

ORIGINAL ARTICLE

Northward genetic penetration across the Himalayas viewed from Sherpa peopleLongli Kang¹, Chuan-Chao Wang², Feng Chen¹, Dali Yao², Li Jin², and Hui Li^{1,2}¹Key Laboratory of High Altitude Environment and Gene Related to Disease of Tibet, Ministry of Education, Tibet University for Nationalities, Xianyang, Shaanxi, China and ²Ministry of Education Key Laboratory of Contemporary Anthropology, School of Life Sciences, Fudan University, Shanghai, China**Abstract**

The Himalayas have been suggested as a natural barrier for human migrations, especially the northward dispersals from the Indian Subcontinent to Tibetan Plateau. However, although the majority of Sherpa have a Tibeto-Burman origin, considerable genetic components from Indian Subcontinent have been observed in Sherpa people living in Tibet. The western Y chromosomal haplogroups R1a1a-M17, J-M304, and F*-M89 comprise almost 17% of Sherpa paternal gene pool. In the maternal side, M5c2, M21d, and U from the west also count up to 8% of Sherpa people. Those lineages with South Asian origin indicate that the Himalayas have been permeable to bidirectional gene flow.

Keywords

Himalayas, mitochondrial DNA, Sherpa, Y chromosome

History

Received 12 December 2013

Accepted 16 February 2014

Published online 11 March 2014

Introduction

The Himalayan mountain range is home to most of the highest peaks of the world and forms a natural barrier separating the Indian subcontinent from the Tibetan Plateau. The Himalayas have a profound effect on the climate of the surrounding areas. They prevent frigid, dry Siberian winds blowing south into the subcontinent, and keep the warm and humid monsoons from traveling northwards into the plateau, which has created the fertile plains in the Indian subcontinent and left the cold and barren snowfield on the Tibetan Plateau.

The completely different terrain and climate have played a vital role in shaping the human migration pattern in this vast region. Linguistically, the Himalayas are thought to serve as the border between the Tibeto-Burman and Indo-European language families (Van Driem, 2001). Some genetic studies have also raised the hypothesis that the Himalayas have formed a natural barrier for gene flow (Cordaux et al., 2004; Metspalu et al., 2004; Qin et al., 2010; Qi et al., 2013; Xue et al., 2006). For instance, Qi et al. (2013) reported that the western and southern Eurasian genetic components only comprise 0.13–2.04% (Y chromosome) and 0.03–0.65% (mtDNA) of 41 geographic populations across the Tibetan Plateau, respectively. It is most interesting that population dispersals across the Himalayas from Indian subcontinent to Tibet Plateau have been very limited, whereas migrations in the opposite direction have occurred periodically (Gayden et al., 2007, 2009, 2013). As in Gayden's report, all three populations from Nepal, Newar, Tamang and the Nepalese in Kathmandu, have the high frequencies of East Asian specific paternal Y chromosome haplogroup O3a2c1*-M134 and its sublineages (21.2%, 86.6% and 19.5%, respectively) (Gayden et al., 2007).

High altitudes, cold climates, and thin air on the Tibetan side have been suggested as the reasons for the observed unidirectional human dispersal pattern (Gayden et al., 2007). Therefore, the question becomes whether we can detect the possible gene flow northward across Himalayas into Tibet Plateau in a population that can physiologically overcome the above difficulties. Sherpa is probably such an ideal population to address this dispute. Sherpa is a Tibeto-Burman population living in the most mountainous region of Tibet of China, Nepal, Bhutan and India. Sherpa people have been called the “workers on Mount Everest” for their great physical attributes in mountaineering and experience at high altitudes. Actually, a lot of papers have discussed the possible genetic mechanisms for Sherpa's adaptation to high altitude (Arai et al., 2002; Droma et al., 2006; Hanaoka et al., 2012; Malacrida et al., 2007; Wu & Kayser, 2006). In this paper, we mainly focus on the origin and migration pattern of Sherpa people. According to the historical literature, Sherpa migrated from the Kham region in eastern Tibet and western Sichuan to the southern foot of the Himalayas (Oppitz, 1968). However, some folktales suggest the Sherpa as descendants of the Tangut Kingdom (1038 to 1227 AD) who fled their homeland in Muyag district to escape Mongol invasion (Gong-Bo, 2011). Here, we use informative Y chromosome and mtDNA markers to give a clue about the northward gene flow across the Himalayas and shed light on the origin of the Sherpa.

Methods**Population samples**

Saliva samples of 183 Sherpa individuals, including 84 males and 99 females, from Zhangmu Town, Shigatse Prefecture, were collected and analyzed in this study. All individuals were adequately informed and gave their informed contents before their participation, with approval from the Ethics Committee of Fudan School of Life Sciences (Fudan University, Shanghai, China).

Correspondence: Hui Li, MOE Key Laboratory of Contemporary Anthropology, School of Life Sciences, Fudan University, Shanghai, China. E-mail: LiHui.Fudan@gmail.com

To obtain a more comprehensive picture of the genetic affiliation of Sherpa to the populations from East Asia and South Asia, Y-chromosome (Cai et al., 2011; Deng et al., 2013; Fornarino et al., 2009; Gan et al., 2008; Gayden et al., 2007; Hammer et al., 2006; He et al., 2012; Kang et al., 2012; Li et al., 2008; Li D et al., 2008; Li D et al., 2013a; Lu et al., 2013; Nonaka et al., 2007; Park et al., 2012; Qi et al., 2013; Sengupta et al., 2006; Shi et al., 2005, 2008; Shi & Su, 2009; Tan et al., 2007; Thanseem et al., 2006; Xue et al., 2006) and mtDNA data (Black et al., 2006; Cordaux et al., 2003; Dancause et al., 2009; Fornarino et al., 2009; Fucharoen et al., 2001; Hill et al., 2006, 2007; Irwin et al., 2008; Jin et al., 2009; Kong et al., 2011; Lertrit et al., 2008; Li et al., 2007; Li D et al., 2013b; Maruyama et al., 2010; Mona et al., 2009; Oota et al., 2002; Peng et al., 2010, 2011; Qi et al., 2013; Tabbada et al., 2010; Tajima et al., 2004; Thanseem et al., 2006; Trivedi et al., 2006; Tsai et al., 2001; Wang et al., 2012; Wen et al., 2004a, 2004b, 2005; Zhao et al., 2009; Zimmermann et al., 2009) were compiled from the literature. The comparative groups included populations speaking Sino-Tibetan, Altaic, Tai-Kadai, Hmong-Mien, Austro-Asiatic, Dravidian and Indo-European languages. Our unpublished data of Muyag, Queyu, Jiarong, Qiang, Horpa and Tibetan from Western Sichuan, and Tibetan from Qinghai, Nagri, Lhasa, Nagqu, Shannan, Nyingchi, Qamdo, and Shigatse were also included in the analyses, although the original data were omitted.

Y chromosome markers

The samples were typed through 75 single nucleotide polymorphisms (SNPs) as listed in the latest Y chromosome phylogenetic tree (Karafet et al., 2008; Yan et al., 2011). The panels were organized as follows: Panel 1 (within Haplogroup O), M175, M119, P203, M110, M268, P31, M95, M176, M122, M324, M121, P201, M7, M134, M117, 002611, P164, L127, KL1; Panel 2 (non-Haplogroup O), M130, P256, M1, M231, M168, M174, M45, M89, M272, M258, M242, M207, M9, M96, P125, M304, M201, M306; Panel 3 (Haplogroup C), M217; Panel 4 (Haplogroup D), P47, N1, P99, M15, M125, M55, M64.1, M116.1, M151, N2, 022457; Panel 5 (Haplogroup N), M214, LLY22g, M128, M46/Tat, P63, P119, P105, P43, M178; Panel 6 (Haplogroup R), M306, M173, M124, M420, SRY10831.2, M17, M64.2, M198, M343, V88, M458, M73, M434, P312, M269, U106/M405; Panel 7 (Haplogroup Q), P36.2.

Those binary markers were hierarchically genotyped by SNaPshot (SNaPshot Multiplex Kit; Applied Biosystems, Carlsbad, CA) and fluorescent allele-specific polymerase chain reaction (PCR). The PCR products were also electrophoresed on a 3730xl Genetic Analyzer (Applied Biosystems, Foster City, CA). Seventeen Y chromosome short tandem repeats (STRs) (DYS19, DYS389I, DYS389II, DYS390, DYS391, DYS392, DYS393, DYS385a, DYS385b, DYS438, DYS439, DYS437, DYS448, DYS456, DYS458, DYS635 and YGATAH4) were amplified using the AmpFISTR Yfiler PCR Amplification kit (Applied Biosystems). Amplified products were separated and detected using the ABI 3730xl Genetic Analyzer (Applied Biosystems) according to the manufacturer's recommended protocol. The data were analyzed using GeneMapper ID version 3.2 (Applied Biosystems, Foster City, CA).

Mitochondrial DNA markers

The HVS-I of the control region was amplified by primers L15974 and H16488 (Yao et al., 2002). Purified polymerase chain reaction products were sequenced using the BigDye terminator cycle sequencing kit and an ABI 3730xl genetic analyzer (Applied Biosystems, Foster City, CA). A SNaPshot assay was used for typing single nucleotide polymorphisms (SNPs) in the coding

regions to confirm haplogroup identity. This assay was designed as a multiplex panel including 21 coding region SNPs and one length variation marker (Qin et al., 2010). Both the HVS-I motif and the coding region variations were used to infer haplogroups according to the classification of Kivisild et al. (2002).

Statistical analyses

Principal component analysis (PCA) was performed using SPSS 18.0 software (SPSS, Chicago, IL). Arlequin 3.5.1.2 was used to calculate the Y-STR Slatkin's genetic distances (Rst) (Excoffier & Lischer, 2010). Neighbor-joining (NJ) unrooted trees based on Rst statistics were carried out using MEGA 5.1 (Tamura et al., 2011). A Markov Chain Monte Carlo analysis of haplogroup structure was carried out using the program Structure 2.3.4 (Pritchard et al., 2000). Networks of mtDNA HVS-I motifs were constructed by the median-joining method (Bandelt et al., 1999) using Network version 4.6 (www.fluxus-engineering.com).

Results

Y chromosomes

According to the nomenclature of Y Chromosome Consortium (YCC) (Karafet et al., 2008; Yan et al., 2011), nine SNP haplogroups were determined from the 84 male individual samples (Figure 1a and Table S1). Haplogroup D1-M15, which is supposed to be the Paleolithic genetic legacy with a wide distribution among most Tibeto-Burman, Tai-Kadai, and Hmong-Mien populations (Shi et al., 2008), is also prevalent in Sherpa (11.90%). Haplogroup D3-P99 and its sublineage D3a-P47 are almost exclusively distributed in Tibeto-Burman populations (Shi et al., 2008), and also found highly frequent in Sherpa (7.14% and 15.48%, respectively). Haplogroup O3a2c1a-M117, one of the three main sublineages of O3, accounts for about 16% of Han Chinese and also exhibits high frequencies in Tibeto-Burman populations (Wang & Li, 2013; Yan et al., 2011). In this study, O3a2c1a-M117 comprises nearly half of Sherpa people (45.24%). The frequencies of another two main components of Sino-Tibetan populations, O3a2c1*-M134 and O3a1c-002611 (Wang et al., 2013; Yan et al., 2011), are negligible in Sherpa (1.19% and 0, respectively). It is particularly noteworthy that Central-South Asia and West Eurasia related haplogroups R1a1a-M17 and J-M304 (Zhong et al., 2011) have also been detected at considerable frequencies in WSC populations, especially R1a1a-M17, which contributes 11.90% of Sherpa.

Figure 2(a) illustrates a PCA plot based on 68 populations, including the Sherpa people in this study and 67 reference populations retrieved from literature. Almost all the Tibeto-Burman populations, including Sherpa, cluster together in the middle left corner of the plot, which is accounted for by the extensive sharing of haplogroup D1-M15, D3-P99, and O3a2c1a-M117 among them. The middle and upper right corner depict the Indo-European, Dravidian and Austro-Asiatic populations in the South Asian Subcontinent, due to the high frequency of haplogroup R, L, and H. The Altaic populations segregate intermediate between the East Asian and South Asian clusters. The Sherpa people slightly tend to deviate from the East Asian cluster owing to its considerable frequency of haplogroup R.

We then used the Y-STR induced R_{st} NJ tree and structure analysis to show the overall clustering pattern of those populations at both the haplogroup and haplotype level. In haplogroup R1a1a-M17, Sherpa is mainly clustered with populations from Afghanistan and India in the NJ tree. However, Sherpa people share most haplotypes with Newar people of Nepal and Brahui people of Pakistan as revealed by the structure plot (Figure 3a). It is worthy to note that only two haplotypes have been identified

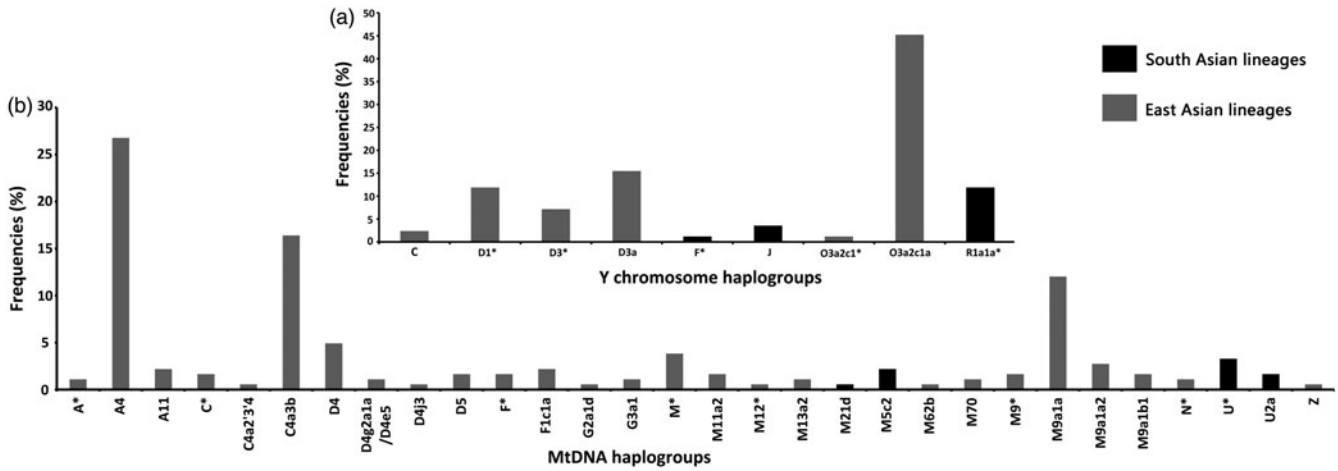


Figure 1. Frequencies of Y chromosome and mitochondrial DNA haplogroups in Sherpa.

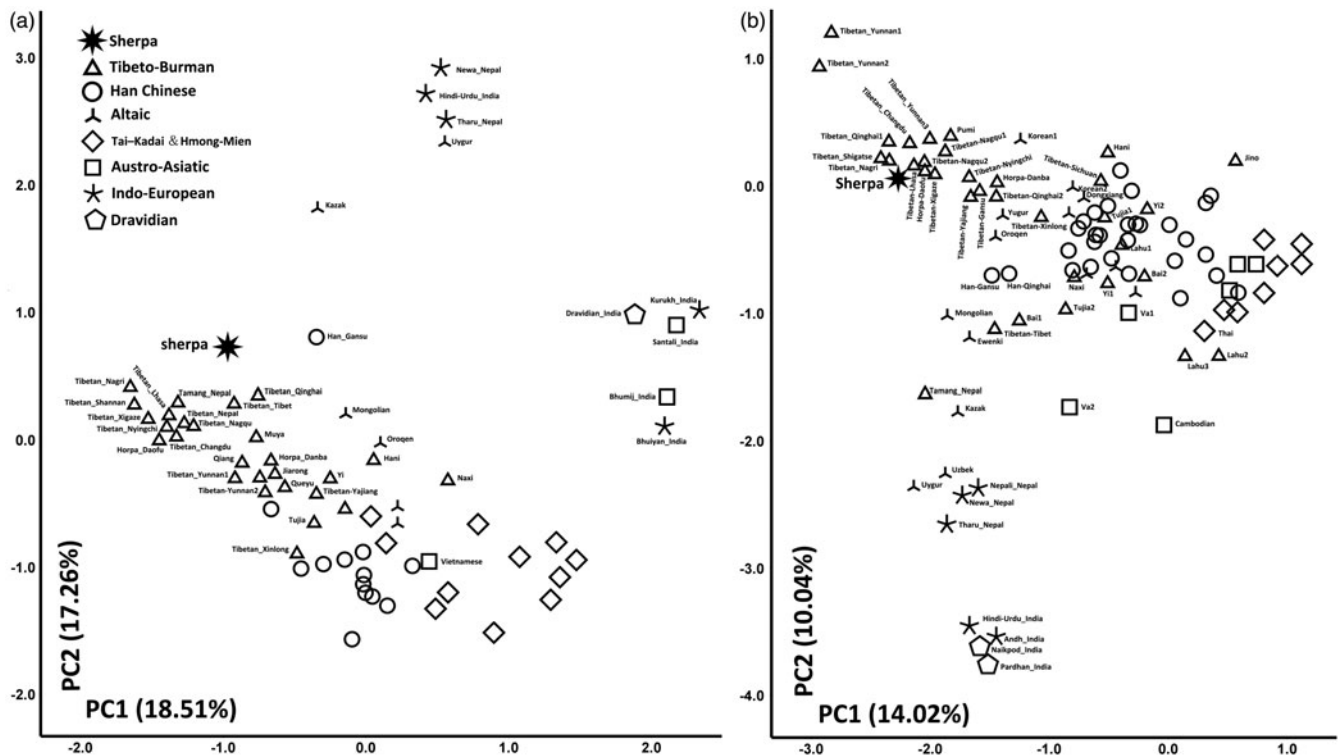


Figure 2. Principal component analysis plots based on frequencies of Y chromosome and mitochondrial DNA haplogroups.

in ten R1a1a-M17 samples of Sherpa, which suggest a founder effect took place when this lineage was involved in Sherpa or this lineage have experienced bottlenecks. In haplogroup O3a2c1a-M117, most of the Tibetan populations cluster tightly together in the NJ tree, along with Sherpa and Tamang of Nepal. However, more haplotypes of Sherpa samples share ancestry with Tibetan and other Tibeto-Burman populations from East Asia other than from Nepal (Figure 3b). Similarly, haplotypes of D1-M15 of Sherpa share ancestry with Tibetan, northwestern Han, and Zhuang (included in Tai-Kadai) populations from East Asia, although Sherpa has tended to be segregated away from the Tibetan cluster in the NJ tree. Similarly with R1a1a-M17, the haplotype diversity of D1-M15 samples in Sherpa is very low, and the haplotypes of Sherpa are only a small subset of those of Tibetan populations, probably also due to the founder effect when Sherpa was formed (Figure 3c). As we have mentioned above, haplogroup D3-P99 and D3a-P47 are almost exclusively distributed in Tibeto-Burman populations; but not only that, the

haplotypes of D3 also show strong similarities among different populations with distinctive and specific seven repeats at locus *DYS392* (Figure 3d and Table S1).

The haplotypes of J-M304 samples have already given us sufficient information to infer their origin and diffusion, although without typing downstream markers for those samples. The J-M304 samples should be probably assigned into J2 haplogroup and the haplotypes of those samples show strong similarities with those of Indian (in Southwest India) (Chennakrishnaiah et al., 2013), Malaysian Indian (Pamjav et al., 2011), and Lebanese samples (Zalloua et al., 2008). One Sherpa sample has been assigned as paragroup F*, which is observed only infrequently and primarily on the Indian subcontinent (Karafet et al., 2008).

Mitochondrial DNA

One-Hundred Eighty samples were assigned to mtDNA haplogroups using a combination of HVS-I sequence motifs and

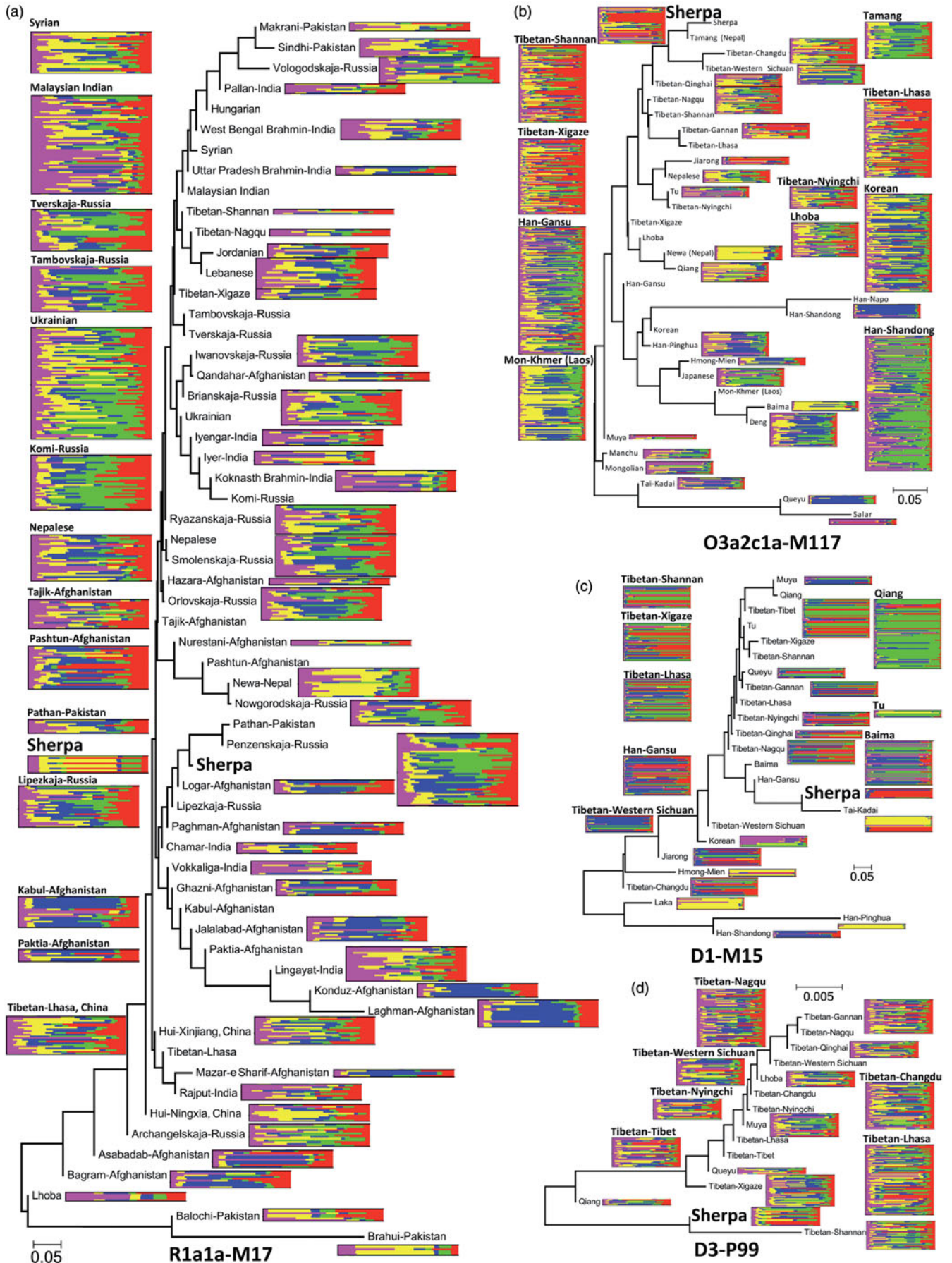


Figure 3. Neighbor-joining unrooted trees based on Y-STR *Rst* genetic distances and structure analysis.

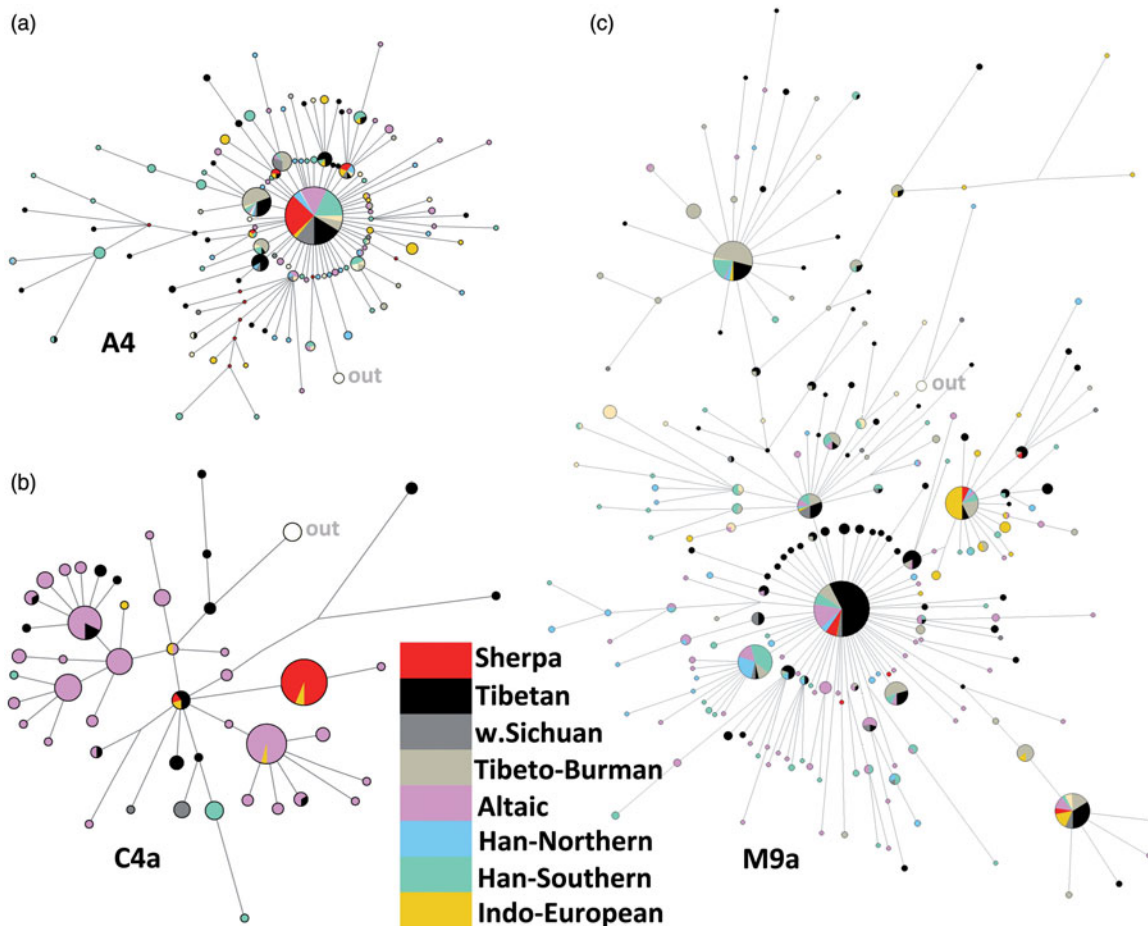


Figure 4. Networks of mitochondrial DNA HVS-I haplotype. Node sizes are proportional to the individual numbers of haplotypes. The length of the lines between nodes is proportional to the mutation steps. Abbreviations: w.Sichuan, Western Sichuan.

SNPs distributed around the coding region of the mtDNA genome. A total of 30 haplogroups or paragroups were identified (Figure 1b and Table S2). The most common mtDNA haplogroups in Sherpa are A4, C4a3b, M9a1a, D4, and U (including U* and U2a), in order of frequency. The majority of the mtDNA lineages belong to eastern Eurasian specific groups, including those from Northeast Asia (A, D4, D5, G, C, and Z) (Derenko et al., 2003, 2007; Tanaka et al., 2004) and Southern China or Southeast Asia (F, M9, M12 and M13) (Li et al., 2007), accounting for 59.02% and 23.50%, respectively. The South Asian lineages (M5c2, M21d, U* and U2a) also comprise 7.65% of Sherpa people, which is consistent with our previous observation (Kang et al., 2013). The considerable South Asian component of mtDNA lineages in Sherpa also corroborates the above Y chromosome evidence that the northward gene flow from the South Asian Subcontinent into Tibet really happened and has played an ignorable role in the formation of Sherpa.

We used a PCA based on the distribution of mtDNA haplogroup frequencies of 100 populations to show the matrilineal genetic patterns (Figure 2b). Most of the Tibeto-Burman populations, including Sherpa, cluster tightly in the upper left corner of the plot. However, the Indo-European and Dravidian populations have been situated in the lower left corner. Similarly with the Y chromosome PCA plot, the Altaic populations also aggregate intermediate between the East Asian Tibeto-Burman cluster and the South Asian groups.

However, the results based on haplogroup frequency comparisons could be misleading because of the quickly changing frequencies of the mtDNA lineages (Lu et al., 2013). A network analysis of individual lineages will most likely offer a better

investigation of maternal relationships among the Sherpa and Himalayan populations. Haplogroup A4, C4a, and M9a comprise more than 60% of Sherpa samples, and the networks of those haplogroups were analyzed based on the HVS-I motif (Figure 4). In haplogroup A4, most haplotypes of Sherpa are shared with Tibeto-Burman, Altaic, and Han Chinese and clustered in the main clade of the network. The Indo-European samples of South Asian Subcontinent are scattered in the terminal nodes throughout the network (Figure 4a), indicating late emergence of A4 into those Indo-European populations. In the network of C4a, nearly all the Sherpa samples cluster together and form a big exclusive clade along with few Nepalese from Katmandu (Figure 4b). Those exclusive haplotypes might represent the ancient component of Sherpa. The initial C4a individuals of Sherpa might have undergone founder events or bottlenecks in their history, and then remained genetically isolated for a long period of time. In the network of M9a, about half Sherpa M9a samples share the root haplotypes with the main ancestral clade, other samples mainly cluster with Indo-European and Tibeto-Burman samples in the terminal small clades (Figure 4c). The star-like networks of M9a and its sublineages have clearly indicated the population expansion of those lineages in Tibeto-Burman populations. The M9a samples of Sherpa and Indo-European populations might probably be results of the expansion of M9a in the Tibet Plateau.

Sherpa also exhibits some basal lineages emerging directly from the root of macrohaplogroup M, such as M5c2, M21d, M62b, and M70. Haplogroup M5, which has been previously observed in the general Andhra Pradesh Tribals, Hindus, and Muslims of Shia, Sunni, Dawoodi Bohra, and Mappala from India, and in Hindus, Tharu, Newar, and Nepalese (in Katmandu) from

Nepal (Eaaswarkhanth et al., 2009; Fornarino et al., 2009; Gayden et al., 2013; Metspalu et al., 2004). Four M5 samples with a same distinguishing HVS-I motif (16,129-16,223-16,240-16,291) have been detected in Sherpa. However, the mutation at site 16,240 is rarely seen, and our database search for the same HVS-I led to only one mtDNA sequence (a Sunni Muslim in North India) (Eaaswarkhanth et al., 2009). Haplogroup M21 has been found at low frequencies in Aboriginal Southeast Asians of Peninsular Malaysia, Philippine, Tharu of Nepal, Cham of Vietnam, and Thai people of Chiang Mai (Fornarino et al., 2009; Hill et al., 2006; Maruyama et al., 2010; Peng et al., 2010; Tabbada et al., 2010; Zimmermann et al., 2009). The one M21d sample of Sherpa shares the similar HVS-I haplotype with Cham people. Haplogroup M62 has been suggested as the genetic relics of the initial Late Paleolithic settlers on the Tibetan Plateau (Zhao et al., 2009). The M62b sample detected in Sherpa shares haplotype with Tibetans in Lhasa, Nyingchi, and Yunnan (Qin et al., 2010; Wen et al., 2004b). Haplogroup M70 has been regarded as Tibetan specific lineage, and the haplotypes of Sherpa M70 samples are also the same with those of Tibetan samples (Gayden et al., 2013).

Discussion

Tibeto-Burman origin of sherpa

About 83% of Sherpa Y chromosomes, including haplogroup C, D, and O, can be assigned an East or Southeast Asian origin. Detailed genetic structures at haplotype level of those lineages reveal strong affinities between Sherpa and Tibeto-Burman populations (especially with Tibetans). From the maternal side, mtDNA lineages that can trace to the East or Southeast Asian origin comprise about 82.5% of Sherpa people, and most of HVS-I haplotypes are shared or close connected with samples of Tibeto-Burman and Altaic populations. The internal homogeneity observed in some lineages suggests a possible founder effect during the origin of Sherpa, especially for Y chromosome haplogroup D1-M15 and O3a2c1a-M117, mtDNA haplogroup A4 and C4a; that is, Sherpa people of those haplogroups are derived from a small number of migrants from a Tibeto-Burman source population.

Northward gene flow across himalayas

Considerable South Asian genetic components have been observed in Sherpa. Y chromosome haplogroup R1a1a-M17, J-M304, F*-M89 comprise almost 17% of Sherpa paternal gene pool. However, in the maternal side, M5c2, M21d, U* and U2a only account for 8% of Sherpa people. The obvious sex-biased gene flow might be caused by physiological difference between male and female, especially under the extreme circumstances on the plateau. Paternal lineage R1a1a-M17 might probably experience severe founder effect or bottlenecks in Sherpa.

In the previous studies, the negligible gene flow from Indian Subcontinent to Tibetan Plateau has suggested a unidirectional human dispersal pattern in the Himalayan mountain range. However, the considerable northward gene flow across Himalayas has been observed in Sherpa's case.

Acknowledgements

Genographic Consortium members: Syama Adhikar¹, Christina J. Adler², Elena Balanovska³, Oleg Balanovsky³, Doron M. Behar⁴, Jaime Bertranpetit⁵, Andrew C. Clarke⁶, David Comas⁵, Alan Cooper², Clio S. I. Der Sarkissian², Matthew C. Dulik⁷, Christoff J. Erasmus⁸, Jill B. Gaieski⁷, Arun Kumar Ganesh Prasad¹, Wolfgang Haak², Angela Hobbs⁸, Asif Javed⁹, Matthew E. Kaplan¹⁰, Begona Martínez-Cruz⁵, Elizabeth A. Matisoo-Smith⁶, MartaMel'e⁵, Nirav C. Merchant¹⁰, R. John Mitchell¹¹, Amanda C. Owings⁷, Laxmi Parida⁹, Ramasamy Pitchappan¹, Daniel E. Platt⁹, Lluís Quintana-Murci¹², Colin Renfrew¹³,

Daniela R. Lacerda¹⁴, Ajay K. Royyuru⁹, Fabricio R. Santos¹⁴, Theodore G. Schurr⁷, Himla Soodyall⁸, David F. Soria Hernanz¹⁵, Pandikumar Swamikrishnan¹⁶, Chris Tyler-Smith¹⁷, Kavitha Valampuri John¹, Arun Varatharajan Santhakumari¹, Pedro Paulo Vieira¹⁸, Janet S. Ziegler¹⁹ and R. Spencer Wells¹⁵.

Affiliations for participants: ¹Madurai Kamaraj University, Madurai, Tamil Nadu, India. ²University of Adelaide, South Australia, Australia. ³Research Centre for Medical Genetics, Russian Academy of Medical Sciences, Moscow, Russia. ⁴Rambam Medical Center, Haifa, Israel. ⁵Universitat Pompeu Fabra, Barcelona, Spain. ⁶University of Otago, Dunedin, New Zealand. ⁷University of Pennsylvania, Philadelphia, Pennsylvania, United States. ⁸National Health Laboratory Service, Johannesburg, South Africa. ⁹IBM, Yorktown Heights, New York, USA. ¹⁰University of Arizona, Tucson, Arizona, USA. ¹¹La Trobe University, Melbourne, Victoria, Australia. ¹²Institut Pasteur, Paris, France. ¹³University of Cambridge, Cambridge, UK. ¹⁴Universidade Federal de Minas Gerais, Belo Horizonte, Minas Gerais, Brazil. ¹⁵National Geographic Society, Washington, District of Columbia, USA. ¹⁶IBM, Somers, New York, USA. ¹⁷The Wellcome Trust Sanger Institute, Hinxton, UK. ¹⁸Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brazil. ¹⁹Applied Biosystems, Foster City, California, USA.

Declaration of interest

The authors report no conflicts of interest. The authors alone are responsible for the content and writing of the paper. This research was partly supported by grants from the National Science Foundation of China (30760097, 30890034 and 31071098), National Outstanding Youth Science Foundation of China (30625016), the Natural Science Foundation of Shanghai (10ZR1402200), Shanghai Commission of Education Research Innovation Key Project (11zz04), and the Genographic Project. L. Kang is supported by China Postdoctoral Science Foundation (200902208), the Key Project of Chinese Ministry of Education (208138), and National "Eleventh Five-Year" Technology Support Program (2007BA/25800); H. Li is supported by Shanghai professional development funding (2010001); and L. Jin is supported by Shanghai Leading Academic Discipline Project (B111) and the Science and Technology Committee of Shanghai Municipality (09540704300).

References

- Arai Y, Tatsumi K, Sherpa NK, Masuyama S, Hasako K, Tanabe N, Takiguchi Y, Kuriyama T. (2002). Impaired oxygenation during sleep at high altitude in Sherpa. *Resp Physiol Neurobiol* 133:131–8.
- Bandelt HJ, Forster P, Röhl A. (1999). Median-joining networks for inferring intraspecific phylogenies. *Mol Biol Evol* 16:37–48.
- Black ML, Dufall K, Wise C, Sullivan S, Bittles AH. (2006). Genetic ancestries in northwest Cambodia. *Ann Hum Biol* 33:620–7.
- Cai X, Qin Z, Wen B, Xu S, Wang Y, Lu Y, Wei L, et al., Genographic Consortium. (2011). Human migration through bottlenecks from Southeast Asia into East Asia during Last Glacial Maximum revealed by Y chromosomes. *PLoS One* 6:e24282.
- Chennakrishnaiah S, Perez D, Gayden T, Rivera L, Regueiro M, Herrera RJ. (2013). Indigenous and foreign Y-chromosomes characterize the Lingayat and Vokkaliga populations of Southwest India. *Gene* 526: 96–106.
- Cordaux R, Saha N, Bentley GR, Aunger R, Sirajuddin SM, Stoneking M. (2003). Mitochondrial DNA analysis reveals diverse histories of tribal populations from India. *Eur J Hum Genet* 11:253–64.
- Cordaux R, Weiss G, Saha N, Stoneking M. (2004). The northeast Indian passageway: A barrier or corridor for human migrations? *Mol Biol Evol* 21:1525–33.
- Dancause KN, Chan CW, Arunotai NH, Lum JK. (2009). Origins of the Moken Sea Gypsies inferred from mitochondrial hypervariable region and whole genome sequences. *J Hum Genet* 54:86–93.
- Deng QY, Wang CC, Wang XQ, Wang LX, Wang ZY, Wu WJ, Li H. (2013). Genetic affinity between the Kam-Sui speaking Chadong and Mulam people. *J Syst Evol* 51:263–70.
- Derenko MV, Grzybowski T, Malyarchuk BA, Dambueva IK, Denisova GA, Czarny J, Dorzhu CM, et al. (2003). Diversity of mitochondrial DNA lineages in South Siberia. *Ann Hum Genet* 67:391–411.
- Derenko M, Malyarchuk B, Grzybowski T, Denisova G, Dambueva I, Perkova M, Dorzhu C, et al. (2007). Phylogeographic analysis of mitochondrial DNA in northern Asian populations. *Am J Hum Genet* 81:1025–41.

- Droma Y, Hanaoka M, Basnyat B, Arjyal A, Neupane P, Pandit A, Sharma D, et al. (2006). Genetic contribution of the endothelial nitric oxide synthase gene to high altitude adaptation in sherpas. *High Alt Med Biol* 7:209–20.
- Eaaswarkhanth M, Haque I, Ravesh Z, Romero IG, Meganathan PR, Dubey B, Khan FA, et al. (2009). Traces of sub-Saharan and Middle Eastern lineages in Indian Muslim populations. *Eur J Hum Genet* 18: 354–63.
- 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* 10:564–7.
- Fornarino S, Pala M, Battaglia V, Maranta R, Achilli A, Modiano G, Torroni A, et al. (2009). Mitochondrial and Y-chromosome diversity of the Tharus (Nepal): A reservoir of genetic variation. *BMC Evol Biol* 9: 154.
- Fucharoen G, Fucharoen S, Horai S. (2001). Mitochondrial DNA polymorphisms in Thailand. *J Hum Genet* 46:115–25.
- Gan RJ, Pan SL, Mustavich LF, Qin ZD, Cai XY, Qian J, Liu CW, et al., Genographic Consortium. (2008). Pinghua Population as an Exception of Han Chinese's Coherent Genetic Structure. *J Hum Genet* 53:303–13.
- 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* 54:216–23.
- 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* 80:884–94.
- Gayden T, Perez A, Persad PJ, Bukhari A, Chennakrishnaiah S, Simms T, Maloney T, et al. (2013). The Himalayas: Barrier and conduit for gene flow. *Am J Phys Anthropol* 151:169–82.
- Gong-Bo ZX. (2011). Comparative study of folklore between the Sherpa on China-Nepal border and the Sharba in Sichuan Songpan. *J Xihua Univ* 4:41–6.
- 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.
- Hanaoka M, Droma Y, Basnyat B, Ito M, Kobayashi N, Katsuyama Y, Kubo K, Ota M. (2012). Genetic variants in EPAS1 contribute to adaptation to high-altitude hypoxia in Sherpas. *PLoS One* 7:e50566.
- He JD, Peng MS, Quang HH, Dang KP, Trieu AV, Wu SF, Jin JQ, et al. (2012). Patrilineal perspective on the Austronesian diffusion in mainland southeast Asia. *PLoS One* 7:e36437.
- Hill C, Soares P, Mormina M, Macaulay V, Clarke D, Blumbach PB, Vizuete-Forster M, et al. (2007). A mitochondrial stratigraphy for Island Southeast Asia. *Am J Hum Genet* 80:29–43.
- Hill C, Soares P, Mormina M, Macaulay V, Meehan W, Blackburn J, Clarke D, et al. (2006). Phylogeography and ethnogenesis of aboriginal Southeast Asians. *Mol Biol Evol* 23:2480–91.
- Irwin JE, Saunier JL, Strouss KM, Diegoli TM, Sturk KA, O'Callaghan JA, Paintner CD, et al. (2008). Mitochondrial control region sequences from a Vietnamese population sample. *Int J Legal Med* 122:257–9.
- Jin HJ, Tyler-Smith C, Kim W. (2009). The peopling of Korea revealed by analyses of mitochondrial DNA and Y-chromosomal markers. *PLoS One* 4:e4210.
- Kang L, Lu Y, Wang C, Hu K, Chen F, Liu K, Li S, et al., Genographic Consortium. (2012). Y-chromosome O3 haplogroup diversity in Sino-Tibetan populations reveals two migration routes into the Eastern Himalayas. *Ann Hum Genet* 76:92–9.
- Kang L, Zheng HX, Chen F, Yan S, Liu K, Qin Z, Liu L, et al. (2013). mtDNA lineage expansions in Sherpa population suggest adaptive evolution in Tibetan Highlands. *Mol Biol Evol* 30:2579–87.
- 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* 18:830–8.
- Kivisild T, Tolk HV, Parik J, Wang Y, Papiha SS, Bandelt HJ, Villems R. (2002). The emerging limbs and twigs of the East Asian mtDNA tree. *Mol Biol Evol* 19:1737–51.
- Kong QP, Sun C, Wang HW, Zhao M, Wang WZ, Zhong L, Hao XD, et al. (2011). Large-scale mtDNA screening reveals a surprising matrilineal complexity in east Asia and its implications to the peopling of the region. *Mol Biol Evol* 28:513–22.
- Lertrit P, Poolswan S, Thosarat R, Sanpachudayan T, Boonyarit H, Chinpaisal C, Suktipat B. (2008). Genetic history of Southeast Asian populations as revealed by ancient and modern human mitochondrial DNA analysis. *Am J Phys Anthropol* 137:425–40.
- Li D, Li H, Ou C, Lu Y, Sun Y, Yang B, Qin Z, Zhou Z, et al. (2008). Paternal genetic structure of Hainan aborigines isolated at the entrance to East Asia. *PLoS One* 3:e2168.
- Li DN, Wang CC, Lu Y, Qin ZD, Yang K, Lin XJ, Li H. (2013a). Three phases for the early peopling of Hainan Island viewed from mitochondrial DNA. *J Syst Evol* 51:671–80.
- Li DN, Wang CC, Yang K, Qin ZD, Lu Y, Lin XJ, Li H. (2013b). Hainan indigenous substitute for genetic lineages of Champa exile Utsat people. *J Syst Evol* 51:287–94.
- Li H, Cai X, Winograd-Cort ER, Wen B, Cheng X, Qin Z, Liu W, et al. (2007). Mitochondrial DNA diversity and population differentiation in southern East Asia. *Am J Phys Anthropol* 134:481–8.
- Li H, Wen B, Chen SJ, Su B, Pramoongjago P, Liu Y, Pan S, et al. (2008). Paternal genetic affinity between western Austronesians and Daic populations. *BMC Evol Biol* 8:146.
- Lu Y, Pan SL, Qin SM, Qin ZD, Wang CC, Gan RJ, Li H. (2013). Genetic evidence for the multiple origins of Pinghua Chinese. *J Syst Evol* 51: 271–9.
- Malacrida S, Katsuyama Y, Droma Y, Basnyat B, Angelini C, Ota M, Danieli GA. (2007). Association between human polymorphic DNA markers and hypoxia adaptation in Sherpa detected by a preliminary genome scan. *Ann Hum Genet* 71:630–8.
- Maruyama S, Nohira-Koike C, Minaguchi K, Nambiar P. (2010). mtDNA control region sequence polymorphisms and phylogenetic analysis of Malay population living in or around Kuala Lumpur in Malaysia. *Int J Legal Med* 124:165–70.
- Metspalu M, Kivisild T, Metspalu E, Parik J, Hudjashov G, Kaldma K, Serk P, et al. (2004). 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.
- Mona S, Grunz KE, Brauer S, Pakendorf B, Castrì L, Sudoyo H, Marzuki S, et al. (2009). Genetic admixture history of eastern Indonesia as revealed by Y chromosome and mitochondrial DNA analysis. *Mol Biol Evol* 26:1865–77.
- Nonaka I, Minaguchi K, Takezaki N. (2007). Y-chromosomal binary haplogroups in the Japanese population and their relationship to 16 Y-STR polymorphisms. *Ann Hum Genet* 71:480–95.
- Oota H, Kitano T, Jin F, Yuasa I, Wang L, Ueda S, Saitou N, Stoneking M. (2002). Extreme mtDNA homogeneity in continental Asian populations. *Am J Phys Anthropol* 118:146–53.
- Oppitz M. (1968). *Geschichte und Sozialordnung der Sherpa* (Vol. 1). Innsbruck: Universitätsverlag Wagner.
- Pamjav H, Zalán A, Béres J, Nagy M, Chang YM. (2011). Genetic structure of the paternal lineage of the Roma People. *Am J Phys Anthropol* 145:21–9.
- Park MJ, Lee HY, Yang WI, Shin KJ. (2012). Understanding the Y chromosome variation in Korea-relevance of combined haplogroup and haplotype analyses. *Int J Legal Med* 126:589–99.
- Peng MS, Palanichamy MG, Yao YG, Mitra B, Cheng YT, Zhao M, Liu J, et al. (2011). Inland post-glacial dispersal in East Asia revealed by mitochondrial haplogroup M9a'b. *BMC Biol* 9:2.
- Peng MS, Quang HH, Dang KP, Trieu AV, Wang HW, Yao YG, Kong QP, Zhang YP. (2010). Tracing the Austronesian footprint in Mainland Southeast Asia: a perspective from mitochondrial DNA. *Mol Biol Evol* 27:2417–30.
- Pritchard JK, Stephens M, Donnelly P. (2000). Inference of population structure using multilocus genotype data. *Genetics* 155: 945–59.
- Qi X, Cui C, Peng Y, Zhang X, Yang Z, Zhong H, Zhang H, et al. (2013). Genetic evidence of Paleolithic colonization and Neolithic expansion of modern humans on the Tibetan Plateau. *Mol Biol Evol* 30: 1761–78.
- Qin Z, Yang Y, Kang L, Yan S, Cho K, Cai X, Lu Y, et al., Genographic Consortium. (2010). A mitochondrial revelation of early human migrations to the Tibetan Plateau before and after the last glacial maximum. *Am J Phys Anthropol* 143:555–69.
- Sengupta S, Zhivotovsky LA, King R, Mehdi SQ, Edmonds CA, Chow CE, Lin AA, et al. (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: 202–21.

- Shi H, Dong YL, Wen B, Xiao CJ, Underhill PA, Shen PD, Chakraborty R, et al. (2005). Y-chromosome evidence of southern origin of the East Asian-specific haplogroup O3-M122. *Am J Hum Genet* 77:408–19.
- Shi H, Su B. (2009). Origin of modern humans in East Asia: Clues from the Y chromosome. *Front Biol China* 4:241–7.
- Shi H, Zhong H, Peng Y, Dong YL, Qi XB, Zhang F, Liu LF, et al. (2008). Y chromosome evidence of earliest modern human settlement in East Asia and multiple origins of Tibetan and Japanese populations. *BMC Biol* 6:45.
- Tabbada KA, Trejaut J, Loo JH, Chen YM, Lin M, Mirazón-Lahr M, Kivisild T, De Ungria MC. (2010). Philippine mitochondrial DNA diversity: A populated viaduct between Taiwan and Indonesia? *Mol Biol Evol* 27:21–31.
- Tajima A, Hayami M, Tokunaga K, Juji T, Matsuo M, Marzuki S, Omoto K, Horai S. (2004). Genetic origins of the Ainu inferred from combined DNA analyses of maternal and paternal lineages. *J Hum Genet* 49:187–93.
- Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. (2011). MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol* 28:2731–9.
- Tan S, Yang M, Yu H, Dong Y, Shou W, Zou J, Tang W, et al. (2007). Y-chromosome polymorphisms define the origin of the Mang, an isolated population in China. *Ann Hum Biol* 34:573–81.
- Tanaka M, Cabrera VM, Gonzalez AM, Larruga JM, Takeyasu T, Fuku N, Guo LJ, et al. (2004). Mitochondrial genome variation in eastern Asia and the peopling of Japan. *Genome Res* 14:1832–50.
- Thanseem I, Thangaraj K, Chaubey G, Singh VK, Bhaskar LV, Reddy BM, Reddy AG, Singh L. (2006). Genetic affinities among the lower castes and tribal groups of India: inference from Y chromosome and mitochondrial DNA. *BMC Genet* 7:42.
- Trivedi R, Sitalaximi T, Banerjee J, Singh A, Sircar PK, Kashyap VK. (2006) Molecular insights into the origins of the Shompen, a declining population of the Nicobar archipelago. *J Hum Genet* 51: 217–26.
- Tsai LC, Lin CY, Lee JC, Chang JG, Linacre A, Goodwin W. (2001). Sequence polymorphism of mitochondrial D-loop DNA in the Taiwanese Han population. *Forensic Sci Int* 119:239–47.
- Van Driem G. (2001). Languages of the Himalayas: An ethnolinguistic handbook of the greater Himalayan region, containing an introduction to the symbiotic theory of language. Leiden: Brill.
- Wang CC, Li H. (2013). Inferring Human History in East Asia from Y Chromosomes. *Investig Genet* 4:11.
- Wang CC, Yan S, Qin ZD, Lu Y, Ding QL, Wei LH, Li S, et al. (2013). Late Neolithic expansion of ancient Chinese revealed by Y chromosome haplogroup O3a1c-002611. *J Syst Evol* 51:280–6.
- Wang D, Su LY, Zhang AM, Li YY, Li XA, Chen LL, Long H, Yao YG. (2012). Mitochondrial DNA copy number, but not haplogroup, confers a genetic susceptibility to leprosy in Han Chinese from Southwest China. *PLoS One* 7:e38848.
- Wen B, Li H, Gao S, Mao X, Gao Y, Li F, Zhang F, et al. (2005). Genetic structure of Hmong-Mien speaking populations in East Asia as revealed by mtDNA lineages. *Mol Biol Evol* 22:725–34.
- Wen B, Li H, Lu D, Song X, Zhang F, He Y, Li F, et al. (2004a). Genetic evidence supports demic diffusion of Han culture. *Nature* 431: 302–5.
- Wen B, Xie X, Gao S, Li H, Shi H, Song X, Qian T, et al. (2004b). Analyses of genetic structure of Tibeto-Burman populations reveals sex-biased admixture in southern Tibeto-Burmans. *Am J Hum Genet* 74:856–65.
- Wu T, Kayser B. (2006). High altitude adaptation in Tibetans. *High Alt Med Biol* 7:193–208.
- Xue Y, Zerjal T, Bao W, Zhu S, Shu Q, Xu J, Du R, et al. (2006). Male demography in East Asia: A north–south contrast in human population expansion times. *Genetics* 172:2431–9.
- Yan S, Wang CC, Li H, Li SL, Jin L, Genographic Consortium. (2011). An updated tree of Y-chromosome Haplogroup O and revised phylogenetic positions of mutations P164 and PK4. *Eur J Hum Genet* 19:1013–15.
- Yao YG, Kong QP, Bandelt HJ, Kivisild T, Zhang YP. (2002). Phylogeographic differentiation of mitochondrial DNA in Han Chinese. *Am J Hum Genet* 70:635–51.
- Zalloua PA, Xue Y, Khalife J, Makhoul N, Debiane L, Platt DE, Royyuru AK, et al. (2008). Y-chromosomal diversity in Lebanon is structured by recent historical events. *Am J Hum Genet* 82:873–82.
- Zhao M, Kong QP, Wang HW, Peng MS, Xie XD, Wang WZ, Jiayang, et al. (2009). Mitochondrial genome evidence reveals successful Late Paleolithic settlement on the Tibetan Plateau. *Proc Natl Acad Sci USA* 106:21230–5.
- Zhong H, Shi H, Qi XB, Duan ZY, Tan PP, Jin L, Su B, et al. (2011). Extended Y chromosome investigation suggests postglacial migrations of modern humans into East Asia via the northern route. *Mol Biol Evol* 28:717–27.
- Zimmermann B, Bodner M, Amory S, Fendt L, Röck A, Horst D, Horst B, et al. (2009). Forensic and phylogeographic characterization of mtDNA lineages from northern Thailand (Chiang Mai). *Int J Legal Med* 123:495–501.

Supplementary material available online

Supplementary Tables 1–2