Forensic Science International: Genetics 29 (2017) e4–e8

Contents lists available at ScienceDirect

Forensic Science International: Genetics

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

Short communication

Forensic characteristics and phylogenetic analysis of Hubei Han population in central China using 17 Y-STR loci

Zheng Wang[a], Weian Du[b,1], Guanglin He[a], Jing Liu[a], Yiping Hou[a,*]

[a] Institute of Forensic Medicine, West China School of Preclinical and Forensic Medicine, Sichuan University, Chengdu 610041, China
[b] School of Forensic Medicine, Southern Medical University, Guangzhou 510515, China

ARTICLE INFO

Article history:
Received 31 December 2016
Received in revised form 12 March 2017
Accepted 17 April 2017
Available online 20 April 2017

Keywords:
Y chromosomal STR
AmpFLSTR Yfiler
Haplotype diversity
phylogenetic analysis

ABSTRACT

Currently, the largest national database within the Y chromosome haplotype reference database (YHRD, https://yhrd.org, release 53) is China, which has approximately 38000 Y chromosomal 17-marker (Yfiler) haplotypes. These haplotype profiles derived from the vast majority of Chinese administrative divisions, but no haplotype data was available for Hubei province, which is located in the Central China region. Herein, 429 unrelated male Chinese Han individuals residing in Hubei province were recruited and genotyped with 17 Y-STR loci. 115 alleles were identified with corresponding allele frequencies spanned from 0.0023 to 07506. The gene diversity (GD) values ranged from 0.3988 at DYS438 to 0.9573 at DYS385a/b. A total of 410 distinct haplotypes were obtained with the overall haplotype diversity (HD) and discrimination capacity (DC) was 0.9995 and 0.9557, respectively. Additionally, genetic relationships along administrative (Han Chinese from different provinces) and ethnic divisions (minority ethnic groups) were analyzed using analysis of molecular variance (AMOVA) tests and visualized by multidimensional scaling plots (MDS). The Han ethnicity including the Hubei Han shows a high genetic homogeneity all across China and significant genetic differences existed between the Hubei Han and some ethnic groups, most prominently for the Kazakhs and the Tibetans.

1. Introduction

Y chromosomal haplotypes consisting of short tandem repeats (STRs) have been widely used in genetic application like forensic evidence examination, historical investigations and genealogical research [1–4]. In the forensic community, Y-STR haplotypes are routinely used to match male perpetrators to evidence, and an appropriate haplotype reference database is necessary to estimate the frequency of specified haplotypes. As the largest and most widely used forensic and general population genetics Y-STR database, the Y-STR Haplotype Reference Database (YHRD, https://yhrd.org, release 53) contains more than 126,000 Yfiler (DYS19, DYS389I, DYS389II, DYS390, DYS391, DYS392, DYS393, DYS385a/b, DYS438, DYS439, DYS437, DYS448, DYS456, DYS458, DYS635 and Y GATA H4) haplotype profiles from various populations and countries around the world [4–6].

Currently, the largest national database within the YHRD is China with approximately 38,000 Yfiler haplotypes derived from local population studies [6]. These haplotypes derived from the vast majority of Chinese administrative divisions, however, no haplotype data was available for Hubei province, which is located in the Central China region. Hubei province with the population of approximately 59.0 million ranked ninth in China, and the dominant ethnic group, Han, accounts for 95.6% of Hubei population (https://en.wikipedia.org/wiki/Hubei).

Consequently, the aim of this study was to provide the first batch of Y-STR haplotype data of Hubei Han population using AmpFLSTR Yfiler PCR amplification kit (Thermo Fisher, CA, USA). Furthermore, genetic similarities between the Hubei Han and previous published Han Chinese residing in different administrative divisions were explored, and genetic distinctions between the Hubei Han and Chinese minority ethnic groups were also explored.

2. Material and Methods

2.1. Sample Collection

Human blood samples were collected with the approval of the Ethics Committee of the Institute of Forensic Medicine, Sichuan

* Corresponding author at: Institute of Forensic Medicine, West China School of Preclinical and Forensic Medicine, Sichuan University, 3-16 Renmin South Road, Chengdu 610041, China.
E-mail address: forensic@scu.edu.cn (Y. Hou).

[1] The author contributed equally to this work and should be considered co-first author.

http://dx.doi.org/10.1016/j.fsigen.2017.04.013
1872-4973/

University. Written informed consent was obtained from each participant. Unrelated blood samples were collected from 429 healthy male Han individuals residing in the Huanggang area of Hubei province, Central China (Fig. 1). The inclusion criteria for participants who have lived in Hubei province for at least three generations were employed.

2.2. DNA extractions and quantification

Genomic DNA was extracted by the QIAamp DNA blood mini kit (Qiagen, Hilden, Germany) according to the manufacturer's protocol. The quantity of the DNA template was determined using Quantifiler Human DNA Quantification Kit (Thermo Fisher) on a 7500 Real-Time PCR System (Thermo Fisher). DNA samples were then normalized to 1.0 ng/μL and stored at −20 °C until amplification.

2.3. DNA amplification and genotyping

Seventeen Y-STRs were co-amplified using the AmpFLSTR Yfiler PCR Amplification kit on a ProFlex 96-well PCR System (Thermo Fisher) in accordance with the manufacturer's instructions. PCR products were separated and analyzed via Applied Biosystems 3130 Genetic Analyzer (Thermo Fisher) following the manufacturer's recommendations. Allele allocation was carried out with GeneMapper ID V3.2 analysis software (Thermo Fisher) using the allelic ladder and the set of bins and panels provided by the kit.

2.4. Statistical analysis

Allele and haplotype frequencies were estimated by the direct counting method. Gene diversity (GD) and haplotype diversity (HD) were calculated using the following formula [7]: $GD = \frac{n}{n-1}\left(1 - \sum p_i^2\right)$, where n denotes the total number of observed allele or haplotypes and pi means the relative frequency of the i-th allele or haplotype, respectively. The match probability (MP) was calculated as the sum of squared haplotype frequencies, whereas the discrimination capacity (DC) was computed as the ratio between the total number of different haplotypes and the number of haplotypes.

Genetic relationships between different groups were quantified by means of R_{ST} [8,9]. The pairwise genetic distance of R_{ST} value and corresponding p value were computed using analysis of molecular variance (AMOVA) on the online tools in YHRD. Genetic similarities and differences were further visualized by multidimensional scaling (MDS) plot. Additionally, genetic homogeneity

Fig. 1. The geographical position of our investigated population, the Huanggang area of Hubei province, Central China, marked with a red star.