Hum Genet (2017) 136:485–497
DOI 10.1007/s00439-017-1759-x

ORIGINAL INVESTIGATION

Revisiting the male genetic landscape of China: a multi-center study of almost 38,000 Y-STR haplotypes

Michael Nothnagel[1] · Guangyao Fan[2] · Fei Guo[3] · Yongfeng He[4] · Yiping Hou[5] · Shengping Hu[6] · Jiang Huang[7] · Xianhua Jiang[8] · Wook Kim[9] · Kicheol Kim[10] · Chengtao Li[11] · Hui Li[12] · Liming Li[12] · Shilin Li[12] · Zhao Li[13] · Weibo Liang[5] · Chao Liu[14] · Di Lu[15] · Haibo Luo[5] · Shengjie Nie[15] · Meisen Shi[16] · Hongyu Sun[17] · Jianpin Tang[18] · Lei Wang[19] · Chuan-Chao Wang[20] · Dan Wang[21] · Shao-Qing Wen[12] · Hongyan Wu[22] · Weiwei Wu[23] · Jiaxin Xing[24] · Jiangwei Yan[25] · Shi Yan[12] · Hongbing Yao[26] · Yi Ye[5] · Libing Yun[5] · Zhaoshu Zeng[27] · Lagabaiyila Zha[28] · Suhua Zhang[11] · Xiufen Zheng[29] · Sascha Willuweit[30] · Lutz Roewer[30]

Received: 8 December 2016 / Accepted: 13 January 2017 / Published online: 30 January 2017

Abstract China has repeatedly been the subject of genetic studies to elucidate its prehistoric and historic demography. While some studies reported a genetic distinction between Northern and Southern Han Chinese, others showed a more clinal picture of small differences within China. Here, we investigated the distribution of Y chromosome variation along administrative as well as ethnic divisions in the mainland territory of the People's Republic of China, including 28 administrative regions and 19 recognized Chinese nationalities, to assess the impact of recent demographic processes. To this end, we analyzed 37,994 Y chromosomal 17-marker haplotype profiles from the YHRD database with respect to forensic diversity measures and genetic distance between groups defined by administrative boundaries and ethnic origin. We observed high diversity throughout all Chinese provinces and ethnicities. Some ethnicities, including most prominently Kazakhs and Tibetans, showed significant genetic differentiation from the Han and other groups. However, differences between provinces were, except for those located on the Tibetan plateau, less pronounced. This discrepancy is explicable by the sizeable presence of Han speakers, who showed high genetic homogeneity all across China, in nearly all studied provinces. Furthermore, we observed a continuous genetic North–South gradient in the Han, confirming previous reports of a clinal distribution of Y chromosome variation and being in notable concordance with the previously observed spatial

Electronic supplementary material The online version of this article (doi:10.1007/s00439-017-1759-x) contains supplementary material, which is available to authorized users.

✉ Michael Nothnagel
michael.nothnagel@uni-koeln.de

[1] Department of Statistical Genetics and Bioinformatics, Cologne Center for Genomics (CCG), University of Cologne, Weyertal 115b, 50931 Cologne, Germany

[2] Department of Public Security Technology, The Center for Forensic Science Research, Railway Police College, Zhengzhou 450053, People's Republic of China

[3] Department of Forensic Medicine, National Police University of China, Shenyang 110854, People's Republic of China

[4] Department of Criminal Investigation, Shaanxi Provincial Public Security Bureau, Xi'an 710016, People's Republic of China

[5] Institute of Forensic Medicine, West China School of Basic Science and Forensic Medicine, Sichuan University, Chengdu 610041, People's Republic of China

[6] Molecular Biology and Forensic Genetics Laboratory, Shantou University Medical College, Shantou, People's Republic of China

[7] Department of Forensic Medicine, Guizhou Medical University, Beijing Road, 9th, Guiyang 550004, People's Republic of China

[8] Liaoning Criminal and Science Technology Research Institute, Shenyang 110032, People's Republic of China

[9] Department of Biological Sciences, Dankook University, Cheonan 330-714, Republic of Korea

[10] Department of Neurology, University of California, San Francisco, San Francisco, CA 94158, USA

[11] Shanghai Key Laboratory of Forensic Medicine, Shanghai Forensic Service Platform, Institute of Forensic Sciences, Ministry of Justice, P.R. China, Shanghai 200063, People's Republic of China

[12] MOE Key Laboratory of Contemporary Anthropology, School of Life Sciences, Fudan University, Shanghai 200438, People's Republic of China

[13] Department of Criminal Investigation, Hebei Provincial Public Security Bureau, Shijiazhuang City 050000, People's Republic of China

distribution of autosomal variation. Our findings shed light on the demographic changes in China accrued by a fast-growing and increasingly mobile population.

Introduction

Y chromosomal haplotypes consisting of short tandem repeats (STRs) are routinely used in forensic laboratories throughout the world to match male trace donors to evidence or to analyze paternal genealogies (Roewer 2009). Calculation of a match probability requires an appropriate database to accurately estimate the probability of randomly drawing the suspect genetic profile from the population pool of DNA profiles (Caliebe and Krawczak 2016). Reference databases aiming to be representative for Y chromosome profiles, however, need to reflect the highly non-uniform distribution of Y-STRs as the principal marker type used in forensics among human world populations (Kayser et al. 2001; Purps et al. 2014). This stratification has been explained by the smaller effective population size of the Y chromosome compared to autosomes or the X chromosome, causing more rapid genetic drift. Moreover, due to the sex-specific inheritance pattern of the Y chromosome, drift can be greatly accentuated by social selection (Jobling 2012). Recent events such as migrations, range expansions or colonizations in historical times affect STR-based haplotypes much stronger than unique-event polymorphisms (SNPs), because the higher mutation rates of STRs can lead to high diversity of their haplotypes even among Y chromosomes that share a recent common ancestor. This property of Y-STRs has provided useful insights into the recent history of populations (Ploski et al. 2002; Weale et al. 2002; Xu et al. 2015; Xue et al. 2005). For example, the dissection of European Y-STR haplotypes into a Western and an Eastern cluster is a strong indication that territories populated by Neolithic population expansions (detectable by Y-SNPs) were converted into stable, delimited habitats by expanding populations which acquired in historical times a unifying language and culture (Kayser et al. 2005; Roewer et al. 2005). Remarkably, this Y chromosomal East–West gradient is in strong contrast with the North–South gradient that has been repeatedly observed for autosomal SNP data (Lao et al. 2008; Nothnagel et al. 2010; Novembre et al. 2008), explicable by differing movement patterns between males and females in the past and by the differing impact of timescales on the population structure in these two genetic systems. Based on a combination of Y-STRs and SNPs, another study found evidence of a northeast China origin of the Xibe who subsequently migrated to northwestern Xinjiang province in the eighteenth century (Powell et al. 2007). Sometimes, when expansions are too recent and have been too fast relative to the subsequent time of habitation, the correlation between genetic variation on the one hand and geography or language on the other can virtually disappear, as recently shown for a large number of autochthonous South American populations (Roewer et al. 2013).

[14] Guangzhou Forensic Science Institute, Guangzhou 510030, People's Republic of China

[15] School of Forensic Medicine, Kunming Medical University, Kunming 650500, People's Republic of China

[16] Center of Cooperative Innovation for Judicial Civilization, Institute of Evidence Law and Forensic Science, China University of Political Science and Law, Ministry of Education, Beijing 100088, People's Republic of China

[17] Department of Forensic Medicine, Zhongshan School of Medicine, Sun Yat-sen University, Guangzhou 510089, People's Republic of China

[18] Department of Forensic Medicine, Guangdong Medical University, Dongguan 523808, People's Republic of China

[19] Department of Forensic Sciences, Police Station of Zhengzhou, Zhengzhou, Henan 450008, People's Republic of China

[20] Department of Archaeogenetics, Max Planck Institute for the Science of Human History, Jena, Germany

[21] Institute of Forensic Medicine and Laboratory Medicine, Jining Medical University, Jining, Shandong, People's Republic of China

[22] Xinxiang Medical University School of Basic Medical, Xinxiang, Henan 453003, People's Republic of China

[23] Institute of Forensic Science, Zhejiang Provincial Public Security Bureau, Hangzhou 310009, People's Republic of China

[24] School of Forensic Medicine, China Medical University, Shenyang, People's Republic of China

[25] Beijing Institute of Genomics, Chinese Academy of Sciences, Beijing 100101, People's Republic of China

[26] Key Laboratory of Evidence Science of Gansu Province, Gansu Institute of Political Science and Law, Lanzhou 730070, People's Republic of China

[27] School of Basic Medical Sciences, Zhengzhou University, Zhengzhou, Henan 450052, People's Republic of China

[28] Forensic Science Department, School of Basic Medical Sciences, Central South University, Changsha 410013, People's Republic of China

[29] Department of Pathology, Department of Surgery, Department of Oncology, University of Western Ontario, Lawson Health Research Institute, London, Canada

[30] Department of Forensic Genetics, Institute of Legal Medicine and Forensic Sciences, Charité-Universitätsmedizin Berlin, Berlin, Germany

A very useful resource to study Y chromosome variation is the Y chromosome haplotype reference database (YHRD; https://yhrd.org). This online resource was set up roughly 15 years ago and contributed massive amounts of empirical data to decipher the Y-specific population differentiation and to quantify the impact of the latter on the frequency estimation process in forensics (Roewer 2016). China, with >41,000 haplotypes derived from local population studies, comprises the largest national database within YHRD (release 52).

China, being a country of large geographic extent as well as of enormous cultural and demographic impact, has repeatedly been the subject of genetic studies to elucidate its prehistoric and historic demography (Chu et al. 1998; Piazza 1998). Many of the more contemporary studies concentrated on Y-linked polymorphisms (Deng et al. 2004; Li et al. 2007; Qi et al. 2013; Wang et al. 2012; Wang and Li 2013; Wen et al. 2004; Xu et al. 2015), whereas others used large sets of autosomal single-nucleotide polymorphisms (Chen et al. 2009; Xu et al. 2009). While classical marker studies reported a genetic distinction between Northern and Southern China, with a boundary corresponding approximately to the Yangtze River (Chu et al. 1998; Xiao et al. 2000), Y chromosome and also autosomal surveys could not find a correspondence between this geographical pattern and the ethnic structure of contemporary China, although substantially different expansion time estimates were obtained for Northern and Southern populations (Xue et al. 2006). Wen et al. already noted similar Y haplogroup frequencies between Northern and Southern Han, being explained by a massive male-dominated movement of Northern Han immigrants into Southern China about 2000 years ago (Wen et al. 2004). The population expansions during the Han dynasty (206 BC–220 AD) changed the demography of the region dramatically and probably overwrote the genetic imprint of more ancient dispersals from Southern to Northern China (Jin and Su 2000) or via a northern route into China (Zhong et al. 2011).

In the present study, we aimed at assessing the effects of recent and ongoing demographic processes in China on Y chromosome variation. To this end, we analyzed those Chinese haplotypes submitted to YHRD where information on 17 commonly used Y-STR markers was available, taking advantage of the largest sample of Y chromosomes collected so far. We calculated and compared forensic diversity indices and explored the genetic variance between provinces, autonomous territories and municipalities as well as between ethnic groups. The question whether Chinese populations tend to become more similar due to a history of admixture in recent times or whether population differentiation along geographic or ethnic boundaries is still being maintained is of importance not only for understanding the past demography of China but also for using and interpreting Y-STR haplotypes in daily forensic practice.

Materials and methods

DNA samples and genotyping

A total of 37,994 unrelated males were sampled with informed consent from 70 distinct population groups living in 28 provinces, metropolitan areas and autonomous regions on the mainland territory of the People's Republic of China. Parts of these samples had been published before (Bai et al. 2013, 2016; Bing et al. 2013; Gao et al. 2016; Guo 2015; Guo et al. 2015, 2016a, b; Han et al. 2016; He and Guo 2013; Hu 2006; Li et al. 2016; Liu et al. 2014; Ou et al. 2015; Shan et al. 2014; Shang and Hu 2015; Shi et al. 2015; Wang et al. 2015, 2016a, b, c; Wu et al. 2011; Xu et al. 2016a, b; Yang et al. 2014; Yao et al. 2016; Zhang et al. 2014), mostly in the form of announcements of population data, while the remaining samples have so far been unpublished. Table S1 provides details on the origin of the samples.

Samples were subsequently typed in 22 participating laboratories and were analyzed for at least 17 Y-STRs (DYS19, DYS389I, DYS389II, DYS390, DYS391, DYS392, DYS393, DYS385a, DYS385b, DYS437, DYS438, DYS439, DYS448, DYS456, DYS458, DYS635, GATAH4) using different commercial kits (AmpFlSTR® *Yfiler*® and *Yfiler*® *Plus* of Thermo Fisher Scientific (Waltham, USA), Powerplex®Y23 of Promega (Madison, USA), AGCU Y24 of AGCU ScienTech Incorporation (Wuxi, China) or STRtyper-27Y of Healthgene Technologies (Ningbo, China)). The PCR products were analyzed using capillary electrophoresis on ABI Genetic Analyzer instruments (Thermo Fisher Scientific), whilst the genotyping results were analyzed using the software GeneMapper® (Thermo Fisher Scientific) and alleles were identified by comparison with the allelic ladder provided in the kit. A nomenclature was used which follows the recommendations by the DNA Commission of the International Society of Forensic Genetics (ISFG) (Gusmao et al. 2006). All genotypes, either published or yet unpublished, were submitted to the online Y Chromosome Haplotype Reference Database (YHRD; https://yhrd.org) for external review. After successful validation, accession numbers were issued and the samples were logged to the YHRD (Table S1). In the present study, allele designations for DYS389I and DYS389II refer to the repeat numbers at individual loci rather than the repeat numbers revealed by the multiplex genotyping method. The present study complies with the ethical principles of the 2000 Helsinki Declaration of the World Medical Association.

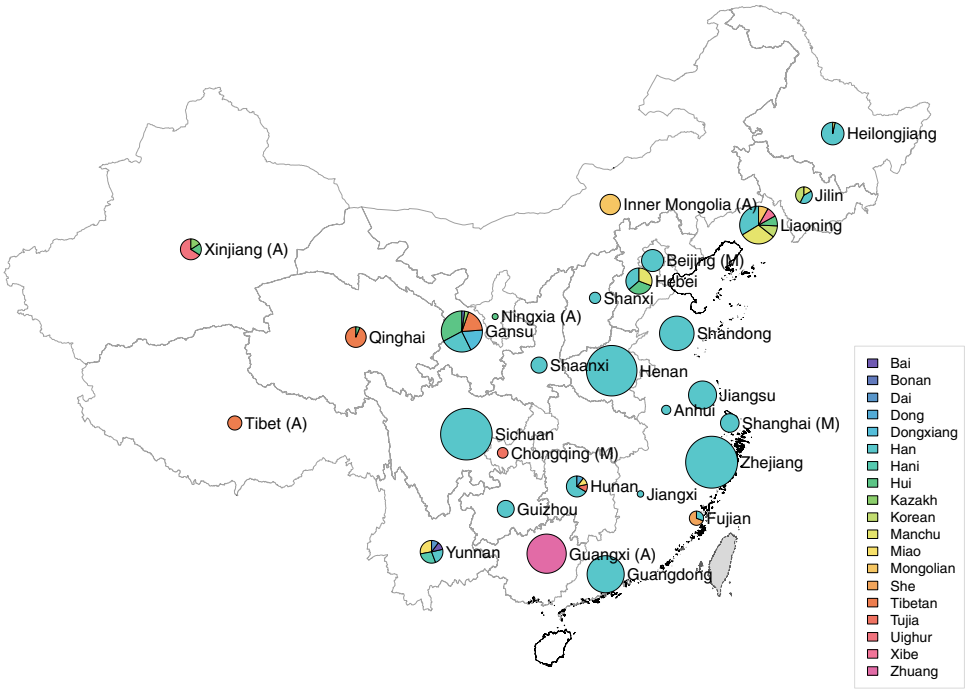

Fig. 1 Size and ethnic composition of Chinese provincial sample sets. *Dark gray line* national border of the People's Republic of China; *light gray lines* boundaries of provinces, metropolitan areas and autonomous regions. *M* Metropolitan area, *A* autonomous region. *Colors code* for the ethnicity of samples

Statistical analysis

Allele and haplotype frequencies were estimated by counting. Following Nei (1987) and Nei and Tajima (1981), gene diversity (GD) and haplotype diversity (HD) were computed as $\mathrm{GD} = \frac{n}{n-1}\left(1 - \sum_a p_a^2\right)$, where n denotes the total number of samples and p_a the relative frequency of the ath allele at the locus (GD) or of the ath haplotype (HD), respectively. The random match probability (MP) was computed as the sum of squared haplotype frequencies, whereas the discrimination capacity (DC) was calculated as the ratio of the number of different haplotypes by the total number of haplotypes. The R software v3.3.1 (R Core Team 2016) was used for statistical analysis unless indicated otherwise, and to create graphs.

Analysis of molecular variance (AMOVA) Genetic relationships between different groups of males were quantified by means of R_{ST}, thereby taking the evolutionary distance between individual Y-STR haplotypes into account (Excoffier and Smouse 1994; Excoffier et al. 1992). For a review of genetic population differentiation measures, see (Balloux and Lugon-Moulin 2002). Groups were defined according to province or ethnicity affiliation, respectively. We used Arlequin v3.5.1.2 (Excoffier and Lischer 2010) to estimate R_{ST}. Significance of $R_{ST} > 0$ was tested using randomization with 1000 replicates per comparison. Samples carrying a deletion at one or more markers were excluded from the analysis of the respective marker set.

Multidimensional scaling analysis Multidimensional scaling (MDS) analysis was based upon pair-wise R_{ST} between sampling sites as estimated with Arlequin (see above), and was carried out with the *sammon* function from the MASS library of R v3.3.1 (R Core Team 2016). Plots in the first two MDS components, C1 and C2, capturing the strongest and second strongest R_{ST}-defined variation, respectively, were generated with *R* using in-house scripts.

Generation of geographic maps Geographic maps were generated in R v3.3.1 (R Core Team 2016) using packages geoR v1.7-5.2 (Ribeiro Jr. and Diggle 2001), sp v1.2-3 (Pebesma and Bivand 2005), BayesX v0.2-9 (Brezger et al. 2005) and plotrix v3.6-3 (Lemon 2006). Geographic information data were obtained from the GADM database of Global Administrative Areas (v2.8, November 2015; http://www.gadm.org/).

Results

Sample characteristics A total of 37,994 samples from 28 provinces, metropolitan areas and autonomous regions, from now on called "provinces" for brevity, of the People's Republic of China were available for analysis (Fig. 1). Sample sizes showed strong variation across provinces (Table 1), ranging from just 62 (Ningxia) to 4699 (Zhejiang) and with a median (inter-quartile range) of 764.5 (425.8–2158.0). Our samples included members of 19 different ethnicities (Table 1), with Han being by far the largest group (26,338;

Table 1 Size and ethnic composition of provincial samples

	Bai	Bonan	Dai	Dong	Dongxiang	Han	Hani	Hui	Kazakh	Korean	Manchu	Miao	Mongolian	She	Tibetan	Tujia	Uighur	Xibe	Zhuang	Total
Anhui						147														147
Beijing (M)						847														847
Chongqing (M)																197				197
Fujian						109								244						353
Gansu		64			555	722		966	93						545					2945
Guangdong						2401														2401
Guangxi (A)																			2687	2687
Guizhou						500														500
Hebei						433		392			366									1191
Heilongjiang						859					30									889
Henan						4448														4448
Hunan				80		511						87				90				768
Inner Mongolia (A)													722							722
Jiangsu						1366														1366
Jiangxi						72														72
Jilin						196				206	81									483
Liaoning						838		217		248	760		206					215		2484
Ningxia (A)								62												62
Qinghai								46							693					739
Shaanxi						450														450
Shandong						2077														2077
Shanghai (M)						602														602
Shanxi						222														222
Sichuan						4634														4634
Tibet (A)															348					348
Xinjiang (A)								138	121								502			761
Yunnan	101		92			205	250					252								900
Zhejiang						4699														4699
Total	101	64	92	80	555	26338	250	1821	214	454	1237	339	928	244	1586	287	502	215	2687	37994

M metropolitan area, *A* autonomous region

Fig. 2 Forensic haplotype diversity measures across Chinese provinces. **a** Haplotype diversity (HD), **b** match probability (MP), **c** discrimination capacity (DC). *Gray-shaded* areas mark provinces without information

69.3%), followed by Zhuang (2687; 7.1%), Hui (1821; 4.8%), Tibetan (1586; 4.2%), Manchu (1237; 3.3%) and Mongolian (928; 2.4%). All other ethnicities had a share of less than 2% of the total sample size. More than half of all ethnicities were restricted to single provinces in our sample set, with only Han (21), Hui (6), Kazakh (2), Korean (2), Manchu (4), Miao (2), Mongolian (2), Tibetan (3) and Tujia (2) being present in more than one provincial sample set. More information on sampling areas, YHRD accession number and reference of previous publication of the data, if applicable, is available from Table S1.

Y chromosomal diversity Mean single-marker gene diversity (GD) ranged between 0.4007 (DYS391) and 0.9620 (DYS385ab), while GD variance was of similar magnitude across all analyzed Chinese provinces, ranging between 0.0126 (DYS385ab) and 0.0733 (DYS438) (Table S2). Haplotype diversity (HD) exceeded 0.9978 in all provinces (mean ± SD 0.994 ± 0.0006), while the related random match probability (MP) did not exceed 0.0069 (0.0032 ± 0.0039) in all provinces except for Ningxia and Jiangxi where it equaled 0.0177 and 0.0139, respectively. Discrimination capacity (DC), defined as the proportion of haplotype singletons, showed considerably more variation, ranging from 0.5367 to 1.0000 (0.8721 ± 0.1108). Thus, while HD and MP showed a largely uniform behavior, DC varied considerably across provinces (Table S2, Fig. 2). However, calculations of these measures were partially based upon small provincial or ethnic sample sizes and should, therefore, be considered with some caution. Very small sample sizes had a strong effect on MP and, to a lesser degree, on HD when performing the calculation either per province or per ethnicity (Figure S1). The impact of sample size was less clear for DC, with a general downward trend per province, but substantial variation and no clear trend between ethnicities (Figure S1). Restricting the analysis to each of those four ethnicities for which samples were available from at least three provinces (Han, Hui, Manchu, Tibetan) yielded considerably more variation across the respective provinces (Figure S2), although this was most likely due to the very disparate sample sizes per province. Bearing in mind the sparse distribution of sampling sites, we refrained from performing geographical interpolation of diversity measures.

Genetic differentiation along administrative divisions (provinces) Calculation of pair-wise R_{ST} values between provinces based on Y chromosomal haplotypes revealed patterns of partial population substructure between provinces (Table S3). The Qinghai and Tibet provinces, both featuring predominantly or even exclusively Tibetan samples, consistently showed substantial differences to all other provinces, with R_{ST} values ranging between 0.08 and 0.28 (mean ± SD 0.21 ± 0.06 and 0.17 ± 0.05, respectively). On the other hand, the Gansu province sample set, which also features a sizeable number of Tibetans, is dominated by members of the Hui, Han and Dongxiang ethnicities and, therefore, showed less pronounced differences to the other provinces. Similarly, while Kazakhs and, to a lesser degree, Uighurs showed substantial substructure with the Han and numerous other ethnicities (see below), the respective provincial sample sets contained several other ethnicities as well, thereby likely diluting the differentiation signal. Correspondingly, Qinghai and Tibet stood out from a cluster of all other provinces in a multidimensional scaling (MDS) analysis of pair-wise R_{ST} between provinces (Fig. 3a).

Genetic differentiation along ethnic divisions (ethnicities) Analysis of pair-wise R_{ST} values between ethnicities revealed a cluster of closely related ethnicities and a few more dissimilar ones. Values for pair-wise R_{ST} ranged from 0.01 continuously to about 0.13 for a majority of ethnicities (Table S4), with the most notable exceptions being Kazakhs and Tibetans. The latter showed substantial substructure with a majority of the other ethnicities ($R_{ST} > 0.15$) and still at least moderate substructure with all others. To a lesser extent, Bai, Bonan, Dai, Dong, Dongxiang, Hani, Korean, Uighurs and Zhuang also showed some substructure with other ethnicities,

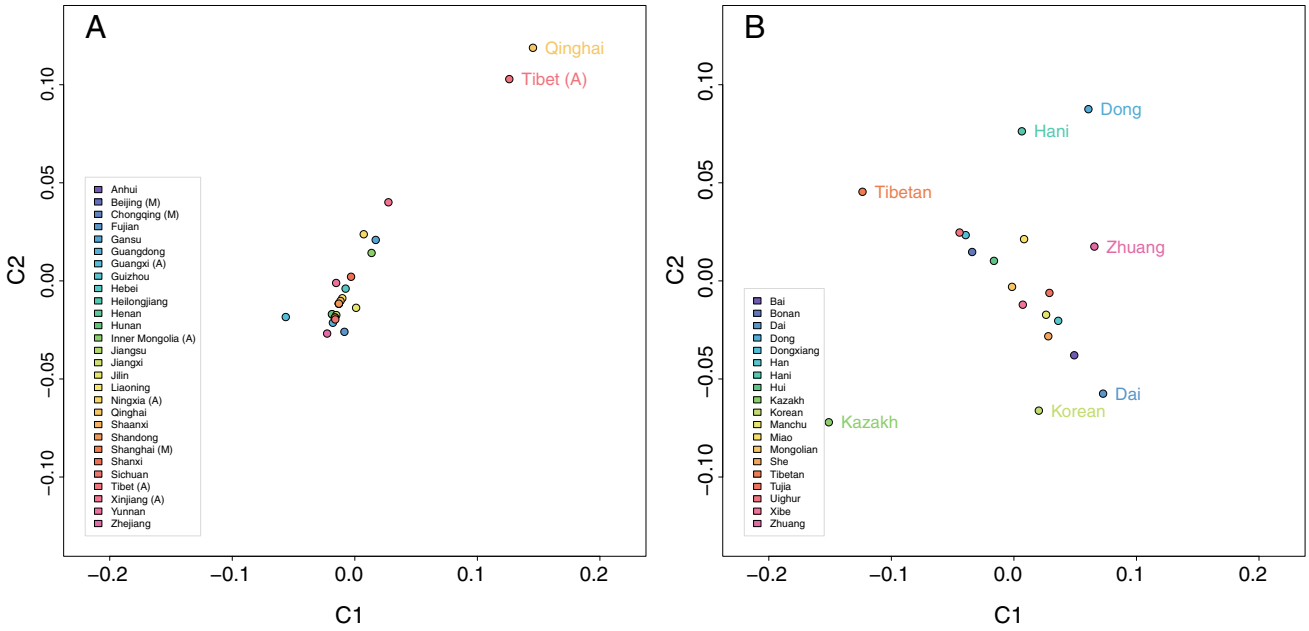

Fig. 3 Genetic distance between administrative and ethnic groups in China. Multidimensional scaling (MDS) *plot* based on R_{ST} with respect to Y chromosomal haplotypes. **a** Provinces (stress: 0.042), **b** ethnicities (stress: 0.044)

as indicated by R_{ST} values above an (arbitrarily chosen) threshold of >0.05 for most ethnicity pairs. These small differences as well as the noted exceptions are also reflected in the MDS plot (Fig. 3b). It should be noted, however, that for some ethnicities, namely Bonan, Dai and Dong, very limited sample sizes (<100; see Table 1) were available, possibly having caused an upward bias in the R_{ST} estimates. Thus, again some caution is advised when interpreting these results.

Genetic differentiation along administrative and ethnic divisions Next, we repeated some analyses while restricting either to a particular ethnicity, most prominently the Han due to their sample size and widespread occurrence across most Chinese provinces, or to a particular province. The Han ethnicity was remarkably homogeneous in their genetic variation across China, with none of the R_{ST} values between pairs of provinces exceeding 0.05 (Table S5). Fujian and Shanxi consistently showed R_{ST} values with all other provinces of around 0.031 and 0.023, respectively, whereas Zhejiang showed such substructure only for a subset of provinces. Accordingly, provincial populations showed little differentiation in an MDS plot, with the exception of Shanxi and the coastal provinces of Fujian and Zhejiang that were slightly detached from the main cluster of provinces (Fig. 4). Of the other ethnicities of which we had samples available from four or more provinces, only the Manchu showed some substructure between provinces, as evidenced by R_{ST} values between 0.06 and 0.09, whilst Hui and Tibetans were more homogenous ($R_{ST} < 0.05$ for all provincial pairs) (Table S6–8).

Given previous reports of a North–South divide in genetic variation in the Han, we also explored if we could confirm these observations with respect to Y chromosomal haplotypes. Remarkably, provincial latitude (Table S1) showed a very good correspondence to provincial position in the MDS plot with respect to the first component, whereas provincial longitude showed virtually no concordance at all (Fig. 5). Correspondingly, we observed a strong correlation of the first MDS component with provincial latitude (Spearman $r = 0.78$, $p = 0.00004$) and very little with longitude ($r = 0.05$, $p = 0.81$). Notably, we did not observe disruptive genetic changes in the Han between Northern and Southern provinces, but rather small differences, as evidenced by small R_{ST} values that broadly follow a North–South gradient. Interestingly, the Heilongjiang province, despite its somewhat remote geographic location in the Northeast of China, assumed a prominent central position next to Liaoning in the MDS plot.

We also investigated if the Han, as the most populous ethnicity that was present in many provincial samples, would show substantial differences to other ethnicities in some provinces. Out of 29 provincial samples, 10 (34.5%) featured more than one ethnicity, namely Fujian (2), Gansu (6), Hebei (3), Heilongjiang (2), Hunan (4), Jilin (3), Liaoning (6), Qinghai (2), Xinjiang (3) and Yunnan (5) (Table 1). The Han showed extensive substructure with Kazakhs in the Gansu province ($R_{ST} = 0.37$), while Dong, Dongxiang, Hani, Miao and Tibetan showed more moderate differences compared to the Han ($0.07 \leq R_{ST} \leq 0.11$) (Table S9). Little

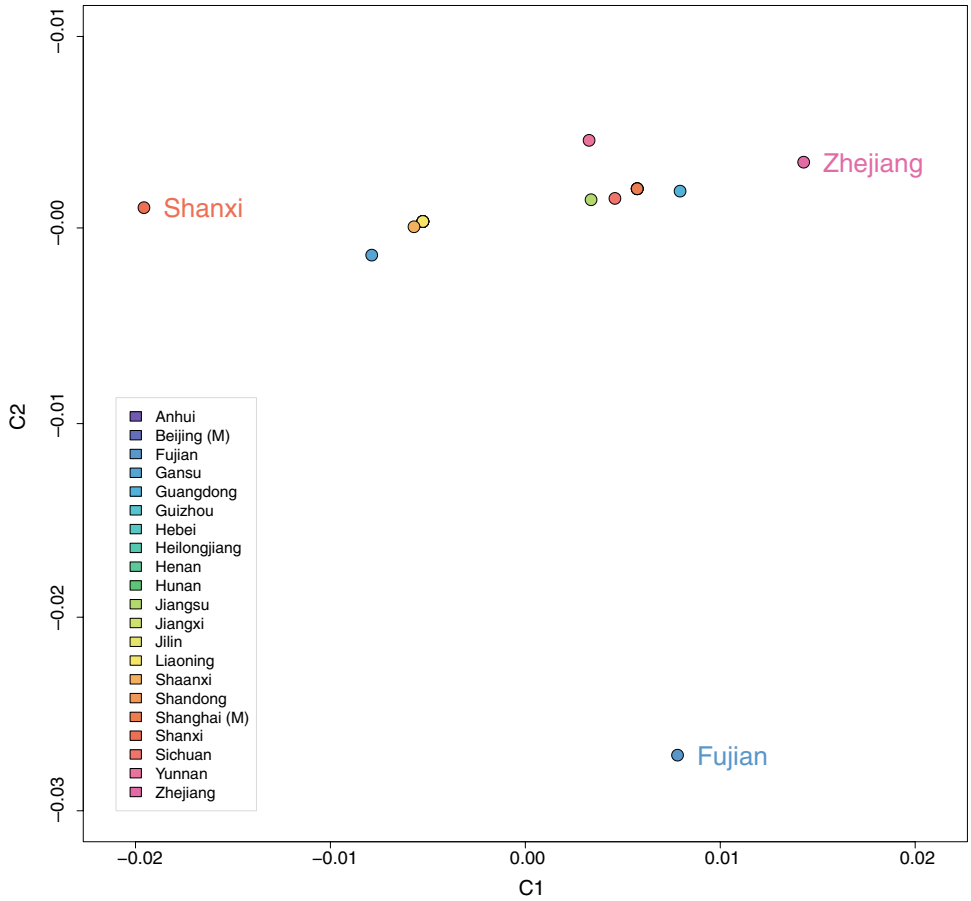

Fig. 4 Genetic distance between Han of different Chinese provinces. Multidimensional scaling (MDS) *plot* based on R_{ST} with respect to Y chromosomal haplotypes in the Han (stress: 0.107)

substructure with the Han ($R_{ST} < 0.05$) was observed for Bai, Bonan, Dai, Hui, Korean, Manchu, Mongolian, She, Tujia and Xibe across all provinces except for three cases where R_{ST} values were slightly above 0.05 (Table S9).

Discussion

In a large population genetic study of China, we investigated the spatial and ethnic distribution of genetic variation in almost 38,000 men by use of 17 Y chromosomal STR markers that are commonly used in both forensic applications and population genetic studies. The large size of most of our provincial samples ensured that locally common patrilines were more comprehensively represented than in previous studies of often much smaller size. We could show that Y-STR haplotype variation is large and highly informative for forensic purposes all across China irrespective of the geographic or ethnic background of the studied population. Our genetic data presented here provide basic information on the country's population differentiation required by the forensic community in China and abroad.

We observed substantial differences in the distribution of Y chromosomal haplotypes between some ethnic groups, most notably Kazakhs and Tibetans, and most other ethnicities. The strong differentiation of the Kazakhs is explicable by a different ancestry as well as repeated migrations and admixture events along the Silk Road in the past 2000 years or so, as has been recently suggested (Mezzavilla et al. 2014), whereas Tibetans' Y-STR haplotype variation may mirror a deep ancestral root that is indicated by the high frequency of the D-M174 haplogroup in Tibetans (Shi et al. 2008; Wang and Li 2013). On a provincial level, however, differences were much less pronounced, with the exceptions of the Tibet and Qinghai provinces. This apparent contradiction is explicable by two additional observations. First, genetic differences within ethnicities were usually of minor magnitude. Most prominently, this applies to the Han, which, despite their by far largest sample size, displayed a remarkable genetic homogeneity. Second, although virtually all ethnicities except for the Han occurred in only one or a few of our provincial sample sets, the Han, and some genetically close ethnicities, usually co-occurred with those ethnicities in the same provincial sample sets, resulting in a dilution of genetic differences between provinces featuring different ethnic groups. The Tibet and Qinghai provinces almost exclusively featured Tibetan samples and, therefore, escaped this diluting effect.

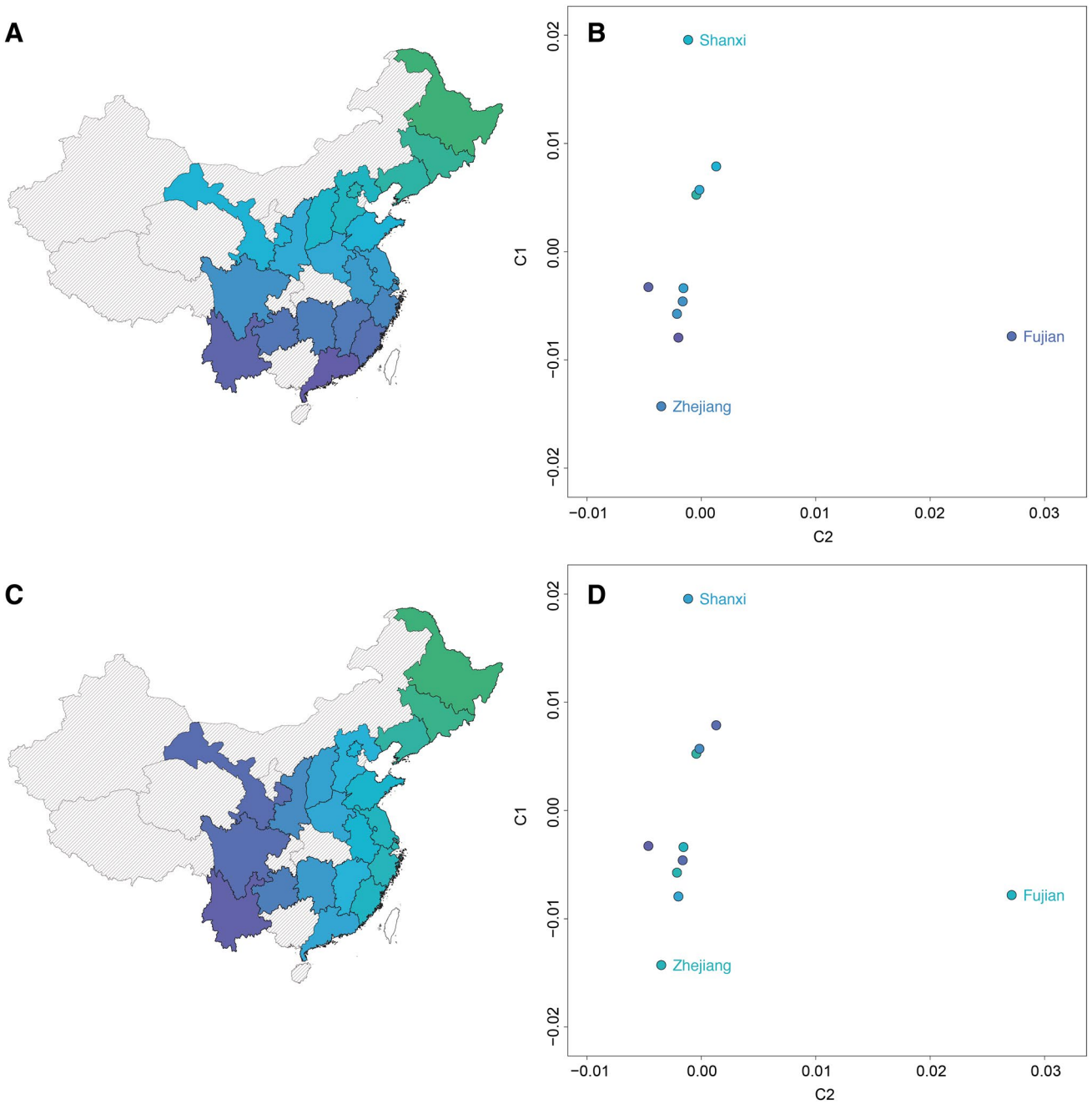

Fig. 5 Correspondence between geographic and MDS position of Han populations. **b**, **d** Multidimensional scaling (MDS) *plot* based on R_{ST} with respect to Y chromosomal haplotypes in the Han (stress: 0.107). **a**, **b** Provincial *color* coding by latitude; **c**, **d** provincial *color* coding by longitude

On the other hand, while, for example, Kazakhs showed substantial genetic differences with respect to the Han and other ethnicities, none of the provincial samples featured a Kazakh majority, resulting in less pronounced inter-provincial differences.

Based on our Y chromosomal STR data, thereby tracing strictly paternal inheritance, we could not confirm previous reports of a genetic distinction between Northern and Southern Han for binary or STR markers (Chu et al. 1998; Xiao et al. 2000; Xue et al. 2006, 2008; Zhao et al. 2015). Instead, we observed continuous gradual differences that generally follow a North–South direction. These gradual changes in Y-STR genetic diversity observed in our study are consistent with a previously suggested demic diffusion, isolation-by-distance model that was based on Y-chromosomal and mtDNA haplogroup assignment, i.e., carrier-ship

of specific alleles at selected single-nucleotide polymorphisms (SNPs), and where males played a more prominent part than females (Wen et al. 2004). Furthermore, they are in accordance with North–South gradients in genetic variation that have been reported for several hundred thousands of autosomal SNPs (Chen et al. 2009; Xu et al. 2009). This is a remarkable observation in two ways. First, unlike in Europe, use of both autosomal SNP and Y chromosomal Y-STR markers revealed concordant patterns of variation, possibly indicating (i) little difference in migration between males and females and (ii) a continuous process of population expansion from the Neolithic to modern times at least for the Eastern and Central parts of China. Given the different "time windows" of both marker types, it is reasonable to conclude that both processes have lasted already for more than two millennia to create concordant patterns of population substructure. Second, the observed latitudinal, but not longitudinal, gradient in Y chromosomal variation resembles previously reported observations using autosomal SNPs from Europe (Lao et al. 2008; Nothnagel et al. 2010; Novembre et al. 2008), likely indicating a similar, albeit younger and slower, population genetic process compared to Europe. The concordant and clinal patterns of genetic variation would also be consistent with a massive population expansion that has been suggested to have started in the Neolithic around ~6 kya (Yan et al. 2014) and to have accelerated around 2000 years ago with the formation of a single Chinese empire after the conquest of several states by the Qin and Han dynasties, not the least resulting in a likely larger effective population size of the Han compared to European populations. In modern times, huge migrations of the Han might have further contributed to a homogeneity of the genetic landscape of China, here to mention the "Crashing into Guandong" (*Chuang Guandong*, starting in the second half of the nineteenth century), when residents of the coastal Shandong province moved to the three northeastern provinces of Heilongjiang, Jilin and Liaoning, or the "Going to the West" (*Zou Xikou*), when people of the central northern Shanxi and Shaanxi provinces migrated to the northern region of Inner Mongolia (seventeenth–twentieth century). More in-depth statistical modeling of our genetic data and inference of likely population historic scenarios, such as expansion times and admixture events, will be subject to future studies.

Conclusion

In summary, our large multi-center study revealed that Y-STR haplotypes are highly variable in China irrespective of the geographic or ethnic background of the studied population. The large size of most regional samples ensured that locally common patrilines were represented to a larger extent than in other, much smaller studies. Characteristic genetic divisions often observed between populations typed for Y-STRs at other continents are largely lacking in most of China. Especially the homogeneity of the Han Chinese population points to a constant process of range expansions and possibly genetic admixture. This process could have accelerated around 2000 years ago with the Qin and Han dynasties that first conquered several states to form a Chinese empire and to unite the whole China. Our genetic data presented here provide basic information on the country's population differentiation required by the forensic community in China and abroad.

Web Resources: R software: http://www.r-project.org/; *geoR, sp, BayesX, plotrix* packages: http://cran.r-project.org/; Arlequin software: http://cmpg.unibe.ch/software/arlequin35/; YHRD: https://yhrd.org/

Compliance with ethical standards

Conflict of interest On behalf of all authors, the corresponding author states that there is no conflict of interest.

Funding Hongbing Yao was supported by the Natural Science Foundation of Gansu Province (1308RJZA190), Scientific Research Project for Colleges of Gansu Province (2014A-085), Evidence Science, Technology and Application for Key Laboratory Foundation of Gansu Province (ZDSYS-Y20) and Gansu Institute of Political Science and Law Foundation Project (GZF2011XQNLW45). Lagabaiyila Zha's work was supported by the Natural Science Foundation for Youth (81302621). Xiufen Zheng received a grant by the National Natural Science Foundation of China (31260267). Hongyu Sun was supported by the National Natural Science Foundation of China (81273347, 81671873), and Fundamental Research Funds for the Central Universities (16ykzd08). Hui Li received funding from MOE Scientific Research Project (113022A), NSFC31671297, MOST2016YFC0900300, and Shanghai Shuguang Project (14SG05). Lei Wang and Zhaoshu Zeng were supported by The Key Scientific Research Project for Higher Institutions of Henan Provincial Education Department (15A180021). Weibo Liang's work was supported by the National Natural Science Foundation of China (81471827, 81671871).

References

Bai R, Zhang Z, Liang Q, Lu D, Yuan L, Yang X, Shi M (2013) Haplotype diversity of 17 Y-STR loci in a Chinese Han population sample from Shanxi Province, Northern China. Forensic Sci Int Genet 7:214–216. doi:10.1016/j.fsigen.2012.10.004

Bai R, Liu Y, Lv X, Shi M, Ma S (2016) Genetic polymorphisms of 17 Y chromosomal STRs in She and Manchu ethnic populations from China. Forensic Sci Int Genet 22:e12–e14. doi:10.1016/j.fsigen.2016.01.011

Balloux F, Lugon-Moulin N (2002) The estimation of population differentiation with microsatellite markers. Mol Ecol 11:155–165

Bing L, Liang W, Pi J, Zhang D, Yong D, Luo H, Zhang L, Lin Z (2013) Population genetics for 17 Y-STR loci(AmpFISTR(R) Y-filerTM) in Luzhou Han ethnic group. Forensic Sci Int Genet 7:e23–e26. doi:10.1016/j.fsigen.2012.11.010

Brezger A, Kneib T, Lang S (2005) BayesX: analyzing Bayesian structured additive regression models. J Stat Softw 14:11. doi:10.18637/jss.v014.i11

Caliebe A, Krawczak M (2016) Probability and Likelihood. In: Amorim A, Budowle B (eds) Handbook of forensic genetics: biodiversity and heredity in civil and criminal investigation. World Scientific, New Jersey

Chen J, Zheng H, Bei JX, Sun L, Jia WH, Li T, Zhang F, Seielstad M, Zeng YX, Zhang X, Liu J (2009) Genetic structure of the Han Chinese population revealed by genome-wide SNP variation. Am J Hum Genet 85:775–785. doi:10.1016/j.ajhg.2009.10.016

Chu JY, Huang W, Kuang SQ, Wang JM, Xu JJ, Chu ZT, Yang ZQ, Lin KQ, Li P, Wu M, Geng ZC, Tan CC, Du RF, Jin L (1998) Genetic relationship of populations in China. Proc Natl Acad Sci USA 95:11763–11768

Deng W, Shi B, He X, Zhang Z, Xu J, Li B, Yang J, Ling L, Dai C, Qiang B, Shen Y, Chen R (2004) Evolution and migration history of the Chinese population inferred from Chinese Y-chromosome evidence. J Hum Genet 49:339–348. doi:10.1007/s10038-004-0154-3

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 10:564–567. doi:10.1111/j.1755-0998.2010.02847.x

Excoffier L, Smouse PE (1994) Using allele frequencies and geographic subdivision to reconstruct gene trees within a species: molecular variance parsimony. Genetics 136:343–359

Excoffier L, Smouse PE, Quattro JM (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. Genetics 131:479–491

Gao T, Yun L, Gao S, Gu Y, He W, Luo H, Hou Y (2016) Population genetics of 23 Y-STR loci in the Mongolian minority population in Inner Mongolia of China. Int J Legal Med 130:1509–1511. doi:10.1007/s00414-016-1433-1

Guo F (2015) Population genetics for 17 Y-STR loci in Mongolian ethnic minority from Liaoning Province, Northeast China. Forensic Sci Int Genet 17:153–154. doi:10.1016/j.fsigen.2015.05.008

Guo F, Zhang L, Jiang X (2015) Population genetics of 17 Y-STR loci in Xibe ethnic minority from Liaoning Province, Northeast China. Forensic Sci Int Genet 16:86–87. doi:10.1016/j.fsigen.2014.12.007

Guo F, Li J, Chen K, Tang R, Zhou L (2016a) Population genetic data for 27 Y-STR loci in the Zhuang ethnic minority from Guangxi Zhuang Autonomous Region in the south of China. Forensic Sci Int Genet (In press)

Guo F, Song L, Zhang L (2016b) Population genetics for 17 Y-STR loci in Korean ethnic minority from Liaoning Province, Northeast China. Forensic Sci Int Genet 22:e9–e11. doi:10.1016/j.fsigen.2016.01.007

Gusmao L, Butler JM, Carracedo A, Gill P, Kayser M, Mayr WR, Morling N, Prinz M, Roewer L, Tyler-Smith C, Schneider PM, Genetics DNACotISoF (2006) 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 157:187–197. doi:10.1016/j.forsciint.2005.04.002

Han Y, Li L, Liu X, Chen W, Yang S, Wei L, Xia M, Ma T, Jin L, Li S (2016) Genetic analysis of 17 Y-STR loci in Han and Korean populations from Jilin Province, Northeast China. Forensic Sci Int Genet 22:8–10. doi:10.1016/j.fsigen.2016.01.003

He J, Guo F (2013) Population genetics of 17 Y-STR loci in Chinese Manchu population from Liaoning Province, Northeast China. Forensic Sci Int Genet 7:e84–e85. doi:10.1016/j.fsigen.2012.12.006

Hu SP (2006) Genetic polymorphism of 12 Y-chromosomal STR loci in the Minnan Han Chinese in Southeast China. Forensic Sci Int 159:77–82. doi:10.1016/j.forsciint.2005.05.017

Jin L, Su B (2000) Natives or immigrants: modern human origin in east Asia. Nat Rev Genet 1:126–133. doi:10.1038/35038565

Jobling MA (2012) The impact of recent events on human genetic diversity. Philos Trans R Soc Lond B Biol Sci 367:793–799. doi:10.1098/rstb.2011.0297

Kayser M, Krawczak M, Excoffier L, Dieltjes P, Corach D, Pascali V, Gehrig C, Bernini LF, Jespersen J, Bakker E, Roewer L, de Knijff P (2001) An extensive analysis of Y-chromosomal microsatellite haplotypes in globally dispersed human populations. Am J Hum Genet 68:990–1018. doi:10.1086/319510

Kayser M, Lao O, Anslinger K, Augustin C, Bargel G, Edelmann J, Elias S, Heinrich M, Henke J, Henke L, Hohoff C, Illing A, Jonkisz A, Kuzniar P, Lebioda A, Lessig R, Lewicki S, Maciejewska A, Monies DM, Pawlowski R, Poetsch M, Schmid D, Schmidt U, Schneider PM, Stradmann-Bellinghausen B, Szibor R, Wegener R, Wozniak M, Zoledziewska M, Roewer L, Dobosz T, Ploski R (2005) Significant genetic differentiation between Poland and Germany follows present-day political borders, as revealed by Y-chromosome analysis. Hum Genet 117:428–443. doi:10.1007/s00439-005-1333-9

Lao O, Lu TT, Nothnagel M, Junge O, Freitag-Wolf S, Caliebe A, Balascakova M, Bertranpetit J, Bindoff LA, Comas D, Holmlund G, Kouvatsi A, Macek M, Mollet I, Parson W, Palo J, Ploski R, Sajantila A, Tagliabracci A, Gether U, Werge T, Rivadeneira F, Hofman A, Uitterlinden AG, Gieger C, Wichmann HE, Ruther A, Schreiber S, Becker C, Nurnberg P, Nelson MR, Krawczak M, Kayser M (2008) Correlation between genetic and geographic structure in Europe. Curr Biol 18:1241–1248. doi:10.1016/j.cub.2008.07.049

Lemon J (2006) Plotrix: a package in the red light district of R. R-NEWS 6:8–12

Li H, Huang Y, Mustavich LF, Zhang F, Tan JZ, Wang LE, Qian J, Gao MH, Jin L (2007) Y chromosomes of prehistoric people along the Yangtze River. Hum Genet 122:383–388. doi:10.1007/s00439-007-0407-2

Li L, Yu G, Li S, Jin L, Yan S (2016) Genetic analysis of 17 Y-STR loci from 1019 individuals of six Han populations in East China. Forensic Sci Int Genet 20:101–102. doi:10.1016/j.fsigen.2015.10.007

Liu Y, Liao L, Gu M, Ye Y (2014) Population genetics for 17 Y-STR loci in a Chinese Han population sample from Mudanjiang city, Northeast China. Forensic Sci Int Genet 13:e16–e17. doi:10.1016/j.fsigen.2014.05.009

Mezzavilla M, Vozzi D, Pirastu N, Girotto G, d'Adamo P, Gasparini P, Colonna V (2014) Genetic landscape of populations along the Silk Road: admixture and migration patterns. BMC Genet 15:131. doi:10.1186/s12863-014-0131-6

Nei M (1987) Molecular evolutionary genetics. Columbia University Press, New York

Nei M, Tajima F (1981) DNA polymorphism detectable by restriction endonucleases. Genetics 97:145–163

Nothnagel M, Lu TT, Kayser M, Krawczak M (2010) Genomic and geographic distribution of SNP-defined runs of homozygosity in Europeans. Hum Mol Genet 19:2927–2935. doi:10.1093/hmg/ddq198

Novembre J, Johnson T, Bryc K, Kutalik Z, Boyko AR, Auton A, Indap A, King KS, Bergmann S, Nelson MR, Stephens M, Bustamante CD (2008) Genes mirror geography within Europe. Nature 456:98–101. doi:10.1038/nature07331

Ou X, Wang Y, Liu C, Yang D, Zhang C, Deng S, Sun H (2015) Haplotype analysis of the polymorphic 40 Y-STR markers in Chinese populations. Forensic Sci Int Genet 19:255–262. doi:10.1016/j.fsigen.2015.08.007

Pebesma EJ, Bivand RS (2005) Classes and methods for spatial data in R. R-NEWS 5(2):9–13. https://cran.r-project.org/doc/Rnews/Rnews_2005-2.pdf

Piazza A (1998) Towards a genetic history of China. Nature 395(636–7):639. doi:10.1038/27071

Ploski R, Wozniak M, Pawlowski R, Monies DM, Branicki W, Kupiec T, Kloosterman A, Dobosz T, Bosch E, Nowak M, Lessig R, Jobling MA, Roewer L, Kayser M (2002) Homogeneity and distinctiveness of Polish paternal lineages revealed by Y chromosome microsatellite haplotype analysis. Hum Genet 110:592–600. doi:10.1007/s00439-002-0728-0

Powell GT, Yang H, Tyler-Smith C, Xue Y (2007) The population history of the Xibe in northern China: a comparison of autosomal, mtDNA and Y-chromosomal analyses of migration and gene flow. Forensic Sci Int Genet 1:115–119. doi:10.1016/j.fsigen.2007.01.015

Purps J, Siegert S, Willuweit S, Nagy M, Alves C, Salazar R, Angustia SM, Santos LH, Anslinger K, Bayer B, Ayub Q, Wei W, Xue Y, Tyler-Smith C, Bafalluy MB, Martinez-Jarreta B, Egyed B, Balitzki B, Tschumi S, Ballard D, Court DS, Barrantes X, Bassler G, Wiest T, Berger B, Niederstatter H, Parson W, Davis C, Budowle B, Burri H, Borer U, Koller C, Carvalho EF, Domingues PM, Chamoun WT, Coble MD, Hill CR, Corach D, Caputo M, D'Amato ME, Davison S, Decorte R, Larmuseau MH, Ottoni C, Rickards O, Lu D, Jiang C, Dobosz T, Jonkisz A, Frank WE, Furac I, Gehrig C, Castella V, Grskovic B, Haas C, Wobst J, Hadzic G, Drobnic K, Honda K, Hou Y, Zhou D, Li Y, Hu S, Chen S, Immel UD, Lessig R, Jakovski Z, Ilievska T, Klann AE, Garcia CC, de Knijff P, Kraaijenbrink T, Kondili A, Miniati P, Vouropoulou M, Kovacevic L, Marjanovic D, Lindner I, Mansour I, Al-Azem M, Andari AE, Marino M, Furfuro S, Locarno L, Martin P, Luque GM, Alonso A, Miranda LS, Moreira H, Mizuno N, Iwashima Y, Neto RS, Nogueira TL, Silva R, Nastainczyk-Wulf M, Edelmann J, Kohl M, Nie S, Wang X, Cheng B et al (2014) A global analysis of Y-chromosomal haplotype diversity for 23 STR loci. Forensic Sci Int Genet 12:12–23. doi:10.1016/j.fsigen.2014.04.008

Qi X, Cui C, Peng Y, Zhang X, Yang Z, Zhong H, Zhang H, Xiang K, Cao X, Wang Y, Ouzhuluobu Basang, Ciwangsangbu Bianba, Gonggalanzi WuT, Chen H, Shi H, Su B (2013) Genetic evidence of paleolithic colonization and neolithic expansion of modern humans on the tibetan plateau. Mol Biol Evol 30:1761–1778. doi:10.1093/molbev/mst093

R Core Team (2016) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/

Ribeiro Jr. PJ, Diggle PJ (2001) geoR: a package for geostatistical analysis. R-NEWS 1(2):15–18. https://cran.r-project.org/doc/Rnews/Rnews_2001-2.pdf

Roewer L (2009) Y chromosome STR typing in crime casework. Forensic Sci Med Pathol 5:77–84. doi:10.1007/s12024-009-9089-5

Roewer L (2016) The Y-Chromosome Haplotype Reference Database (YHRD)—publicly available reference and research datasets for the forensic interpretation of Y-chromosome STR profiles. In: AMORIM A, Budowle B (eds) Handbook of forensic genetics: biodiversity and heredity in civil and criminal investigation. World Scientific, New Jersey

Roewer L, Croucher PJ, Willuweit S, Lu TT, Kayser M, Lessig R, de Knijff P, Jobling MA, Tyler-Smith C, Krawczak M (2005) Signature of recent historical events in the European Y-chromosomal STR haplotype distribution. Hum Genet 116:279–291. doi:10.1007/s00439-004-1201-z

Roewer L, Nothnagel M, Gusmao L, Gomes V, Gonzalez M, Corach D, Sala A, Alechine E, Palha T, Santos N, Ribeiro-Dos-Santos A, Geppert M, Willuweit S, Nagy M, Zweynert S, Baeta M, Nunez C, Martinez-Jarreta B, Gonzalez-Andrade F, Fagundes de Carvalho E, da Silva DA, Builes JJ, Turbon D, Lopez Parra AM, Arroyo-Pardo E, Toscanini U, Borjas L, Barletta C, Ewart E, Santos S, Krawczak M (2013) Continent-wide decoupling of Y-chromosomal genetic variation from language and geography in native South Americans. PLoS Genet 9:e1003460. doi:10.1371/journal.pgen.1003460

Shan W, Ablimit A, Zhou W, Zhang F, Ma Z, Zheng X (2014) Genetic polymorphism of 17 Y chromosomal STRs in Kazakh and Uighur populations from Xinjiang, China. Int J Legal Med 128:743–744. doi:10.1007/s00414-013-0948-y

Shang J, Hu SP (2015) Haplotype data of 23 Y-chromosome markers in Minnan Han Chinese and comparison with those of 12 Y-chromosome markers. J Huazhong Univ Sci Technol Med Sci 35:456–463. doi:10.1007/s11596-015-1453-y

Shi H, Zhong H, Peng Y, Dong YL, Qi XB, Zhang F, Liu LF, Tan SJ, Ma RZ, Xiao CJ, Wells RS, 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. doi:10.1186/1741-7007-6-45

Shi M, Liu Y, Zhang J, Bai R, Lv X, Ma S (2015) Analysis of 24 Y chromosomal STR haplotypes in a Chinese Han population sample from Henan Province, Central China. Forensic Sci Int Genet 17:83–86. doi:10.1016/j.fsigen.2015.04.001

Wang CC, Li H (2013) Inferring human history in East Asia from Y chromosomes. Investig Genet 4:11. doi:10.1186/2041-2223-4-11

Wang C, Yan S, Hou Z, Fu W, Xiong M, Han S, Jin L, Li H (2012) Present Y chromosomes reveal the ancestry of Emperor CAO Cao of 1800 years ago. J Hum Genet 57:216–218. doi:10.1038/jhg.2011.147

Wang D, Liu F, Kong L, Yuan Z, Chen G, Ye J (2015) Population data of 17 Y-STR haplotypes in Jining Han population from Shandong province, East China. Forensic Sci Int Genet 19:47–49. doi:10.1016/j.fsigen.2015.05.017

Wang H, Mao J, Xia Y, Bai X, Zhu W, Peng D, Liang W (2016a) Genetic polymorphisms of 17 Y-chromosomal STRs in the Chengdu Han population of China. Int J Legal Med. doi:10.1007/s00414-016-1511-4

Wang L, Chen F, Kang B, Zheng H, Zhao Y, Li L, Zeng Z (2016b) Genetic population data of Yfiler Plus kit from 1434 unrelated Hans in Henan Province (Central China). Forensic Sci Int Genet 22:e25–e27. doi:10.1016/j.fsigen.2016.02.009

Wang Y, Zhang YJ, Zhang CC, Li R, Yang Y, Ou XL, Tong DY, Sun HY (2016c) Genetic polymorphisms and mutation rates of 27 Y-chromosomal STRs in a Han population from Guangdong Province, Southern China. Forensic Sci Int Genet 21:5–9. doi:10.1016/j.fsigen.2015.09.013

Weale ME, Weiss DA, Jager RF, Bradman N, Thomas MG (2002) Y chromosome evidence for Anglo-Saxon mass migration. Mol Biol Evol 19:1008–1021

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 (2004) Genetic evidence supports demic diffusion of Han culture. Nature 431:302–305. doi:10.1038/nature02878

Wu W, Pan L, Hao H, Zheng X, Lin J, Lu D (2011) Population genetics of 17 Y-STR loci in a large Chinese Han population from Zhejiang Province, Eastern China. Forensic Sci Int Genet 5:e11–e13. doi:10.1016/j.fsigen.2009.12.005

Xiao CJ, Cavalli-Sforza LL, Minch E, Du RF (2000) Geographic distribution maps of human genes in China. Yi Chuan Xue Bao 27:1–6

Xu S, Yin X, Li S, Jin W, Lou H, Yang L, Gong X, Wang H, Shen Y, Pan X, He Y, Yang Y, Wang Y, Fu W, An Y, Wang J, Tan J, Qian J, Chen X, Zhang X, Sun Y, Zhang X, Wu B, Jin L (2009) Genomic dissection of population substructure of Han Chinese and its implication in association studies. Am J Hum Genet 85:762–774. doi:10.1016/j.ajhg.2009.10.015

Xu H, Wang CC, Shrestha R, Wang LX, Zhang M, He Y, Kidd JR, Kidd KK, Jin L, Li H (2015) Inferring population structure and demographic history using Y-STR data from worldwide populations. Mol Genet Genom 290:141–150. doi:10.1007/s00438-014-0903-8

Xu J, Li L, Wei L, Nie Z, Yang S, Xia M, Ma T, Sun H, Zhao X, Ping Y, Zhou H, Xue F, Zhao Z, Jin L, Li S (2016a) Genetic analysis of 17 Y-STR loci in Han population from Shandong Province in East China. Forensic Sci Int Genet 22:e15–e17. doi:10.1016/j.fsigen.2016.01.016

Xu S, Yang S, Yang M, Jia D, Han X, Wang W, Jin L, Li L, Li S (2016b) Analysis of Y-chromosome short tandem repeat loci on 1082 Nantong Han individuals in eastern China. Forensic Sci Int Genet 23:e18–e19. doi:10.1016/j.fsigen.2016.04.011

Xue Y, Zerjal T, Bao W, Zhu S, Lim SK, Shu Q, Xu J, Du R, Fu S, Li P, Yang H, Tyler-Smith C (2005) Recent spread of a Y-chromosomal lineage in northern China and Mongolia. Am J Hum Genet 77:1112–1116. doi:10.1086/498583

Xue Y, Zerjal T, Bao W, Zhu S, Shu Q, Xu J, Du R, Fu S, Li P, Hurles ME, Yang H, Tyler-Smith C (2006) Male demography in East Asia: a north-south contrast in human population expansion times. Genetics 172:2431–2439. doi:10.1534/genetics.105.054270

Xue F, Wang Y, Xu S, Zhang F, Wen B, Wu X, Lu M, Deka R, Qian J, Jin L (2008) A spatial analysis of genetic structure of human populations in China reveals distinct difference between maternal and paternal lineages. Eur J Hum Genet 16:705–717. doi:10.1038/sj.ejhg.5201998

Yan S, Wang CC, Zheng HX, Wang W, Qin ZD, Wei LH, Wang Y, Pan XD, Fu WQ, He YG, Xiong LJ, Jin WF, Li SL, An Y, Li H, Jin L (2014) Y chromosomes of 40% Chinese descend from three Neolithic super-grandfathers. PLoS One 9:e105691. doi:10.1371/journal.pone.0105691

Yang Y, Yuan W, Guo F, Jiang X (2014) Population data of 17 Y-STR loci in Nanyang Han population from Henan Province, Central China. Forensic Sci Int Genet 13:145–146. doi:10.1016/j.fsigen.2014.07.013

Yao HB, Wang CC, Tao X, Shang L, Wen SQ, Zhu B, Kang L, Jin L, Li H (2016) Genetic evidence for an East Asian origin of Chinese Muslim populations Dongxiang and Hui. Sci Rep 6:38656. doi:10.1038/srep38656

Zhang S, Tian H, Wang Z, Zhao S, Hu Z, Li C, Ji C (2014) Development of a new 26plex Y-STRs typing system for forensic application. Forensic Sci Int Genet 13:112–120. doi:10.1016/j.fsigen.2014.06.015

Zhao YB, Zhang Y, Zhang QC, Li HJ, Cui YQ, Xu Z, Jin L, Zhou H, Zhu H (2015) Ancient DNA reveals that the genetic structure of the northern Han Chinese was shaped prior to 3,000 years ago. PLoS One 10:e0125676. doi:10.1371/journal.pone.0125676

Zhong H, Shi H, Qi XB, Duan ZY, Tan PP, Jin L, Su B, Ma RZ (2011) Extended Y chromosome investigation suggests postglacial migrations of modern humans into East Asia via the northern route. Mol Biol Evol 28:717–727. doi:10.1093/molbev/msq247