Research Article

Late Neolithic expansion of ancient Chinese revealed by Y chromosome haplogroup O3a1c-002611

[1]Chuan-Chao WANG   [2]Shi YAN   [1]Zhen-Dong QIN   [1]Yan LU   [1]Qi-Liang DING
[1]Lan-Hai WEI   [1]Shi-Lin LI   [1]Ya-Jun YANG   [1,2]Li JIN   [1]Hui LI[*]
the Genographic Consortium[†]

[1](*State Key Laboratory of Genetic Engineering and MOE Key Laboratory of Contemporary Anthropology, School of Life Sciences, Fudan University,*
Shanghai 200433, China)
[2](*Group of Computational Genetics, CAS-MPG Partner Institute for Computational Biology, Shanghai Institutes for Biological Sciences,*
Chinese Academy of Sciences, Shanghai 200031, China)

Abstract Y chromosome haplogroup O3-M122 is the most prevalent haplogroup in East Asia, and provides an ideal tool for dissecting primary dispersals of the East Asians. Most of the sub-haplogroups of O3-M122 have been sufficiently investigated except for O3a1c-002611, despite its great prevalence and huge population, especially in Han Chinese. In this study, we identified 508 individuals with haplogroup O3a1c-002611 out of 7801 males from 117 East and Southeast Asian populations, typed at two newly discovered downstream Y-SNP markers and ten commonly used Y-STRs. Defined by SNPs IMS-JST002611 (in short, 002611), F11, and F238, three lineages internal to haplogroup O3a1c-002611 have distinct geographical distributions. Furthermore, Y-STR diversity shows a general south-to-north decline, which is consistent with the prehistorically northward migration of the other O3-M122 lineages. The northward migration of haplogroup O3a1c-002611 started about 13 thousand years ago (KYA). The expansions of subclades F11 and F238 in ancient Han Chinese began about 5 and 7 KYA immediately after the separation between the ancestors of the Han Chinese and Tibeto-Burman.
Key words Han Chinese, Neolithic Age, population expansion, Y chromosome.

Y chromosomes, with combination of archeology and linguistics, provide a powerful tool for reconstruction of the history of human primary colonization and dispersals in East Asia. Haplogroup O-M175 is a dominant component of the East Asian Y-chromosome gene pool, accounting for 75% of the total paternal lineages of Chinese (Yan et al., 2011). Previous studies of East Asian-specific haplogroup O1-M119, O2-M95, and O3-M122 suggested the southern origin of northern populations and the northward migration after the initial Paleolithic peopling of East Asia (Su et al., 1999, 2000a, 2000b; Wen et al., 2004a, 2004b; Shi et al., 2005). Particularly, the haplogroup O3-M122 is highly prevalent in almost all the populations of the Sino-Tibetan linguistic family. For instance, the frequencies of the three main subhaplogroups of O3-M122: O3a1c-002611, O3a2c1*-M134, and O3a2c1a-M117 in Han Chinese are 16.9%, 11.4%, and 16.3%, respectively (Yan et al., 2011). The systematic

sampling and genotyping of O3 lineage provided clues about the origin and dispersal of Sino-Tibetan populations (Shi et al., 2005). The age of O3a2c1-M134 (including M117) was calculated to be 20–40 thousand years (KY) for the Chinese M134 and 5.2–5.9 KY for the Tibeto-Burman M134. The time estimation suggested that the ancestors of the Sino-Tibetan population reached the upper and middle Yellow River basin about 20–40 KY ago (KYA). Around 6 KYA, the Chinese subfamily split away from Tibeto-Burman, and then eventually inhabited the whole of China (Su et al., 2000b; Shi et al., 2005). However, the dispersal and expansion patterns of Han Chinese remain unclear because haplogroup O3a1c-002611 with the highest frequency in Han Chinese has not been studied (Yan et al., 2011), and no downstream markers have yet been described which could distinguish Han Chinese O3 chromosomes from Tibeto-Burman. Here, to address this shortcoming, we analyzed more than 7800 male samples across East Asia, using the newly discovered 002611-downstream SNPs: F11 (rs17276393) and F238 (Hg18-12607511, A > G; Supplementary Doc. S1). We also examined the STR diversity of the associated O3a1c-002611

subclades to better understand the demographic history of Han Chinese. We found that F11 and F238 are characteristic of Neolithic dispersals and expansions, or more recent events of gene flow of Han Chinese.

1   Material and methods

We analyzed a total of 7801 unrelated males, including 6668 males from 95 East Asian populations and 1133 males from 31 Southeast Asian populations (Supplementary Doc. S1). All subjects signed informed consent. Our study was approved by the Ethics Committee of the Fudan School of Life Sciences before commencement. We first genotyped two Y-chromosome markers: M175 and M122. The

M122-derived individuals were then subjected to further typing of three biallelic markers, which define three subgroups: O3a1c*-002611, O3a1c1-F11, and O3a1c2-F238. Those binary markers were hierarchically genotyped by SNaPshot (ABI SNaPshot Multiplex Kit, Carlsbad, CA, USA) and PCR products were electrophoresed on a 3730xl Genetic Analyzer (Applied Biosystems, Carlsbad, CA, USA). We also typed 10 commonly used Y-STR markers: DYS19, DYS389I, DYS389II, DYS390, DYS391, DYS392, DYS393, DYS437, DYS438, and DYS439 using fluorescence-labeled primers (Applied Biosystems) and then running denatured PCR products on ABI 3730.

The Y-chromosome SNP data were used to analyze the frequency distribution of O3a1c-002611 and its subclades in worldwide populations by
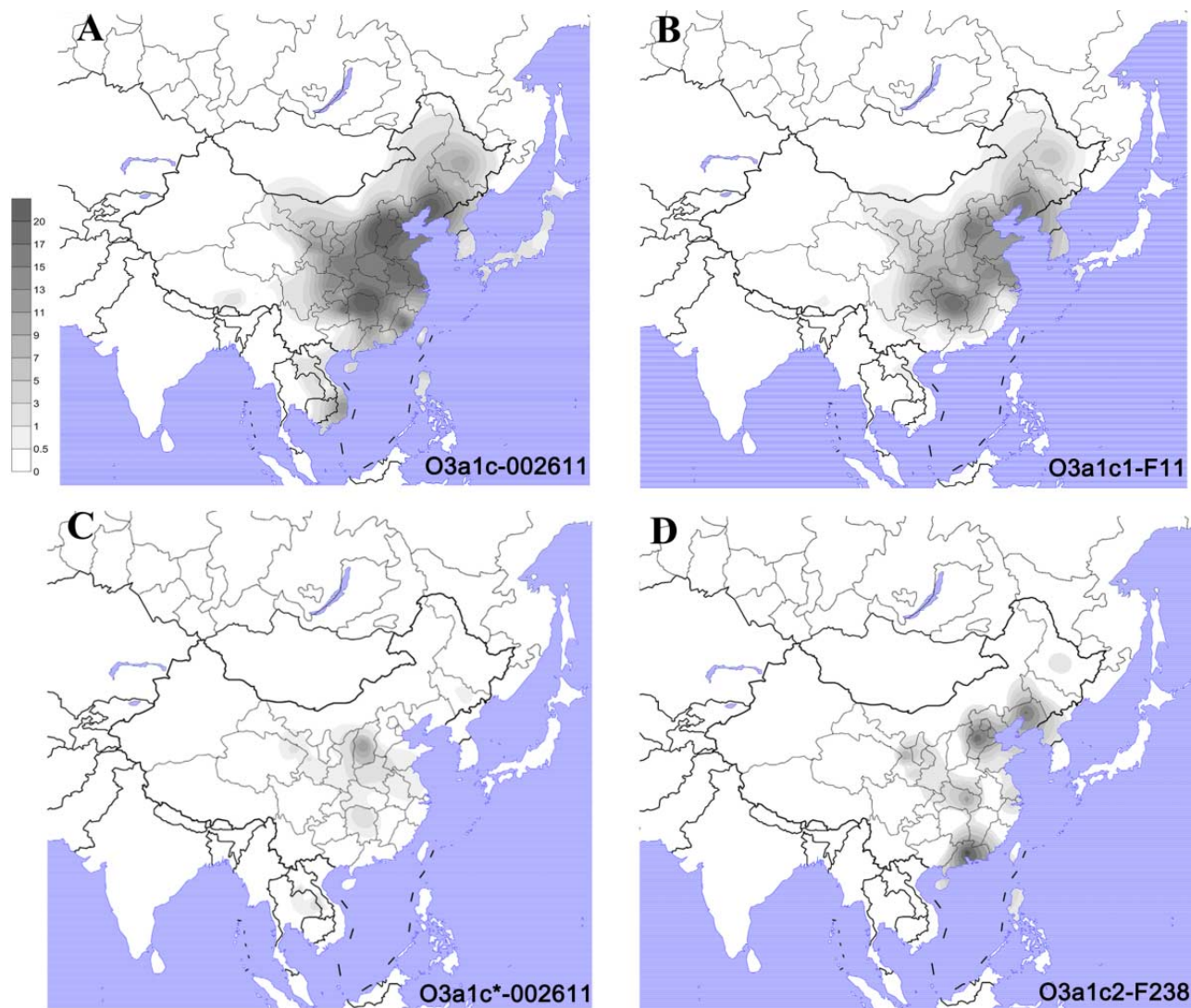

Fig. 1. **A**, Frequency distribution of haplogroup O3a1c-002611 and **B–D**, its subhaplogroups in worldwide populations. The maps were created by using the Kriging algorithm of the Surfer 9.11 package based on Google maps.

generation of contour maps using Surfer 8.0 Software (Golden Software, Golden, CO, USA). The Y-STR data were used to construct the median-joining networks using the program NETWORK 4.6.0.0 (Fluxus Engineering) (Bandelt et al., 1999) and to calculate the average gene diversity based on 10 STR loci by Arlequin 3.11 (Excoffier et al., 2005). The ages of STR variation of the O3a1c-002611 and its subhaplogroups were estimated using the STR data by use of the SNP-STR coalescence method (Zhivotovsky, 2001; Ramakrishnan & Mountain, 2004; Zhivotovsky et al., 2004; Sengupta et al., 2006). The Y-STR mutation rate of each locus we used was following Goedbloed et al. (2009).

2   Results and Discussion

Haplogroup O3a1c-002611 is prevalent in different ethnic groups in China and Southeast Asia (Fig. 1; Table S1), including Vietnam (14.29%), Sichuan of southwestern China (Han, 14.60%), Jilin of northeastern China (Korean, 9.36%), Inner Mongolia (Mongolian, 6.58%), and Gansu of northwestern China (Baima, 7.35%; Han, 11.30%), but is absent or found at very low frequencies in Indonesia and Oceania (Karafet et al., 2010). However, the subhaplogroups of O3a1c-002611 have a geographically pronounced distribution as shown in Fig. 1. Haplogroup O3a1c2, defined by F238, is restricted to China, especially Han Chinese and not detected in samples from Laos. Haplogroup

O3a1c1-F11 is the most widespread subhaplogroup, extending from Laos in Southeastern Asia all the way to Northeastern China. However, the frequencies of haplogroup O3a1c1-F11 in different populations vary dramatically. High frequencies of haplogroup O3a1c1-F11 are observed in Han Chinese, Korean, and Mongolian populations. Relatively low frequencies of haplogroup O3a1c1-F11 are observed in Tibetan populations.

To estimate STR gene diversity, we grouped the populations based on geographical regions (Table S2). A general south-to-north STR diversity decline of haplogroup O3a1c-002611 was observed (Table 1). The populations in Southwest China and Laos have the highest diversity, such as Tibetan and Southwestern Han. The Manchu population also shows a relatively high diversity. However, this is probably because of recent extensive admixture between the Manchu with Northern Asian and Han Chinese populations (Wang, 1994). In contrast, Northeastern Han and Mongolian populations display relatively low diversities. The diversities of Northwestern Han, Central Han, and Eastern Han are almost the same. The STR diversity pattern suggests a probable Southeast Asian origin, and then a northward migration of haplogroup O3a1c-002611. The diversity of O3a1c1-F11 also exhibits almost the same pattern.

We then estimated the ages of STR variation based on the 10 STR data. The STR variation ages for haplogroups O3a1c, O3a1c*, O3a1c1, and O3a1c2

Table 1   Average gene diversity of haplogroup O3a1c-002611 and its subhaplogroups

Population	Average gene diversity		
	O3a1c-002611	O3a1c1-F11	O3a1c2-F238
Eastern Han	0.447 ± 0.244	0.384 ± 0.215	0.423 ± 0.241
Central Han	0.456 ± 0.259	0.337 ± 0.203	–
Northwestern Han	0.472 ± 0.257	0.388 ± 0.218	0.427 ± 0.242
Northeastern Han	0.378 ± 0.227	0.333 ± 0.207	–
Southwestern Han	0.480 ± 0.271	0.471 ± 0.270	–
Altaic	0.418 ± 0.243	0.281 ± 0.185	–
Korean	0.442 ± 0.255	0.378 ± 0.226	–
Baima	0.436 ± 0.266	0.382 ± 0.245	–
Qiangic	0.396 ± 0.242	0.331 ± 0.212	–
Tibetan	0.496 ± 0.281	0.458 ± 0.272	–
Laos	0.486 ± 0.278	0.456 ± 0.274	–

Eastern Han: Shandong, Jiangsu, Shanghai, Zhejiang, Jiangxi, Anhui of China.
Central Han: Hunan, Henan, Hebei, Hubei, Shanxi, Beijing of China.
Northeastern Han: Heilongjiang, Jilin, Liaoning of China.
Northwestern Han: Gansu, Xinjiang of China.
Southwestern Han: Sichuan, Guangxi, Yunnan of China.
Laos: Bo, Lao, Laven, PhuThai, Rien, Sô, Tai Mène from Laos.
Altaic: Tu (Qinghai), Manchu (Jilin and Liaoning), Mongolian (Inner Mongolia) from China.
Korean: Korean from Jilin, China.
Baima: Baima from Gansu, China.
Qiangic: Muyag, Queyu, Qiang from Sichuan, China.
Tibetan: Tibetans from Qinghai, Tibet, and Gansu of China.
–, missing data.

were estimated as 12.8 ± 1.3, 9.08 ± 2.7, 12.1 ± 1.6, and 7.1 ± 1.0 KY, respectively. In combination with the above analysis of STR diversity pattern, the mutations of 002611 and F11 seemed to occur almost at the same time, about 12 KYA in the Southeast Asia populations, and then started northward migration into East Asia. The haplogroup O3a1c* might contain subclades that have not been identified, as the relatively

old age and complex network (Fig. 2) of the haplogroup was observed.

To give a clear dispersal pattern of the subclades of 002611, we estimated the ages of haplogroups O3a1c1-F11 and O3a1c2-F238 in Han Chinese. The STR variation age for O3a1c1-F11 in Han Chinese was 4.8 ± 1.0 KY, reflecting a very recent population expansion of Han Chinese, which can also be observed
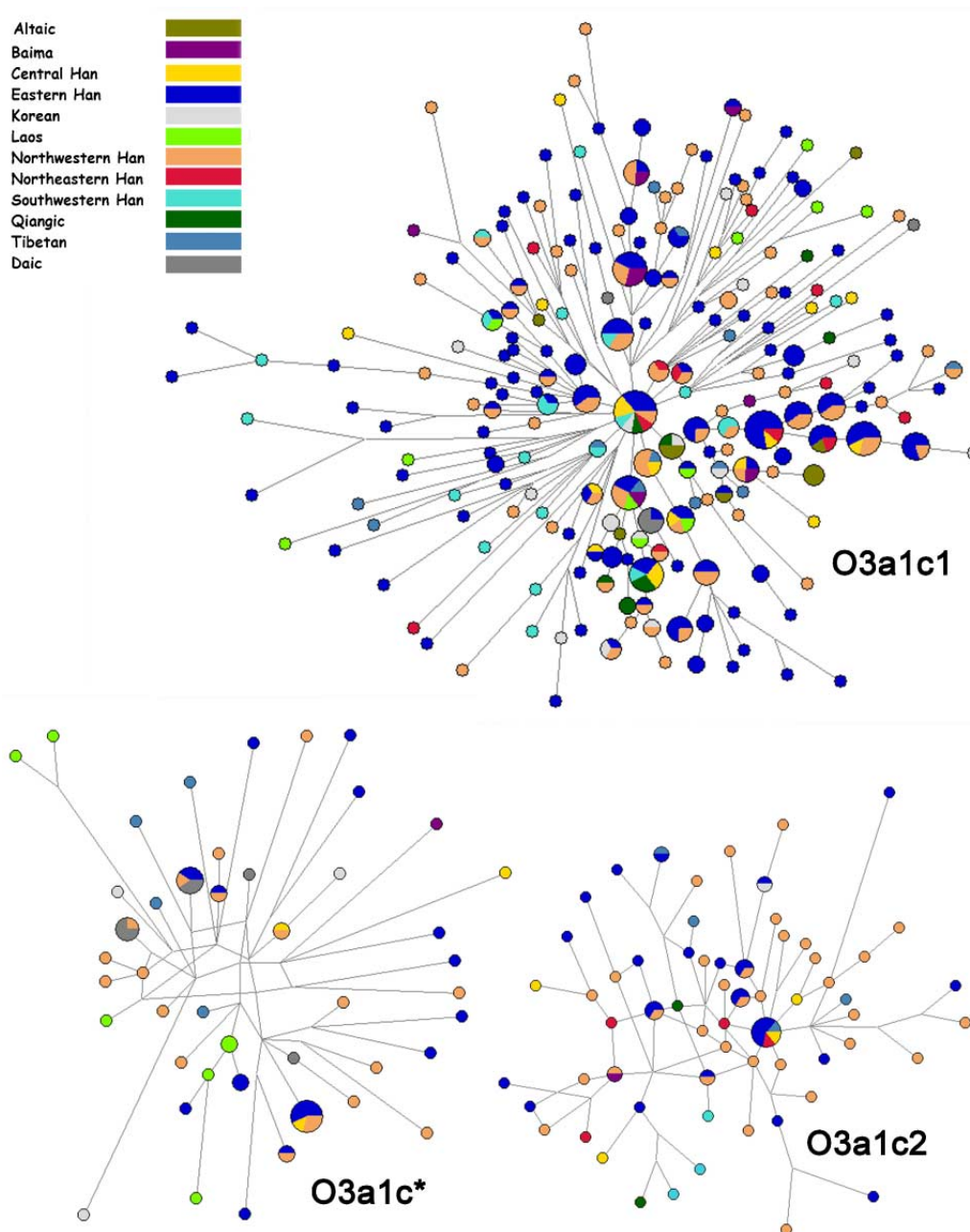

Fig. 2. Networks of subhaplogroup O3a1c*-002611 + , F11-, F238-; O3a1c1-F11 and O3a1c2-F238. Sample size for each population within haplogroup O3a1c-002611: Altaic, 18; Baima, 10; Central Han, 24; Daic, 2; Eastern Han, 209; Korean, 19; Laos, 18; Northwestern Han, 134; Northeastern Han, 14; Qiangic, 11; Southwestern Han, 24; Tibetan, 20.

in the star-like network of haplogroup O3a1c1-F11 (Fig. 2). However, the STR age for O3a1c2-F238 in Han Chinese was 7.2 ± 1.1 KY, which is almost the same as the age estimated in all the populations studied and is consistent with the recent origin given by the network of O3a1c2-F238 (Fig. 2). The absence of O3a1c2-F238 in the populations of Laos and the very low frequencies in Tibetan and other populations, except the Han Chinese of China, suggest that the mutation F238 probably occurred in Proto-Han-Chinese about 7.0 KYA. As the STR diversities of O3a1c2-F238 between Northwestern Han and Eastern Han are nearly equal, it is difficult to tell exactly where this mutation originated.

In conclusion, our genetic evidence of Y chromosome haplogroup O3a1c-002611 outlines a clear picture of the dispersal and expansion patterns of Han Chinese in the late Neolithic Age. Haplogroup O3a1c-002611 and O3a1c1-F11 started their northward migration about 12 KYA from Southeast Asia, along with other O3-M122 lineages, and reached the upper and middle Yellow River basin. About 7 KYA, haplogroup O3a1c2-F238 originated in the ancestors of modern Sino-Tibetan populations. About 6 KYA, the Han Chinese split from the Proto-Sino-Tibetan, and started their migration to the east and south (Su et al., 2000b). About 5 KYA, haplogroup O3a1c1-F11 experienced rapid expansion, probably in the Eastern Han Chinese, with recent gene flow with surrounding populations and eventually became prevalent in different ethnic groups in East Asia.

References

Bandelt HJ, Forster P, Rohl A. 1999. Median-joining networks for inferring intraspecific phylogenies. Molecular Biology and Evolution 16: 37–48.

Excoffier L, Laval G, Schneider S. 2005. Arlequin (version 3.0): An integrated software package for population genetics data analysis. Evolutionary Bioinformatics Online 1: 47–50.

Goedbloed M, Vermeulen M, Fang RN, Lembring M, Wollstein A, Ballantyne K, Lao O, Brauer S, Krüger C, Roewer L, Lessig R, Ploski R, Dobosz T, Henke L, Henke J, Furtado MR, Kayser M. 2009. Comprehensive mutation analysis of 17 Y-chromosomal short tandem repeat polymorphisms included in the AmpFlSTR Yfiler PCR amplification kit. International Journal of Legal Medicine 123: 471–482.

Karafet TM, Hallmark B, Cox MP, Sudoyo H, Downey S, Lansing JS, Hammer MF. 2010. Major east-west division underlies Y chromosome stratification across Indonesia. Molecular Biology and Evolution 27: 1833–1844.

Ramakrishnan U, Mountain JL. 2004. Precision and accuracy of divergence time estimates from STR and SNPSTR variation. Molecular Biology and Evolution 21: 1960–1971.

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

Shi H, Dong YL, Wen B, Xiao CJ, Underhill PA, Shen PD, Chakraborty R, Jin L, Su B. 2005. Y-chromosome evidence of southern origin of the East Asian-specific haplogrou p O3-M122. American Journal of Human Genetics 77: 408–419.

Su B, Jin L, Underhill P, Martinson J, Saha N, McGarvey ST, Shriver MD, Chu J, Oefner P, Chakraborty R, Deka R. 2000a. Polynesian origins: Insights from the Y chromosome. Proceedings of the National Academy of Sciences USA 97: 8225–8228.

Su B, Xiao C, Deka R, Seielstad MT, Kangwanpong D, Xiao J, Lu D, Underhill P, Cavalli-Sforza L, Chakraborty R, Jin L. 2000b. Y chromosome haplotypes reveal prehistorical migrations to the Himalayas. Human Genetics 107: 582–590.

Su B, Xiao J, Underhill P, Deka R, Zhang W, Akey J, Huang W, Shen D, Lu D, Luo J, Chu J, Tan J, Shen P, Davis R, Cavalli-Sforza L, Chakraborty R, Xiong M, Du R, Oefner P, Chen Z, Jin L. 1999. Y-chromosome evidence for a northward migration of modern humans into Eastern Asia during the last Ice Age. American Journal of Human Genetics 65: 1718–1724.

Wang ZH. 1994. History of nationalities in China. Beijing: China Social Science Press.

Wen B, Li H, Lu D, Song X, Zhang F, He Y, Li F, Gao Y, Mao X, Zhang L, Qian J, Tan J, Jin J, Huang W, Deka R, Su B, Chakraborty R, Jin L. 2004a. Genetic evidence supports demic diffusion of Han culture. Nature 431: 302–305.

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. 2004b. Analyses of genetic structure of Tibeto-Burman populations reveals sex-biased admixture in southern Tibeto-Burmans. American Journal of Human Genetics 74: 856–865.

Yan S, Wang C-C, Li H, Li S-L, Jin L. Genographic Consortium. 2011. An updated tree of Y-chromosome Haplogroup O and revised phylogenetic positions of mutations P164 and PK4. European Journal of Human Genetics 19: 1013–1015.

Zhivotovsky LA. 2001. Estimating divergence time with the use of microsatellite genetic distances: Impacts of population growth and gene flow. Molecular Biology and Evolution 18: 700–709.

Zhivotovsky LA, Underhill PA, Cinnio lu C, Kayser M, Morar B, Kivisild T, Scozzari R, Cruciani F, Destro-Bisol G, Spedini G, Chambers GK, Herrera RJ, Yong KK, Gresham D, Tournev I, Feldman MW, Kalaydjieva L. 2004. The effective mutation rate at Y chromosome short tandem repeats, with application to human population-divergence time. American Journal of Human Genetics 74: 50–61.

Supplementary Material

The following supplementary materials are available online for this article at http://onlinelibrary.wiley.com/doi/10.1111/j.1759-6831.2012.00244.x/suppinfo:

Doc. S1. Population information and primers designed for newly typed SNPs.

Table S1. Frequencies of haplogroup O3a1c-002611 and its subhaplogroups.

Table S2. Detailed sample information, Y-SNP and Y-STR genotype data.

Appendix I Members of the Genographic Consortium

Janet S. ZIEGLE (Applied Biosystems, Foster City, CA, USA)

Li JIN (Fudan University, Shanghai, China)

Hui LI (Fudan University, Shanghai, China)

Shilin LI (Fudan University, Shanghai, China)

Pandikumar SWAMIKRISHNAN (IBM, Somers, NY, USA)

Laxmi PARIDA (IBM, Yorktown Heights, NY, USA)

Daniel E. PLATT (IBM, Yorktown Heights, NY, USA)

Ajay K. ROYYURU (IBM, Yorktown Heights, NY, USA)

Lluis QUINTANA-MURCI (Institut Pasteur, Paris, France)

R. John MITCHELL (La Trobe University, Melbourne, Australia)

Marc HABER (Lebanese American University, Beirut, Lebanon)

Pierre A. ZALLOUA (Lebanese American University, Beirut, Lebanon)

Syama ADHIKARLA (Madurai Kamaraj University, Madurai, India)

ArunKumar GANESHPRASAD (Madurai Kamaraj University, Madurai, India)

Ramasamy PITCHAPPAN (Madurai Kamaraj University, Madurai, India)

Varatharajan Santhakumari ARUN (Madurai Kamaraj University, Madurai, India)

R. Spencer WELLS (National Geographic Society, Washington DC, USA)

Himla SOODYALL (National Health Laboratory Service, Johannesburg, South Africa)

Elena BALANOVSKA (Research Centre for Medical Genetics, Russian Academy of Medical Sciences, Moscow, Russia)

Oleg BALANOVSKY (Research Centre for Medical Genetics, Russian Academy of Medical Sciences, Moscow, Russia)

Chris TYLER-SMITH (The Wellcome Trust Sanger Institute, Hinxton, UK)

Daniela R. LACERDA (Universidade Federal de Minas Gerais, Belo Horizonte, Brazil)

Fabrício R. SANTOS (Universidade Federal de Minas Gerais, Belo Horizonte, Brazil)

Pedro Paulo VIEIRA (Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brazil)

Jaume BERTRANPETIT (Universitat Pompeu Fabra, Barcelona, Spain)

David COMAS (Universitat Pompeu Fabra, Barcelona, Spain)

Begoña MARTÍNEZ-CRUZ (Universitat Pompeu Fabra, Barcelona, Spain)

David F. SORIA-HERNANZ (Universitat Pompeu Fabra, Barcelona, Spain)

Christina J. ADLER (University of Adelaide, Adelaide, Australia)

Alan COOPER (University of Adelaide, Adelaide, Australia)

Clio S. I. Der SARKISSIAN (University of Adelaide, Adelaide, Australia)

Wolfgang HAAK (University of Adelaide, Adelaide, Australia)

Matthew E. KAPLAN (University of Arizona, Tucson, AZ, USA)

Nirav C. MERCHANT (University of Arizona, Tucson, AZ, USA)

Colin RENFREW (University of Cambridge, Cambridge, UK)

Andrew C. CLARKE (University of Otago, Dunedin, New Zealand)

Elizabeth A. MATISOO-SMITH (University of Otago, Dunedin, New Zealand)

Matthew C. DULIK (University of Pennsylvania, Philadelphia, PA, USA)

Jill B. GAIESKI (University of Pennsylvania, Philadelphia, PA, USA)

Amanda C. OWINGS (University of Pennsylvania, Philadelphia, PA, USA)

Theodore G. SCHURR (University of Pennsylvania, Philadelphia, PA, USA)

Miguel G. VILAR (University of Pennsylvania, Philadelphia, PA, USA)