Research Article

Genetic Polymorphisms and Forensic Efficiencies of a Set of Novel Autosomal InDel Markers in a Chinese Mongolian Group

Wenqing Zhang, Xiaoye Jin, Yijie Wang, Tingting Kong, Wei Cui, Chong Chen, Yuxin Guo, and Bofeng Zhu

1Key Laboratory of Shaanxi Province for Craniofacial Precision Medicine Research, College of Stomatology, Xi’an Jiaotong University, Xi’an, China
2Clinical Research Center of Shaanxi Province for Dental and Maxillofacial Diseases, College of Stomatology, Xi’an Jiaotong University, Xi’an, China
3College of Forensic Medicine, Xi’an Jiaotong University Health Science Center, Xi’an, China
4Department of Forensic Genetics, School of Forensic Medicine, Southern Medical University, Guangzhou, China

Correspondence should be addressed to Bofeng Zhu; zhubofeng7372@126.com

Received 16 July 2019; Accepted 17 October 2019; Published 8 January 2020

Copyright © 2020 Wenqing Zhang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Insertion/deletion (InDel) markers have been treated as a prospective and helpful aid in the fields of forensic human identifications and biogeography origin researches for the past few years. In this study, we analyzed genetic polymorphisms and forensic efficiencies of 35 InDels in a novel multiplex PCR-InDel panel in a Chinese Mongolian group. All these 35 InDel loci were observed to conform to Hardy–Weinberg equilibrium and linkage equilibrium. The mean values of expected heterozygosity and observed heterozygosity were 0.4788 and 0.4852, respectively. Besides, the interpopulation differentiations and genetic distributions based on 35 InDels found that the Chinese Mongolian group might have closer genetic relationships and similar population genetic structures with East Asian populations.

1. Introduction

InDels are length polymorphisms resulting from the insertion or deletion of one or more nucleotides in the genome [1]. In 2002, Weber et al. firstly identified and characterized 2000 human biallelic InDels, which differed greatly in the lengths of the observed alleles, and they also emphasized the usefulness of InDels in genetic researches because of their richness and ease of analysis [2]. Since then, in more and more published studies, InDels have been used for a variety of purposes [3, 4]. InDels have many strengths in forensic analyses: firstly, they are widely distributed across the human genome and commonly display small amplicons which are conducive to the analyses of degraded or dated samples; secondly, the mutation rates of InDels are lower compared with short tandem repeat (STR) loci; thirdly, they have no microvariant products, which could make them more applicable for the interpretation of the mixture. Additionally, they could also serve as ancestry-informative markers (AIMs) for characterizing population substructure and performing biogeographical origin analyses [5–7]. In recent years, more and more studies have found that InDels could be useful in human identification [1], mixed stain identification [8], and so on.

Mongolian, in terms of population size, is the tenth largest ethnic group in China, distributed in Gansu, Qinghai Provinces, Xinjiang Uygur, and Inner Mongolia Autonomous Regions. Some Mongolians also dwell in Liaoning, Jilin, Heilongjiang, and other provinces. The language of Mongolian group belongs to the Altaic family. The main religion of the Mongolian people is Buddhism (http://www.paulnoll.com/China/Minorities/min-Mongolian.html). Nowadays, genetic analyses of the Chinese Mongolian group mainly focused on STR loci, such as 19 X-STR loci [9] and 19 autosomal STRs [10], 22 autosomal STR loci [11], 12 X-STR [12], and 27 Y-STR [13];
besides, Jin et al. used 48 single nucleotide polymorphism (SNP) loci to study genetic relationships among continental populations and Chinese populations including Mongolian group [14]. However, to date, few studies on autosomal InDels in the Chinese Mongolian ethnic group have been conducted.

Previously, we developed a novel multiplex PCR-InDel panel for forensic individual identifications in the Chinese Kazak group and reference populations from East Asia [15]. Here, genetic distributions and forensic efficiencies of these InDels in the Chinese Mongolian group were further investigated. Besides, heat maps of fixation index ($F_{st}$) and Nei’s genetic distances ($D_A$ distances), principal component analysis (PCA), phylogenetic reconstruction, population clustering analysis of the studied Mongolian ethnic group and other reference populations were also constructed to explore their genetic relationships.

2. Materials and Methods

2.1. Subjects and Sample Collection. We collected a total of 110 bloodstain samples from unrelated healthy Mongolian individuals in China. All subjects signed the written informed consent prior to sampling. This study obtained the approval of the Ethics Committees of Xi’an Jiaotong University Health Science Center and Southern Medical University, China.

2.2. PCR Amplification and InDel Genotyping. In this study, the PCR amplification of 35 InDel loci was conducted on a GeneAmp PCR system 9700 thermal cycler (Applied Biosystems, Foster City, CA, USA), following previous descriptions [15]. Then, the PCR amplification products were separated and detected by capillary electrophoresis on the ABI 3500 Genetic Analyzer (Applied Biosystems, Foster City, CA, USA). The allele typing was performed by GeneMapper v3.2 software (Applied Biosystems, Foster City, CA, USA).

2.3. Reference Populations. The reference populations of this study were the intercontinental populations [16] and Kazak group [15] reported before, which were African (7 populations), American (4 populations), East Asian (5 populations), European (5 populations), South Asian (5 populations), and Chinese Kazak group. Detailed information of these continental populations are as follows: African populations include ACB, ASW, ESN, GWD, LWK, MSL, and YRI; American populations include CLM, MXL, PEL, and PUR; East Asian populations include CDX, CHB, CHS, JPT, and KHV; European populations include CEU, FIN, GBR, IBS, and TSI; South Asian populations include GIH, ITU, PJL, STU, and BEB. Detailed information of these reference populations was presented in Table 1. And the geographic localization of the Chinese Mongolian and other reference populations was shown in Figure S1.

2.4. Statistical Analysis. The allelic frequency distributions and forensic statistical parameters which included the values of observed heterozygosity (Ho), expected heterozygosity (He), polymorphism information content (PIC), discrimination power (DP), probability of exclusion (PE), match probability (MP), typical paternity index (TPI), and $P$ values for Hardy–Weinberg equilibrium tests ($P$ values) and linkage disequilibrium (LD) analyses of 35 InDel loci in Mongolian group were calculated by STRAF online program (version 1.0.5) [17]. A map showing population distributions was plotted by R software (version 3.4.5) (https://www.r-project.org/). $D_A$ distances and $F_{st}$ values among the studied Mongolian group and other reference populations were calculated by DISPAN program [18] and Genepop software (version 4.0) [19], respectively. Then heat maps of $D_A$ and $F_{st}$ values of these populations were conducted on heatmap package (version 1.0.12) by R software (version 3.4.5). The PCA of the studied Mongolian group and 27 compared populations was generated by MVSP software (version 3.1). Moreover, PCA of these populations at individual level was conducted by PLINK software (version 1.9) [20], which was visualized by ggplot 2 package (version 3.2.0) of R software (version 3.4.5). Beyond that, a multidimensional scaling (MDS) plot [21] was conducted by SPSS software (version 23.0). Additionally, a phylogenetic tree based on $D_A$ distances was established by MEGA software (version 6.06) [22]. The population genetic structure analyses were evaluated by using the STRUCTURE software (version 2.3.4.) [23] and CLUMPP software (version 1.1.2). The appropriate $K$ value was assessed by the Structure Harvester online tool (version 0.6.94) (http://taylor0.biology.ucla.edu/structureHarvester/).

3. Results

3.1. Hardy–Weinberg Equilibrium Tests and Forensic Statistical Parameters of 35 InDels. The $P$ values for Hardy–Weinberg equilibrium tests and forensic statistical parameters of 35 InDels were shown in Table 2. From Table 2, the significant differences from $P$ values at the 35 InDels were not found after a Bonferroni correction ($P = 0.05/35 = 0.0014$). The PIC values ranged from 0.1190 (rs3054057) to 0.3750 (rs4024564), 0.1160964, rs61681053, and rs10556291), with the mean value of 0.3609. The He and Ho values ranged from 0.1276 (rs3054057) to 0.5023 (rs10556291), and from 0.1182 (rs3054057) to 0.6273 (rs3028455), with the mean values of 0.4788 and 0.4852, respectively. The values of MP, DP, PE, and TPI varied from 0.3522 (rs4024564) to 0.7757 (rs3054057), 0.2243 (rs3054057) to 0.6478 (rs4024564), 0.0114 (rs3054057) to 0.3249 (rs3028455), and 0.5670 (rs3054057) to 1.3415 (rs3028455), respectively. Additionally, the frequencies of insertion alleles (+) and deletion alleles (−) ranged from 0.3682 (rs371194629) to 0.9318 (rs3054057), and from 0.0682 (rs3054057) to 0.6318 (rs371194629), respectively. The forensic parameters of cumulative PE and combined DP values of 35 InDel loci were 0.99925 and 0.9999999999999904, respectively.
3.2. Linkage Disequilibrium Analyses of 35 InDels. LD tests of these 35 InDel loci in the Chinese Mongolian group were calculated by STRAF online program (version 1.0.5). As shown in Table S1, pairwise InDels were observed to conform to linkage equilibrium after applying a Bonferroni correction ($P = 0.05/595 = 0.000084$), indicating that these 35 InDel loci were mutually independent in the studied Mongolian ethnic group.

3.3. Interpopulation Differentiations Based on 35 InDels. Absolute values of insertion allelic frequency differences ($\delta$) between the studied Mongolian group and other reference populations were given in Table S2. Results showed that the studied Mongolian group had relatively low $\delta$ values ($<0.1$) with East Asian populations and Kazak group at most loci in comparisons with other reference populations. And then, genetic distances ($D_A$) of the studied group and other reference populations were calculated using allelic frequencies of 35 InDel loci, as shown in Figure 1(a) and Table S3. As one of the most generally used genetic distances, $D_A$ distance has been used to measure genetic differences of different populations. It is developed on the assumption that genetic drift and mutation events finally lead to genetic differences [24]. As shown in Figure 1(a), the population names at the bottom axis and right vertical axis of the triangle corresponded to the paired populations of each block, which represented the $D_A$ values of the paired populations. The different color scaling on the upper right corner of the plot showed the $D_A$ values ranged from 0 to 0.06 which varied from orange to red. Different colors stood for different levels of $D_A$ values: orange for low $D_A$ values and red for high $D_A$ values. Orange color blocks between the Chinese Mongolian and the Kazak groups as well as five East Asian populations (CDX, CHB, CHS, JPT, and KHV) were observed, indicating that they had close genetic relationships. However, the colors of the blocks between the Mongolian group and some African populations were nearly red, meaning that there were relatively large $D_A$ values between the Chinese Mongolian and these African populations. In addition, we also generated a heat map based on the $Fst$ values of the pairwise populations to further measure population differentiations, as shown in Figure 1(b) and Table S4. Likewise, the population genetic relationships were reflected by the depth of each block’s color, which changed from deep green to blue. The closer the color was to deep green, the lower the $Fst$ value was, indicating that the genetic differences of pairwise populations were the smaller. We also found that the blocks of the Chinese Mongolian group and five East Asian and Kazak groups showed deep green colors while the Chinese Mongolian group and other intercontinental populations showed light green or blue colors, showing that genetic differentiations between Chinese Mongolian and the five East Asian populations as well as Kazak group were smaller compared to the other reference intercontinental populations.

3.4. Principal Component Analysis and Multidimensional Scaling. The genetic relationships between the Chinese

<table>
<thead>
<tr>
<th>Populations</th>
<th>Abbreviations</th>
<th>Continents</th>
<th>Sample sizes</th>
<th>Sources</th>
</tr>
</thead>
<tbody>
<tr>
<td>African Caribbean in Barbados</td>
<td>ACB</td>
<td>Africa</td>
<td>96</td>
<td>1000 genomes project phase 3</td>
</tr>
<tr>
<td>African Ancestry in Southwest USA</td>
<td>ASW</td>
<td>Africa</td>
<td>61</td>
<td>1000 genomes project phase 3</td>
</tr>
<tr>
<td>Esan in Nigeria</td>
<td>ESN</td>
<td>Africa</td>
<td>99</td>
<td>1000 genomes project phase 3</td>
</tr>
<tr>
<td>Gambian in Western Division, Mandinka</td>
<td>GWD</td>
<td>Africa</td>
<td>113</td>
<td>1000 genomes project phase 3</td>
</tr>
<tr>
<td>Luhya in Webuye, Kenya</td>
<td>LWK</td>
<td>Africa</td>
<td>99</td>
<td>1000 genomes project phase 3</td>
</tr>
<tr>
<td>Mende in Sierra Leone</td>
<td>MSL</td>
<td>Africa</td>
<td>85</td>
<td>1000 genomes project phase 3</td>
</tr>
<tr>
<td>Yoruba in Ibadan, Nigeria</td>
<td>YRI</td>
<td>Africa</td>
<td>108</td>
<td>1000 genomes project phase 3</td>
</tr>
<tr>
<td>Colombian in Medellin, Colombia</td>
<td>CLM</td>
<td>America</td>
<td>94</td>
<td>1000 genomes project phase 3</td>
</tr>
<tr>
<td>Mexican ancestry in Los Angeles, California</td>
<td>MXL</td>
<td>America</td>
<td>64</td>
<td>1000 genomes project phase 3</td>
</tr>
<tr>
<td>Peruvian in Lima, Peru</td>
<td>PEL</td>
<td>America</td>
<td>85</td>
<td>1000 genomes project phase 3</td>
</tr>
<tr>
<td>Puerto Rican in Puerto Rico</td>
<td>PUR</td>
<td>America</td>
<td>104</td>
<td>1000 genomes project phase 3</td>
</tr>
<tr>
<td>Chinese Dai in Xishuangbanna, China</td>
<td>CDX</td>
<td>East Asia</td>
<td>93</td>
<td>1000 genomes project phase 3</td>
</tr>
<tr>
<td>Han Chinese in Beijing, China</td>
<td>CHB</td>
<td>East Asia</td>
<td>103</td>
<td>1000 genomes project phase 3</td>
</tr>
<tr>
<td>Southern Han Chinese, China</td>
<td>CHS</td>
<td>East Asia</td>
<td>105</td>
<td>1000 genomes project phase 3</td>
</tr>
<tr>
<td>Japanese in Tokyo, Japan</td>
<td>JPT</td>
<td>East Asia</td>
<td>104</td>
<td>1000 genomes project phase 3</td>
</tr>
<tr>
<td>Kinh in Ho Chi Minh City, Vietnam</td>
<td>KHV</td>
<td>East Asia</td>
<td>99</td>
<td>1000 genomes project phase 3</td>
</tr>
<tr>
<td>Utah residents with Northern and Western European ancestry</td>
<td>CEU</td>
<td>Europe</td>
<td>99</td>
<td>1000 genomes project phase 3</td>
</tr>
<tr>
<td>Finnish in Finland</td>
<td>FIN</td>
<td>Europe</td>
<td>99</td>
<td>1000 genomes project phase 3</td>
</tr>
<tr>
<td>British in England and Scotland</td>
<td>GBR</td>
<td>Europe</td>
<td>91</td>
<td>1000 genomes project phase 3</td>
</tr>
<tr>
<td>Iberian populations in Spain</td>
<td>IBS</td>
<td>Europe</td>
<td>107</td>
<td>1000 genomes project phase 3</td>
</tr>
<tr>
<td>Toscani in Italy</td>
<td>TSI</td>
<td>Europe</td>
<td>107</td>
<td>1000 genomes project phase 3</td>
</tr>
<tr>
<td>Gujarati Indian in Houston, TX, USA</td>
<td>GH</td>
<td>South Asia</td>
<td>103</td>
<td>1000 genomes project phase 3</td>
</tr>
<tr>
<td>Indian Telugu in the UK</td>
<td>ITU</td>
<td>South Asia</td>
<td>102</td>
<td>1000 genomes project phase 3</td>
</tr>
<tr>
<td>Punjabi in Lahore, Pakistan</td>
<td>PJI</td>
<td>South Asia</td>
<td>96</td>
<td>1000 genomes project phase 3</td>
</tr>
<tr>
<td>Sri Lankan Tamil in the UK</td>
<td>STU</td>
<td>South Asia</td>
<td>102</td>
<td>1000 genomes project phase 3</td>
</tr>
<tr>
<td>Bengali in Bangladesh</td>
<td>BEB</td>
<td>South Asia</td>
<td>86</td>
<td>1000 genomes project phase 3</td>
</tr>
<tr>
<td>Chinese Kazak</td>
<td>XJK</td>
<td>From China</td>
<td>510</td>
<td>Our lab</td>
</tr>
</tbody>
</table>
Mongolian and the other compared populations were explored using PCA by the MVSP software (version 3.1). The advantage of the PCA is that it allows graphical representation of multidimensional data with reduced number of dimensions [25]. As shown in Figure 2(a), different continental populations formed the corresponding population clusters which were in line with their geographical origins; however, four admixture American populations were distributed among the European, South Asian, and East Asian populations. Furthermore, we also found that the Chinese Mongolian group was adjacent to five East Asian populations, indicating that the Chinese Mongolian group had closer genetic relationships with these East Asian populations than the other reference intercontinental populations. Moreover, PCA of the studied group and other reference populations at individual level was performed using PLINK software (version 1.90). As shown in Figure S2, one point represented a sample, and seven different colors represented five different intercontinental populations and Kazak group as well as the studied Mongolian group. Obtained results revealed that the distributions of some Mongolian individuals were overlapped with the East Asian, Kazak, South Asian, American, and European populations, while African populations separated from them into an independent cluster. The PCA result of individual level was due to the smaller differences in allele frequencies of these 35 InDel loci between the Mongolian group and these four reference continental populations, whereas the larger differences in allele frequencies of these loci between the Mongolian group and African populations in this study.

For further validation, a MDS plot based on pairwise Fst values of these populations was generated as shown in Figure 2(b). Similar population distribution patterns were observed in MDS, implying that the Chinese Mongolian and East Asian as well as Chinese Kazak populations had relatively close genetic ties.
Figure 1: The heat maps of pairwise $D_A$ (a) and pairwise $Fst$ (b) values of the Chinese Mongolian group and the 27 reference populations based on the 35 same InDel loci using the R software.

Figure 2: Continued.
3.5. Phylogenetic Analysis among the Chinese Mongolian Ethnic Group and 27 Reference Populations. The purpose of phylogenetic analysis is to intuitively infer or evaluate the relationships among different populations [26]. Populations with lower genetic distances commonly form a branch on the phylogenetic tree. We constructed a phylogenetic tree of the Chinese Mongolian and other reference populations by MEGA software (version 6.06). As shown in Figure 3, two main branches could be observed: seven African populations formed a branch; East Asian, South Asian, American, European, Chinese Kazak, and the studied Chinese Mongolian groups formed another branch. At the second branch, five European populations clustered together; four American populations gathered as a subbranch; five South Asian populations gathered as another subbranch; the studied Chinese Mongolian group firstly formed the subbranch with five East Asian populations, and then followed by the Chinese Kazak group, revealing that the Chinese Mongolian group had smaller genetic diferentiations with these East Asian and Kazak populations.

3.6. Population Genetic Structure Analysis among the Studied Mongolian Group and 27 Reference Populations. In this study, a population clustering analysis method was used to reflect ancestral proportion memberships of the Chinese Mongolian and 27 compared populations with the number of hypothetical populations (K) which were assumed from 2 to 7 by using the STRUCTURE software (version 2.3.4.) and CLUMPP software (version 1.1.2). Then the appropriate K value was estimated by Structure Harvester (http://taylor0. biology.ucdavis.edu/structureHarvester/), as shown in Figure S3. Results showed that the appropriate K value was 3 for the population data set used in this study according to the appropriate K value standard in a previous report [27]. Clustering analyses of these populations were displayed in Figure 4, and the population names were marked at the top of the figure. When populations are far apart in geographic distances, individuals of these populations generally have different membership coefficients in deductive clustering. Clustering analyses showed that the color compositions of the Chinese Mongolian group were more similar to those of the East Asian populations than those of other intercontinental populations at K = 2 - 7, which further indicated that the population structure of the Chinese Mongolian and East Asian populations was similar.

4. Discussion

In this study, we assessed genetic polymorphisms and forensic efficiencies of 35 InDels in the Chinese Mongolian group. Moreover, genetic relationships between the studied group and the other reference populations were explored based on these 35 InDels. Obtained cumulative PE and combined DP values of 35 InDel loci were 0.99925 and 0.9999999999999904 in the Chinese Mongolian group, indicating that these 35 InDel loci can be used as a valid tool for forensic individual identifications and as an assistant system for paternity testing. The results of PCA, phylogenetic tree, and structure analysis showed that the genetic diferentiations between Chinese Mongolian group and the East Asian populations were smaller than those between the Chinese Mongolian group and the other reference populations.

Genghis Khan established the Mongol Empire in the 13th century, which was a successful nomadic nation. The Mongol Empire’s territorial expansion promoted cultural exchanges between Asia and Europe, which had a
remarkable influence on the genetic structure of the Eurasian people [28]. Xinjiang Mongolian is a subgroup of the Oirats, which is a branch of Mongolian (https://en.wikipedia.org/wiki/Mongols). Wei et al. stated that the Xinjiang Mongolian group had close genetic relationships with Uyghur, Xibe, and other Chinese populations [10], which was consistent with the historical record of Xinjiang Mongolian geographical migration. A previous study showed that the genetic structure of Mongolian was similar to that of CHB, JPT, and other East Asian populations [29]; besides, our result was also consistent with the results of Mei et al. which showed that Kazak group had a closer genetic relationship with the Mongolian group [30].
Our study validated the forensic applicability of these 35 InDel loci in the Xinjiang Mongolian group. In order to further carry out the population study and explain the origin of the Mongolian group, it is necessary to further evaluate the genetic characteristics of the Mongolian group by using other genetic markers like AIM, mitochondrial markers, and so on.

5. Conclusion

This study evaluated forensic efficiencies of a set of novel 35 InDels and assessed the genetic structure of the Chinese Mongolian group based on these InDels. The results of forensic value evaluation indicated that this system of 35 InDels was efficient enough to forensic human identifications in the Mongolian group. And the results of the population genetic analyses indicated that the genetic relationships between the Chinese Mongolian and East Asian populations were relatively close, followed by Kazak group. In a word, these results enrich the Mongolian group data and lay the basis of forensic applications of these 35 InDels in the Mongolian group.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that there are no conflicts of interest.

Acknowledgments

This study was supported by the National Natural Science Foundation of China (81525015) and Guangdong Province Universities and Colleges Pearl River Scholar Funded Scheme (GDUPS, 2017).

Supplementary Materials

Supplementary Figure 1: the geographic localization of the Chinese Mongolian and 27 reference groups. Supplementary Figure 2: principle component analysis of the Chinese Mongolian group and the 27 reference populations at individual levels based on PLINK software (version 1.90) and ggplot2 package (version 3.2.0) of R software (version 3.4.5). The African individuals are represented by blue dots, the American individuals are represented by deep green dots, the East Asian individuals are represented by light green dots, the European individuals are represented by red dots, the Kazak individuals are represented by grey dots, the South Asian individuals are represented by purple dots, and the studied Mongolian individuals are represented by yellow dots. Supplementary Figure 3: the estimated probability value (a) and Delta K (b) at each K value. Supplementary Table 1: the pairwise \( D_A \) values of the Chinese Mongolian group and the 27 reference populations based on a panel of 35 InDels. Supplementary Table 2: the pairwise \( Fst \) values of the Chinese Mongolian group and the 27 reference populations based on a panel of 35 InDels. (Supplementary Materials)

References


