed using the Kriging linear model in Surfer 16.0.3 (Golden Software, LLC). The coalescence age or time to the most recent common ancestor (TMRCA) of basal mtDNA haplogroup was estimated using ρ statistics method as described previously (Soares et al. 2009).

Statement And Declarations

Ethics approval and consent to participate

All individuals gave informed consent before the blood collection. This study was approved by the Nepal Health Research Council, Kathmandu, Nepal and Institutional Ethical Committee of CSIR-Centre for Cellular and Molecular Biology, Hyderabad, India. The relevant protocol was explained to each individual, and an informed consent written in Nepali was got by either signature or fingerprint if the subject could not write.

Consent for publications

Not applicable

Funding

The authors declare that no funds or grants were received during the preparation of this manuscript.

Competing interests

The authors declare that they have no competing interests.

Authors' information

Affiliations

Central Department of Biotechnology, Tribhuvan University, Kirtipur, Nepal

Rajdip Basnet, Nagendra P. Awasthi, Isha Pradhan, Krishna Das Manandhar and Tilak Shrestha

CSIR-Centre for Cellular and Molecular Biology, Hyderabad, Telangana, India

Rakesh Tamang, Alla G Reddy, Deepak Kashyap and Kumarasamy Thangaraj

Birbal Sahni Institute of Palaeosciences, Lucknow, India

Niraj Rai

School of Medicine, Deakin University, Geelong, Australia

Nagendra P. Awasthi

Department of Zoology, University of Calcutta, Kolkata, India

Rakesh Tamang

Australian Centre for Disease Preparedness, 5 Portarlington Rd, Geelong VIC 3219, Australia

Pawan Parajuli

Cytogenetics laboratory, Department of Zoology, Banaras Hindu University, Varanasi, India

Gyaneshwer Chaubey

Centre for DNA Fingerprinting and Diagnostics, Hyderabad, India

Kumarasamy Thangaraj

Contributions

K.T., T.S., and N.R. conceived and designed the experiments. R.B., N.A., P.P, and I.P. collected the samples and performed the experiments. D.K. contributed to experiments. R.B., and N.R. analyzed, interpreted the data and wrote the manuscript. R.B., R.T., I.P. and N.A. contributed to data analysis. G.C. and K.D.M. edited the manuscript. The authors read and approved the final manuscript.

Availability of data and materials

Accession Numbers: Sequence data for the 139 mtDNA whole-genome sequences have been deposited in GenBank nucleotide core database under accession numbers OM638153 to OM638291.

Acknowledgements

We are grateful to all the volunteers for providing their blood samples in this project. We would like to thank Mr. Sandesh Shrestha, Mr. Ram Singh Dhama, and Mr. Ganga Prasad Phuyal for their help during the sample collection. K.T. was supported by J.C Bose Fellowship, SERB, and CSIR, Ministry of Science and Technology, Government of India

References

1. Asari M et al (2007) Utility of haplogroup determination for forensic mtDNA analysis in the Japanese population. 9:237–240
2. Banerjee N, Sharma JL (1969) Neolithic tools from Nepal and Sikkim. 9: 53 – 8

3. Bhandari S et al (2015) Genetic evidence of a recent Tibetan ancestry to Sherpas in the Himalayan region. *Sci Rep* 5:16249. doi: 10.1038/srep16249
4. Chandrasekar A et al (2009) Updating phylogeny of mitochondrial DNA macrohaplogroup m in India: dispersal of modern human in South Asian corridor. *PLoS ONE* 4:e7447. doi: 10.1371/journal.pone.0007447
5. Chaubey G et al (2008) Phylogeography of mtDNA haplogroup R7 in the Indian peninsula. *BMC Evol Biol* 8:227. doi: 10.1186/1471-2148-8-227
6. Chaubey G et al (2011) Population genetic structure in Indian Austroasiatic speakers: the role of landscape barriers and sex-specific admixture. *Mol Biol Evol* 28:1013–1024. doi: 10.1093/molbev/msq288
7. Chen FH et al (2015) Agriculture facilitated permanent human occupation of the Tibetan Plateau after 3600 B.P. *Science* 347:248–250. doi: 10.1126/science.1259172
8. Cole AM et al (2017) Genetic structure in the Sherpa and neighboring Nepalese populations. *BMC Genomics* 18:102. doi: 10.1186/s12864-016-3469-5
9. Comas D et al (2004) Admixture, migrations, and dispersals in Central Asia: evidence from maternal DNA lineages. *Eur J Hum Genet* 12:495–504. doi: 10.1038/sj.ejhg.5201160
10. Corvinus G (1996) The prehistory of Nepal after 10 years of research. 14: 43–55
11. Dhital MR (2015) *Geology of the Nepal Himalaya: regional perspective of the classic collided orogen.* Springer
12. Duong NT et al (2018) Complete human mtDNA genome sequences from Vietnam and the phylogeography of Mainland Southeast Asia. *Sci Rep* 8:11651. doi: 10.1038/s41598-018-29989-0
13. Eltsov N, Volodko N (2009) mtPhyl-software tool for human mtDNA analysis and phylogeny reconstruction
14. Excoffier L, Lischer HE (2010) Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Mol Ecol Resour* 10:564–567. doi: 10.1111/j.1755-0998.2010.02847.x
15. Fedorova SA et al (2013) Autosomal and uniparental portraits of the native populations of Sakha (Yakutia): implications for the peopling of Northeast Eurasia. *BMC Evol Biol* 13:127. doi: 10.1186/1471-2148-13-127
16. Fornarino S et al (2009) Mitochondrial and Y-chromosome diversity of the Tharus (Nepal): a reservoir of genetic variation. *BMC Evol Biol* 9:154. doi: 10.1186/1471-2148-9-154
17. Fujimoto A et al (2008) A replication study confirmed the EDAR gene to be a major contributor to population differentiation regarding head hair thickness in Asia. *Hum Genet* 124:179–185. doi: 10.1007/s00439-008-0537-1
18. Gayden T et al (2007) The Himalayas as a directional barrier to gene flow. *Am. J. Hum. Genet.* 80. doi: 10.1086/516757

19. Gayden T et al (2013) The Himalayas: barrier and conduit for gene flow. *Am J Phys Anthropol* 151:169–182. doi: 10.1002/ajpa.22240
20. Gneccchi-Ruscione GA et al (2017) The genomic landscape of Nepalese Tibeto-Burmans reveals new insights into the recent peopling of Southern Himalayas. *Sci Rep* 7:15512. doi: 10.1038/s41598-017-15862-z
21. Jeong C et al (2016) Long-term genetic stability and a high-altitude East Asian origin for the peoples of the high valleys of the Himalayan arc. *Proc Natl Acad Sci U S A* 113:7485–7490. doi: 10.1073/pnas.1520844113
22. Kamberov YG et al (2013) Modeling recent human evolution in mice by expression of a selected EDAR variant. *Cell* 152:691–702. doi: 10.1016/j.cell.2013.01.016
23. Kang L et al (2013) mtDNA lineage expansions in Sherpa population suggest adaptive evolution in Tibetan highlands. *Mol Biol Evol* 30:2579–2587. doi: 10.1093/molbev/mst147
24. Kang L et al (2016) MtDNA analysis reveals enriched pathogenic mutations in Tibetan highlanders. *Sci Rep* 6:31083. doi: 10.1038/srep31083
25. Kloss-Brandstätter A et al (2011) HaploGrep: a fast and reliable algorithm for automatic classification of mitochondrial DNA haplogroups. *Hum Mutat* 32:25–32
26. Kovach W (1999) MVSP—A multivariate statistical package for Windows, ver. 3.1. Kovach Computing Services, Pentraeth, Wales, UK 137
27. Kutanan W et al (2017) Complete mitochondrial genomes of Thai and Lao populations indicate an ancient origin of Austroasiatic groups and demic diffusion in the spread of Tai-Kadai languages. *Hum Genet* 136:85–98. doi: 10.1007/s00439-016-1742-y
28. Li Y-C et al (2019a) Neolithic millet farmers contributed to the permanent settlement of the Tibetan Plateau by adopting barley agriculture. *Natl Sci Rev* 6:1005–1013. doi: 10.1093/nsr/nwz080
29. Li YC et al (2015) Ancient inland human dispersals from Myanmar into interior East Asia since the Late Pleistocene. *Sci Rep* 5:9473. doi: 10.1038/srep09473
30. Li YC et al (2019b) River Valleys Shaped the Maternal Genetic Landscape of Han Chinese. *Mol Biol Evol* 36:1643–1652. doi: 10.1093/molbev/msz072
31. Malla KP (1981) Linguistic archaeology of the Nepal Valley: A preliminary report
32. Negi N et al (2016) The paternal ancestry of Uttarakhand does not imitate the classical caste system of India. *J Hum Genet* 61:167–172. doi: 10.1038/jhg.2015.121
33. Oota H et al (2001) Human mtDNA and Y-chromosome variation is correlated with matrilineal versus patrilineal residence. *Nat Genet* 29:20–21. doi: 10.1038/ng711
34. Oppitz M (1974) Myths and facts: Reconsidering some data concerning the clan history of the Sherpas Available at: <http://www.dspace.cam.ac.uk/handle/1810/227209>
35. Palanichamy MG et al (2015) West Eurasian mtDNA lineages in India: an insight into the spread of the Dravidian language and the origins of the caste system. *Hum Genet* 134. doi: 10.1007/s00439-015-1547-4

36. Peng MS et al (2011) Inland post-glacial dispersal in East Asia revealed by mitochondrial haplogroup M9a'b. *BMC Biol.* 9. doi: 10.1186/1741-7007-9-2
37. Qi X et al (2013) Genetic evidence of paleolithic colonization and neolithic expansion of modern humans on the Tibetan plateau. *Mol Biol Evol* 30:1761–1778. doi: 10.1093/molbev/mst093
38. Richards M et al (2002) In search of geographical patterns in European mitochondrial DNA. *Am J Hum Genet* 71:1168–1174. doi: 10.1086/342930
39. Rieder MJ et al (1998) Automating the identification of DNA variations using quality-based fluorescence re-sequencing: analysis of the human mitochondrial genome. *Nucleic Acids Res* 26:967–973. doi: 10.1093/nar/26.4.967
40. Sharma DR (1988) Archaeological remains of the Dang valley. 88: 8–15
41. Silva M et al (2017) A genetic chronology for the Indian Subcontinent points to heavily sex-biased dispersals. *BMC Evol Biol* 17:88. doi: 10.1186/s12862-017-0936-9
42. Soares P et al (2009) Correcting for purifying selection: an improved human mitochondrial molecular clock. *Am J Hum Genet* 84:740–759. doi: 10.1016/j.ajhg.2009.05.001
43. Summerer M et al (2014) Large-scale mitochondrial DNA analysis in Southeast Asia reveals evolutionary effects of cultural isolation in the multi-ethnic population of Myanmar. *BMC Evol Biol* 14:17. doi: 10.1186/1471-2148-14-17
44. Tamang R et al (2018) Reconstructing the demographic history of the Himalayan and adjoining populations. *Hum Genet* 137:129–139. doi: 10.1007/s00439-018-1867-2
45. Thangaraj K et al (2005) Reconstructing the origin of Andaman Islanders. *Science* 308:996. doi: 10.1126/science.1109987
46. Thangaraj K et al (2006) In situ origin of deep rooting lineages of mitochondrial Macrohaplogroup 'M' in India. *BMC Genomics* 7:151. doi: 10.1186/1471-2164-7-151
47. THANGARAJ K et al (2002) CAG Repeat Expansion in the Androgen Receptor Gene Is Not Associated With Male Infertility in Indian Populations. 23:815–818. 10.1002/j.1939-4640.2002.tb02338.x
48. van Oven M, Kayser M (2009) Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. *Hum Mutat* 30:E386–E394. doi: 10.1002/humu.20921
49. Wang HW et al (2012) Revisiting the role of the Himalayas in peopling Nepal: insights from mitochondrial genomes. *J Hum Genet* 57:228–234. doi: 10.1038/jhg.2012.8
50. Whelpton J (2005) *A history of Nepal*. Cambridge University Press
51. Zhang C et al (2017) Differentiated demographic histories and local adaptations between Sherpas and Tibetans. *Genome Biol* 18:115. doi: 10.1186/s13059-017-1242-y
52. Zhang X et al (2013) Analysis of mitochondrial genome diversity identifies new and ancient maternal lineages in Cambodian aborigines. *Nat Commun* 4:2599. doi: 10.1038/ncomms3599
53. Zhao M (2009) Mitochondrial genome evidence reveals successful Late Paleolithic settlement on the Tibetan Plateau. *Proc. Natl. Acad. Sci. USA* 106. doi: 10.1073/pnas.0907844106

Figures

Figure 1

Map of Nepal and the approximate geographic locations of the Nepalese populations. Numbers refer to the populations (1. Newar; 2. Nepali-mix; 3. Tamang; 4. Tharu; 5. Magar; 6. Sherpa; and 7. Brahmin). More details regarding the populations are displayed in Table 1.

Figure 2

Principal component analysis among the Nepalese and 127 other Asian populations. Non-Nepalese populations were mostly grouped by language families. The first (PC1) and the second (PC2) Principal components explain 24.37% and 21.4% of the genetic variance, respectively.

Figure 3

EDAR 1540C allele frequency distribution. (A) Geographic distribution of the EDAR 1540C allele frequency worldwide. The Map was generated using the Kriging linear model of Surfer 16.0.3 (Golden Software, LLC). (B) Geographic distribution of the EDAR 1540C allele frequency in different groups of Asia. The frequency is shown in proportion to the bubble size.

Figure 4

Diversity and distribution pattern of haplogroup Z. (A) Phylogenetic tree of sub-clade Z3 based on complete mtDNA sequences. Symbol * represents the new haplogroup defined in this study. (B) Network of sub-haplogroup Z3 based on HVR I sequence, color-coded by geographic origin. (C) Phylogenetic tree of haplogroup Z7 and Z8 based on complete mtDNA sequences. (D) Spatial frequency distribution map of haplogroup Z. Populations (Additional file 2: Table S4) having a frequency of $Z > 0$ were included in the analysis and were marked by the blue dots on the contour map.

Figure 5

Diversity and distribution pattern of haplogroup F1. (A) Network of haplogroup F1 (F1d, F1d1, F1c1a2a, F1g, and F1g1) complete mtDNA sequences, color-coded by geographic origin. (B) network of sub-

haplogroup F1d1, color-coded by geographic origin. **(C, D)** Spatial frequency distribution map of sub-haplogroup F1d and F1c in Asia. Out of 308 Asian populations (Additional File 2: Table S4), only the populations having a frequency of $F1d > 0$ were included in the analysis and were marked by the blue dots in the Map.

Figure 6

Maximum Parsimonious tree of haplogroup A17 and newly defined haplogroup A27. Complete phylogenetic tree based on 52 complete mtDNA sequences is given in the Additional File 1: Figure S10

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [AdditionalFile1.pdf](#)
- [AdditionalFile2.xlsx](#)