

Short Communication

Y-chromosome O3 Haplogroup Diversity in Sino-Tibetan Populations Reveals Two Migration Routes into the Eastern Himalayas

Longli Kang^{1,2†}, Yan Lu^{1†}, Chuanchao Wang¹, Kang Hu², Feng Chen², Kai Liu³, Shilin Li^{1,4}, Li Jin^{1,4*}, Hui Li^{1,4*} and The Genographic Consortium[‡]

¹MOE Key Laboratory of Contemporary Anthropology, School of Life Sciences and Institutes of Biomedical Sciences, Fudan University, Shanghai 200433, China

²Key Laboratory of High Altitude Environment and Genes Related to Diseases of Tibet Autonomous Region, School of Medicine, Tibet University for Nationalities, Xianyang, Shaanxi 712082, China

³Tibet College for Vocational Technologies, Lhasa, Tibet 850000, China

⁴CMC Institute of Health Sciences, Taizhou, Jiangsu 225300, China

Summary

The eastern Himalayas are located near the southern entrance through which early modern humans expanded into East Asia. The genetic structure in this region is therefore of great importance in the study of East Asian origins. However, few genetic studies have been performed on the Sino-Tibetan populations (Luoba and Deng) in this region. Here, we analyzed the Y-chromosome diversity of the two populations. The Luoba possessed haplogroups D, N, O, J, Q, and R, indicating gene flow from Tibetans, as well as the western and northern Eurasians. The Deng exhibited haplogroups O, D, N, and C, similar to most Sino-Tibetan populations in the east. Short tandem repeat (STR) diversity within the dominant haplogroup O3 in Sino-Tibetan populations showed that the Luoba are genetically close to Tibetans and the Deng are close to the Qiang. The Qiang had the greatest diversity of Sino-Tibetan populations, supporting the view of this population being the oldest in the family. The lowest diversity occurred in the eastern Himalayas, suggesting that this area was an endpoint for the expansion of Sino-Tibetan people. Thus, we have shown that populations with haplogroup O3 moved into the eastern Himalayas through at least two routes.

Keywords: East Asian, population genetics, human migration, Y chromosome

Introduction

The peopling of the Himalayas may be quite late relative to other areas of eastern Eurasia, being settled only in the past 5000–7000 years (Huang, 1994). This region is now occupied by many populations speaking Sino-Tibetan languages. Tibetan speakers, including Monba, Bhutanese, Sikkimese, and northern tribes in Nepal (Lewis, 2009), reside in the western Himalayas, and can trace their origins back to the

Tibetan expansion in the recent 2000 years (Huang, 1994). In the eastern Himalayas between Tibet and Assam, two main populations reside in the region, the Luoba (synonym of Adi) and Deng (synonym of Mishmi) (Kang et al., 2010). However, we know very little about how and when the Luoba and Deng arrived in this region. The languages of these two populations belong to the North Assam branch of the Sino-Tibetan language family, and thus provide no clear evidence for their origins. Therefore, we investigated genetic diversity in these populations to clarify their origins and biological relationships to other Sino-Tibetan groups.

The best genetic systems to use for tracing population history are the Y chromosome (Jin & Su, 2000; Underhill et al., 2000; Jobling & Tyler-Smith, 2003) and the mitochondrial DNA (Wallace, 1994). The mitochondrial DNA data from the region of the Tibetan Plateau and the Himalayas have revealed

[†]These authors contributed equally to this paper.

[‡]Lists of participants and affiliations appear in the Appendix.

*Corresponding authors: Hui Li & Li Jin, Fudan School of Life Sciences, 220 Handan Road, Shanghai 200433, China. Tel.: +86-21-55664574; Fax: +86-21-55664885; E-mail: LiHui.Fudan@gmail.com

that there have been multiple population expansions into the plateau (Fornarino et al. 2009; Qin et al., 2010), although these data have provided limited information about the migration routes into the Himalayas. Therefore, we attempted to elucidate the population origins of the Luoba and Deng by studying their Y-chromosome diversity, since the Y chromosome has diversified into dozens of haplogroups among the world populations (Y-Chromosome-Consortium, 2002).

Among these paternal lineages, haplogroup O3 is the dominant haplogroup in Sino-Tibetan populations (Shi et al., 2005), and therefore, is the most useful paternal lineage for studying the expansion history of Sino-Tibetan populations. In this paper, we investigated Y-chromosome variation in the Luoba and Deng people, and analyzed the phylogeography of haplogroup O3, to explore the migration routes of Sino-Tibetan people into the eastern Himalayas. Our data suggest at least two migration routes from North China into this region.

Materials and Methods

Population Samples

The saliva samples collected and analyzed in this study include 90 Deng from Zayü County of Nyingtri Prefecture, and 130 Luoba from Mainling County of Nyingtri Prefecture. All individuals gave their informed contents before their participation in the study. This study was approved by the Ethics Committee of Fudan University, School of Life Sciences.

To obtain a more comprehensive picture of the genetic affiliation of the Himalayan populations to groups from East Asia, Y-chromosome data from 75 populations were compiled from the literature (Shi et al., 2005; Gayden et al., 2007; Gan et al., 2008; Shi & Su, 2009) (Table S1). The comparative groups included populations speaking Sino-Tibetan, Altaic, Tai-Kadai, Hmong-Mien, Austro-Asiatic, and Indo-European languages. Unpublished data from North Assam provided by the Genographic Consortium South Asian Regional Center were also included in the analyses, although the original data were omitted.

Y-Chromosome Genotyping

The samples were typed through seven panels of 75 single nucleotide polymorphisms (SNPs), as listed in the latest Y-chromosome phylogenetic tree (Karafet et al., 2008). 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 (rs17269396), KL1 (rs17276338); 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.

These binary markers were hierarchically genotyped with a SNaPshot[®] Multiplex Kit (Applied Biosystems, Foster City, CA) and fluorescent allele-specific PCR (Lehmann & Romano, 2005). PCR products (fragments) were read on a 3730xl Genetic Analyzer (Applied Biosystems).

In addition, 14 short tandem repeat (STR) polymorphisms (STRs: DYS19, DYS385a, DYS385b, DYS388, DYS389I, DYS389II, DYS390, DYS391, DYS392, DYS393, DYS426, DYS437, DYS438, DYS439) were typed using fluorescently labeled primers for PCR amplification. The allelic data were read on a 3100 Genetic Analyzer (Applied Biosystems).

All samples were successfully typed for SNPs, but 29 Luoba and 21 Deng samples were not successfully typed for STRs. Therefore, our STR analyses were based on the rest of the 61 Luoba and the 109 Deng samples.

Statistical Analyses

Haplogroups were assigned to each sample according to the SNP haplotyping following the most updated Y-chromosome nomenclature (Karafet et al., 2008). The haplogroup diversity (H) is calculated as:

$$H = \frac{N}{N-1} \left(1 - \sum_i x_i^2 \right)$$

where x_i is the (relative) haplogroup frequency of each haplogroup in the sample and N is the sample size (Nei & Tajima, 1981). F_{st} values (Nei & Kumar, 2000) for populations were estimated for each subhaplogroup (namely, O3-M134 and O3-M117) of haplogroup O3 based on STR diversities, using SPSS 15.0 software (SPSS, Chicago, USA). Multidimensional scaling plots were constructed using F_{st} values and SPSS. The phylogenetic relationships among STR haplotypes were determined by constructing networks with Network 4.6 software (Polzin & Daneschmand, 2003), and ages of the clades were also estimated in the networks using an evolutionary mutation rate of 0.00069 (Zhitovovskiy et al., 2004). The average haplotype diversity within each population was estimated from STR data using ARLEQUIN 3.0 software (Excoffier et al., 2005) and plotted onto a geographic map using Surfer 7.0 (Golden Software, Inc., Golden, CO, USA).

Table 1 Y-SNP haplogroup frequencies (%) of the eastern Himalayans.

Population	C3	D*	D1	D3	J	N1*	O*	O3*	O3a3c*	O3a3c1*	Q	R*	R1*	R1a1	Haplogroup diversity
Luoba	M217	M174	M15	P99	M304	LLY22g	M175	M122	M134	M117	M242	M207	M173	M17	0.7608
Deng	1.11		1.11	1.11			1.11	1.11		31.11	63.33				0.5072

Analyses of the molecular variations (AMOVA) were also done using ARLEQUIN 3.0 software.

Genetic components of each individual were estimated by STRUCTURE 2.2 (Pritchard et al., 2000) based on STR data. If the data consisted of STR data of multiple Y-chromosome SNP haplogroups, then the estimated structure would presumably reflect something about the Y-chromosome SNP tree and not population structure per se. Therefore, we did STRUCTURE analyses on the STR data of O3-M117 and O3-M134 separately.

In these analyses, reference populations were classified into seven groups: Northwestern Tibeto-Burman (Tibetan), Northeastern Tibeto-Burman (Qiang), Southwestern Tibeto-Burman (northeastern India), Southeastern Tibeto-Burman (South China and Indo-China), Han Chinese, Northern Asian (Altaic), and Southeast Asian (Hmong-Mien, Tai-Kadai, and Austro-Asiatic). The resemblances of any of these groups to the Deng or Luoba were then examined with the methods mentioned above.

Results

SNP Haplogroups

The pattern of Y-chromosome haplogroup diversity in a population can often provide a clear overview of its origin. The two Himalayan populations, Luoba and Deng, exhibited quite different patterns of Y-chromosome variation (Table 1). In the Deng, O3 was the dominant haplogroup and C, D, and N appeared as minor lineages. This pattern was similar to that seen in many other Sino-Tibetan populations (Shi et al., 2008). However, the Luoba showed a more complicated pattern, having high frequencies of not only O but also N and D, and also low frequencies of J, R, and Q. High frequencies of O and D are commonly seen in Tibetans (Qian et al., 2000). However, haplogroup N does not typically occur at high frequency in East Asians but is observed in Uralic populations from northern Europe (Rootsi et al. 2007). The Luoba has a haplogroup diversity of more than three times than that of the Deng. Low diversity mostly indicates a bottleneck effect. However, here two populations might both have gone through bottlenecks, while later population admixtures

might have raised the diversity of Luoba. Therefore, based on the SNP haplogroup patterns, the Luoba had more complicated origins than the Deng, and were genetically closer to Tibetans, whereas the Deng appeared very similar to other Sino-Tibetan populations from South China.

Because haplogroup D might have been present in Tibetan populations prior to the Last Glacial Age (Shi et al., 2008), it may not be directly linked to the history of Sino-Tibetan expansions. Conversely, haplogroup O3 is quite relevant for the genetic history study of the Sino-Tibetan populations because of its ubiquity in groups speaking these languages. Thus, the subsequent analyses in this paper focus mostly on the STR diversity within this haplogroup.

Cluster Analyses

To assess the relationships among the haplogroup O3 Y-chromosomes from different populations and explore the origin of O3 in the Himalayas, we performed multiple clustering analyses with the STR haplotypes from haplogroups O3-M117 and O3-M134, two of the major subbranches of O3 (Table S2). In our reference populations, haplogroup O3-M117 was more widely distributed among the populations than O3-M134, thus, more populations were included in the analyses for haplogroup O3-M117.

The multidimensional scaling plot of F_{ST} estimates for O3-M117 (Fig. 1A) distinguished the Tibetan populations (upper left side) from the Qiang populations (lower right side). The Luoba populations were positioned much closer to Tibetans, while the Deng populations showed similarities to the Qiang.

For the other reference samples, the distribution was more discrete but clusters were still observed. The southwestern Tibeto-Burman populations from India were similar to the Luoba and close to Tibetans, while the southeastern Tibeto-Burman populations from China were close to the Qiang and Deng. In the MDS plot of F_{ST} estimates for O3-M134 Y chromosomes (Fig. 1A), similar patterns were found, with the Deng being close to the Qiang and the Luoba being close to Tibetans.

Results of AMOVA between the Deng or Luoba and each of the other population groups are displayed in Fig. 1B. In both O3-M117 and O3-M134 haplogroups, the differences between the Deng and the Luoba were pronounced. AMOVA

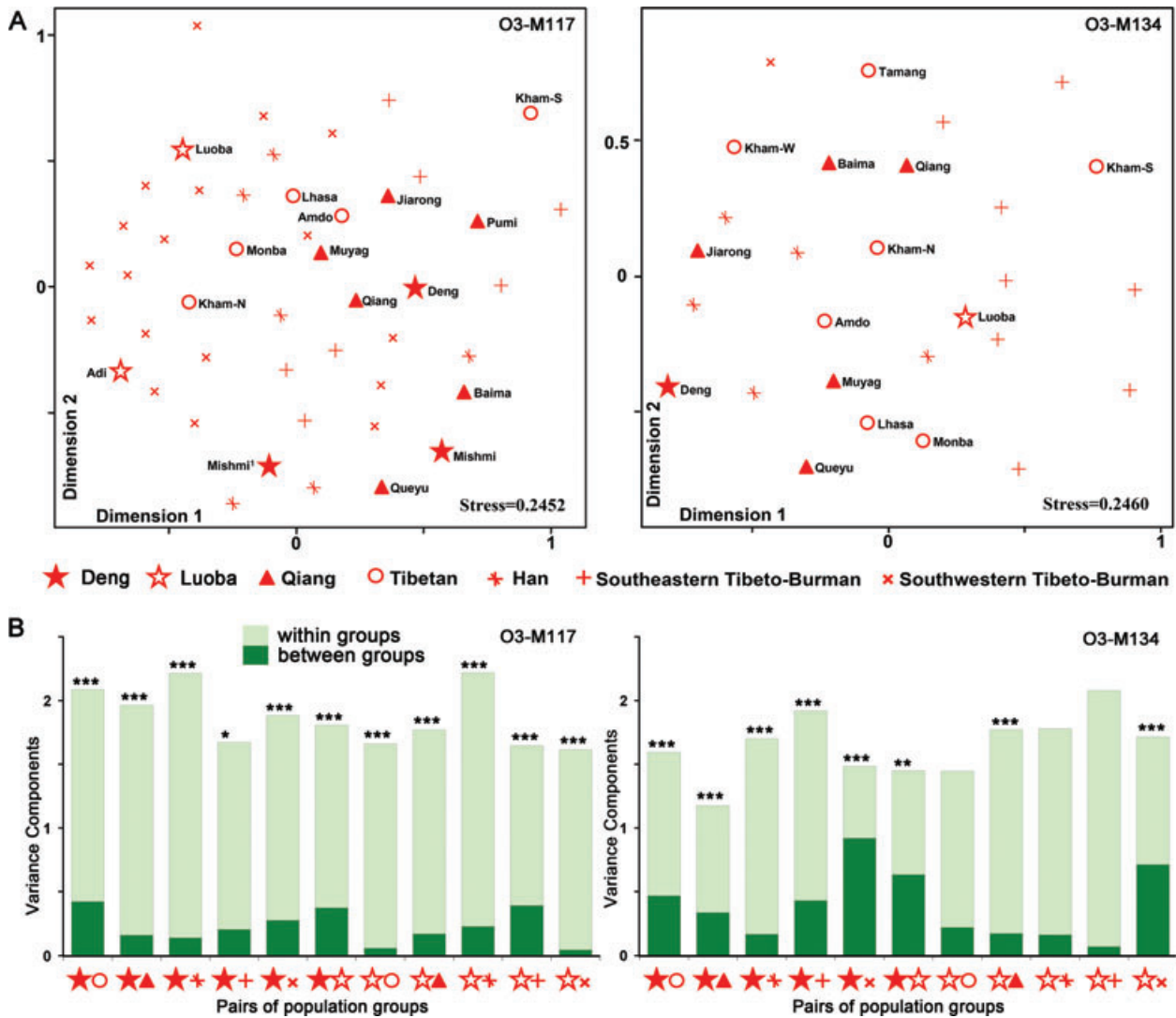


Figure 1 Clustering analyses based on the STR data of Y haplogroup O3. (A) Multidimensional plots for the Sino-Tibetan populations. (B) Pairwise AMOVA between the Deng or Luoba and other population groups. Variance components are displayed as a histogram. Significance levels: **P* < 0.05; ***P* < 0.01; ****P* < 0.00001.

also grouped the Deng to Qiang and Han, and the Luoba to Tibetans and Southwestern Tibeto-Burman.

We also performed network analyses for the two haplogroups (Fig. S1). However, the networks have low resolution in distinguishing the populations from each other. We also noted that most of the Deng O3-M134 haplotypes were shared with those of the Han Chinese, and the Luoba O3-M117 haplotypes were shared with those of the Tibetans. The time was also estimated in the networks. The haplotype data of the Luoba were too discrete for time estimate, while some Deng haplotypes concentrated in several clades provided opportunity for estimate. Two clades were 2113 ± 798 (Clade 1)

and 2052 ± 1122 (Clade 2) years old, respectively, agreeing with the results of the history studies (Huang, 1994).

The STRUCTURE software can extract some genetic components out of a set of population data, and estimate the frequency of each component for each individual sample. When we set eight components for the East Asian samples of O3-M117, the genetic structures of Luoba and Deng were distinctly different (Fig. S2). The total samples were roughly divided into two groups. Luoba was very similar to the Tibetans and the southwestern Tibeto-Burman populations, while Deng was similar to the Qiang, Han Chinese, and the Southeast Asians. In the structure of O3-M134 (Fig. S3),

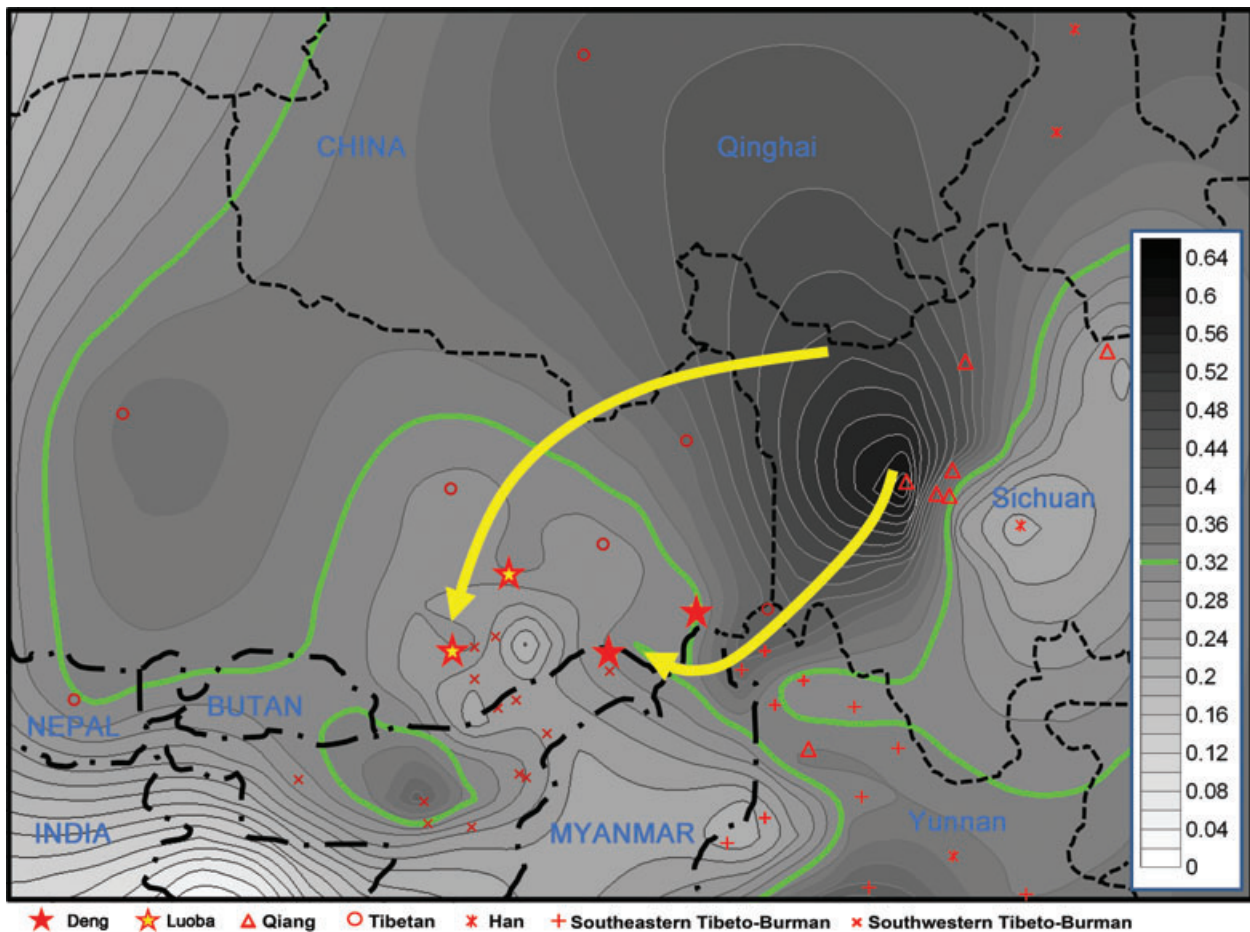


Figure 2 Average gene diversity map and the possible migration routes into the eastern Himalayas.

although there were not distinctly divided groups, the Deng and Luoba were still well distinguished. The pattern of Deng was most similar to some Han Chinese samples.

Average Gene Diversity

Usually, gene diversity is associated with the antiquity and size of a particular population (Nei & Kumar, 2000). We therefore estimated the gene diversity of each population using the STR data from haplogroup O3-M117, which appeared in most of the populations being compared. Geographically, we observed a single center with the highest gene diversity in the area of the Qiang people (Fig. 2). This result suggested that the Qiang might be the oldest population from the Sino-Tibetan family, at least with respect to O3-M117. The gene diversity also gradually declined along two routes in the eastern Himalayas. The first moved through Qinghai and Tibet in a counterclockwise direction, while the second passed through Yunnan in a clockwise direction. The Luoba and the Deng reside at the end of these two routes, respectively.

Discussion

Origin of Sino-Tibetan Populations

Haplogroup O3 is the dominant paternal haplogroup in Sino-Tibetan populations (Shi et al., 2005), and therefore, is more informative in revealing the history of those populations than other haplogroups. There are around 20 subhaplogroups within haplogroup O3, among which O3-M117 occurs at the highest frequency (Yan et al., 2011). In fact, O3-M117 is the only O3 subhaplogroup appearing in certain Sino-Tibetan populations from the border of Yunnan and Tibet (Fig. 2). For this reason, it is crucial to analyze haplogroup O3-M117 Y chromosomes when studying the genetic origin of Sino-Tibetan populations, especially those from southeastern Tibet (the eastern Himalayas).

A potential place for the origin of all Sino-Tibetan populations is the region east of Tibet, as we found the highest STR diversity of O3-M117 in the Qiang population from this area. The ancient tales of the Han Chinese clearly trace

their origin to the Qiang people (Wang, 1994). Archaeological findings also suggested that the Yangshao Culture of approximately 7000 years ago had its origin in the region of the Qiang people (Liu, 2005; Zhao et al., 2011). For members of the Tibeto-Burman linguistic subfamily, most populations had their previous names of “certain branches of Qiang,” which was also recorded in ancient Chinese books (Ge, 1985). Therefore, our genetic evidence supports data from historical and archaeological studies that indicate the Qiang group to be the origin of the Sino-Tibetan expansion.

Two Migration Routes into the Eastern Himalayas

The migration(s) of Sino-Tibetan populations into the eastern Himalayas might be much more recent than into other regions of their present distribution, as quite low STR diversity was found in the populations living in the eastern Himalayas. Our genetic analysis suggests that there were at least two routes of migrations into the region. The populations in the western part of the eastern Himalayas, including the Luoba and most of the Sino-Tibetan populations from northeastern India, were closely related to Tibetans. The populations from the eastern part of the region, including the Deng and some populations near northern Myanmar, were related to the Qiang and other southeastern Sino-Tibetan populations. Thus, our data suggest that the Sino-Tibetan populations in the eastern Himalayas arose through different expansion events based on different sets of paternal lineages.

The routes indicated in the map of Fig. 2 were just a rough estimate. A smoothing approach was used for the no data areas, which is a limitation of the map (Rendine et al., 1999; Sokal et al., 1999). Fortunately, population data along the routes are relatively prevalent in this study.

Distinct Histories for Other Haplogroups and Genetic Markers

The history of O3 lineages did not reveal the full histories of the Sino-Tibetan populations. As we have shown, other Y-chromosome haplogroups are present in our population samples. These haplogroups may tell different stories of the population origins. For example, haplogroup N might have a very long history in this region since ancestral East Asians first entered East Asia from this region, which was related to the origin of the Uralic populations (Rootsi et al., 2007) and worthy of further detailed studies. In addition, given that haplogroup D is a common paternal lineage in Tibet and Japan (Qian et al., 2000; Su et al., 2000; Wen et al., 2004; Hammer et al., 2006), we need to examine whether this haplogroup was carried together with haplogroup O3 from the Tibetans

into the Luoba or whether it had been in this region before O3 arrived. More interestingly, the presence of haplogroup J, Q, and R lineages in the Luoba suggest some migration or gene flow with populations from western or northern Eurasia. Therefore, more detailed studies are definitely required to reveal those missing histories at one of the entrances of East Asia.

Other genetic markers than Y chromosomes have also been involved in revealing the peopling of the Tibet and Himalayas, including mitochondrial DNA, autosomal STRs, and immunoglobulins. The mitochondrial DNA showed that Tibetans have multiple origins, including a Paleolithic origin from North Asia and a Neolithic origin from the Qiangs (Qin et al., 2010). The Nepalese have mitochondrial lineages from Sino-Tibetan, Indian, and western Eurasian origins (Fornarino et al. 2009). However, no data of mitochondria have been reported for the Luoba and Deng. Autosomal STRs revealed that the Tibetan populations are all similar to each other, while Luoba and Deng are quite unique with pronounced signals of founder effects (Kang et al., 2010). The data of immunoglobulins showed a similarity between the Tibetans and the North Asians including the Northern Han Chinese (Matsumoto, 1988), and no data of the Luoba and Deng have been reported. Altogether, genetic studies for the eastern Himalayas are quite limited, and the present data of other genetic markers all agree with the conclusion of our study.

Acknowledgements

This research was partly supported by grants from the National Science Foundation of China (30760097, 30890034, 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

- Excoffier, L., Laval, G. & Schneider, S. (2005) Arlequin (version 3.0): An integrated software package for population genetics data analysis. *Evol Bioinform Online* **1**, 47–50.
- Fornarino, S., Pala, M., Battaglia, V., Maranta, R., Achilli, A., Modiano, G., Torroni, A., Semino, O. & Santachiara-Benerecetti, S. A. (2009) Mitochondrial and Y-chromosome diversity of the Tharus (Nepal): A reservoir of genetic variation. *BMC Evol Biol* **9**, 154.

- Gan, R. J., Pan, S. L., Mustavich, L. F., Qin, Z. D., Cai, X. Y., Qian, J., Liu, C. W., Peng, J. H., Li, S. L., Xu, J. S., Jin, L., Li, H. & The Cenographic Consortium. (2008) Pinghua Population as an Exception of Han Chinese's Coherent Genetic Structure. *J Hum Genet* **53**, 303–313.
- Gayden, T., Cadenas, A. M., Regueiro, M., Singh, N. B., Zhivotovsky, L. A., Underhill, P. A., Cavalli-Sforza, L. L., & Herrera, R. J. (2007) The Himalayas as a directional barrier to gene flow. *Am J Hum Genet* **80**, 884–894.
- Ge, L. (1985) The historical relationship between ancient Qiang and Tibetans. *J Sun Yat-sen Univ* **2**, 92–101.
- Hammer, M. F., Karafet, T. M., 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.
- Huang, W. (1994) The prehistoric human occupation of the Qinghai-Xizang plateau. *Göttinger Geographische Abhandlungen* **95**, 201–219.
- Jin, L. & Su, B. (2000) Natives or immigrants: Modern human origin in east Asia. *Nat Rev Genet*, **1**, 126–133.
- Jobling, M. A. & Tyler-Smith, C. (2003) The human Y chromosome: An evolutionary marker comes of age. *Nat Rev Genet* **4**, 598–612.
- Kang, L., Li, S., Gupta, S., Zhang, Y., Liu, K., Zhao, J., Jin, L. & Li, H. (2010) Genetic structures of the Tibetans and the Deng people in the Himalayas viewed from autosomal STRs. *J Hum Genet* **55**, 270–277.
- Karafet, T. M., Mendez, F. L., Meilerman, M. B., Underhill, P. A., Zegura, S. L. & Hammer, M. F. (2008) New binary polymorphisms reshape and increase resolution of the human Y chromosomal haplogroup tree. *Genome Res* **18**, 830–838.
- Lehmann, E. L. & Romano, J. P. (2005) *Testing Statistical Hypotheses*. New York: Springer.
- Lewis, M. P. (ed.). (2009) *Ethnologue: Languages of the World, Sixteenth edition*. Dallas, TX: SIL International.
- Liu, L. (2005) *The Chinese Neolithic Trajectories to Early States*. Cambridge, UK: Cambridge University Press.
- Matsumoto, H. (1988) Characteristics of Mongoloid and neighboring populations based on the genetic markers of human immunoglobulins. *Hum Genet* **80**, 207–218.
- Nei, M. & Kumar, S. (2000) *Molecular Evolution and Phylogenetics*. New York: Oxford University Press.
- Nei, M. & Tajima, F. (1981) DNA polymorphism detectable by restriction endonucleases. *Genetics* **97**, 145–163.
- Polzin, T. & Daneschmand, S. V. (2003) On Steiner trees and minimum spanning trees in hypergraphs. *Operations Res Let* **31**, 12–20.
- Pritchard, J. K., Stephens, M. & Donnelly, P. (2000) Inference of population structure using multilocus genotype data. *Genetics* **155**, 945–959.
- Qian, Y., Qian, B., Su, B., Yu, J., Ke, Y., Chu, Z., Shi, L., Lu, D., Chu, J. & Jin, L. (2000) Multiple origins of Tibetan Y chromosomes. *Hum Genet* **106**, 453–454.
- Qin, Z., Yang, Y., Kang, L., Yan, S., Cho, K., Cai, X., Lu, Y., Zheng, H., Zhu, D., Fei, D., Li, S., Jin, L., & Li, H. (2010) A mitochondrial revelation of early human migrations to the Tibetan Plateau before and after the last glacial maximum. *Am J Phys Anthropol* **143**, 555–569.
- Rendine, S., Piazza, A., Menozzi, P. & Cavalli-Sforza, L. L. (1999) A problem with synthetic maps: Reply to Sokal et al. *Hum Biol* **71**, 15–25.
- Roots, S., Zhivotovsky, L. A., Baldovic, M., Kayser, M., Kutuev, I. A., Khusainova, R., Bermisheva, M. A., Gubina, M., Fedorova, S. A., Ilumae, A. M., Khusnutdinova, E. K., Voevoda, M. I., Osipova, L. P., Stoneking, M., Lin, A. A., Ferak, V., Parik, J., Kivisild, T., Underhill, P. A. & Villemers, R. (2007) A counter-clockwise northern route of the Y-chromosome haplogroup N from Southeast Asia towards Europe. *Eur J Hum Genet* **15**, 204–211.
- Shi, H., Dong, Y. L., Wen, B., Xiao, C. J., Underhill, P. A., Shen, P. D., Chakraborty, R., Jin, L. & Su, B. (2005) Y-chromosome evidence of southern origin of the East Asian-specific haplogroup O3-M122. *Am J Hum Genet* **77**, 408–419.
- Shi, H. & Su, B. (2009) Origin of modern humans in East Asia: Clues from the Y chromosome. *Front Biol China* **4**, 241–247.
- Shi, H., Zhong, H., Peng, Y., Dong, Y. L., Qi, X. B., Zhang, F., Liu, L. F., Tan, S. J., Ma, R. Z., Xiao, C. J., Wells, R. S., Jin, L., & Su, B. (2008) Y chromosome evidence of earliest modern human settlement in East Asia and multiple origins of Tibetan and Japanese populations. *BMC Biol* **6**, 45.
- Sokal, R. R., Oden, N. L., & Thomson, B. A. (1999) A problem with synthetic maps. *Hum Biol* **71**, 1–13.
- Su, B., Xiao, C., Deka, R., Seielstad, M. T., Kangwanpong, D., Xiao, J., Lu, D., Underhill, P., Cavalli-Sforza, L., Chakraborty, R. & Jin, L. (2000) Y chromosome haplotypes reveal prehistorical migrations to the Himalayas. *Hum Genet* **107**, 582–590.
- Underhill, P. A., Shen, P., Lin, A. A., Jin, L., Passarino, G., Yang, W. H., Kauffman, E., Bonne-Tamir, B., Bertranpetit, J., Francalacci, M., Ibrahim, M., Jenkins, T., Kidd, J. R., Mehdi, S. Q., Seielstad, M. T., Wells, R. S., Piazza, A., Davis, R. W., Feldman, M. W., Cavalli-Sforza, L. L. & Oefner, P. J. (2000) Y chromosome sequence variation and the history of human populations. *Nat Genet* **26**, 358–361.
- Wallace, D. C. (1994) Mitochondrial DNA sequence variation in human evolution and disease. *Proc Natl Acad Sci U S A* **91**, 8739–8746.
- Wang, Z. H. (1994) *History of Nationalities in China (in Chinese)*. Beijing, China: Social Sciences Press.
- Wen, B., Xie, X., Gao, S., Li, H., Shi, H., Song, X., Qian, T., Xiao, C., Jin, J., Su, B., Lu, D., Chakraborty, R. & Jin, L. (2004) Analyses of genetic structure of Tibeto-Burman populations reveals sex-biased admixture in southern Tibeto-Burmans. *Am J Hum Genet* **74**, 856–865.
- Y-Chromosome-Consortium. (2002) A nomenclature system for the tree of human Y-chromosomal binary haplogroups. *Genome Res* **12**, 339–348.
- Yan, S., Wang, C. C., Li, H., Li, S. L. & Jin, L. (2011) An updated tree of Y chromosome Haplogroup O and revised phylogenetic position of mutations P164 and PK4. *Eur J Hum Genet* **19**, 1013–1015.
- Zhao, Y. B., Li, H. J., Li, S. N., Yu, C. C., Gao, S. Z., Xu, Z., Jin, L., Zhu, H. & Zhou, H. (2011) Ancient DNA evidence supports the contribution of Di-Qiang people to the Han Chinese gene pool. *Am J Phys Anthropol* **144**, 258–268.
- Zhivotovsky, L. A., Underhill, P. A., Cinnioglu, C., Kayser, M., Morar, B., Kivisild, T., Scozzari, R., Cruciani, F., Destro-Bisol, G., Spedini, G., Chambers, G. K., Herrera, R. J., Yong, K. K., Gresham, D., Tournev, I., Feldman, M. W., & Kalaydjieva, L. (2004) The effective mutation rate at Y chromosome short tandem repeats, with application to human population-divergence time. *Am J Hum Genet* **74**, 50–61.

Appendix

The Genographic Consortium

Syama Adhikarla¹, Christina J. Adler², Elena Balanovska³, Oleg Balanovsky³, Doron M. Behar⁴, Jaume 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, Begoña Martínez-Cruz⁵, Elizabeth A. Matisoo-Smith⁶, Marta Melé⁵, 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, Fabrício 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.

Supporting Information

Additional supporting information may be found in the online version of this article:

Table S1 General information of the population samples analyzed in this paper.

Table S2 SNP haplogroups and STR diversities of Luoba and Deng.

Figure S1 STR networks for the O3 samples of the East Asians.

Figure S2 O3-M117 STR structure of the East Asians estimated by STRUCTURE software.

Figure S3 O3-M134 STR structure of the East Asians estimated by STRUCTURE software.

As a service to our authors and readers, this journal provides supporting information supplied by the authors. Such materials are peer-reviewed and may be re-organised for online delivery, but are not copy-edited or typeset. Technical support issues arising from supporting information (other than missing files) should be addressed to the authors.

Received: 4 August 2011

Accepted: 25 October 2011