Journal of Human Genetics (2012) 57, 228–234

www.nature.com/jhg

ORIGINAL ARTICLE

Revisiting the role of the Himalayas in peopling Nepal: insights from mitochondrial genomes

Hua-Wei Wang[1,2], Yu-Chun Li[1], Fei Sun[3], Mian Zhao[1,4], Bikash Mitra[2,5], Tapas Kumar Chaudhuri[5], Pasupati Regmi[6], Shi-Fang Wu[1,4], Qing-Peng Kong[1,4] and Ya-Ping Zhang[1,2,4]

Himalayas was believed to be a formidably geographical barrier between South and East Asia. The observed high frequency of the East Eurasian paternal lineages in Nepal led some researchers to suggest that these lineages were introduced into Nepal from Tibet directly; however, it is also possible that the East Eurasian genetic components might trace their origins to northeast India where abundant East Eurasian maternal lineages have been detected. To trace the origin of the Nepalese maternal genetic components, especially those of East Eurasian ancestry, and then to better understand the role of the Himalayas in peopling Nepal, we have studied the matenal genetic composition extensively, especially the East Eurasian lineages, in Nepalese and its surrounding populations. Our results revealed the closer affinity between the Nepalese and the Tibetans, specifically, the Nepalese lineages of the East Eurasian ancestry generally are phylogenetically closer with the ones from Tibet, albeit a few mitochondrial DNA haplotypes, likely resulted from recent gene flow, were shared between the Nepalese and northeast Indians. It seems that Tibet was most likely to be the homeland for most of the East Eurasian in the Nepalese. Taking into account the previous observation on Y chromosome, now it is convincing that bearer of the East Eurasian genetic components had entered Nepal across the Himalayas around 6 kilo years ago (kya), a scenario in good agreement with the previous results from linguistics and archeology.
Journal of Human Genetics (2012) **57,** 228–234; doi:10.1038/jhg.2012.8; published online 22 March 2012

Keywords: mtDNA; Nepalese; origin

INTRODUCTION

Understanding how the ancestors of modern humans had successfully settled and adapted to areas with extreme conditions, for example, highlands, continues to be a hot issue of wide interest. For instance, a recent study on mitochondrial genome information revealed that the modern humans had successfully colonized the Tibetan Plateau since the Late Pleistocene.[1] As a neighbor of Tibet, Nepal is located at the southern piedmont of the Himalayas and is bordered by India and China. Similar to the Tibetans, the Nepalese are also famous for their high-altitude living conditions. In consistent with the language affiliation of the Nepalese populations (viz. Indo-European and Tibeto-Burman),[2–3] recent studies on the Nepalese have detected substantial genetic contribution from South Asians, East and West Eurasians,[4,5] it becomes evident that the genetic landscape of the Nepalese has been largely shaped by later immigrants from the neighboring regions.[4–6] For instance, South Asian genetic components (for example, Y-chromosome haplogroups H-M52*, H-M69*, H1-M82*, H1-M370*, R1a1-M198 and R2-M124;[4,6] mitochondrial DNA (mtDNA) haplogroups M31, M33, M35, M38, R6 and R30[5,7]) are prevalent in the Nepalese, reflecting extensive connections between Nepal and India; whereas the East Eurasian influence on the Nepalese is surprisingly substantial, as manifested by the presence of Y-chromosome haplogroup O3-M117[4,5] and mtDNA haplogroups B5, D4 and G.[5,7] As the Himalayas acts as a formidably geographical barrier between Nepal and East Asia (Tibet), how these East Eurasian components dispersed to Nepal becomes an issue of hot debate.

Recently, by comparing the Y-chromosome lineages between the Nepalese and the Tibetan populations, Gayden *et al.* have proposed that the East Eurasian genetic components had been introduced into Nepal from Tibet directly,[4,6] a scenario seemingly in agreement with the hypothesized retreat of Baric speakers.[8] However, considering the close vicinity between Nepal and northeast India in geography, a possibility that these genetic components of the East Eurasian ancestry might trace their origins to northeast India could not be ruled out completely. Indeed, abundant East Eurasian mtDNA lineages have already been detected in the northeast Indian populations,[9–12] raising

[1]State Key Laboratory of Genetic Resources and Evolution, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming, Yunnan Province, China; [2]Laboratory for Conservation and Utilization of Bio-Resources & Key Laboratory for Microbial Resources of the Ministry of Education, Yunnan University, Kunming, PR China; [3]Department of Parasitology, Medical Teaching Institute, Xinxiang Medical University, Xinxiang, China; [4]KIZ/CUHK Joint Laboratory of Bioresources and Molecular Research in Common Diseases, Kunming, China; [5]Cellular Immunology Laboratory, Department of Zoology, University of North Bengal Siliguri, Siliguri, India and [6]Little Angel's College Hattiban, Lalitpur, Nepal
Correspondence: Dr Q-P Kong or Dr Y-P Zhang, State Key Laboratory of Genetic Resources and Evolution, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming 650223, Yunnan Province, China.
E-mail: kongqp@yahoo.com.cn or zhangyp1@263.net.cn
Received 30 September 2011; revised 5 January 2012; accepted 8 January 2012; published online 22 March 2012

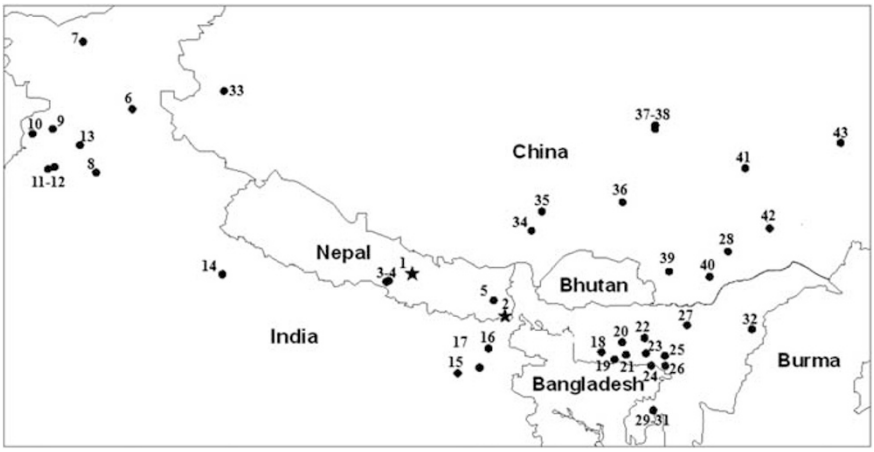

Figure 1 Sampling locations of the populations analyzed in this study. The Nepalese populations collected in this study are highlighted with solid pentacles, and the black dots represent the populations retrieved from the literature.

Table 1 Information of the 43 populations analyzed in this study

Population	*Code*	*Location*	*Country*	*Group*	*Sample size*	*Reference*
Nepal_Kat	1	Kathmandu	Nepal	Nepalese	200	This study
Nepal_Mix	2	Eastern Nepal	Nepal	Nepalese	46	This study
Tharu_CI	3	Central Terai	Nepal	Nepalese	57	Fornarino *et al.*[5]
Tharu_CII	4	Central Terai	Nepal	Nepalese	76	Fornarino *et al.*[5]
Tharu_E	5	Eastern Terai	Nepal	Nepalese	40	Fornarino *et al.*[5]
Kanet	6	Himachal	India	Northwest Indian	37	Metspalu *et al.*[11]
Kashmir	7	Kashmir	India	Northwest Indian	19	Kivisild *et al.*[14]
Lobana	8	Punjab	India	Northwest Indian	65	Kivisild *et al.*[14]
Brahmin	9	Punjab	India	Northwest Indian	26	Metspalu *et al.*[11]
Jat Sikh	10	Punjab	India	Northwest Indian	30	Metspalu *et al.*[11]
Kshatriya	11	Punjab	India	Northwest Indian	34	Metspalu *et al.*[11]
Scheduled caste	12	Punjab	India	Northwest Indian	20	Metspalu *et al.*[11]
Sikh	13	Punjab	India	Northwest Indian	40	Cordaux *et al.*[9]
Tharu_UP	14	Uttar Pradesh	India	North Indian	38	Metspalu *et al.*[11]; Kivisild *et al.*[14]
Kharia	15	Bihar	India	North Indian	39	Kumar *et al.*[17]; Banerjee *et al.*[16]
Bhumij	16	Bihar	India	North Indian	60	Kumar *et al.*[17]; Banerjee *et al.*[16]
Santhal	17	Bihar	India	North Indian	45	Kumar *et al.*[17]
Garo	18	Meghalaya	India	Northeast Indian	76	Reddy *et al.*[12]
Lyngnga	19	Meghalaya	India	Northeast Indian	74	Reddy *et al.*[12]
Nongtrai	20	Meghalaya	India	Northeast Indian	27	Reddy *et al.*[12]
Maram	21	Meghalaya	India	Northeast Indian	60	Reddy *et al.*[12]
Bhoi	22	Meghalaya	India	Northeast Indian	29	Reddy *et al.*[12]
Khynriam	23	Meghalaya	India	Northeast Indian	82	Reddy *et al.*[12]
War_Khas	24	Meghalaya	India	Northeast Indian	29	Reddy *et al.*[12]
Pnar	25	Meghalaya	India	Northeast Indian	51	Reddy *et al.*[12]
War_Jaint	26	Meghalaya	India	Northeast Indian	17	Reddy *et al.*[12]
Adi	27	Assam	India	Northeast Indian	45	Cordaux *et al.*[9]
Apa_1	28	Arunachal Pradesh	India	Northeast Indian	26	Cordaux *et al.*[9]
Apa_2	29	Tripura	India	Northeast Indian	21	Cordaux *et al.*[9]
Nishi	30	Tripura	India	Northeast Indian	44	Cordaux *et al.*[9]
Tipperah	31	Tripura	India	Northeast Indian	20	Roychoudhury *et al.*[15]
Naga	32	Nagaland	India	Northeast Indian	43	Cordaux *et al.*[9]
Tib_Ng	33	Ngari	China	Tibetan	46	Qin *et al.*[13]
Tib_R1	34	Shigatse	China	Tibetan	59	Qin *et al.*[13]
Tib_R2	35	Shigatse	China	Tibetan	220	Zhao *et al.*[1]
Tib_L	36	Lhasa	China	Tibetan	59	Qin *et al.*[13]
Tib_N1	37	Nagqu	China	Tibetan	168	Zhao *et al.*[1]
Tib_N2	38	Nagqu	China	Tibetan	58	Qin *et al.*[13]
Tib_S	39	Shannan	China	Tibetan	56	Qin *et al.*[13]
Monba	40	Nyingchi	China	Tibetan	51	Qin *et al.*[13]
Tib_Ny	41	Nyingchi	China	Tibetan	53	Qin *et al.*[13]
Lhoba	42	Shannan	China	Tibetan	20	Qin *et al.*[13]
Tib_C	43	Chamdo	China	Tibetan	61	Qin *et al.*[13]

the possibility that the East Eurasian influence could trace its origin back to the dispersal of Tibeto-Burman people who arrived at Nepal *via* northeast India.[5] Therefore, it is plausible that the bearer of the East Eurasian genetic components might have arrived at Nepal either from Tibet directly (across the Himalayas) or from northeast India instead. To test both scenarios and get deeper insights into the origin of the Nepalese, a simple way is to compare the phylogenetic affinity of the East Eurasian lineages, respectively, observed in Nepal, northern

Table 2 The genetic distance between populations from Nepal, China (Tibet), northern India based on the East Eurasian matrilineal components

	Nepalese	*northwest Indian*	*north Indian*	*northeast Indian*	*Tibetan*
Nepalese	0.00				
northwest Indian	0.03	0.00			
north Indian	0.12	0.17	0.00		
northeast Indian	0.04	0.05	0.12	0.00	
Tibetan	0.02	0.03	0.13	0.03	0.00

Table 3 Admixture analysis of the Nepalese by comparing with its potential parental populations

	Hybrid Nepalese population				
Parental group	*Kathmandu*	*eastern Nepal*	*Tharu_CI*	*Tharu_CII*	*Tharu_E*
Tibetan	0.58 ± 0.16	0.45 ± 0.17	0.87 ± 0.23	0.75 ± 0.19	0.12 ± 0.24
northeast Indian	0.24 ± 0.14	0.17 ± 0.14	0.00 ± 0.19	0.00 ± 0.16	0.12 ± 0.20
northwest Indian	0.19 ± 0.12	0.39 ± 0.13	0.00 ± 0.17	0.20 ± 0.14	0.56 ± 0.18
north Indian	0.01 ± 0.06	0.00 ± 0.07	0.13 ± 0.09	0.06 ± 0.07	0.19 ± 0.09

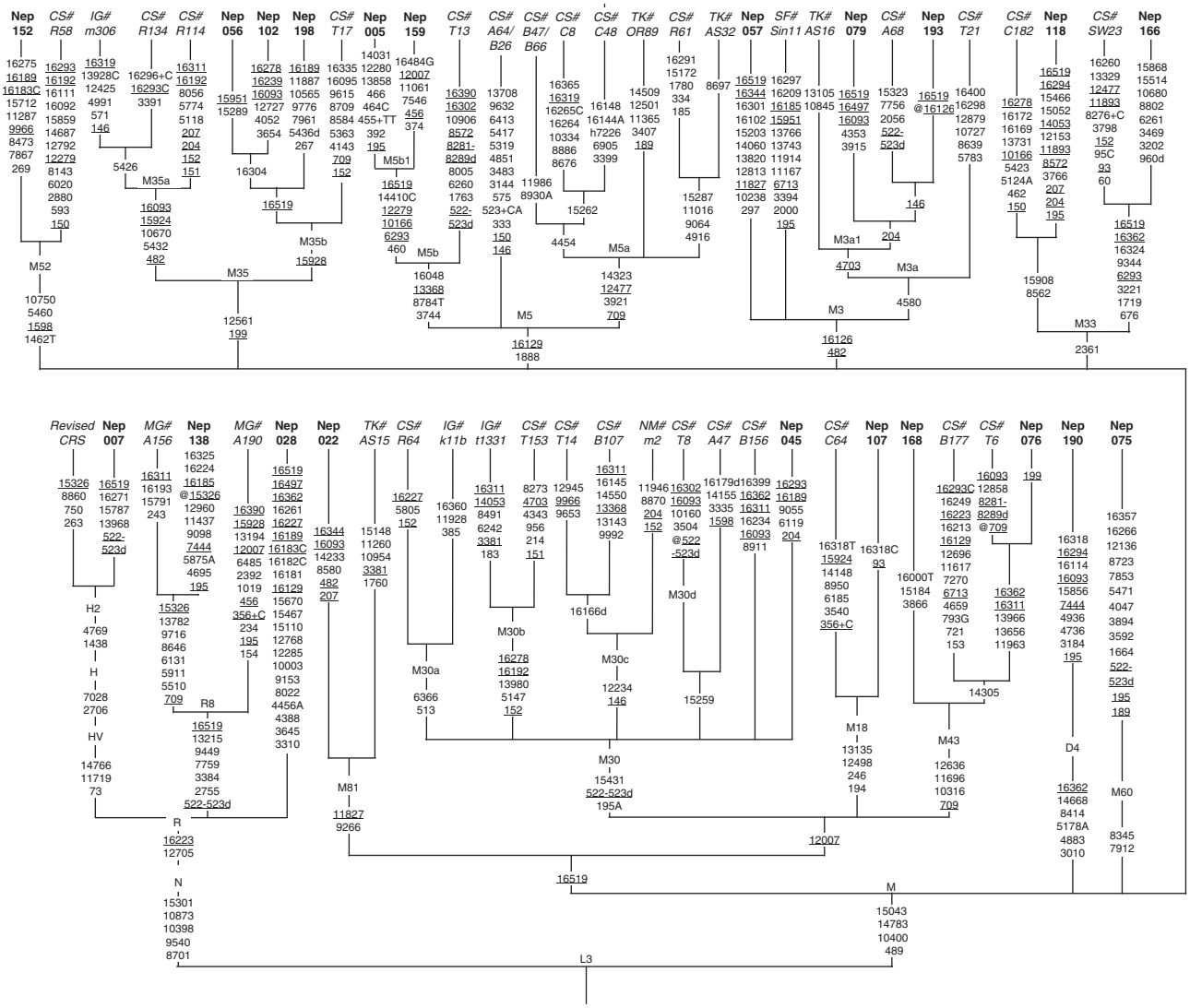

Figure 2 Reconstructed mtDNA tree of the completely sequenced representatives of the major Nepalese mtDNA lineages. The phylogenetic tree was reconstructed on the basis of 58 mitochondrial genomes, among which 21 mtDNAs were sampled from Katmandu, Nepal and generated in this study, whereas the rest 37 related mtDNAs were collected from the literature.[5,26,27,31–33] Mutations are recorded according to the rCRS.[19] Suffixes A, C, G, and T indicate transversions, "d" signifies a deletion and a plus (+) signs an insertion; recurrent mutations are underlined. The prefix "h" indicates heteroplasmy and "@" highlights back mutation. The length polymorphisms (for example, 309+C, 309+2C, 315+C and 315+2C) are ignored during the tree reconstruction. The reconstruction of highly recurrent mutations (for example, 16519, and the insertion/deletion of "CA" repeats in region 514–523) is tentative at best.

India (including northeast, north and northwest India) and Tibet at higher molecular resolution, and which becomes feasible with the recently released Tibetan mtDNA data.[1,13]

To achieve this objective, mtDNA variation within a number of 246 Nepalese individuals from Kathmandu and eastern Nepal were collected and studied in this study (Supplementary Table S1, Supplementary Material online), and the recently released mtDNA data from Nepal[5] and the neighboring regions, especially those from Tibet,[1,13] northeast and northwest India[9–12,14–17] were also considered. To further understand the Nepalese mtDNA landscape, a total of 21 representative individuals were selected for completely mtDNA genome sequencing in order to unambiguously determine their phylogenetic status (Figure 2).

MATERIALS AND METHODS

Sample collection

Blood samples of 246 unrelated individuals were collected from Kathmandu and eastern Nepal with informed consent (Figure 1 and Table 1), and this project was approved by the Ethics Committee at Kunming Institute of Zoology, Chinese Academy of Sciences. The geographical locations of the populations are displayed in Figure 1.

DNA amplification, sequencing and quality control

Total DNA was isolated with the standard phenol/chloroform method and stored at −80 °C. An mtDNA segment (spanning nucleotide position 16024–16569/1–576 and covering the whole-mtDNA control region) was amplified and sequenced as fully described in our previous works.[1,18] Mutations are recorded by comparing with the revised Cambridge reference sequence (rCRS).[19] All the individuals were allocated into specific haplogroup based on their control-region information; the assignments were further confirmed by typing additional diagnostic coding-region mutations according to the reconstructed phylogenetic trees of East Asian,[1,20–25] South Asian[5,7,12,26–33] and Southeast Asian[34–38] (Supplementary Table S1, Supplementary Material online). For the mtDNA sample of interest, entire genome was amplified and sequenced as described elsewhere.[1,18] To avoid any potential problems in mtDNA data quality, necessary quality control measures[21] and some caveats[39] were followed as described previously. The new mtDNA sequences reported in this study have been deposited in GenBank under accession numbers JF742217–JF742461 (for control-region sequences) and JF742196–JF742216 (for whole-mitochondrial genomes).

Data analysis

The principle component analysis (PCA) was conducted based on the East Eurasian haplogroup frequencies as described previously[40] (Supplementary Table S2, Supplementary Material online). Genetic distances (Table 2) were estimated by using the package Arlequin 3.11.[41] Admixture estimation was performed by the Weighted Least Squares (WLS) Method using the Statistical Package for the Social Sciences (SPSS) 14.0 software (Table 3).[42] The reduced median networks of haplogroups of interest were constructed by using the network 4.510 program (http://www.fluxus-engineering.com/sharenet.htm) and adjusted manually (Figures 4 and 5).[43] The ages of the lineages G2a2 (further defined by mutation 16193 and named G2a2 tentatively) and M9a1a2a (characterized by mutations 16145, 16316 and a back mutation at site 16362, and designated as M9a1a2a) (Table 4) were estimated by using the ρ statistic[44,45] with the suggested calibration rates.[44,46]

RESULTS AND DISCUSSION

On the basis of the combined information from control-region and partial coding-region segments, the majority (96.34%; 237/246) of the Nepalese mtDNAs could unambiguously be allocated into the defined haplogroups of East Eurasian (36.59%; 90/246),[1,20–25] South Asian (51.63%; 127/246)[5,7,11,12,26–31] and West Eurasian ancestries (8.13%; 20/246)[26,31,47,48] (Supplementary Table S2, Supplementary Material online). It is apparent that the genetic components of East Eurasian (36.59%) and South Asian (51.63%) ancestry have comprised the vast majority of the Nepalese gene pool (Supplementary Table S1, Supplementary Material online),[5] and this pattern remains almost stable for both the East Eurasian (45.11%; 189/419) and South Asian (47.49%; 199/419) components after taking into account the recently reported Nepalese mtDNA data.[5] As for the 21 samples with ambiguously phylogenetic status, completely sequencing their mtDNA genomes revealed that virtually all of these samples in fact belong to the already defined haplogroups, such as M3, M5, M18, M30, M35, M43, D4, R8 and M60. Of note is that, beside two singular branches identified in this study, we also defined a novel haplogroup characterized by variations 9266 and 11827, which was named M81 here (Figure 2).

To get more insights into the origin of the East Eurasian maternal components observed in the Nepalese and therefore test the two competing scenarios about how these components had been introduced into Nepal,[4,5,8] we focused on the phylogenetic affinity between the East Eurasian haplogroups identified in the Nepalese and those from the Tibetan, northeast and northwest Indian populations. Figure 3 illustrates the principle component analysis plot of the 43 populations under study, which was constructed based merely on the East Eurasian lineages. Among the five Nepalese populations under study, three clustered with the Tibetans (Figure 3). After we considered all the Nepalese regional populations as a whole and calculated its

Table 4 Estimated ages of haplogroups G2a2 and M9a1a2a based on the calibration rates proposed in Forster *et al.*[44] and Soares *et al.*[46]

		Soares et al.'s rate[46]		*Forster et al.'s rate*[44]	
Haplogroup	N	ρ ± σ	T (ky)	ρ ± σ	T (ky)
G2a2	23	0.30 ± 0.20	5.74 (1.98–9.49)	0.30 ± 0.20	6.14 (2.12–10.16)
M9a1a2a	20	0.30 ± 0.19	5.65 (2.13–9.18)	0.30 ± 0.19	6.05 (2.28–9.83)

Abbreviation: ky, kilo years.
The ρ and σ values were obtained by considering the substitutions in segment 16090–16365 as fully described previously.[44,46]

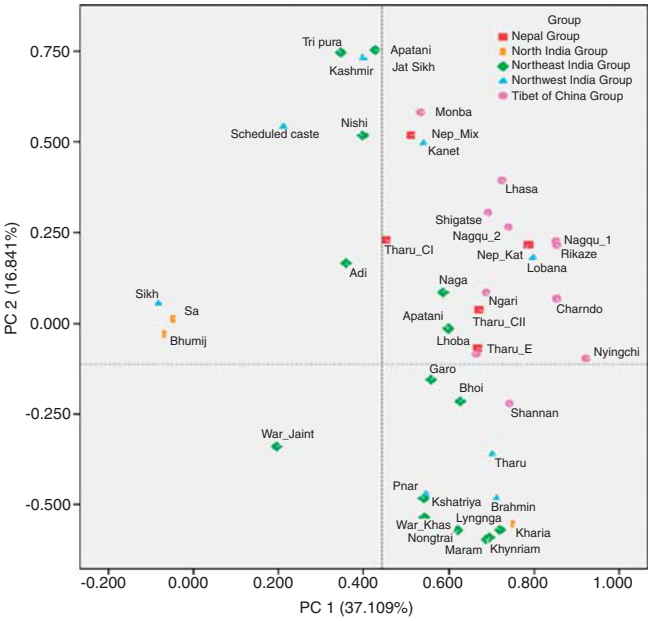

Figure 3 Principle component analysis (PCA) of the populations under study.

Fst value with the populations from its neighboring regions, the smallest genetic distance was observed between the Nepalese and the Tibetans (Table 2). By taking the Tibetans and northern Indians as the parental populations, the results of the admixture estimation analysis revealed that the Tibetans made major contribution to virtually all Nepalese populations (except for the eastern Tharu population; Table 3). Afterwards, we further compared the phylogenetic affinity of the East Eurasian lineages observed in Nepalese (including haplogroups A11, C, G2a, M9a, F1c and Z; Figures 4 and 5) with those from the neighboring regions, for example, Tibet, northeast and northwest India, by means of median networks.[43] On the basis of the constructed networks (Figures 4 and 5), several features could be observed: (1) the Nepalese share some basal or internal haplotypes with the Tibetans; (2) the Nepalese harbor a number of unique haplotypes at the terminal level, most of which branched off directly from the nodes occupied almost exclusively by the Tibetan lineages and (3) only a few haplotypes are shared sporadically between the Nepalese and the northern Indians. Taken together, the Nepalese lineages of East Eurasian ancestry generally show much closer affinity with the ones from Tibet, albeit a few mtDNA haplotypes, likely resulted from recent gene flow, were shared between the Nepalese and northern (including northeast) Indians (Figures 4 and 5).

Even though we focused on the East Eurasian lineages identified in the Nepalese populations, we did observe a number of the Nepalese-specific haplotypes, strongly suggesting their rather ancient origin and most plausibly *de novo* differentiation in Nepal. To get some hint at the arrival time of the lineages, we have focused on two clades from haplogroups G2a and M9a1a2 simply because both clades contain the Nepalese haplotypes at their terminal branch or basal node and likely have differentiated in Nepal; estimating their ages would then help to

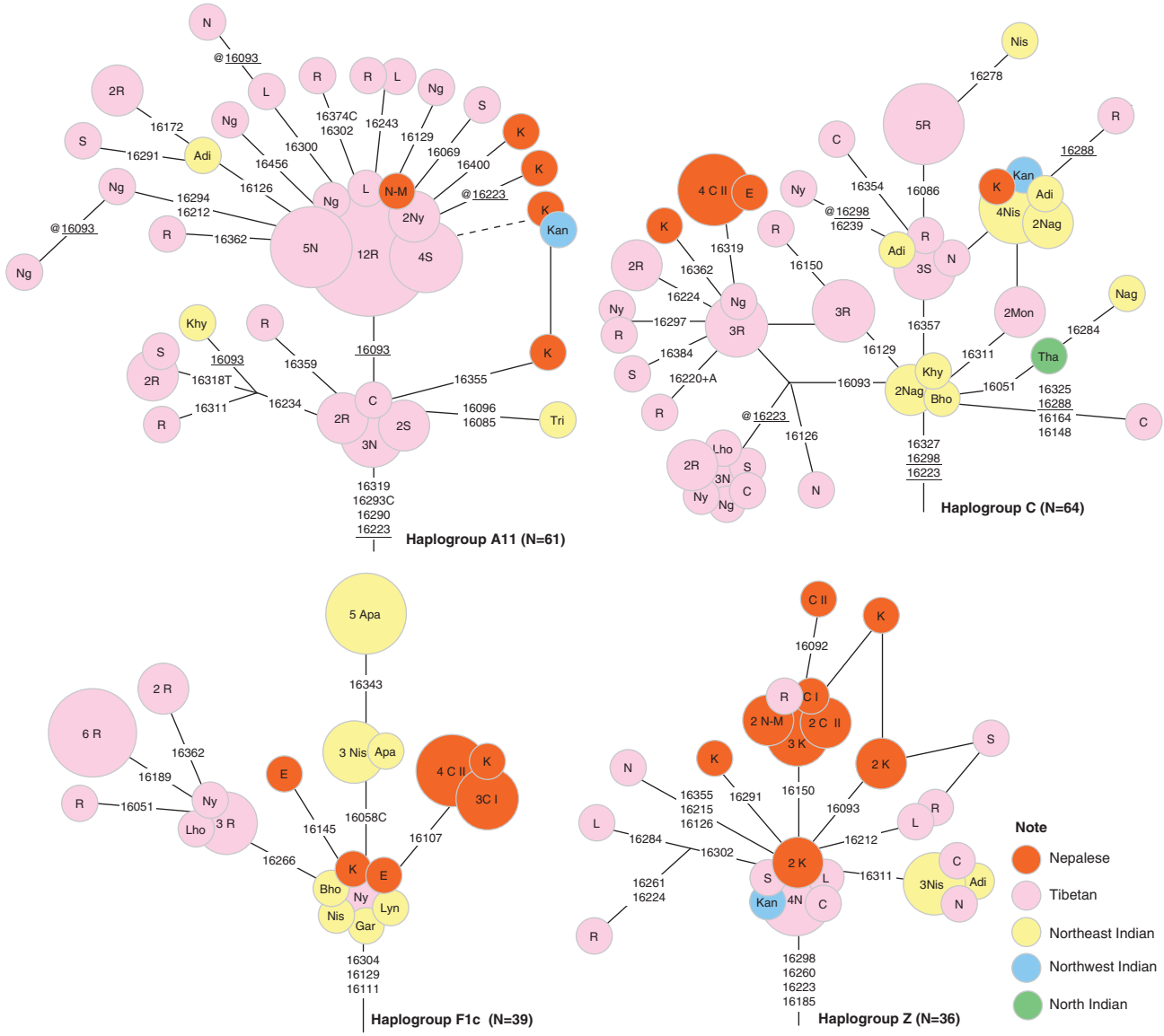

Figure 4 Median networks of haplogroups A11, C, F1c and Z. The reduced median networks of haplogroups of interest were constructed by using the network 4.510 program (http://www.fluxus-engineering.com/sharenet.htm) and adjusted manually according to Bandelt *et al.*[43] The data used here were collected from this study (Supplementary Table S1) and the literature (Table 1). The sequence variation used for network construction was confined to segment 16047–16497. Suffixes T, C and G refer to transversions; recurrent mutations are underlined and "@" denotes a reverse mutation, 16193+C was omitted in the median network. Code in the circles refers to the population abbreviation as displayed in Supplementary Table S2.

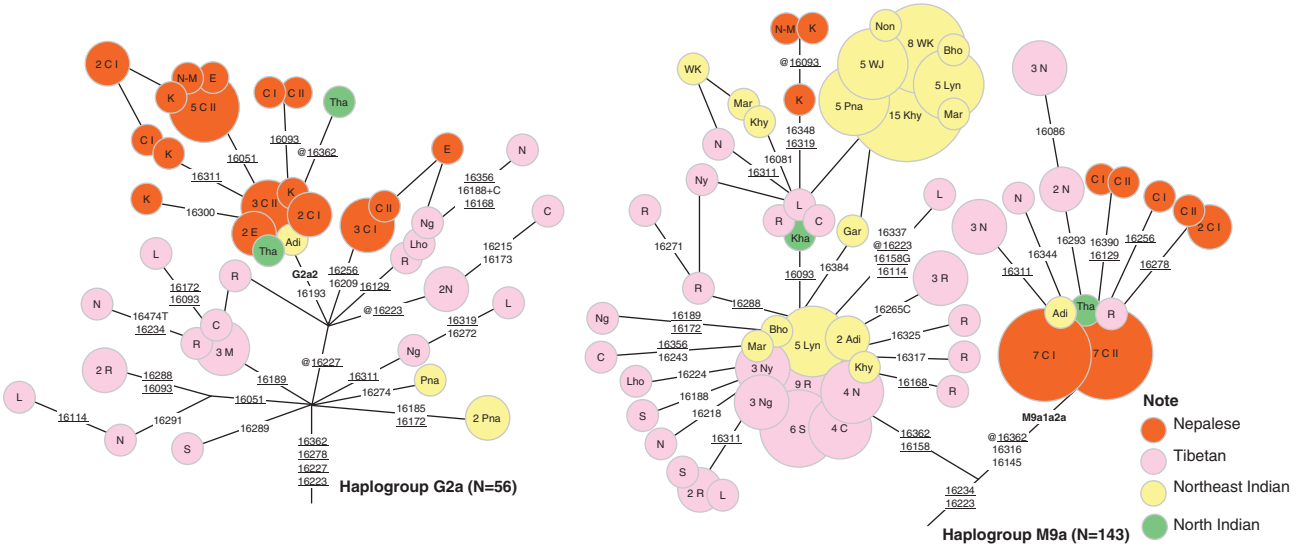

Figure 5 Median networks of haplogroups G2a2 and M9a1a2a. For more information, see Figure 4.

date the arrival time of the migration from Tibet. In fact, time estimation results revealed that haplogroups G2a2 and M9a1a2a have very similar ages of ~5.7 kya, and this age becomes a little older (~6 kya) when calibration rate proposed by Forster *et al.*[44] was used. To this end, the very similar ages of both haplogroups, which likely had *in situ* differentiated in Nepal, strongly suggest that the bearers of these East Eurasian maternal components would have arrived at Nepal no later than 5.7 kya (Table 4). In retrospect, previous work has suggested that the maternal genetic components from the northern East Eurasian was introduced into Tibet around 8.2 kya,[1] and our time estimation results fit this dating frame very well. It is then conceivable that the settlement of Nepal by the bearer of the East Eurasian genetic components occurred likely before 5.7 kya, a result in good agreement with the archeological findings reporting shared the Neolithic features between Nepal and Tibet (references therein).[49]

Previous studies have observed substantial East Eurasian genetic components in the Nepalese populations;[4,5] however, it remains controversial whether the East Eurasian lineages have been introduced into Nepal from Tibet directly (across the Himalayas)[4,6] or *via* northeast India.[5,8,50] By extensively analyzing the mtDNA variation in Nepal, Tibet, northern India populations, our observations, based on the principle component analysis, Fst and admixture estimation, revealed the closer genetic affinity between the Nepalese and the Tibetans, and this result was further substantiated by the median networks, (Figures 4 and 5) in which most of the Nepalese mtDNAs prevalent among northern Asian populations shared the haplotypes with the Tibetans at root level or branched off directly from the nodes consisting almost exclusively of the Tibetan lineages. Our results strongly suggest that most of the East Eurasian maternal components identified in the Nepalese were introduced directly from Tibet,[4,6] and the time estimation results further date that this peopling scenario plausibly occurred about 6 kya. Indeed, this inference seems to be in striking accordance with the historically recorded passes (such as the Kodari and Rasuwa Passes), which bridged the Nepalese and the Tibetans since the ancient time.[3] However, the observed gene flow from northeast India suggests genetic contribution, albeit limited, from this region, a scenario echoing the proposed inland dispersal route.[50] In this spirit, our findings complete the understanding of the origin of the Nepalese and the way how the East Eurasian genetic components had been introduced into Nepal. Taking into account the previous observation on Y chromosome,[4] now it is convincing that the East Eurasian had entered Nepal across the Himalayas around 6 kya, a scenario in good agreement with the previous findings from linguistics and archeology.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGEMENTS

We thank Wen-Zhi Wang, Chun-Ling Zhu, Yang Yang, Shi-Kang Gou and Tao Sha for their technical assistant. We are grateful to the volunteers for providing the blood samples to the project. This work was supported by grants from the Natural Science Foundation of Yunnan Province (No. 2011FB106) and the National Natural Science Foundation of China (NSFC, Nos. 30900797 and 30621092).

1 Zhao, M., Kong, Q. P., Wang, H. W., Peng, M. S., Xie, X. D., Wang, W. Z. *et al.* Mitochondrial genome evidence reveals successful Late Paleolithic settlement on the Tibetan Plateau. *Proc. Natl. Acad. Sci. USA* **106,** 21230–21235 (2009).
2 Wang, H. W. *Nepal* (Social Sciences Documentation Publishing House, Beijing, China, 2004).
3 Xue, K. Q., Zhao, C. Q., Sun, S. H., Ge, W. J., Liu, Y., Xing, G. C. *et al. Concise Encyclopaedia Of South Asia and Central Asia* (China Social Science Press, Beijing, China, 2004).
4 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,** 884–894 (2007).
5 Fornarino, S., Pala, M., Battaglia, V., Maranta, R., Achilli, A., Modiano, G. *et al.* Mitochondrial and Y-chromosome diversity of the Tharus (Nepal): a reservoir of genetic variation. *BMC Evol. Biol.* **9,** 154 (2009).
6 Gayden, T., Mirabal, S., Cadenas, A. M., Lacau, H., Simms, T. M., Morlote, D. *et al.* Genetic insights into the origins of Tibeto-Burman populations in the Himalayas. *J. Hum. Genet.* **54,** 216–223 (2009).
7 Thangaraj, K., Chaubey, G., Kivisild, T., Rani, D. S., Singh, V. K., Ismail, T. *et al.* Maternal footprints of Southeast Asians in North India. *Hum. Hered.* **66,** 1–9 (2008).
8 Su, B., Xiao, C. J., Deka, R., Seielstad, M. T., Kangwanpong, D., Xiao, J. H. *et al.* Y chromosome haplotypes reveal prehistorical migrations to the Himalayas. *Hum. Genet.* **107,** 582–590 (2000).
9 Cordaux, R., Saha, N., Bentley, G. R., Aunger, R., Sirajuddin, S. M. & Stoneking, M. Mitochondrial DNA analysis reveals diverse histories of tribal populations from India. *Eur. J. Hum. Genet.* **11,** 253–264 (2003).
10 Cordaux, R., Weiss, G., Saha, N. & Stoneking, M. The Northeast Indian passageway: a barrier or corridor for human migrations? *Mol. Biol. Evol.* **21,** 1525–1533 (2004).

11 Metspalu, M., Kivisild, T., Metspalu, E., Parik, J., Hudjashov, G., Kaldma, K. *et al.* Most of the extant mtDNA boundaries in south and southwest Asia were likely shaped during the initial settlement of Eurasia by anatomically modern humans. *BMC Genet.* **5,** 26 (erratum 26:41) (2004).
12 Reddy, B. M., Langstieh, B. T., Kumar, V., Nagaraja, T., Reddy, A. N. S., Meka, A. *et al.* Austro-Asiatic tribes of Northeast India provide hitherto missing genetic link between South and Southeast Asia. *PLoS ONE* **2,** e1141 (2007).
13 Qin, Z. D., Yang, Y. J., Kang, L. L., Yan, S., Cho, K., Cai, X. Y. *et al.* A mitochondrial revelation of early human migrations to the Tibetan Plateau before and after the last glacial maximum. *Am. J. Phys. Anthropol.* **143,** 555–569 (2010).
14 Kivisild, T., Bamshad, M. J., Kaldma, K., Metspalu, M., Metspalu, E., Reidla, M. *et al.* Deep common ancestry of Indian and western-Eurasian mitochondrial DNA lineages. *Curr. Biol.* **9,** 1331–1334 (1999).
15 Roychoudhury, S., Roy, S., Basu, A., Banerjee, R., Vishwanathan, H., Usha Rani, M. V. *et al.* Genomic structures and population histories of linguistically distinct tribal groups of India. *Hum. Genet.* **109,** 339–350 (2001).
16 Banerjee, J., Trivedi, R. & Kashyap, V. K. Mitochondrial DNA control region sequence polymorphism in four indigenous tribes of Chotanagpur plateau, India. *Forensic Sci. Int. Genet.* **149,** 271–274 (2005).
17 Kumar, V., Langstieh, B. T., Madhavi, K. V., Naidu, V. M., Singh, H. P., Biswas, S. *et al.* Global patterns in human mitochondrial DNA and Y-chromosome variation caused by spatial instability of the local cultural processes. *PLoS Genet.* **2,** e53 (2006).
18 Wang, H. W., Jia, X. Y., Ji, Y. L., Kong, Q. P., Zhang, Q. J., Yao, Y. G. *et al.* Strikingly different penetrance of LHON in two Chinese families with primary mutation G11778A is independent of mtDNA haplogroup background and secondary mutation G13708A. *Mutat. Res.* **643,** 48–53 (2008).
19 Andrews, R. M., Kubacka, I., Chinnery, P. F., Lightowlers, R. N., Turnbull, D. M. & Howell, N. Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. *Nat. Genet.* **23,** 147 (1999).
20 Kivisild, T., Tolk, H. V., Parik, J., Wang, Y. M., Papiha, S. S., Bandelt, H. J. *et al.* The emerging limbs and twigs of the East Asian mtDNA tree. *Mol. Biol. Evol.* **19,** 1737–1751 (2002).
21 Kong, Q. P., Yao, Y. G., Sun, C., Bandelt, H. J., Zhu, C. L. & Zhang, Y. P. Phylogeny of east Asian mitochondrial DNA lineages inferred from complete sequences. *Am. J. Hum. Genet.* **73,** 671–676 (2003).
22 Tanaka, M., Cabrera, V. M., González, 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).
23 Kong, Q. P., Bandelt, H. J., Sun, C., Yao, Y. G., Salas, A., Achilli, A. *et al.* Updating the East Asian mtDNA phylogeny: a prerequisite for the identification of pathogenic mutations. *Hum. Mol. Genet.* **15,** 2076–2086 (2006).
24 Derenko, M., Malyarchuk, B., Grzybowski, T., Denisova, G., Dambueva, I., Perkova, M. *et al.* Phylogeographic analysis of mitochondrial DNA in northern Asian populations. *Am. J. Hum. Genet.* **81,** 1025–1041 (2007).
25 Kong, Q. P., Sun, C., Wang, H. W., Zhao, M., Wang, W. Z., Zhong, L. *et al.* Large-scale mtDNA screening reveals a surprising matrilineal complexity in East Asia and its implications to the peopling of the region. *Mol. Biol. Evol.* **28,** 513–522 (2011).
26 Palanichamy, M. G., Sun, C., Agrawal, S., Bandelt, H. J., Kong, Q. P., Khan, F. *et al.* Phylogeny of mitochondrial DNA macrohaplogroup N in India, based on complete sequencing: implications for the peopling of South Asia. *Am. J. Hum. Genet.* **75,** 966–978 (2004).
27 Sun, C., Kong, Q. P., Palanichamy, M. G., Agrawal, S., Bandelt, H. J., Yao, Y. G. *et al.* The dazzling array of basal branches in the mtDNA macrohaplogroup M from India as inferred from complete genomes. *Mol. Biol. Evol.* **23,** 683–690 (2006).
28 Thangaraj, K., Chaubey, G., Singh, V. K., Vanniarajan, A., Thanseem, I., Reddy, A. G. *et al. In situ* origin of deep rooting lineages of mitochondrial Macrohaplogroup 'M' in India. *BMC Genomics* **7,** 151–156 (2006).
29 Chandrasekar, A., Kumar, S., Sreenath, J., Sarkar, B. N., Urade, B. P., Mallick, S. *et al.* Updating phylogeny of mitochondrial DNA macrohaplogroup m in India: dispersal of modern human in South Asian corridor. *PLoS ONE* **4,** e7447 (2009).
30 Thangaraj, K., Chaubey, G., Kivisild, T., Reddy, A., Singh, V. K., Rasalkar, A. A. *et al.* Reconstructing the origin of Andaman Islanders. *Science* **308,** 996 (2005).
31 Kivisild, T., Shen, P. D., Wall, D., Do, B., Sung, R., Davis, K. *et al.* The role of selection in the evolution of human mitochondrial genomes. *Genetics* **172,** 373–387 (2006).
32 Maca-Meyer, N., González, A. M., Larruga, J. M., Flores, C. & Cabrera, V. M. Major genomic mitochondrial lineages delineate early human expansions. *BMC Genet.* **2,** 13 (2001).
33 Ingman, M. & Gyllensten, U. Mitochondrial genome variation and evolutionary history of Australian and New Guinean aborigines. *Genome Res.* **13,** 1600–1606 (2003).
34 Macaulay, V., Hill, C., Achilli, A., Rengo, C., Clarke, D., Meehan, W. *et al.* Single, rapid coastal settlement of Asia revealed by analysis of complete mitochondrial genomes. *Science* **308,** 1034–1036 (2005).
35 Hill, C., Soares, P., Mormina, M., Macaulay, V., Meehan, W., Blackburn, J. *et al.* Phylogeography and ethnogenesis of aboriginal Southeast Asians. *Mol. Biol. Evol.* **23,** 2480–2491 (2006).
36 Hill, C., Soares, P., Mormina, M., Macaulay, V., Clarke, D., Blumbach, P. B. *et al.* A mitochondrial stratigraphy for island southeast Asia. *Am. J. Hum. Genet.* **80,** 29–43 (2007).
37 Peng, M. S., Quang, H. H., Dang, K. P., Trieu, A. V., Wang, H. W., Yao, Y. G. *et al.* Tracing the Austronesian Footprint in Mainland Southeast Asia: A Perspective from Mitochondrial DNA. *Mol. Biol. Evol.* **27,** 2417–2430 (2010).
38 Tabbada, K. A., Trejaut, J., Loo, J. H., Chen, Y. M., Lin, M., Mirazón-Lahr, M. *et al.* Philippine mitochondrial DNA diversity: a populated viaduct between Taiwan and Indonesia? *Mol. Biol. Evol.* **27,** 21–31 (2010).
39 Kong, Q. P., Salas, A., Sun, C., Fuku, N., Tanaka, M., Zhong, L. *et al.* Distilling artificial recombinants from large sets of complete mtDNA genomes. *PLoS ONE* **3,** e3016 (2008).
40 Yao, Y. G., Kong, Q. P., Bandelt, H. J., Kivisild, T. & Zhang, Y. P. Phylogeographic differentiation of mitochondrial DNA in Han Chinese. *Am. J. Hum. Genet.* **70,** 635–651 (2002).
41 Excoffier, L., Laval, G. & Schneider, S. Arlequin (version 3.0): an integrated software package for population genetics data analysis. *Evol. Bioinform. Online* **1,** 47–50 (2005).
42 Long, J. C., Williams, R. C., McAuley, J. E., Medis, R., Partel, R., Tregellas, W. M. *et al.* Genetic variation in Arizona Mexican Americans: estimation and interpretation of admixture proportions. *Am. J. Phys. Anthropol.* **84,** 141–157 (1991).
43 Bandelt, H. J., Macaulay, V. & Richards, M. Median networks: speedy construction and greedy reduction, one simulation, and two case studies from human mtDNA. *Mol. Phylogenet. Evol.* **16,** 8–28 (2000).
44 Forster, P., Harding, R., Torroni, A. & Bandelt, H. J. Origin and evolution of Native American mtDNA variation: a reappraisal. *Am. J. Hum. Genet.* **59,** 935–945 (1996).
45 Saillard, J., Forster, P., Lynnerup, N., Bandelt, H. J. & Nørby, S. mtDNA variation among Greenland Eskimos: the edge of the Beringian expansion. *Am. J. Hum. Genet.* **67,** 718–726 (2000).
46 Soares, P., Ermini, L., Thomson, N., Mormina, M., Rito, T., Röhl, A. *et al.* Correcting for purifying selection: an improved human mitochondrial molecular clock. *Am. J. Hum. Genet.* **84,** 740–759 (2009).
47 Finnilä, S., Lehtonen, M. S. & Majamaa, K. Phylogenetic network for European mtDNA. *Am. J. Hum. Genet.* **68,** 1475–1484 (2001).
48 Herrnstadt, C., Elson, J. L., Fahy, E., Preston, G., Turnbull, D. M., Anderson, C. *et al.* Reduced-median-network analysis of complete mitochondrial DNA coding-region sequences for the major African, Asian, and European haplogroups. *Am. J. Hum. Genet.* **70,** 1152–1171 (2002).
49 Sharmai, D. R. Archaeological Remains of the Dang Valley. *Ancient Nepal.* **88,** 8–15 (1988).
50 Peng, M. S., Palanichamy, M. G., Yao, Y. G., Mitra, B., Cheng, Y. T., Zhao, M. *et al.* Inland post-glacial dispersal in East Asia revealed by mitochondrial haplogroup M9a'b. *BMC Biol.* **9,** 2 (2011).

Supplementary Information accompanies the paper on Journal of Human Genetics website (http://www.nature.com/jhg)