**Analysis of Y-chromosome short tandem repeat loci on 1,082 Nantong Han
individuals in eastern China**

Sixian Xu^{1†}, Shuping Yang^{1†}, Min Yang^{1,2}, Dongtao Jia², Xuefeng Han³, Wei Wang⁴,
Li Jin¹, Liming Li^{1*}, Shilin Li^{1*}

¹Ministry of Education Key Laboratory of Contemporary Anthropology and State Key
Laboratory of Genetic Engineering, Collaborative Innovation Center for Genetics and
Development, School of Life Sciences, Fudan University, Shanghai, China

²Nantong Bureau of Public Safety, Jiangsu, China

³Department of Radiology, Yangpu Hospital, Tongji University School of Medicine,
Shanghai, China

⁴Jilin Province Cancer Hospital, Changchun, China

[†]The two authors contributed equally to the work.

***Corresponding author:**

Dr. Liming Li, School of Life Sciences, Fudan University,
2005 Songhu Road, Shanghai 200438, China

Tel: 8602151630602

Fax: 8602151630607

Email: lee6013282@gmail.com

Dr. Shilin Li, School of Life Sciences, Fudan University,
2005 Songhu Road, Shanghai 200438, China

Tel: 8602151630602

Fax: 8602151630607

Email: lishilin@fudan.edu.cn

Analysis of Y-chromosome short tandem repeat loci on 1,082 Nantong Han individuals in eastern China

Dear Editor,

Nantong is a prefecture-level city in Jiangsu Province in eastern China. It is located on the north bank of the Yangtze River. According to the 2010 census, the population of Nantong had reached 7.3 million, with Han Chinese as the largest ethnic group (99.7%). Nantong was historically known as Tongzhou City. For hundreds of years, it was a central avenue between northern and southern China for both land and water travel. According to the historical records, there were several waves of migration into Nantong from both northern and southern China, which gave the area its four dialects [1]. It is here speculated that the Han Chinese living in Nantong may also be a combination of the northern and southern China genetically.

In the present study, a total of 1,082 peripheral blood samples were collected from unrelated male Han individuals who had given informed consent. All participants' families had lived in the Nantong area for more than three generations. The samples were gathered from three counties in Nantong Tongzhou (n = 473), Haian (n = 518) and Gangzha (n = 91) (Figure S1). The phenol-chloroform method was used to extract genomic DNA and polymerase chain reaction (PCR) was performed with the AmpF/STR Yfiler™ PCR Amplification Kit (Thermo Fisher Scientific Company, Carlsbad, USA) in the GeneAmp PCR System 9700 (Thermo Fisher Scientific Company). A total of 17 Y-STR loci were analyzed, including DYS19, DYS389I, DYS389II,

DYS385a/b, DYS390, DYS391, DYS392, DYS393, DYS437, DYS438, DYS439, DYS448, DYS456, DYS458, DYS635, and YGATAH4. PCR products were separated by capillary electrophoresis in ABI PRISM 3130xL Genetic Analyzer (Thermo Fisher Scientific Company). GeneMapper ID software v3.2 (Thermo Fisher Scientific Company) was used for genotype assignment. DNA typing and assignment of nomenclature were based on ISFG recommendations [2].

Haplotype diversity was estimated using the following formula: $HD = n(1 - \sum P_i^2)/(n - 1)$, where n is the number of samples, P_i is the frequency of the i th haplotype. Gene diversity was estimated for each Y-STR locus, and the formula was similar to the one used for HD: $GD = n(1 - \sum p_i^2)/(n - 1)$, where p_i is the frequency of the i th allele) [3].

The Nantong population was divided into 3 sub-parts based on the collecting location: Tongzhou, Haian and Gangzha. Pairwise Rst (P -values, 10,000 permutations) between the populations of the present study and other reports involving Chinese populations using Arlequin v3.5 were estimated [4]. Because of the complex repeat structure, DYS385 a/b was excluded from the analysis. Because DYS389 is a special marker that has two different values at the same location, the two values were designated “DYS389I” and “DYS389b” (=DYS389II-DYS389I) for the analysis. A total of 9,660 reference samples were obtained from earlier reports, including twelve Chinese Han (Anhui [5], Jiangsu [5], Jiangxi [5], Shandong [5], Shanghai [5], Zhejiang_1 [5], Gansu [6], Nanyang [7], Luzhou [8], Mudanjiang [9], Shanxi [10], and Zhejiang [11]) and nine other Chinese ethnic populations (Mongolian [12], Xibe [13],

Lassa Tibetan [14], Qinghai Tibetan [15], Kazakh [16], Uyghur [16], Manchu [17], Taiwan-Paiwan [18], and Tujia [19]). The results were visualized by multidimensional scaling plot (MDS) using R v3.1.2 (<http://www.r-project.org>). A neighbor-joining phylogenetic tree was constructed for the Nantong population and the reference populations based on a distance matrix of Rst using Mega 5.10 [20].

A total of 879 different haplotypes were observed among 1,082 individuals of Nantong Han population, of which 777 were unique, 54 were shared by two individuals, 26 by three individuals, 8 by four individuals, 7 by five, 2 by six, 1 by seven, 3 by eight and 1 by nine (Table S1). Haplotype diversity reached 0.9993, and gene diversities ranged from 0.2992 (DYS438) to 0.9636 (DYS385a/b).

Table S2 shows the Rst values calculated to measure genetic distances between 17 Yfiler haplotypes of 24 populations that showed statistical significance. Results showed that Nantong had no significant difference from most Han populations including those of Anhui, Jiangsu, Jiangxi and Shanghai after Bonferroni correction ($P > 0.05/276 = 0.000181$, 276 was the number of tests), but they did show significant differences between those of Shanxi, Gansu and Luzhou. Results also showed Nantong Han to differ significantly from most non-Han Chinese ethnic populations, including Mongolian, Xibe, Lassa Tibetan, Qinghai Tibetan, Kazakh, Uyghur and Taiwan-Paiwan. The Manchu and Tujia ethnic populations showed the smallest genetic distance from Nantong Han populations of any Chinese minority ethnic groups. This is probably because their ancestors mixed more with Han Chinese during their early settlement. Results showed that the Nantong Tongzhou, Haian and

Gangzha subpopulations showed distinct and significant differences from several populations like those of Shandong, Zhejiang, Mudanjiang and Nanyang. The Gangzha subpopulation was more similar to Zhejiang. Haian was a little more similar to Shandong, Nanyang and Mudanjiang. Tongzhou showed significant differences from Haian. The same conclusion could be observed in Figure S2. Generally speaking, Nantong Han are genetically more similar to other Han populations (especially southern ones), and it was found to be distinct from other non-Han populations.

Previous studies have shown Chinese Han populations to cluster mainly in northern and southern Han [21]. In the present study, 17 Y-STR loci were analyzed on 3 Han subpopulations collected from Nantong, which was historically an important channel between the northern and southern China. As shown in Figure S3, the populations were divided into two groups according to their geographical locations including northern and southern China. Although all of Nantong subpopulations were located on the southern bunch, Haian was more similar to northern Han than to Tongzhou and Gangzha.

In conclusion, the haplotype frequencies and genetic parameters of 17 Y-STR loci of the Nantong Han population is here reported. The data provide a reference for a Y-STR database of Nantong and may be valuable for population genetic analysis.

These data were submitted to YHRD, and the accession numbers were obtained: Nantong, China [Han], n=1,082, YA004138.

This study was supported by grants from the National Science Foundation of

China (31271338), Project of Chinese Ministry of Education (113022A), Twelve-Five Science and Technology Support Program (2012ZX10001003-002-023) and National High Technology Research and Development Program (2012AA021802).

The study follows the new guidelines for publication of population data requested by the journal [22].

REFERENCES

1. Huang Y. On Essential Features of Nantong's History and Culture. *Journal of Nantong Teachers College (Social Sciences Edition)* 2004, 20(4). (In Chinese)
2. Gusmao L, Butler JM, Carracedo A, Gill P, Kayser M, Mayr WR, Morling N, Prinz M, Roewer L, Tyler-Smith C *et al*: DNA Commission of the International Society of Forensic Genetics (ISFG): An update of the recommendations on the use of Y-STRs in forensic analysis. *Forensic Sci Int* 2006, 157(2-3):187-197.
3. Nei M: Molecular evolutionary genetics. New York: Columbia University Press; 1987, pp. 176-179.
4. Excoffier L, Lischer HEL: Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Mol Ecol Resour* 2010, 10(3):564-567.
5. Li L., Yu G., Li S., Jin L., Yan S. Genetic analysis of 17 Y-STR loci from 1019 individuals of six Han populations in East China. *Forensic Sci Int Genet* 2015, 20:101-102.
6. Li L., Ma T., Xu J., Yang Y., Yang S., Xia M. *et al*. Genetic analysis of 17 Y-STR loci in Han population from Gansu province, northwestern China. *Forensic Sci Int Genet*, 2015, 19:134-135.
7. Yang YY, Yuan WL, Guo F, Jiang XH: Population data of 17 Y-STR loci in Nanyang Han population from Henan Province, Central China. *Forensic Sci Int-Genet* 2014, 13:145-146.
8. Bing L, Liang WB, Pi JH, Zhang DM, Yong D, Luo HB, Zhang LS, Lin Z: Population genetics for 17 Y-STR loci(AmpFISTR (R) Y-filer™) in Luzhou Han ethnic group. *Forensic Sci Int-Genet* 2013, 7(2):E23-E26.
9. Liu Y, Liao LC, Gu MB, Ye Y: Population genetics for 17 Y-STR loci in a Chinese Han population sample from Mudanjiang city, Northeast China. *Forensic Sci Int-Genet* 2014, 13:E16-E17.
10. Bai RF, Zhang Z, Liang QZ, Lu D, Yuan L, Yang X, Shi MS: Haplotype diversity of 17 Y-STR loci in a Chinese Han population sample from Shanxi Province, Northern China. *Forensic Sci Int-Genet* 2013, 7(1):214-216.
11. Wu WW, Pan LP, Hao HL, Zheng XT, Lin JF, Lu DJ: Population genetics of 17 Y-STR loci in a large Chinese Han population from Zhejiang Province, Eastern China. *Forensic Sci Int-Genet* 2011, 5(1):E11-E13.

12. Guo F: Population genetics for 17 Y-STR loci in Mongolian ethnic minority from Liaoning Province, Northeast China. *Forensic Sci Int-Genet* 2015, 17:153-154.
13. Guo F, Zhang L, Jiang XH: Population genetics of 17 Y-STR loci in Xibe ethnic minority from Liaoning Province, Northeast China. *Forensic Sci Int-Genet* 2015, 16:86-87.
14. Li KL, Kai L, Yuming M: Y chromosome STR haplotypes of Tibetan living Tibet lassa. *Forensic SciInt* 2007, 172(1):79-83.
15. Zhu BF, Wu YM, Shen CM, Yang TH, Deng YJ, Xun X, Tian YF, Yan JC, Li T: Genetic analysis of 17 Y-chromosomal STRs haplotypes of Chinese Tibetan ethnic group residing in Qinghai province of China. *Forensic SciInt* 2008, 175(2-3):238-243.
16. Shan WJ, Ablimit A, Zhou WJ, Zhang FC, Ma ZH, Zheng XF: Genetic polymorphism of 17 Y chromosomal STRs in Kazakh and Uighur populations from Xinjiang, China. *Int J Legal Med* 2014, 128(5):743-744.
17. He J, Guo F: Population genetics of 17 Y-STR loci in Chinese Manchu population from Liaoning Province, Northeast China. *Forensic Sci Int-Genet* 2013, 7(3):E84-E85.
18. Wu FC, Ho CW, Pu CE, Hu KY, Willuweit S, Roewer L, Liu DH: Y-chromosomal STRs haplotypes in the Taiwanese Paiwan population. *Int J Legal Med* 2011, 125(1):39-43.
19. Yang YR, Jing YT, Zhang GD, Fang XD, Yan JW: Genetic analysis of 17 Y-chromosomal STR loci of Chinese Tujia ethnic group residing in Youyang Region of Southern China. *Leg Med* 2014, 16(3):173-175.
20. Hall BG. Building phylogenetic trees from molecular data with MEGA. *Mol Biol Evol* 2013, 30:1229-1235.
21. Xu S, Yin X, Li S, Jin W, Lou H, Yang L. *et al.* Genomic dissection of population substructure of Han Chinese and its implication in association studies. *Am J Hum Genet* 2009, 85: 762-774.
22. Carracedo A, Butler JM, Gusmao L, Linacre A, Parson W, Roewer L, Schneider PM: Update of the guidelines for the publication of genetic population data. *Forensic Sci Int-Genet* 2014, 10:A1-A2.