

SHORT REPORT

Refined phylogenetic structure of an abundant East Asian Y-chromosomal haplogroup O*-M134

Chao Ning^{1,5}, Shi Yan^{2,5}, Kang Hu³, Yin-Qiu Cui^{*,1} and Li Jin^{*,2,4}

The human Y-chromosome haplogroup O-M134 is one of the most abundant paternal lineages in East Asian populations, comprising ~13% of Han Chinese males, and also common in Kazakh, Korean, Japanese, Thai and so on. Despite its considerable prevalence, its current substructure is poorly resolved with only one downstream marker (M117) previously investigated. Here we address this deficiency by investigating some single-nucleotide polymorphisms (SNPs) previously reported being potentially associated with O-M134 based on high-throughput DNA-sequencing data. Using a panel of 1301 Chinese males we first identified 154 haplogroup O-M134 subjects. We then investigated the phylogenetic structure within this haplogroup using 10 SNPs (F444, F629, F3451, F46, F48, F209, F2887, F3386, F1739 and F152). Two major branches were identified, O-M117 and O-F444 and the latter was further divided into two main subclades, O-F629 and O-F3451, accounting for 10.84 and 0.92% of the Han Chinese, respectively. This update of O-M134 diversification permits better resolution of male lineages in population studies of East Asia.

European Journal of Human Genetics (2016) 24, 307–309; doi:10.1038/ejhg.2015.183; published online 26 August 2015

INTRODUCTION

Y chromosome (NRY) is now a widely used tool to trace the origin and migration events of modern humans.^{1–5} Detailed characterization of Y-chromosome phylogeny has permitted researchers to reveal previously inaccessible details of human demographic history.^{6–8} Y-chromosome haplogroup O-M175 is the prominent lineage in East Asia populations, comprises ~75% of the male populations in China^{2,9–11} and ~87% in Southeast Asia.^{12–14} O-M134 is one of the most frequent sub-lineages of O-M175, which represent ~30 million people in East Asia (Supplementary Figure S1). The frequency of O*-M134(xM117) varies in different regions, namely China (~11.4%),¹² Japan (~3.6%),¹⁵ Korea (~9.6%),¹⁶ Mongolia (~4.1%),¹⁷ Thailand (~11.6%)¹⁸ and Khasi in India (~16.6%).¹⁷ This paragroup also has been found at high frequency in some Kazakhs (26.1%).¹⁹ Despite its abundance and wide distribution, the phylogeny of O-M134 has not been adequately resolved with respect to O-002611²⁰ and O-M95.²¹ To date, the only marker investigated in literatures internal to O-M134 has been M117, and this was not sufficient to resolve the phylogeny of the populations belonging to this haplogroup.²² New data from high-throughput sequencing allowed the identification of more explicit branching structure within the entire Y-chromosome tree.²³ More than 200 potential single-nucleotide polymorphisms (SNPs) of haplogroup O-M134 were discovered by hybridization capture and Illumina sequencing of 22 Han Chinese male Y chromosomes.²³ Using 10 SNPs (F444, F629, F3451, F46, F48, F209, F2887, F3386, F1739 and F152), we genotyped

>1300 Chinese males and updated the phylogenetic structure of haplogroup O-M134.

MATERIALS AND METHODS

We collected a total of 1301 blood samples of unrelated males with the approval of the Ethics Committee of Biological Research at Fudan University and with the signed informed consent of the donors. DNA was extracted, and genotyped at M117 and M134. Samples belonging to O*-M134(xM117) were then genotyped at F444, F629, F3451, F46, F48, F209, F2887, F3386, F1739 and F152 using the SNaPshot multiplex kit (ABI, Carlsbad, CA, USA). This information was submitted to dbSNP (<http://www.ncbi.nlm.nih.gov/projects/SNP/>), and the assigned accession numbers are ss1712295747–ss1712295756. The amplification and extension primers are listed in Supplementary Table S1. We also typed 17 commonly used Y-STR (short tandem repeat) markers: DYS19, DYS389I, DYS389b, DYS390, DYS391, DYS392, DYS393, DYS437, DYS438, DYS439, DYS448, DYS456, DYS458, DYS635, H4, DYS385a and DYS385b using fluorescence labeled primers (AmpFLSTR Yfiler PCR Amplification Kit, Life Technologies, Carlsbad, CA, USA). The results were read from a 3730 sequencer (Supplementary Table S2), and ages of each clade was estimated from the median-joining networks within Network 4.6.1.1 (Fluxus) using the rho statistic. Both the evolutionary (6.9×10^{-4} per STR per generation) and the observed genealogical (2.1×10^{-3} per STR per generation) mutation rate were applied in this study,^{24,25} assuming a generation time of 25 years. A Contour map was drawn based on the frequency distribution of Haplogroup O*-M134(xM117) generated in this study and previously published data (Supplementary Table S3) using the

¹School of Life Sciences, Jilin University, Changchun, China; ²Ministry of Education, Key Laboratory of Contemporary Anthropology and Center for Evolutionary Biology, School of Life Sciences and Institutes of Biomedical Sciences, Fudan University, Shanghai, China; ³Key Laboratory of High Altitude Environment and Gene Related to Disease of Tibet Ministry of Education, Tibet University for Nationalities, Xianyang, China; ⁴Chinese Academy of Sciences Key Laboratory of Computational Biology, CAS-MPG Partner Institute for Computational Biology, SIBS, CAS, Shanghai, China

*Correspondence: Dr Y-Q Cui, School of Life Sciences, Jilin University, 2699 Qianjing Street, Changchun, Jilin 130012, China. Tel: +86 431 85155345; Fax: +86 431 85155127; E-mail: cuiyq@jlu.edu.cn or: Professor Dr L Jin, Ministry of Education, Key Laboratory of Contemporary Anthropology and Center for Evolutionary Biology, School of Life Sciences, Fudan University, Shanghai 200438, China. Tel: +86 21 65642800; Fax: +86 21 55663714; E-mail: lijn.fudan@gmail.com

⁵These authors contributed equally to this work.

Received 31 December 2014; revised 2 July 2015; accepted 5 July 2015; published online 26 August 2015

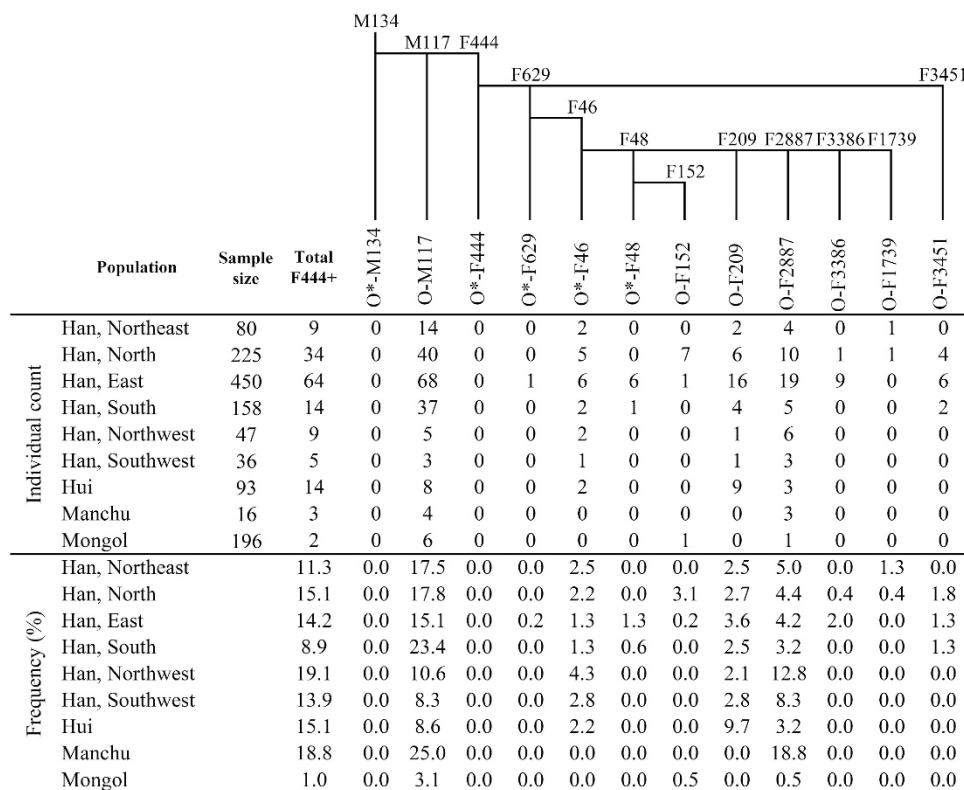


Figure 1 Refined phylogenetic tree of Haplogroup O*-M134, and the frequency of its sub-haplogroups genotyped in this study. 'East' refers to the samples origins from the provinces of Jiangsu, Zhejiang, Anhui and Shanghai. 'South' refers to provinces of Guangdong, Hunan, Hubei and Fujian. 'North' refers to provinces of Hebei, Henan, Shandong, Shanxi, Tianjin and Beijing. 'Northwest' refers to provinces of Shaanxi and Gansu. 'Southwest' refers to Guizhou, Sichuan and Chongqing. 'Northeast' refers to provinces of Jilin, Heilongjiang and Inner Mongolia.

Kriging procedure with the aid of the Surfer 8.0 Software (Golden Software, Golden, CO, USA).

RESULTS

Of the 1301 male individuals, 154 (~12%) were identified as belonging to O*-M134(xM117), in agreement with previous studies of East Asian populations.^{9,12,22} We genotyped 10 SNPs (F444, F629, F3451, F46, F48, F209, F2887, F3386, F1739 and F152) in these O*-M134(xM117) individuals, and used results to refine the O-M134 phylogenetic tree (Figure 1). We investigate the recently described O-F444 subclade within O-M134. This subgroup is parallel to O-M117, and all the novel SNPs were found in subgroup O-F444, with no individuals falling in Haplogroup O*-M134(xM117). This indicates two distinct founders within the O-M134 membership of mainland China.²² Furthermore, all the O-F444 samples tested were derived at F629 or F3451, and these two subclades had different geographic distributions (Supplementary Figure S1). Clade O-F3451 was confined to Han Chinese in the North, South and East, whereas clade O-F629 was found in all the populations of this study. The majority of this clade belong to a lineage marked by the F46 mutation (Figure 1), and the remaining part is O*-F629(xF46). In total, five parallel SNPs (F48, F209, F2887, F3386 and F1739) were assigned to downstream the O-F46 branch, the star-like structure of the O-F46 lineage indicating a male population expansion in East Asia.²³ Of the six O-F46 lineages (O*-F46, O-F48, O-F209, O-F2887, O-F3386, O-F1739), the two haplogroups, O-F209 and O-F2887, were by far the

most abundant in the Han Chinese in this study (~3 and ~4.2%, respectively).

The STR data (Supplementary Table S2) allowed us to estimate the coalescence times of the O-F444 subclades in East Asians. Despite long-standing debates about the accuracy of STRs for time estimation,^{25,26} the times proposed here provide at least a frame of reference (Table 1), more accurate times will be confirmed in further study on dense sequencing. Two sets of commonly used mutation rates were applied in this study, one evolutionary rate (6.9×10^{-4} per STR per generation) and the other genealogical rate (2.1×10^{-3} per STR per generation). The ages estimated using evolutionary rate is three times higher than using genealogical rate. This phenomenon has been observed before and was explained by selection on deleterious mutations.²⁷ However, owing to the better fit of genealogical rate than mutation rate with sequence-based time estimations,^{25,28} we pose our time here on genealogical rate. The coalescent time of O-F444 in the Han Chinese was 7.99 ± 1.36 kya (thousand years ago). The TMRCA (time of most recent common ancestor) of the O-F46 lineage estimated was 7.76 ± 1.32 kya, which was within the time estimated from high-throughput sequencing on Y chromosomes SNPs.²³

DISCUSSION

Haplogroup O-M134 occurs most frequently in East Asians.^{9,29} The genotyping of 10 potential SNPs under Haplogroup O-M134 enables us to refine and update the phylogeny of this lineage, and revealed that the internal structure of this haplogroup may be used to resolve the position of a large proportion of individuals in mainland China.

Table 1 Calculated coalescence times (thousand years) of Y-chromosome haplogroup O-F444 and its sub-haplogroups in Han Chinese

Haplogroup	Coalescence time (based on evolutionary rate ^a)	Coalescencetime (based on genealogical rate ^b)
O-F444	24.3±4.15	7.99±1.36
O-F629	17.7±3.12	5.82±1.07
O-F46	23.3±3.98	7.76±1.32
O-F48	16.2±1.76	5.33±0.58
O-F152	5.04±1.51	1.65±0.50
O-F209	14.4±2.16	4.70±0.71
O-F2287	17.2±2.70	5.67±0.89
O-F3386	9.16±2.11	3.01±0.74
O-F1739	—	—
O-F3451	24.9±3.91	8.17±1.28

— represents not calculated owing to small sample size.

^aEvolutionary rate: 0.00069 per locus per 25 years (see Zhivotovsky LA *et al*²⁵).

^bGenealogical rate: 0.0021 per locus per locus per 25 years (see Ge JY *et al*²⁴).

This enables more fine-scaled estimations of patrilineal origin and will also be useful for forensic studies.

The general outline structure of O-M134 consists of bifurcations, with the exception of the star-like structure of O-F46 and O-F8, indicating demographic expansion.²³ So far, estimation of TMRCA was performed following the stepwise model, a model fit well for Y STRs. However, owing to the uncertainty mutation rate for STR markers and possible deviation from stepwise model, the time estimated from STRs should be regarded just as a reference, despite its prevalence in time calculation. The time estimated from STRs should be regard just as a reference, despite its prevalence in time calculation. The estimated TMRCA of the O-F46 lineage was 7.76 ± 1.32 kya, which is a little different but within the time estimated from high-throughput sequencing of Y-chromosome SNPs.²³ This date is a little after the shift from a hunter-gather subsistence to intensive agriculture, for example, the Yangshao Culture (6.9–4.9 kya) in the Central Yellow Basin, and the Majiayao Culture (6.0–4.9 kya) in the Upper Yellow River Basin.³⁰ Yan *et al*²³ suggested that crop harvests may have provided a more stable source of food than hunting and gathering, enabling populations to reach greater densities.

Despite the prevalence of O-F444 across East Asian and Southeast Asian populations,^{12,15–19} this haplogroup has also been found with moderate frequency in Qiangic populations (~5%).²⁸ The Qiangic populations might participate in the establishing of the Sino-Tibetan populations based on recent genetic evidences.^{17,18,28} Detailed characterization of O-F444 may provide a broader framework of Sino-Tibetan populations.

In sum, our study has greatly improved the resolution of East Asian Y chromosomes. With the more widespread use of high-throughput DNA sequencing has decreased markedly the time and expense of DNA sequencing. The discovery of novel bi-allelic markers will permit improved resolution of the Y-chromosome phylogenetic tree, and help us understand better the evolutionary history of populations in East Asia.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGEMENTS

This research was supported by the grants from the National Science Foundation of China (Grant No. 31271338 and 31371266). We thank Professor Erika Hagelberg for help in editing this manuscript.

- 1 Jobling MA, Tyler-Smith C: Fathers and sons: the Y chromosome and human evolution. *Trends Genet* 1995; **11**: 499–456.
- 2 Su B, Xiao CJ, Deka R *et al*: Y chromosome evidence for a northward migration of modern humans into East Asia during the last Ice Age. *Am J Hum Genet* 1999; **65**: 1718–1724.
- 3 Chiaroni J, Underhill PA, Cavalli-Sforza LL: Y chromosome diversity, human expansion, drift, and cultural evolution. *Proc Natl Acad Sci USA* 2009; **106**: 20174–20179.
- 4 Underhill PA, Kivisild T: Use of Y chromosome and mitochondrial DNA population structure in tracing human migrations. *Annu Rev Genet* 2007; **41**: 539–564.
- 5 Karafet TM, Mendez FL, Meilerman MB, Underhill PA, Zegura SL, Hammer MF: New binary polymorphisms reshape and increase resolution of the human Y chromosomal haplogroup tree. *Genome Res* 2008; **18**: 830–838.
- 6 Jin L, Su B: Natives or immigrants: modern human origin in East Asia. *Nat Rev Genet* 2000; **1**: 126–133.
- 7 Sykes B, Irven C: Surnames and the Y chromosome. *Am J Hum Genet* 2000; **66**: 1417–1419.
- 8 Jobling MA, Tyler-Smith C: The human Y chromosome: an evolutionary marker comes of age. *Nat Rev Genet* 2003; **4**: 598–612.
- 9 Shi H, Dong YL, Wen B *et al*: Y-chromosome evidence of southern origin of the East Asian -specific haplogroup O3-M122. *Am J Hum Genet* 2005; **77**: 408–419.
- 10 Wen B, Li H, Lu D *et al*: Genetic evidence supports demic diffusion of Han culture. *Nature* 2004; **431**: 302–305.
- 11 Zhong H, Shi H, Qi X-B *et al*: Extended Y chromosome investigation suggests postglacial migrations of modern humans into East Asia via the northern route. *Mol Biol Evol* 2011; **28**: 717–727.
- 12 Cai X, Qin Z, Wen B *et al*: Human migration through Bottlenecks from Southeast Asia into East Asia during last glacial maximum revealed by Y chromosomes. *PLoS One* 2011; **6**: e24282.
- 13 Li H, Wen B, Chen S-J *et al*: Paternal genetic affinity between Western Austronesians and Daic populations. *BMC Evol Biol* 2008; **8**: 146.
- 14 Karafet TM, Hallmark B, Cox MP *et al*: Major East-West division underlies Y chromosome stratification across Indonesia. *Mol Biol Evol* 2010; **27**: 1833–1844.
- 15 Nonaka I, Minaguchi K, Takezaki N: Y-chromosomal binary haplogroups in the Japanese population and their relationship to 16 Y-STR polymorphisms. *Ann Hum Genet* 2007; **71**: 480–495.
- 16 Jin H-J, Tyler-Smith C, Kim W: The peopling of Korea revealed by analyses of mitochondrial DNA and Y-chromosomal markers. *PLoS One* 2009; **4**: e4210.
- 17 Kang L, Lu Y, Wang C *et al*: Y-chromosome O3 haplogroup diversity in Sino-Tibetan populations reveals two migration routes into the Eastern Himalayas. *Ann Hum Genet* 2012; **76**: 92–99.
- 18 Su B, Xiao CJ, Deka R *et al*: Y chromosome haplotypes reveal prehistorical migrations to the Himalayas. *Hum Genet* 2000; **107**: 582–590.
- 19 Dulik MC, Osipova LP, Schurr TG: Y-chromosome variation in Altaian Kazakhs reveals a common paternal gene pool for Kazakhs and the influence of Mongolian expansions. *PLoS One* 2011; **6**: e17548.
- 20 Wang C-C, Yan S, Qin Z-D *et al*: Late Neolithic expansion of ancient Chinese revealed by Y chromosome haplogroup O3a1c-002611. *J Syst Evol* 2013; **51**: 280–286.
- 21 Zhang X, Kampuansai J, Qi X *et al*: An updated phylogeny of the human Y-chromosome lineage O2a-M95 with novel SNPs. *PLoS One* 2014; **9**: e101020.
- 22 Yan S, Wang CC, Li H, Li SL, Jin L, Genographic C: An updated tree of Y-chromosome haplogroup O and revised phylogenetic positions of mutations P164 and PK4. *Eur J Hum Genet* 2011; **19**: 1013–1015.
- 23 Yan S, Wang C-C, Zheng H-X *et al*: Y Chromosomes of 40% Chinese descend from three Neolithic super-grandfathers. *PLoS One* 2014; **9**: e105691.
- 24 Ge JY, Budowle B, Aranda XG, Planz JV, Eisenberg AJ, Chakraborty R: Mutation rates at Y chromosome short tandem repeats in Texas populations. *Forensic Sci Int Genet* 2009; **3**: 179–184.
- 25 Zhivotovsky LA, Underhill PA, Cinnioglu C *et al*: The effective mutation rate at Y chromosome short tandem repeats, with application to human population-divergence time. *Am J Hum Genet* 2004; **74**: 50–61.
- 26 Di Giacomo F, Luca F, Popa LO *et al*: Y chromosomal haplogroup J as a signature of the post-neolithic colonization of Europe. *Hum Genet* 2004; **115**: 357–371.
- 27 Penny D: Evolutionary biology - relativity for molecular clocks. *Nature* 2005; **436**: 183–184.
- 28 Wang C-C, Wang L-X, Shrestha R *et al*: Genetic structure of Qiangic populations residing in the Western Sichuan corridor. *PLoS One* 2014; **9**: e103772.
- 29 Xue YL, Zejal T, Bao WD *et al*: Male demography in East Asia: a north-south contrast in human population expansion times. *Genetics* 2006; **172**: 2431–2439.
- 30 The Institute of Archaeology Chinese Academy of Social Sciences. *Chinese Archaeology—Neolithic*. Social Sciences Press: Beijing, China, 2010.