

Materials and methods

Sample preparation

We used the self-developed multiplex amplification panel to type 43 autosomal InDel loci and Amelogenin gene to identify gender in 301 unrelated healthy donors collected from Chinese Han population residing in Beijing city. All the participants signed written informed consents and stated that they were no blood relationships with each other. The ethics committee of Xi'an Jiaotong University Health Science Center and Southern Medical University approved all the processes including the sample collections, experimental design and so on during this research (Approval No. XJTULAC201). Table S1 lists the detailed information of Han Chinese from Beijing and 26 reference populations from the 1000 Genomes Project phase III.

Multiplex PCR amplification and corresponding InDel genotyping

The self-developed panel including 43 autosomal InDel loci, along with a genetic marker was amplified on the GeneAmp PCR System 9700 (Applied Biosystems, Foster City, USA). The total volume of PCR amplification system was 20 μ L. Subsequently, the PCR products were isolated and detected on ABI 3500xL Genetic Analyzer (Thermo Fisher Scientific, South San Francisco, USA). Then, the genotyping results for 43 InDel loci and Amelogenin locus were obtained by GeneMapper ID-X software v.1.3 (Thermo Fisher Scientific, South San Francisco, USA).

Statistical analyses

The P values of Hardy-Weinberg equilibrium (HWE), and linkage disequilibrium (LD) tests of 43 InDels in Han Chinese from Beijing were calculated by Genopop v.3.1 software. Insertion and deletion allele frequencies, discrimination power (PD), match probability (PM), probability of exclusion (PE), observed heterozygosity (H_o) and polymorphism information content (PIC) were calculated by Modified Powerstat spreadsheet. Expected heterozygosity (H_e) was calculated by STRAF software v.1.0.5. Heatmaps of insertion allele frequencies and intercontinental fixation index (F_{ST}) values based on the 43 InDel loci were conducted by R software v.4.0.5. Nei's genetic distance (D_A) and F_{ST} values based on 19 overlapped InDel loci were evaluated by DISPAN and Arlequin software v.3.5, respectively. Phylogenetic tree of Han Chinese from Beijing and 26 reference populations from the 1000 Genomes Project phase III was performed by Interactive Tree of Life on-line site (<https://itol.embl.de/itol.cgi>). Principal component analyses (PCA) were constructed by Snipper online website (<http://mathgene.usc.es/snipper/index.php>) and R software

v.4.0.5 on individual and population levels, respectively. The population genetic structure analyses were evaluated by STRUCTURE software v.2.3.4, and STRUCTUR plot at optimal K value was visualized by distruct software v.1.1. Then, Pairwise F_{ST} and P values of locus-by-locus comparisons between Han Chinese from Beijing and 26 reference populations were measured by Arlequin v.3.5.