Legal Medicine 31 (2018) 49–58

Contents lists available at ScienceDirect

Legal Medicine

journal homepage: www.elsevier.com/locate/legalmed

The genetic diversity and applicability assessment of autosomal STRs among Chinese populations by a novel Fixation Index and Nei's index

Caiyong Yin[a], Chuwei Deng[a], Xiaoqin Qian[b], Huijie Huang[a], Yanfang Yu[a], Li Hu[a], Peng Chen[a,*], Feng Chen[a,*]

[a] *Department of Forensic Medicine, Nanjing Medical University, Nanjing, Jiangsu 210029, China*
[b] *Institute of Forensic Medicine, West China School of Basic Science and Forensic Medicine, Sichuan University, Chengdu, Sichuan 610041, China*

ARTICLE INFO

Keywords:
Forensic science
STR
Fixation Index
Chinese Han
Chinese minorities
Population differentiation

ABSTRACT

The population-specific F_{ST} in STR loci of Chinese populations has not been focused on. Here, we genotyped 19 STRs in 530 unrelated healthy individuals of Xuzhou Han population, and collected data of 30,308 samples from 32 Hans and 50 minorities nationwide. The population-specific β_i and locus-specific β_{il} were calculated to evaluate the applicable value of STRs. Next, we generated the genetic structure of various ethnic populations by Neighbor-Joining tree and Multidimensional Scaling plot based on pairwise Nei's distances. We found that TH01 and TPOX possessed high ability in discriminating populations which may be reled to the mutation rate of these STRs. Additionally, our data indicated that Chinese Han was homogenous and the population-specific β_is of northern Hans were generally smaller than southern Hans ($p > 0.05$). We concluded that population-specific F_{ST} for autosomal STR loci could be used to reveal the unique genetic characteristics and thus uncover the genetic relationship among Chinese populations.

1. Introduction

China is a heterogeneous country with 56 nationalities. Approximately 92% of the total population is the Han nationality and the rest derives from different minorities, and the top three minorities with less population are Zhuang (0.12%), Hui (0.07%) and Manchu (0.07%) (http://www.stats.gov.cn/tjsj/pcsj/rkpc/6rp/indexch.htm). Previous studies find the significant difference between northern and southern populations of China with means of archaeological analysis [1] and genetic evidences on microsatellites [2], Y-SNP [3] and autosomal SNP [4]. Nonetheless, some researchers suggest no proofs to support this view by analyzing mtDNA [5] and virus [6]. Therefore, population heterogeneity of the populations of China hasn't been reached an agreement from the genetic perspective. The difficulty in structure reconstruction of Chinese population is also embodied in the patterns of Chinese cultural pluralism. The linguistic affinities and culture differences are evidenced not only by different nationalities, but also by different areas and different time of the identical nationalities [7]. These issues cause difficulties in origin inference and dispersal model construction of Chinese.

Fixation Index (F_{ST}) is an important ramification of F-statistics, which is firstly introduced in 1951 [8]. F_{ST} is described as the estimator of the coancestry coefficient within a population [9] and in the study by Reynolds et al., they concluded that F_{ST} is a monotonically increasing function of divergence time and pointed that the formula is an estimator of divergence time [10]. In 2012, Bruce et al. emphasized that the evolutionary framework allowed F_{ST} to be regarded as a measure of the time since the sampled populations diverged from an ancestral population [11]. The more divergence time of one population means the longer process of differentiation this population has undergone. Other derivative study also supported the significance of F_{ST} in defining the population ancestor and divergence time [12]. In the latest research concerning population-specific F_{ST} [13], the β_is among populations worldwide are reported. It shows that under the large data volume, the values of Asian populations, relatively greater than Caucasians, both lie between that of African and Native American groups, which nicely indicate the ancient degree of the studied populations. The population-specific F_{ST} intuitively shows the level of genetic diversities among

---

* Corresponding authors.
E-mail addresses: fchen@njmu.edu.cn (F. Chen), chenpeng@njmu.edu.cn (P. Chen).

https://doi.org/10.1016/j.legalmed.2017.12.012
Received 11 September 2017; Received in revised form 4 December 2017; Accepted 27 December 2017

populations and population age. Therefore, as a new tool for analyzing origin inference, population-specific F_{ST} begins to take shape. However, the value and significance of β_is in studying population structure and migration of Chinese populations have not been reported.

In this research, we aimed to exploit the genetic structure of Chinese nationalities in aspect of population-specific F_{ST}. Firstly, we genotyped 530 Hans living in Xuzhou and obtained a representative of Jiangsu Han population with other three populations in Jiangsu province reported before [14]. Next, we collected the genotyping files from reported populations covering the whole China areas, and employed F statistics to estimate the population-specific F_{ST} of these populations. The detailed distribution of population-specific β_is and locus-specific β_{il}s were discovered. Also, we constructed Neighbor-Joining (N-J) tree and Multidimensional Scaling (MDS) plot of various Chinese ethnic groups to further validate the findings from β_is.

2. Materials and methods

2.1. Samples collection and DNA extraction

We collected blood samples of 530 unrelated and healthy individuals within formed consent whose families have been living in Xuzhou, Jiangsu Province, China for at least three generations. The whole-blood samples were anti-coagulated with EDTA and then frozen at −80 °C before use. To extract genomic DNA, the Chelex-100 was used as described previously [15]. The study was carried out in according to the human and ethical research principles of Nanjing Medical University and approved by the ethics committee in Nanjing Medical University.

2.2. STR genotyping

Nineteen STR loci and a sex determination gene (Amelogenin gene) were co-amplified using the Goldeneye™ DNA ID system 20A kit (Goldeneye Ltd, Beijing, China) in the GeneAmp PCR system 9700 with a gold-plate silver or silver block (Applied Biosystems, Foster City, CA, USA). The amplified products were separated on ABI3130XL DNA Genetic Analyzer (Applied Biosystems, USA) and genotyped with the GeneMapper ID-X software (Applied Biosystems, Foster City, CA, USA).

2.3. Statistical analysis

The values of forensic efficiency and statistical parameters include observed heterozygosity (Ho), expected heterozygosity (He), power of discrimination (PD), polymorphism information content (PIC), and power of exclusion (PE). Genetic parameters of interest and the exact chi-square test for the Hardy-Weinberg equilibrium (HWE) were calculated using the modified PowerStates spread-sheet (version 1.2) [16]. The estimated population-specific β_i and locus-specific β_{il} were calculated using “HIERFSTAT” package of R (3.3.2) while “easyGGplot2” package was used to perform violin plots. Phylip 3.69 program was used to calculate Nei’s genetic distances [17] and to draw a Neighbor-Joining (N-J) tree of 53 populations. Next, N-J trees were visualized with the TreeView software. In addition, “MASS” package was employed to perform Multidimensional Scaling plot (MDS) and we achieved the values of initial stress to evaluate the quality of MDS. Available published populations included in the present study were shown in Supplementary Table 1.

3. Results and discussion

3.1. Allelic diversities of Xuzhou Han population

The allele frequencies, forensic efficiencies, and statistical parameters were shown in Supplementary Table 2. The PD ranged from 0.8123 (TH01) to 0.9860 (Penta E) with the value of 0.99999999999999999999999884 for combined probability of discrimination (CPD). The PE ranged from 0.2954 (TPOX) to 0.8535 (Penta E) with the value of 0.99999999378 for combined probability of exclusion (CPE). This panel containing 19 STRs was proved to be of high values in practical application. The He had the range from 0.6038 (TPOX) to 0.9283 (Penta E). The increase of PIC values from 0 to 1 implied the ascending of polymorphism. In this research, all STR loci presented high polymorphisms (PIC > 0.5) with the best at Penta E (0.9139). After Bonferroni correction, no significant deviation was observed from HWE when the level of significance was set as 0.05/19 ≈ 0.0026.

3.2. Estimated locus-specific F_{ST} of Chinese Han and minorities

While the STR data of Qinghai and Tibetan Hans were unavailable, we analyzed the genotyping data covering CODIS-13 in remaining 32 Han populations, in which Taizhou, Yangzhou, Huai’an Hans [14], together with Xuzhou Han represented Jiangsu Han group. The estimated locus-specific β_{il} for each province-harbored Han population were listed in Table 1. A violin plot was illustrated to compare the calculated locus-specific β_{il} applying to 32 Han groups from different administrated regions. It showed high polymorphisms of β_is’ distribution among 19 STRs (Fig. 1a). STRs were arranged along corresponding mutation rates (gradually increasing from left to right according to our previous work [18]). It was obvious that distributions of the TH01 and TPOX loci were close to normal distribution. The distributions of other STRs were skewed of different degree, especially for D12S391, D6S1043, and Penta E. The loci in which alleles distributed normally were determined to possess high ability in discriminating Chinese Han populations.

Then, we focused on the determination of locus-specific β_{il} of 19 STRs in all 55 Chinese ethnic minorities. No published STR data from the Gaoshan, Tajik, Tatar, Hezhen, and Moinba ethnic groups was available to date. We combined 5076 samples from 5 representative Han populations (Shaanxi, Henan, Changchun, Guangdong, and Sichuan Hans) and 11,034 samples from 50 ethnic minorities (Supplementary Table 3), all labelled with linguistic affinities, to increase the accuracy of estimated locus-specific β_{il} for the representative ethnic groups of Chinese (Table 2). The distributions of β_{il} values at ethnic minorities (Fig. 1b) were significantly different from those in Fig. 1a, mainly manifested in the modes, density and state of distributions, primarily caused by the higher genetic diversity in 51 ethnic populations than Han groups. It was observed that distributions of more STR loci tended to be normal, such as D7S820 and vWA. Nonetheless, the TH01 and TPOX loci remained unchanged roughly. As defined [19], the more fusiform patterns, in other words, the more normal distributions always indicate higher ability of population discrimination. However, the distributions of some STRs like D6S1043 and Penta E still remained skewed distribution.

Further, we summarized the potential rules underlying the relationship between STR diversity and according power of population discrimination in the context of F statistics. According to our previous studies on mutation rate of STRs as well as others, STRs with low ability of point mutation were proved here to exert high efficiency in discriminating ethnic groups [18,20–23], which was exemplified by TH01 and TPOX with the lowest mutability (mutation rate = 0.0000 and 0.0011, respectively). In this research, the intricate inter-population genetic variation was proved to be easily distinguished by allele frequencies’ changes within established set of alleles in each microsatellite evaluated by F statistics, instead of that of new-born alleles caused by mutation. Also, it demonstrated that though STRs with high mutability played positive roles in forensic cases [18,20–23], they were negative in population differentiation analysis, as well the phylogenetic structure reconstruction.