A comprehensive exploration of the genetic legacy and forensic features of Afghanistan and Pakistan Mongolian-descent Hazara

Article *in* Forensic Science International: Genetics · June 2019

DOI: 10.1016/j.fsigen.2019.06.018

CITATIONS
0

READS
253

11 authors, including:

Guanglin He
Sichuan University
73 PUBLICATIONS **210** CITATIONS
SEE PROFILE

Atif Adnan
China Medical University (PRC)
28 PUBLICATIONS **47** CITATIONS
SEE PROFILE

Allah Rakha
University of Health Sciences Lahore
42 PUBLICATIONS **161** CITATIONS
SEE PROFILE

Ivy Hui-Yuan Yeh
Nanyang Technological University
32 PUBLICATIONS **103** CITATIONS
SEE PROFILE

Contents lists available at ScienceDirect

Forensic Science International: Genetics

journal homepage: www.elsevier.com

Short communication

A comprehensive exploration of the genetic legacy and forensic features of Afghanistan and Pakistan Mongolian-descent Hazara

Guanglin He [a, b, ***, 1], Atif Adnan [c, 1], Allah Rakha [d], Hui-Yuan Yeh [e], Mengge Wang [a], Xing Zou [a], Jianxin Guo [b], Muhammad Rehman [f], Abulhasan Fawad [f], Pengyu Chen [g, h, **], Chuan-Chao Wang [b, *]

[a] Institute of Forensic Medicine, West China School of Basic Science and Forensic Medicine, Sichuan University, Chengdu, 610041, China
[b] Department of Anthropology and Ethnology, Institute of Anthropology, Xiamen University, Xiamen, 361005, China
[c] Department of Forensic Genetics and Biology, School of Forensic Medicine, China Medical University, Shenyang, 110122, China
[d] Department of Forensic Sciences, University of Health Sciences Lahore, 54000, Pakistan
[e] School of Humanities, Nanyang Technological University, Nanyang, 639798, Singapore
[f] Forensic Medicine Directorat, Ministry of Public Health, Kabul, Afghanistan
[g] Center of Forensic Expertise, Affiliated Hospital of Zunyi Medical University, Zunyi, 563099, Guizhou, China
[h] School of Forensic Medicine, Zunyi Medical University, Zunyi, 563099, Guizhou, China

ARTICLE INFO

Keywords:
Hazara
Bouyei
Indel
Admixture
Genetic structure
Forensic characteristics

ABSTRACT

Afghanistan and Pakistan are rich with a complex landscape of culture, linguistics, ethnicity and genetic legacy at the crossroads between Indian-Subcontinent and Central Asia. Hazara people have historically been suggested to be Mongolian decedents but seldom been genetically studied. To dissect the genetic structure and explore the forensic characteristics of Hazara people, we first genotyped 30 Insertion/deletion (Indel) markers in 468 samples from 2 aboriginal Hazara populations from Afghanistan and Pakistan, and 100 East Asian comparative Bouyei samples using the Investigator® DIPplex kit. Subsequently, we carried out a comprehensive population genetic analysis from four different datasets: 8895 30-Indel genotype data from 51 populations, 15,895 30-Indel allele frequency data from 98 populations, 1048 genotypes of 993 STRs and Indels from 53 HGDP populations and 2068 whole-genomes (621,799 single nucleotide polymorphisms) from 165 worldwide Human origin reference populations, to further unravel the genetic complexity between Hazara and worldwide human populations using various statistical analysis. We find that 30 Indels are in accordance with HWE, and informative and polymorphic in both Central Asians Hazara and East Asian Bouyei populations. The forensic combined probability of exclusion is larger than 0.9943 and the cumulative power of discrimination is larger than 0.99999999999936. These forensic parameters show the high level of diversity, which makes the Indel amplification system suitable for forensic routine work and may be used as a supplementary assay for routine forensic investigation. The results from pairwise genetic distances, MDS, PCA, and phylogenetic relationship reconstruction demonstrate that present-day Hazaras are genetically closer to the Turkic-speaking populations (Uyghur, Kazakh, and Kyrgyz) residing in northwest China than with other Central/South Asian populations and Mongolian. Outgroup and admixture f_3, f_4, f_4-ratio,

* Corresponding author.
** Corresponding author at: Center of Forensic Expertise, Affiliated Hospital of Zunyi Medical University, Zunyi, 563099, Guizhou, China.
*** Corresponding author at: Institute of Forensic Medicine, West China School of Basic Science and Forensic Medicine, Sichuan University, Chengdu, 610041, China.
Email addresses: Guanglinhescu@163.com (G. He); pychenfs@163.com (P. Chen); wang@xmu.edu.cn (C-C Wang)
[1] The author contributed equally to this work and should be considered co-first author.

https://doi.org/10.1016/j.fsigen.2019.06.018
Received 19 March 2019; Received in revised form 23 June 2019; Accepted 23 June 2019
Available online xxx
1872-4973/

qpWave, and *qpAdm* results further demonstrate that Hazara shares more alleles with East Asians than with other Central Asians and carries 57.8% Mongolian-related ancestry. Overall, our findings suggest that Hazaras have experienced genetic admixture with the local or neighboring populations and formed the current East-West Eurasian admixed genetic profile after their separation from the Mongolians.

1. Introduction

Pakistan and Afghanistan are two linguistically and ethnically diverse countries located at the crossroad of South Asia, Central Asia and East Asia (Figure S1). Pakistan has a population size over 212 million and is the world's sixth-most populous country, which is dominated by at least eighteen ethnic groups. The ethnic landscape here includes six major ethnic groups (including Sindhis, Punjabis, Pathans, Muhajirs, Baloch and Kashmiris) and exceeds twelve minor ethnic groups (Brahuis, Saraikis, Hazara, Burusho and others). Afghanistan, one landlocked country, is located at the northern and western boundary of Pakistan with a population size exceeding 31 million. Ethnic groups in Afghanistan mainly include Pashtun, Tajik, Hazara, Uzbek, Aimak, Turkmen, Baloch, and others.

Too many demic diffusion events existing in Pakistan and Afghanistan have shaped their population dynamics and genetic legacy [1–3]. Archaeological and archaeogenetic evidence indicated that anatomically modern human habitation in here dates back to the middle Paleolithic Era (at least 50,000 years ago). Subsequent agriculture expansion from Anatolia also played an important influence on the formation of the genetic pool, due to massive population admixture and turnover happening among the hunter-gatherers, pastoralists, and farmers in the Neolithic times. Recent historic or prehistoric events, such as the Indus Valley Civilization, Origins and Prosperity of Zoroastrianism, Alexander the Great of Macedon's invasion, Mauryas Empire, Muslim Arabs Conquests, Mongol Conquest, Silk Road Trade, First Anglo-Afghan War and Soviet-Afghan War, also have infused with new genetic materials into Indian Subcontinent and promoted genetic admixture and assimilation. Population genetic and demographical histories of ethnic groups in these areas are interesting and important in the historical investigation, molecular anthropology, medical genetics, and forensic sciences. Enigmatic Kalash people have no ancestry inherited from their claimed Greek ancestors and are now considered as an isolated population after they experienced a strong genetic bottleneck, some extent ancient genetic drift and divergence and specific natural selection [3]. Pakistan Parsi genetically derived from the ancient Iranian and subsequently assimilated by the Indian subcontinent's population [4]. There are other interesting genetic admixture events, such as Pakistan Makranis derived from the African Diaspora via Indian Ocean Slave Trade [2] and precise genetic legacy of Zoroastrians [1].

New genetic legacy introduced by the Mongol Conquest is another interesting part in the gene pool of Indian-Subcontinent and Eurasia. The present-day Hazara populations are historically recovered as deriving from the historical expansion of Mongolians during the Siege of Bamyan (1221), who was thought as the direct descendants of the army of Genghis Khan's Mongol Empire (1162–1227). There are approximately 7~8 million Hazara people residing in Afghanistan, Pakistan, Iran, European, Australia, Canada, and Indonesia. Hazara is the third-largest ethnic group in Afghanistan with a population size of over 2.84 million and also a large minority group with over 0.65 million people in Pakistan. Morphological similarities and differences in the facial bone structures and other physical attributes found by physical anthropologists combined with their language and culture resemblance with Central Asian Turkic speakers and East Asian Mongolian speakers suggest that Hazara may have genetic admixture from the inner Asian Turkic and Mongol people. The expansion of Mongolia Empire played an important role in the conformation of the profound legacy on the linguistic, social, cultural and genetic diversity of ethnic groups residing in Eurasia. Uniparental mitochondrial and Y-chromosomal evidence also supported Hazara's genetic affiliation with Mongolian populations [5–7]. The genetic materials of Mongolian descendants have been extensively investigated with the use of uniparental genetic variations, however, much less is known about the autosomal insertion/deletion (Indel) markers in the Hazara population. Besides, the population structure, forensic characteristics, genetic admixture history of the Hazara population remain largely uncharacterized. Here, we also included the East Asian Bouyei as the comparative studied populations. Bouyei is 11th largest populations among 56 officially recognized ethnic groups. The language they used is belong to the Tai-Kadai language family. There are approximately 2.5 million Bouyei people living in the South China. However, the population genetic and forensic characteristics of this population are still in its infancy.

In forensic paternity testing and individual identification, short tandem repeat (STR) polymorphisms have been regarded as the gold standard for decades. However, long amplified fragments (ranging from 100 to 500 base pairs) and stutter peaks have restricted its usefulness in the forensic degraded and mixture cases due to unexplainable and confusing genotyping results. Besides, the high mutation rate of STR (10^{-3}~10^{-4} per generation), which derived by the short-term evolutionary processes, has prevented it from being the widely-using tool in the forensic family investigation and evolutionary genetic studies [8]. Lower-mutated single nucleotide polymorphism (SNP) seems to become the new favorite in forensic DNA labs, but the genotyping technique of Snapshot mini-sequencing is time-consuming and laborious [9]. Recently, forensic genetic researchers have paid considerable attention to the Insertion/Deletion polymorphisms. Indel markers harbor the advantageous features from both SNP and STR (no stutter, lower mutation rate, and shorter amplicon size) [10,11]. To promote the popularity of Indel marker in the forensic applications and provide more polymorphic and informative discrimination system, Investigator® DIPplex kit consisting of 30 autosomal Indels [12] and AGCU InDel 50 kit [13] were subsequently developed and validated. However, the forensic features and corresponding reference database have only been investigated in the European, American, African and East Asian, the forensic allele frequency distribution and forensic statistical parameters

of Indel markers in South Asian or Central Asian populations remain uncharacterized [10–12,14–48].

Nevertheless, the uniparentally inherited mitochondrial and Y-chromosomal variations just provide a sex-biased and partial view of Hazara population history. No genetic studies have appropriately conducted to investigate the autosomal ancestry sources of Mongolian-descent admixture in Hazara population. Thus, the aim of this study is to strengthen our knowledge of the genetic structure, geographic origins, forensic features of Central Asian Mongolian-descent Hazara population using various population genetic statistical analyses. We provide the first batch of population data of Indel markers in 468 Afghanistan and Pakistan Hazara individuals and 100 uninvestigated Chinese Tai-Kadai-speaking Bouyei people using the Investigator® DIPplex kit. Subsequently, comprehensive population genetic studies from four different datasets: 30-Indel genotype data of 8895 individuals from 51 populations, 30-Indel allele frequency data of 15,895 individuals from 98 populations, and 1048 genotypes of 993 STRs, Indels from 53 Human Genome Diversity Project (HGDP) populations and 2068 whole-genomes (621,799 single nucleotide polymorphisms, SNPs) from 165 worldwide Human origin reference populations, were performed to further unravel the genetic complexity between Hazara and worldwide human populations using varies of statistical analyses.

2. Materials and methods

2.1. Sample collection and DNA preparation

A total of 468 unrelated Hazara individuals and 100 Bouyei individuals were collected with the informed written consent (Figure S1 and Fig. 1A). Hazara samples included 221 male individuals residing in Bamyan province, one of the dominant Hazara populating regions in Afghanistan, and 247 unrelated Hazara individuals (125males and 122 females) living in the Hazara Town in Quetta located in Balochistan in Pakistan. Bouyei individuals comprised 53 males and 47 females residing in the Qiannan Bouyei and Miao Autonomous Region in Guizhou Province, Southwest China. This study was considered and endorsed by the institutional review boards of Zunyi Medical University and Xiamen University. Human genomic DNA from Pakistan samples was isolated using the ReliaPrep™ Blood gDNA Miniprep System (Promega, Madison, USA) and blood stain samples from Afghanistan and Guizhou Province were extracted using the PureLink Ge-

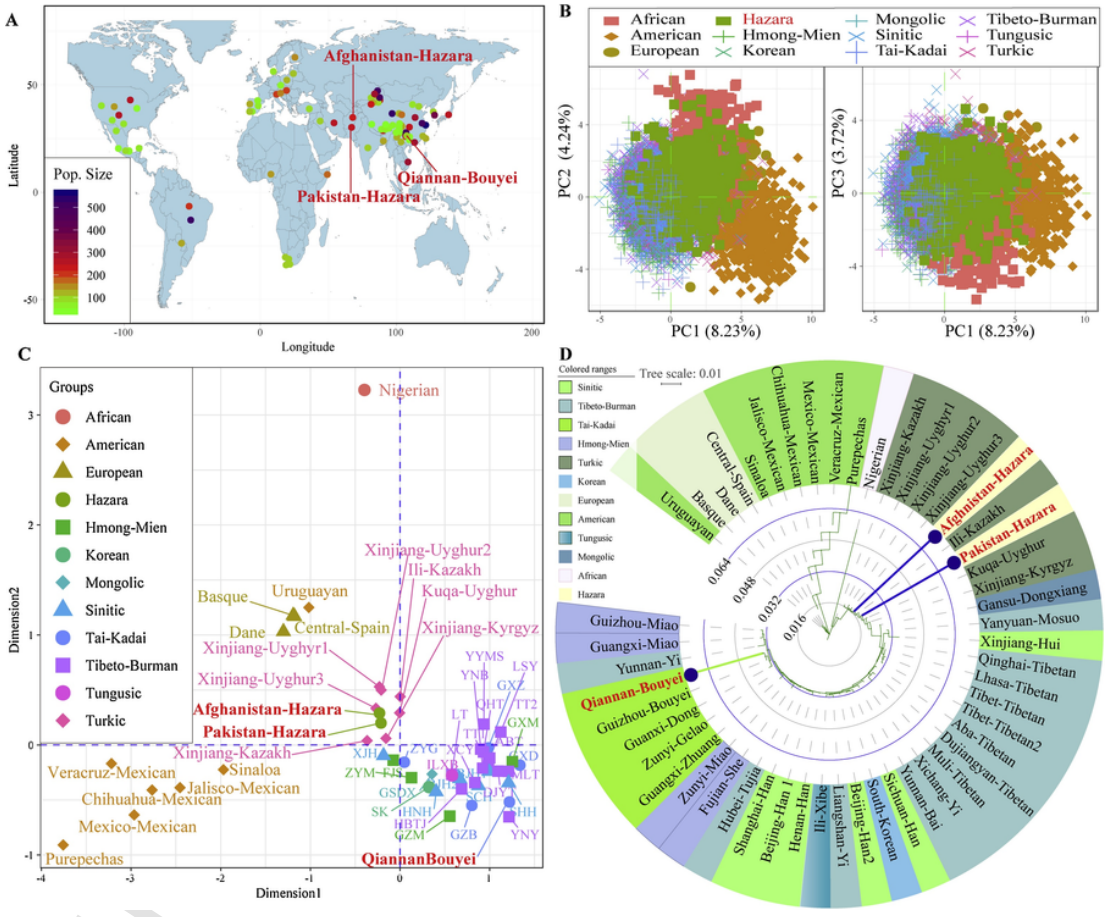

Fig. 1. Genetic relationships among 8895 unrelated individuals from 51 worldwide human populations based on raw genotype data of 30 Indels included in the Investigator® DIPplex kit. (A). Geographic positions and population size of Qiannan Bouyei Pakistan and Afghanistan Hazara populations and all reference populations used in the Indel-based population comparisons. (B). Principal component analysis among 8895 individuals from four continental regions. (C). Multidimensional scaling plots display the genetic affinity between three investigated populations and other 48 reference populations. (D). The neighbor-joining tree showing the phylogenetic relationships between Hazara, Bouyei populations and reference populations on the basis of the pairwise Fst standard genetic distance.

nomic DNA Mini Kit (Thermo Fisher Scientific, Wilmington, DE, USA). We utilized the Nanodrop 2000 (Thermo Fisher Scientific) to measure the DNA concentration based on the manufacturer's instructions.

2.2. PCR amplification and capillary electrophoresis

30 autosomal Indels and one sex-determinate gene were simultaneously amplified using a modified 10ul reaction on the ProFlex 96-Well PCR System (Thermo Fisher Scientific) according to the instruction of the Investigator® DIPplex kit. We used 0.24 μl multi Taq2 DNA polymerase, 1.0 μl template DNA, 2 μl primer mix and 2 μl reaction mix, and appropriated ddH_2O in each PCR reaction. We employed the following thermal cycling parameters: initial enzyme activation for 4 min at 94 °C, targeted fragments amplification for 30 cycles of at 94 °C for 30 s, 61 °C for 120 s and 72 °C for 75 s, and followed by the final extension at 68 °C for 10 min. For PCR amplified products isolation and detection, we used the capillary electrophoresis implemented in the ABI 3130 Genetic Analyzer (Applied Biosystems, Foster City, CA, USA) with DNA size standard 550 (BTO) (Qiagen, Germany) and HiDi Formamide (Applied Biosystems). Allele nomenclature was carried out using the GeneMapper v3.2 software (Applied Biosystems) on the basis of instructions from this software and recommendations of this amplification kit. We used the ddH_2O as the negative control sample and 9947A cell line as the positive control sample in each batch of amplification and capillary electrophoresis.

2.3. Reference datasets and data merging

To comprehensively and deeply dissect the genetic structure and reconstruct the phylogenetic relationship between the Hazara and worldwide human populations, we employed four different datasets (namely DatasetI, DatasetII, DatasetIII, and DatasetIV) [10–12,14–49]. We first merged our newly genotyped data with previously published data to form the first set of data. DatasetI was used to perform population genetic analysis based on the raw genotype data of 30 Indels in 8895 individuals from 51 worldwide human populations, including two Hazara groups, one African population, seven American populations, three European populations, 38 East Asian populations from eight language families or groups of Hmong-Mien, Korean, Mongolic, Sinitic, Tai-Kadai, Tibeto-Burman, Tungusic, and Turkic. We subsequently combined our allele frequency distribution of one Bouyei and two Hazara populations with 95 publicly available population data from Africa, South and North America, Europe and Asia (Fig. 1A). This set of data consisting of genetic variations from 15,895 individuals is regarded as datasetII. In addition, raw genotype data of 993 markers (783 STRs and 210 Indels) in 1048 subjects from 53 worldwide populations (including one Pakistan Hazara population) including the Foundation Jean Dausset's Human Genome Diversity Project and Centre d'Etude du Polymorphisme Humain (HGDP-CEPH) were downloaded from the publicly available databases, which is referred as DatasetIII [49]. Finally, 2068 genome-wide SNP genotypes including 621,799 SNPs from 14 Hazara individuals and other 2054 individuals from 165 worldwide populations genotyped using Affymetrix Human Origin array were employed as the DatasetIV.

2.4. Statistical analysis

Statistical parameters of forensic interest, including power of discrimination (PD), power of exclusion (PE), polymorphism information content (PIC), match probability (PM) and typical paternity index (TPI), and allelic frequency of 30 Indels in the Afghanistan and Pakistan Hazara populations were estimated using the online software of STR analysis for Forensics (STRAF) [50]. Estimation of Linkage Disequilibrium (LD) and Hardy-Weinberg Equilibrium (HWE) were carried out using the Arlequin v.3.5 [51]. Genetic diversity indexes, consisting of observed heterozygosity (Ho) and expected heterozygosity (He), of 30 studied loci were also calculated utilizing the Arlequin v.3.5 [51].

We employed two typical pairwise genetic distances (Fst and Nei) to explore the genetic similarities and differences between the targeted populations and other reference populations [52,53]. Pairwise Fst genetic distances between Qiannan Bouyei, Afghanistan and Pakistan Hazara populations and other 48 reference populations in the DatasetI were calculated using the STRAF [50]. The Fst values between Hazara population and other 52 HGDP populations in the DatasetIII were estimated using Arlequin v.3.5 [51]. Pairwise Nei standard genetic distances between three studied populations and other 95 worldwide populations in the DatasetII were assessed using the Phylogeny Inference Packages (PHYLIP) version 3.5 [51]. We first performed a principal component analysis (PCA) among 51 worldwide populations based on the raw genotype using the STRAF [50] and 98 populations based on the allelic frequency distributions using the Multivariate Statistical Package (MVSP) version 3.22 software [54] to explore the population genetic structure and relationship among targeted and reference populations. Then, we used the multidimensional scaling plots (MDS) instrumented in the IBM SPSS Statistics 21 [55] and applied the neighbor-joining algorithm instrumented in the Molecular Evolutionary Genetics Analysis Version 7.0 (Mega 7.0) [56] to further discover the patterns of genetic affinity and reconstruct the phylogenetic relationships. We followingly estimated the individual ancestry components using the Structure version 2.3.4.21 based on the genotype data under the 'LOCPRIOR' and 'correlated allele frequencies' models [57]. Finally, we used three-population-test of *admixture-f_3 (A, B; Hazara)* to explore the admixture source populations of Hazara and *outgroup-f_3 (Hazara, X; Yoruba)* to explore the genetic affinity between Hazara and other reference populations from the DatasetIV. A, B and X represent the Human Origin reference populations. Finally, we used the *qpWave* to validate the potential ancestry populations of Hazara, and used *f_4-ratio* and *qpAdm* to estimate the admixture proportion [58].

3. Results

3.1. Genetic diversity, population genetic features and forensic characterization based on the 30 Indels

We first estimated the status of HWE and LD of 30 included Indel markers in Qiannan Bouyei, Afghanistan and

Pakistan Hazara populations. As shown in Tables S1-6, no deviation from HWE is observed in three studied populations after the Bonferroni Correction ($p > 0.0167$). Departure from LD is identified in the only pair of HLD77 and HLD93 ($p = 0.0000$) in Pakistan Hazara population when the Bonferroni correction is applied ($p > 0.0001$). Allelic frequency and forensic parameters of 30 Indel markers are presented in Figure S2A and Tables S1-3. For insertion allele, the corresponding allele frequency ranges from 0.2753 at HLD39 locus to 0.7443 at HLD64 locus in Hazara and 0.0950 to 0.9450 in Bouyei group. All studied loci are highly heterozygous in Hazara population with the average Ho of 0.4680 in Pakistan Hazara, 0.4671 in Afghanistan Hazara and 0.408 in Qiannan Bouyei. The individual values of Ho and He respectively vary from 0.3846 to 0.5425 and 0.3815 to 0.5011 in Hazara, and 0.1100 to 0.6200 and 0.1045 to 0.5025 in Bouyei, respectively. The indexes of PM, PD and PE span from 0.3397 to 0.8042, 0.1958 to 0.6603, 0.0100 to 0.3156, respectively. PIC and TPI values span from 0.0985 to 0.3750, 0.5618 to 1.3158, respectively. The combined probabilities of Discrimination (CPD) and cumulative powers of exclusion (CPE) are 0.99999999999936 and 0.9943 in Pakistan Hazara, 0.99999999999937 and 0.99514 in Afghanistan Hazara, and 0.999999999913 and 0.9907 in Qiannan Bouyei. We also compared the deletion allele frequency difference among 98 worldwide human populations (Figure S2B). Allele frequency distribution in the Hazara population is close to the patterns of previously investigated Turkic-speaking populations. And the patterns of allele frequency divergence of Qiannan Bouyei are consistent with geographically close populations, such as Guangxi Zhuang and Guizhou Bouyei. Overall, analysis results from allele frequency and forensic statistical parameters demonstrate that all investigated 30-Indel markers are more informative and polymorphic in both Afghanistan and Pakistan Hazara populations than East Asian Bouyei, suggesting the 30-Indel commercial amplification system is suitable for using as a powerful supplementary tool in forensic paternity identification and individual discrimination in Hazara population and East Asians.

3.2. Geographic affinity of Afghanistan and Pakistan Hazara and Qiannan Bouyei populations via raw-data of DIPplex system

The genetic distance of F_{st} indexes between two Hazara populations and 48 worldwide populations was calculated and presented in Table S7 and visualized as the heatmap in Figure S2C. Small intra-population differentiation among two Hazara populations is observed ($F_{st} = 0.0018$). For inter-population differentiation, the smallest genetic distance with Pakistan Hazara is to be noted at Xinjiang Kazakh ($F_{st} = 0.0026$), followed by Ili Kazakh ($F_{st} = 0.0027$), Xinjiang Kyrgyz (Fst = 0.0036) and four Xinjiang Uyghur populations. Significant genetic differences are identified between Pakistan Hazara and American or African populations. Due to strong genetic affinity identified between Pakistan Hazara and Afghanistan Hazara, we observed similar patterns of genetic similarity between Afghanistan Hazara and other reference populations: genetic affinity exists between Afghanistan Hazara and Turkic-speaking populations (Uyghur, Kyrgyz, and Kazakh) and genetic divergence with American and African populations. Qiannan Bouyei has the closest genetic relationship with Guizhou Bouyei (0.0012) and Miao (0.0027). Top three components of individual-level PCA explain 16.19% variances from the total variations (Fig. 1B), Hazara populations are placed closer to East Asian populations and located approximately intermediate position among Asian, African and American populations in the two-dimensional PC plots. Qiannan Bouyei is generally overlapped with East Asians.

To further have an insight at the intra- and inter-population variation among Hazara and DatasetI reference populations, two-dimensional scaling plots based on the pairwise Fst genetic distance matrix were constructed (Fig. 1C). African Nigerian is isolated with others and situated within the upper position. Five Americans are placed in the left lower position, Europeans with South American Uruguayan are localized in the central position, and East Asians except for Turkic-speaking populations are located at the right lower position. Two Hazara populations are clustered closely and grouped with Kazakh, Uyghur and Kyrgyz populations, which are placed in the intermediate position between European and East Asian. Bouyei is grouped closely with Shanghai Han and Yunnan Yi. Phylogenetic clustered results correspond to the continental geographical origin (Fig. 1D). Pakistan Hazara is first clustered with Kuqa Uyghur and then clustered with Ili Kazakh and Afghanistan Hazara. Qiannan Bouyei was first clustered with Yunnan Yi and then grouped with Chinese Tai-Kadai populations. We subsequently dissected individual ancestry components among 8895 individuals using the model-based algorithm and inferred the detailed ancestry compositions under the predefined ancestry source populations from 2 to 9 ($K = 2^{-}9$). As shown in Fig. 2, with K increasing from 2 to 5, American, African, European, Tibeto-Burman dominant and East Asian dominant ancestries subsequently appear in the individual ancestry vertical line. When the referred ancestry source populations continue to increase to 9, different populations derived their ancestry from at least two predefined ancestry populations. We can still observe the similar proportional ancestry of each individual from the same continental groups or language families. In the optimized five predefined ancestry sources inferred from the Structure Harvester, Afghanistan Hazara is composed of 46.50% ancestry from European and 37.40% from East Asian. The 7.90% Tibeto-Burman dominant ancestry, 7.20% African ancestry and 1.00% American ancestry were also identified in Afghanistan Hazara. Pakistan Hazara harbors 48.40% ancestry from European and 31.80% from East Asian, as well as 10.70% from Tibeto-Burman dominant ancestry, 8.20% from African ancestry and 1.00% American ancestry. Qiannan Bouyei harbors 91.5% ancestry component from East Asian.

3.3. Genetic heterogeneity and phylogenetic relationships between Hazara, Bouyei groups and 95 worldwide populations based on the allele frequency distribution of 30 Indels

In addition to exploring the genetic similarities and differences under the genetic variations of the entire populations being genotyped via the Investigator® DIPplex kit, datasetII consisting of 15,895 individuals from 98 worldwide human populations (Fig. 1A) was employed to conducted the second population genetic analysis using pairwise Nei standard genetic distance, multidirectional scal-

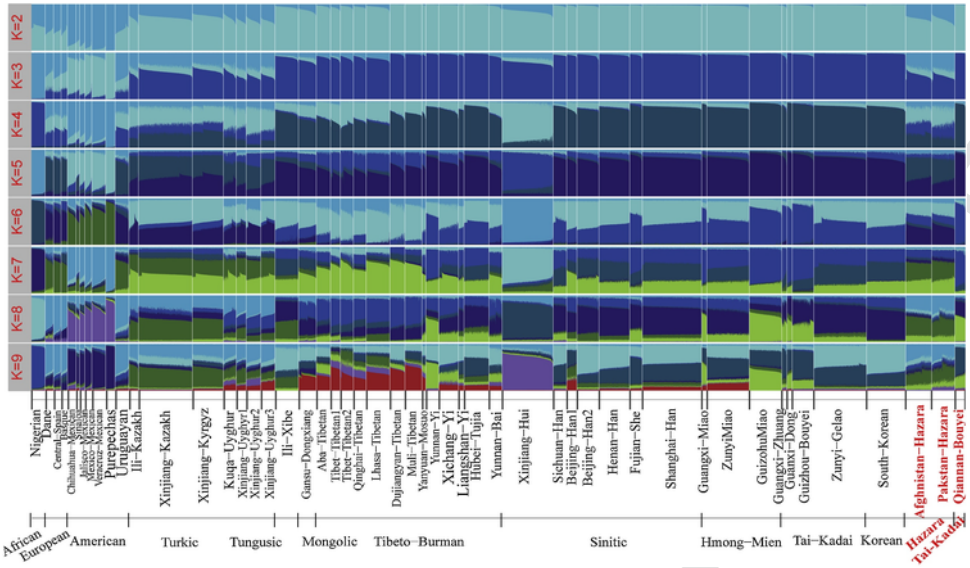

Fig. 2. Individual genetic ancestry components inferred from the Structure results among 8895 individuals from 51 populations based on the genetic variations of 30 Indels with a predefined population ranging from 2 to 9 (k = 2~9). The optimized k value is 5 using the Structure Harvester.

ing plots, population-level principal components analysis and phylogenetic relationship reconstruction. Genetic distance measures of Nei between the three investigated groups and other 95 reference populations were assessed and listed in Fig. 3 and Table S8. Strong genetic affinity is observed among two Hazaras (Nei = 0.0032). Hazara populations show the closest and large similar genetic distances with Kazakh populations, followed by the Uyghur, Kyrgyz, Dongxiang, and Hui residing in the northwestern region of China (Fig. 3A–B). Tai-Kadai-speaking Bouyei is genetically closest to the geographically close groups, subsequently followed by Tibeto-Burman and Turkic-speaking populations (Fig. 3C). Patterns of genetic similarity inferred from heatmap of Nei genetic distance are consistent with the results revealed by the pairwise Fst fixation index (Figure S3). Population-level PCA was first carried out based on the allelic frequency correlation (Figure S4) to assess the genetic relationship. Top five components have extracted

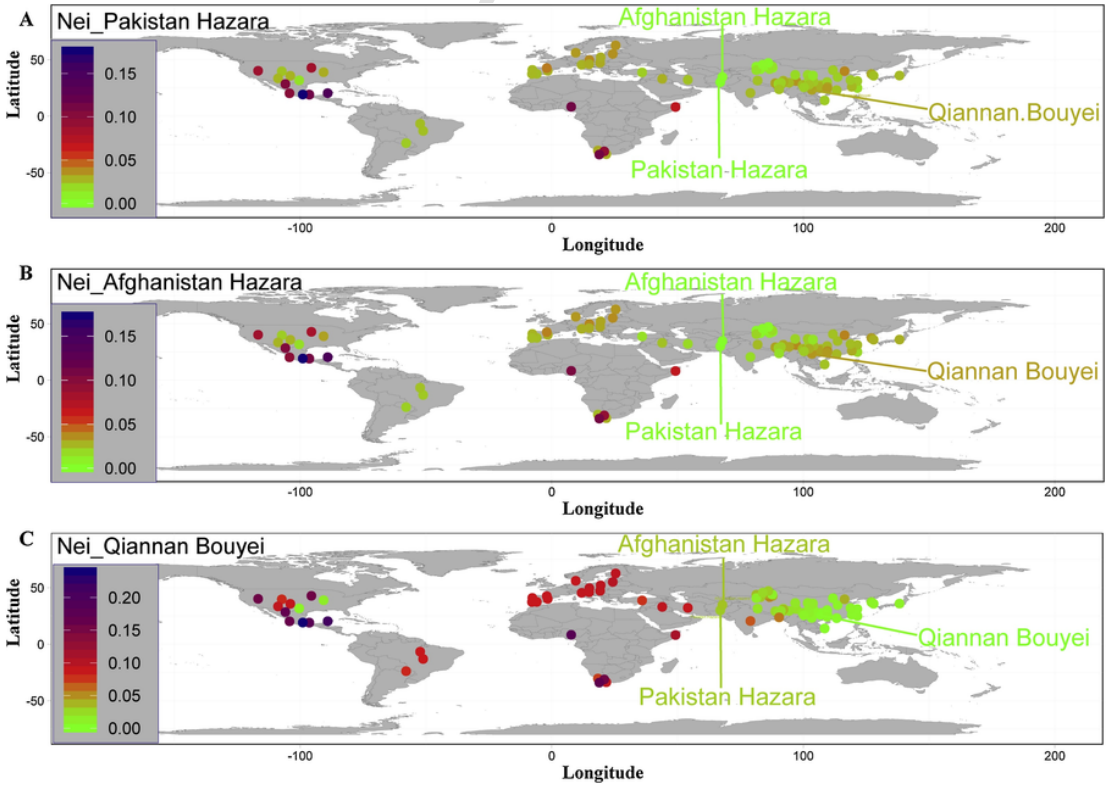

Fig. 3. The pairwise Nei's genetic distances between Pakistan Hazara (A), Afghanistan Hazara (B), Qiannan Bouyei (C) and worldwide reference populations calculated based on 30 Indel variants.

87.714% variations from the total variance. The first principal component explains 55.357% (62.400% of the top five) separating the American and East Asian populations from other groups. The second principal component accounts for 18.948% variance (21.359% of the top five) distinguishing African and American populations from others. East Asian populations from Hmong-Mien, Tai-Kadai, Japanese, Korean, Mongolic, Sinitic, Tungusic, and Tibeto-Burman-speaking language groups are clustered together closely and localized at the leftmost end of the x-axis. European populations are grouped together with the South and West Asian Indo-European speakers, and three South American populations are located in the intermediate position between North American and African populations. It is interesting to find that two Hazara groups are clustered closely with Turkic-speaking populations from northwest China rather than local or neighboring Indo-European speakers, indicating the similar genetic profile between Hazara, Uyghur, Kazakh, and Kyrgyz due to the genetic assimilation in the past two hundred years. Qiannan Bouyei is close to Guizhou Miao (Figure S4). To further illuminate the patterns of genetic relationship between Hazara, Bouyei, and others based on the genetic distance variance, we constructed a two-dimensional scaling plot (Figure S4E). Dimension1 takes into account the East Asian, Central Asian, European and North American genetic differentiation cline. Dimension2 reflects the African, Eurasian and North American genetic differentiation cline. Phylogenetic relationship reconstruction results are presented in Fig. 4. Five apparent clades are identified: North American, European, African, Turkic-speaking and Tibeto-Burman-speaking populations. Two admixed clades are simultaneously detected: one consisting of South/West Asians and South American, and the other composing of East Asian populations with the exception of Turkic and Tibeto-Burman speakers. Our results reveal the strong correlation between genetic affinity and geography and linguistic affiliation.

3.4. Genetic affinity of Hazara under the genetic background of HGDP-CEPH population genetic variations of 993 STRs/Indels

Genetic relationships between Hazara, Uyghur, and Mongolian in the context of worldwide human populations were finally dissected employing the genotype data of 993 STRs/Indels from 50 populations included in the HEDP-CEPH. Measures of genetic distances are presented in Table S9, which shows the Hazara's closest affinity is to Uyghur (Fst = 0.0033), and the second and third closest affinities are to local Pathan (Fst = 0.0088) and Mongolian (Fst = 0.0099), followed by Central/South Burusho and East

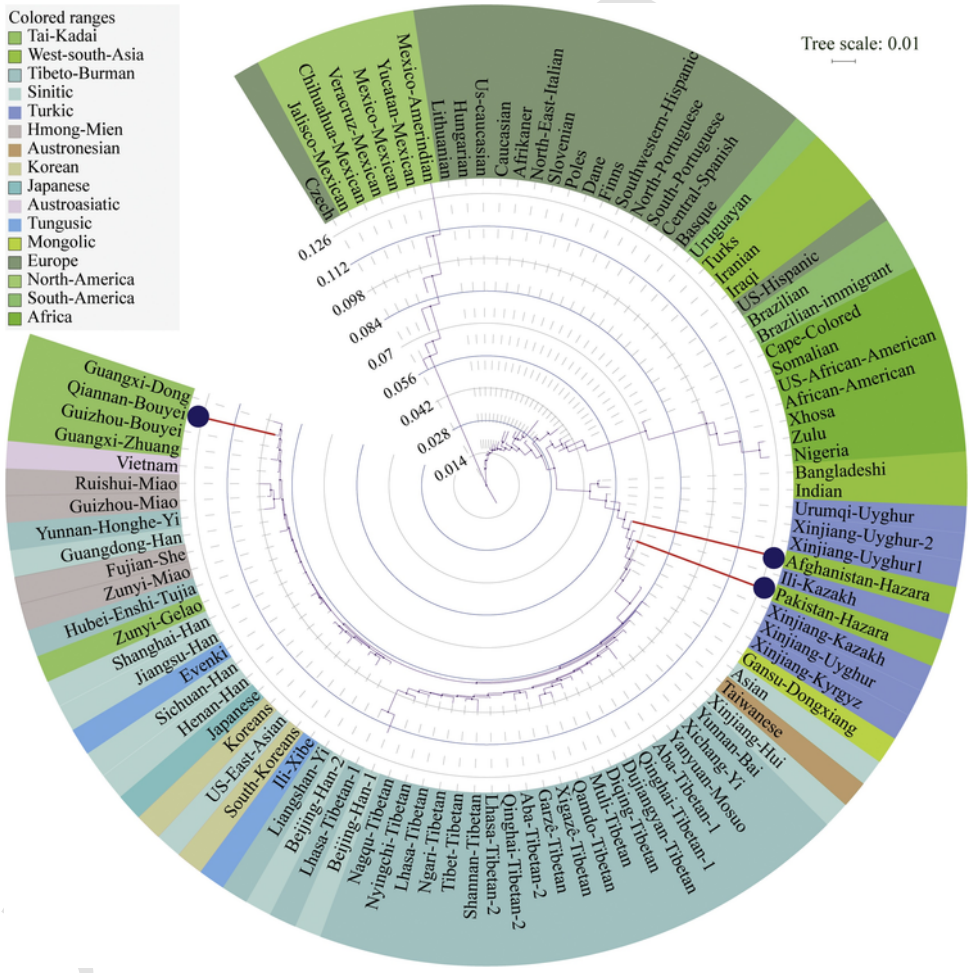

Fig. 4. Phylogenetic relationship shows the genetic similarities and differences between two Hazara populations, one Bouyei and other 95 worldwide human populations from Africa, Europe, West/South Asia, East Asia and America based on the allele frequency correlation of 30 Indels.

Asian Xibe and Daur. The genetic makeup of Hazara is distinct to American Surui (Fst = 0.1621), followed by American Karitiana, Pima and Colombian. Mongolian has the closest genetic relationship with Tujia, followed by other northern East Asian populations of Daur, Hezhen, Xibe, and northern Han. The patterns of genetic similarity and difference between Uyghur and their worldwide reference populations are similar to the findings of the Hazara population. Two-dimensional MDS plots put Hazara and Uyghur in the intermediate position between East Asian and the Middle East or European groups, but closer to the Central/South Asian than to East Asian populations (Fig. 5A). African, American and the isolated Kalash population have a distinct relationship with others, as reported by the recent whole-genome high-density genetic variation study [3]. We then reconstructed a phylogenetic relationship tree based on the neighbor-joining algorithm using the pairwise Fst distance matrix calculated from the genetic variations from 993 polymorphic markers (Fig. 5B). All 53 worldwide populations are generally clustered into six genetic affinity clades: American, East Asian, Oceanian, African, Central/South Asian, and European and the Middle East populations. Hazara is first grouped with Uyghur and then grouped with other East Asian populations, but not clustered with geographical neighboring South Asian populations. To dissect the Hazara ancestry component under the genetic variation in the DatasetIII, we conducted the Structure analysis with predefined ancestry populations from 2 increasing to 13 (Fig. 5C). At K = 2, American and East Asian populations are separated from the other populations by sharing the same color. Unique African, European, East Asian, and Oceanian dominant ancestry com-

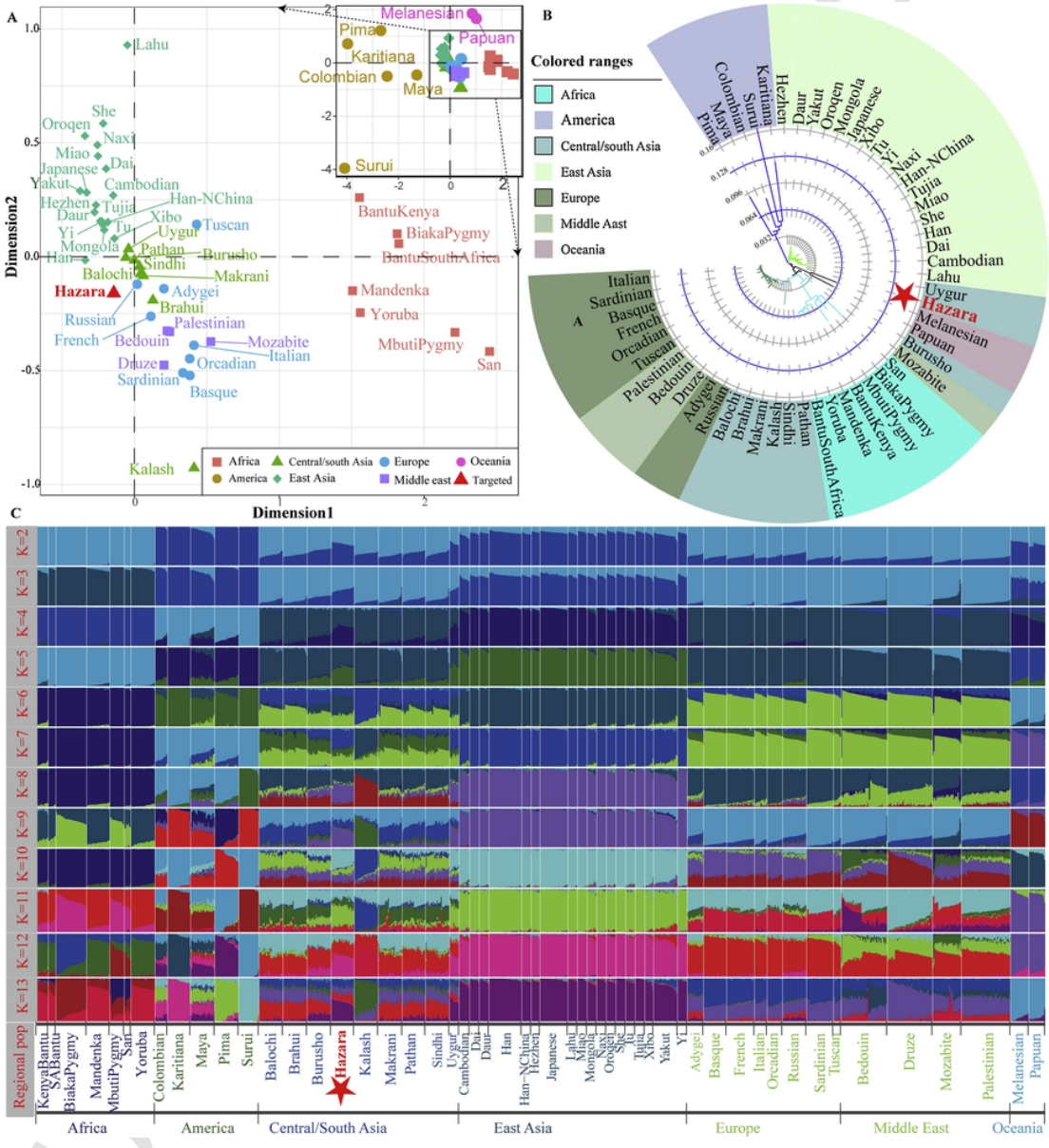

Fig. 5. Genetic heterogeneity and homogeneity between Hazara and other 52 populations from HEPH-HGDP based on the combined genetic variations of 783 microsatellites and 210 insertion/deletion polymorphisms. (A). The genetic affinity among 56 worldwide populations. (B). The phylogenetic relationship constructed using the neighbor-joining algorithm based on the pairwise Fst genetic distance. (C). Genetic ancestry among 1048 individuals from 53 populations inferred from the results of structure using 993 informative and polymorphic markers.

ponents subsequently emerge from K = 2 to K = 5. We estimated the k = 5 as the best K value. We start to observe within-population substructures in ancestry component assigning when the inferred ancestry populations larger than 5 (K > 5). African populations keep homogeneous when K < 8, but population substructures are identified with K values ranging from 9 to 13, such as Pygmy populations (Biaka and Mbuti) showing their unique component at K = 9. Oceanian and East Asian persistently keep homogenous even with larger K values. Central/South Asian and Middle East populations have variable ancestry components mainly deriving from geographical adjacent East Asian and European populations. We observe 48.20% European related ancestry, 48.87% East Asian related ancestry, 0.17% African related ancestry, 0.47% Oceanian related ancestry, and 2.30% American related ancestry in Hazara population. We note that the African and Oceanian related ancestry may not be reliable due to the very low percentages. Similar ancestry composition is observed in the Uyghur population deriving 46.32% ancestry from European and 51.45% from East Asian populations. The Mongolian population derives 87.66% ancestry from East Asians, 9.27% from Europeans and 3.07% from others. We find that Hazara shows a more similar genetic profile with East-West Eurasian admixed Uyghur population than with Mongolians. These observed results suggest that Pakistan Hazara might have had genetic close-fitting contact with European or adjacent admixed populations following their separation from the Mongolians.

3.5. Fine-scale genetic structure and admixture history of Hazara referred from whole-genome SNPs

We finally dissect the fine-scale genetic structure of Hazara using 61,6938 SNPs from 165 worldwide populations (DatasetIV) and provided formal tests for genetic admixture using ADMIXTOOLs. The genetic affinity between Hazara and other references revealed by Outgroup f_3 *(X, Hazara; Yoruba)* demonstrates that Hazara shares more alleles with East Eurasians than West Eurasians (Fig. 6A). Due to the aforementioned correlation between Hazara, Uyghur, and Mongolian, we subsequently estimated the shared genetic drift between Uyghur, Mongolian and others using the f_3 *(Mongolian, X; Yoruba)* and f_3 *(Uyghur, X; Yoruba)*. Different patterns of the shared alleles are observed (Fig. 6B˜C). Significant negative values of the *admixture-*f_3 *(A, B; Hazara)* statistics are observed between the references respectively from European and Asians (Table S10), which further suggests Hazara carries both Asian and European ancestry.

Allele-sharing f_4 *(X, Mongolian; Hazara, Yoruba)* (Fig. 7) were also performed. Significant departures of the positive f_4 values suggest an excess of allele sharing between Hazara and testing population X, and negative values indicate more shared alleles with Mongolian. Our results demonstrate that present-day Hazara shares more genetic components with Mongolians than with other worldwide populations except for few populations in northeast Asia and Siberia, for example, Oroqen, Ulchi, Nganasan, Hezhen, Daur, Korean, and Yakut. We then used f_4-ratio, *qpWave,* and *qpAdm* to estimate the admixture proportions in Hazara. The f_4-ratio statistics in the forms of f_4*(Eskimo, Yoruba; X, Australian)/(Eskimo, Yoruba; Mongolian, Aus-*

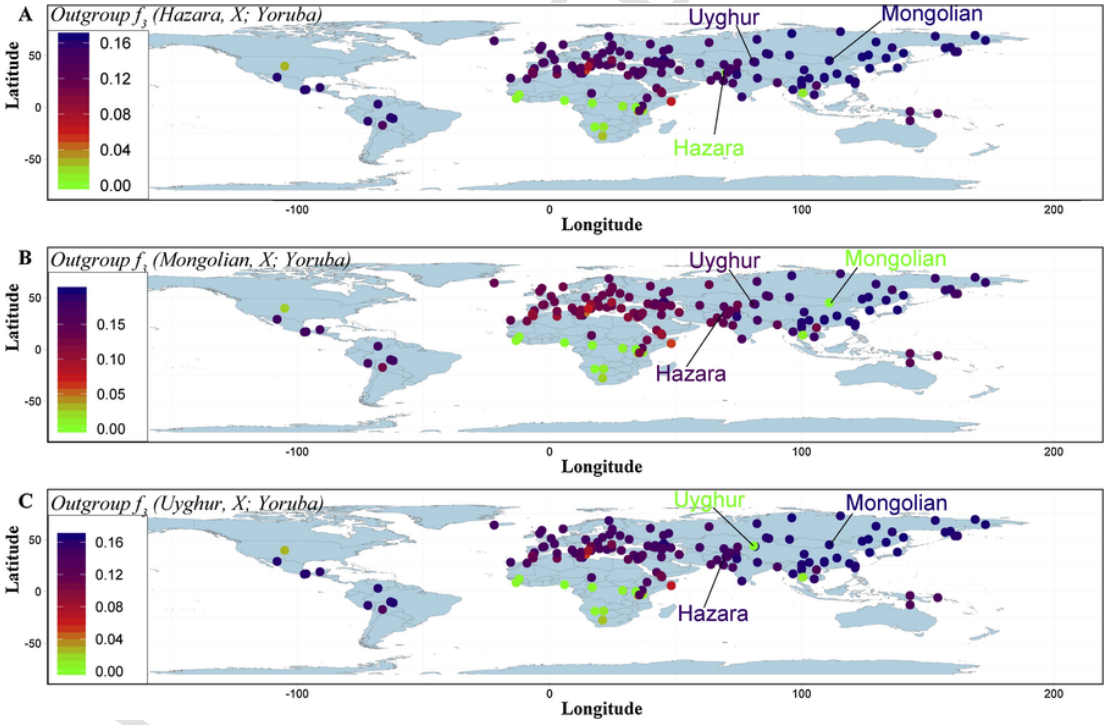

Fig. 6. Outgroup-f_3 results based on the genetic variations of 621,799 single nucleotide polymorphisms. (A). *Outgroup-*f_3 statistics values of form f_3 *(Hazara, X; Yoruba)*. (B) *Outgroup-*f_3 statistics values of form f_3 *(Mongolian, X; Yoruba)*. (C) *Outgroup-*f_3 statistics values of form f_3 *(Uyghur, X; Yoruba)*.

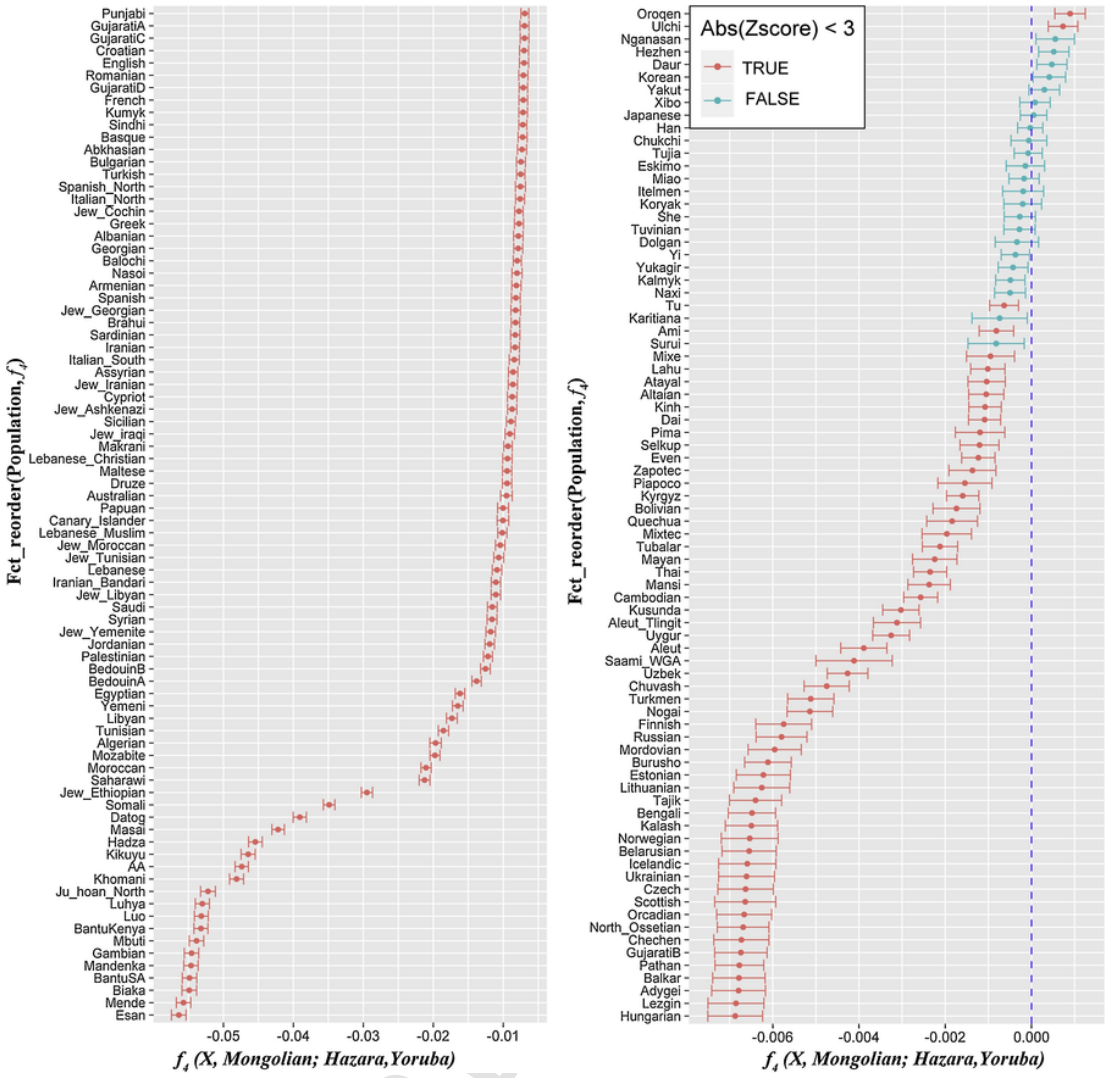

Fig. 7. *D statistics* results of form *D*(X, Mongolian; Hazara, Yoruba) showing the shared genetic shift between Hazara or Mongolian and Testing populations based on 621,799 genetic variants.

tralian) and f_4*(Nganasan, Yoruba; X, Papuan)/(Nganasan, Yoruba; Mongolian, Papuan)* respectively reveal that Hazara harbors 45.48% (Z-score = 20.794) and 46.94% Mongolian-related ancestry (Z-score = 25.371) (Tables S11˜12). We used Uyghur and Hazara as the targeted populations, French and Mongolian as the source groups, and Yoruba, Mbuti, Australian and Papuan as the outgroups to perform the *qpWave* and *qpAdm* to find the lowest number of ancestry streams and explore the proportions of ancestry. Our *qpWave* results suggest two ancestry streams in Hazara (rank = 1, $p = 0.638$) and Uyghur (rank = 1, $p = 0.0311$) regarding to the above four outgroups. We have concluded that Hazara carries 40.2% French-related ancestry and 59.8% Mongolian-related ancestry and Uyghur brings 42.2% French-related ancestry and 57.8% Mongolian-related ancestry.

4. Discussion

Hazara populations residing in Afghanistan and Pakistan are regarded as the descendants of soldiers of the Mongolia Empire according to the historical recordings and Y-chromosomal haplogroup analysis [5,6]. Despite many genetic studies on the population history reconstruction of worldwide ethnolinguistically and geographically distinct populations [1–4,59,60], the genetic history and forensic characteristics of the Hazara population remain unclear. We here report the genotypes of 30 Insertion/deletion (Indel) markers in 468 samples from 2 aboriginal Hazara populations from Afghanistan and Pakistan and 100 samples from East Asians Tai-Kadai-speaking populations. Our study sheds light on the genetic origin, structure, and relationship of Pakistan and Afghanistan Hazara populations and Chinese Bouyei using four comprehensive genetic variation datasets. Hazara populations are placed between East Asian and European clusters in our PCA, MDS, phylogenetic tree, suggesting their admixed genetic profile between Western Eurasian and Eastern Eurasian. We propose that self-declared Mongolian-descendants of Hazara people are an admixed population deriving ancestry from both Western Eurasian and Eastern Eurasian. Genetic ancestry dissection of the Hazara population indicates that they have received more genetic influences from the surrounding populations and experienced the different history of population admixture and assimilation comparing with Mongolians after splitting from the

common ancestor. The supporting patterns observed in our Structure results based on the 30 Indels polymorphisms in 50 worldwide populations and 993 STRs or Indels polymorphisms from 53 HGDP populations can be explained either by the recent admixture from different ancestry sources (such as admixed genetic profile observed in Uyghur and American-Africans) or by the sharing ancestry from their common ancestor before their divergence under different evolutionary forces (likely genetic materials among Russians residing in the Siberia and northeast Asian Yakuts) [49]. Combined with the population migration history, historical recordings and previous uniparentally inherited markers [5,6,61], current genetic landscape of Afghanistan and Pakistan Hazara populations derived from recent admixture between East Asian and European or Central Asian admixed populations. Our subsequent whole-genome genetic evidence further supports the East-West Eurasian genetic admixture in Hazara via three-population and four-population testing.

Our population genetic studies find the stronger genetic affinity between Hazara and Turkic-speaking populations in Central Asia, for instances, Uyghur, Kyrgyz, Kazakh, rather than between Hazara and local Afghanistan and Pakistan populations or present-day Mongolians. The complex historical population migration and admixture events have shaped the current-day interesting genetic landscape in Central and South Asia. For example, the Uyghur is also a typical admixed population harboring the western and eastern anthropometric traits. Xu et al. once found that Xinjiang Uyghur derived about 60% ancestry from West Eurasians and 40% ancestry from East Eurasians based on the entire variations of Chromosome 21. Then they further dissected admixed ancestry proportions based on the high-density whole-genome SNP variations that HGDP northern Uyghur has 53% derived from East Asian and 47% from West-Eurasians, and PanAsian southern Uyghur has 48% East Asian ancestry and 52% West-Eurasians ancestry [62,63]. The fine-scale ancestry makeup revealed the multiple-way contacts occurring in bronze age introducing four ancestry sources in present-day Uyghurs: 15–17% from Siberian, 29–47% from East Asian, 12–20% from South Asian and 25–37% from European [64]. Our study based on the f_4-*ratio* testing supporting the aforementioned ancestry proportion of Uyghur, which carries 41.5% French-related ancestry and 58.5% Mongolian ancestry. The complex genetic profile observed in Hazara people can be explained by their genetic contacts with adjacent neighbors. Hazara genetic characteristics are corresponding well to historical recordings and linguistic affiliation supports their Mongolian origin with a long-term (approximately one millennium) of contact and exchange with Central or South Asian ethnic groups. Recently, one whole-genome sequencing project of Mongolian populations revealed apparent population stratification among geographically or culturally diverse tribes [7]. Thus, more comprehensive population history of genetic dynamics, admixture divergence of Hazara, as well as clearer genetic relationships among Hazara, Uyghur, and Mongolian, are needed to be dissected and reconstructed based on the whole-genome sequencing or high-density genotyping data from more representative modern and ancient samples from geographically/culturally diverse populations.

5. Conclusion

We provide the first batch of crucial Indel resource and forensic reference dataset in Central/South Asian Hazara populations and Chinese Tai-Kadai-speaking Bouyei, which will facilitate the understanding of forensic features and the widely-using of Indel-based amplification system in the Central and South Asians and East Asians. Our findings from the forensic measures indicate that all 30 investigated Indels markers are informative and polymorphic in Pakistan and Afghanistan Hazara populations and Bouyei group, suggesting those markers can be used as a powerful supplementary tool for forensic paternity and personal identification in the Asians. Mongolian-descent Hazara people are an admixed population deriving about half ancestry from East Asians and another half from West Eurasians. Results from the comprehensive population genetic studies via the pairwise genetic distances, MDS, PCA, phylogenetic tree and formal testing for admixture in ADMIXTOOLs demonstrate that Afghanistan and Pakistan Hazara population are genetically closer to the Turkic-speaking Uyghur, Kazakh and Kyrgyz than to their local or adjacent neighbors.

Author contribution

C.W., P.C. and G.H. designed this study,
A.A. collected the samples,
P.C., X.Z., M.W., and J.G. conducted the experiment,
G.H. wrote the manuscript,
G.H., A.A., M.R., and A.F. analyzed the results,
C.W., H.Y., and A.R. modified the manuscript.
All authors reviewed the manuscript.

Compliance with ethical standards

This study is conducted in accordance with the standards of the Declaration of Helsinki and approved by the ethical review board of Sichuan University and Xiamen University. All samples are obtained from participants with informed written consent.

Competing financial interests

None.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (31801040), Nanqiang Outstanding Young Talents Program of Xiamen University (X2123302), and Fundamental Research Funds for the Central Universities (ZK1144).

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:https://doi.org/10.1016/j.fsigen.2019.06.018.

References

[1] S. Lopez, M.G. Thomas, L. van Dorp, N. Ansari-Pour, S. Stewart, A.L. Jones, E. Jelinek, L. Chikhi, T. Parfitt, N. Bradman, M.E. Weale, G. Hellenthal, The genetic legacy of Zoroastrianism in Iran and India: insights into population structure, gene flow, and selection, Am. J. Hum. Genet. 101 (3) (2017) 353–368.

[2] R. Laso-Jadart, C. Harmant, H. Quach, N. Zidane, C. Tyler-Smith, Q. Mehdi, Q. Ayub, L. Quintana-Murci, E. Patin, The genetic legacy of the indian ocean slave trade: recent admixture and post-admixture selection in the makranis of Pakistan, Am. J. Hum. Genet. 101 (6) (2017) 977–984.

[3] Q. Ayub, M. Mezzavilla, L. Pagani, M. Haber, A. Mohyuddin, S. Khaliq, S.Q. Mehdi, C. Tyler-Smith, The Kalash genetic isolate: ancient divergence, drift, and selection, Am. J. Hum. Genet. 96 (5) (2015) 775–783.

[4] G. Chaubey, Q. Ayub, N. Rai, S. Prakash, V. Mushrif-Tripathy, M. Mezzavilla, A.K. Pathak, R. Tamang, S. Firasat, M. Reidla, M. Karmin, D.S. Rani, A.G. Reddy, J. Parik, E. Metspalu, S. Rootsi, K. Dalal, S. Khaliq, S.Q. Mehdi, L. Singh, M. Metspalu, T. Kivisild, C. Tyler-Smith, R. Villems, K. Thangaraj, "Like sugar in milk": reconstructing the genetic history of the Parsi population, Genome Biol. 18 (1) (2017) 110.

[5] M. Haber, D.E. Platt, M. Ashrafian Bonab, S.C. Youhanna, D.F. Soria-Hernanz, B. Martinez-Cruz, B. Douaihy, M. Ghassibe-Sabbagh, H. Rafatpanah, M. Ghanbari, J. Whale, O. Balanovsky, R.S. Wells, D. Comas, C. Tyler-Smith, P.A. Zalloua, C. Genographic, Afghanistan's ethnic groups share a Y-chromosomal heritage structured by historical events, PLoS One 7 (3) (2012), e34288.

[6] L. Quintana-Murci, R. Chaix, R.S. Wells, D.M. Behar, H. Sayar, R. Scozzari, C. Rengo, N. Al-Zahery, O. Semino, A.S. Santachiara-Benerecetti, A. Coppa, Q. Ayub, A. Mohyuddin, C. Tyler-Smith, S. Qasim Mehdi, A. Torroni, K. McElreavey, Where west meets east: the complex mtDNA landscape of the southwest and Central Asian corridor, Am. J. Hum. Genet. 74 (5) (2004) 827–845.

[7] H. Bai, X. Guo, N. Narisu, T. Lan, Q. Wu, Y. Xing, Y. Zhang, S.R. Bond, Z. Pei, Y. Zhang, D. Zhang, J. Jirimutu, D. Zhang, X. Yang, M. Morigenbatu, L. Zhang, B. Ding, B. Guan, J. Cao, H. Lu, Y. Liu, W. Li, N. Dang, M. Jiang, S. Wang, H. Xu, D. Wang, C. Liu, X. Luo, Y. Gao, X. Li, Z. Wu, L. Yang, F. Meng, X. Ning, H. Hashenqimuge, K. Wu, B. Wang, S. Suyalatu, Y. Liu, C. Ye, H. Wu, K. Leppala, L. Li, L. Fang, Y. Chen, W. Xu, T. Li, X. Liu, X. Xu, C.R. Gignoux, H. Yang, L.C. Brody, J. Wang, K. Kristiansen, B. Burenbatu, H. Zhou, Y. Yin, Whole-genome sequencing of 175 Mongolians uncovers population-specific genetic architecture and gene flow throughout North and East Asia, Nat. Genet. (2018).

[8] K. Slooten, F. Ricciardi, Estimation of mutation probabilities for autosomal STR markers, Forensic Sci. Int. Genet. 7 (3) (2013) 337–344.

[9] B. Mehta, R. Daniel, C. Phillips, D. McNevin, Forensically relevant SNaPshot((R)) assays for human DNA SNP analysis: a review, Int. J. Legal Med. 131 (1) (2017) 21–37.

[10] B. Zhu, Q. Lan, Y. Guo, T. Xie, Y. Fang, X. Jin, W. Cui, C. Chen, Y. Zhou, X. Li, Population genetic diversity and clustering analysis for Chinese Dongxiang group with 30 autosomal InDel loci simultaneously analyzed, Front. Genet. 9 (2018) 279.

[11] T. Xie, Y. Guo, L. Chen, Y. Fang, Y. Tai, Y. Zhou, P. Qiu, B. Zhu, A set of autosomal multiple InDel markers for forensic application and population genetic analysis in the Chinese Xinjiang Hui group, Forensic Sci. Int. Genet. 35 (2018) 1–8.

[12] S. Turrina, G. Filippini, D. De Leo, Forensic evaluation of the Investigator DIPplex typing system, Forensic Sci. Int. Genet. Suppl. Ser. 3 (1) (2011) e331–e332.

[13] L. Chen, W. Du, W. Wu, A. Yu, X. Pan, P. Feng, C. Feng, C. Li, L. Xu, C. Liu, C. Liu, Developmental Validation of a Novel Six-Dye Typing System with 47 A-InDels and 2 Y-InDels, Forensic Sci. Int. Genet. (2019).

[14] L. Wang, M. Lv, D. Zaumsegel, L. Zhang, F. Liu, J. Xiang, J. Li, P.M. Schneider, W. Liang, L. Zhang, A comparative study of insertion/deletion polymorphisms applied among Southwest, South and Northwest Chinese populations using Investigator((R)) DIPplex, Forensic Sci. Int. Genet. 21 (2016) 10–14.

[15] B.L. LaRue, J. Ge, J.L. King, B. Budowle, A validation study of the Qiagen Investigator DIPplex(R) kit; an INDEL-based assay for human identification, Int. J. Legal Med. 126 (4) (2012) 533–540.

[16] L. Poulsen, M.S. Farzad, C. Borsting, C. Tomas, V. Pereira, N. Morling, Population and forensic data for three sets of forensic genetic markers in four ethnic groups from Iran: persians, Lurs, Kurds and Azeris, Forensic Sci. Int. Genet. 17 (2015) 43–46.

[17] S. Akhteruzzaman, S.A. Das, I. Hosen, A. Ferdous, Genetic polymorphism of 30 InDel markers for forensic use in Bangladeshi population, Forensic Sci. Int. Genet. Suppl. Ser. 4 (1) (2013) e348–e349.

[18] Y.L. Wei, C.J. Qin, H. Dong, J. Jia, C.X. Li, A validation study of a multiplex INDEL assay for forensic use in four Chinese populations, Forensic Sci. Int. Genet. 9 (2014) e22–5.

[19] T.J.B. Ferreira Palha, E.M. Ribeiro Rodrigues, G.C. Cavalcante, A. Marrero, I.R. de Souza, C.J. Seki Uehara, C.H.A. Silveira da Motta, D. Koshikene, D.A. da Silva, E.F. de Carvalho, G. Chemale, J.M. Freitas, L. Alexandre, R.T.F. Paranaiba, M.P. Soler, S. Santos, Population genetic analysis of insertion-deletion polymorphisms in a Brazilian population using the Investigator DIPplex kit, Forensic Sci. Int. Genet. 19 (2015) 10–14.

[20] F. Reis, C. Vieira da Silva, A. Amorim, V. Bogas, T. Ribeiro, M.J. Porto, H. Afonso Costa, Genetic characterization of the Brazilian immigrant population in Lisboa with InDel genetic markers (P), Forensic Sci. Int. Genet. Suppl. Ser. 6 (2017) e327–e328.

[21] P. Martin, O. Garcia, B. Heinrichs, I. Yurrebaso, A. Aguirre, A. Alonso, Population genetic data of 30 autosomal indels in Central Spain and the Basque Country populations, Forensic Sci. Int. Genet. 7 (2) (2013) e27–30.

[22] M. Fondevila, C. Phillips, C. Santos, R. Pereira, L. Gusmao, A. Carracedo, J.M. Butler, M.V. Lareu, P.M. Vallone, Forensic performance of two insertion-deletion marker assays, Int. J. Legal Med. 126 (5) (2012) 725–737.

[23] A. Zidkova, A. Horinek, V. Kebrdlova, M. Korabecna, Application of the new insertion-deletion polymorphism kit for forensic identification and parentage testing on the Czech population, Int. J. Legal Med. 127 (1) (2013) 7–10.

[24] S.L. Friis, C. Borsting, E. Rockenbauer, L. Poulsen, S.F. Fredslund, C. Tomas, N. Morling, Typing of 30 insertion/deletions in Danes using the first commercial indel kit--Mentype(R) DIPplex, Forensic Sci. Int. Genet. 6 (2) (2012) e72–4.

[25] A.M. Neuvonen, J.U. Palo, M. Hedman, A. Sajantila, Discrimination power of investigator DIPplex loci in Finnish and Somali populations, Forensic Sci. Int. Genet. 6 (4) (2012) e99–102.

[26] L. Hong, X.G. Wang, S.J. Liu, Y.M. Zhang, O.U. Xue-Ling, Y. Chen, W.H. Chen, Hong-yu Sun, Genetic Polymorphisms of 30 Indel Loci in Guangdong Han Population, J. Sun Yat-sen Univ. (Med. Sci.) 34 (2) (2013) 299–304.

[27] C. Shen, B. Zhu, T. Yao, Z. Li, Y. Zhang, J. Yan, B. Wang, X. Bie, F. Tai, A 30-InDel assay for genetic variation and population structure analysis of Chinese Tujia Group, Sci. Rep. 6 (2016) 36842.

[28] Z. Kis, A. Zalan, A. Volgyi, Z. Kozma, L. Domjan, H. Pamjav, Genome deletion and insertion polymorphisms (DIPs) in the Hungarian population, Forensic Sci. Int. Genet. 6 (5) (2012) e125–6.

[29] C. Tomas, L. Poulsen, K. Drobnic, V. Ivanova, J. Jankauskiene, D. Bunokiene, C. Borsting, N. Morling, Thirty autosomal insertion-deletion polymorphisms analyzed using the Investigator((R)) DIPplex Kit in populations from Iraq, Lithuania, Slovenia, and Turkey, Forensic Sci. Int. Genet. 25 (2016) 142–144.

[30] M. Nunotani, T. Shiozaki, N. Sato, S. Kamei, K. Takatsu, T. Hayashi, M. Ota, H. Asamura, Analysis of 30 insertion-deletion polymorphisms in the Japanese population using the Investigator DIPplex(R) kit, Leg. Med. (Tokyo) 17 (6) (2015) 467–470.

[31] Z. Li, J. Xu, P. Chen, C. Yin, L. Hu, H. Huang, R. Li, D. Jia, F. Chen, Forensic efficiency and genetic divergence of 30 autosomal InDels in Chinese Han population from Jiangsu province, Forensic Sci. Int. Genet. 37 (2018) e17–e19.

[32] E.H. Kim, H.Y. Lee, I.S. Yang, W.I. Yang, K.J. Shin, Population data for 30 insertion-deletion markers in a Korean population, Int. J. Legal Med. 128 (1) (2014) 51–52.

[33] A. Carvalho, M.F. Pinheiro, Population data of 30 insertion/delection polymorphisms from a sample

taken in the North of Portugal, Int. J. Legal Med. 127 (1) (2013) 65–67.

[34] W. Pepinski, M. Abreu-Glowacka, M. Koralewska-Kordel, E. Michalak, K. Kordel, A. Niemcunowicz-Janica, M. Szeremeta, M. Konarzewska, Population genetics of 30 INDELs in populations of Poland and Taiwan, Mol. Biol. Rep. 40 (7) (2013) 4333–4338.

[35] H. Jian, L. Wang, H. Wang, X. Bai, M. Lv, W. Liang, Population genetic analysis of 30 insertion-deletion (INDEL) loci in a Qinghai Tibetan group using the Investigator DIPplex Kit, Int. J. Legal Med. (2018).

[36] Z. Wang, S. Zhang, S. Zhao, Z. Hu, K. Sun, C. Li, Population genetics of 30 insertion-deletion polymorphisms in two Chinese populations using Qiagen Investigator(R) DIPplex kit, Forensic Sci. Int. Genet. 11 (2014) e12–4.

[37] G. Martinez-Cortes, M. Garcia-Aceves, A.F. Favela-Mendoza, J.F. Munoz-Valle, J.S. Velarde-Felix, H. Rangel-Villalobos, Forensic parameters of the Investigator DIPplex kit (Qiagen) in six Mexican populations, Int. J. Legal Med. 130 (3) (2016) 683–685.

[38] K.M. Seong, J.H. Park, Y.S. Hyun, P.W. Kang, D.H. Choi, M.S. Han, K.W. Park, K.W. Chung, Population genetics of insertion-deletion polymorphisms in South Koreans using Investigator DIPplex kit, Forensic Sci. Int. Genet. 8 (1) (2014) 80–83.

[39] C.V. da Silva, S. Matos, H.A. Costa, P. Morais, R.M. Dos Santos, R. Espinheira, J.C. Santos, A. Amorim, Genetic portrait of south Portugal population with InDel markers, Forensic Sci. Int. Genet. 7 (4) (2013) e101–3.

[40] Y. Guo, C. Shen, H. Meng, Q. Dong, T. Kong, C. Yang, H. Wang, R. Jin, B. Zhu, Population differentiations and phylogenetic analysis of Tibet and Qinghai tibetan groups based on 30 InDel loci, DNA Cell Biol. 35 (12) (2016) 787–794.

[41] M. Saiz, F. Andre, N. Pisano, N. Sandberg, B. Bertoni, S. Pagano, Allelic frequencies and statistical data from 30 INDEL loci in Uruguayan population, Forensic Sci. Int. Genet. 9 (2014) e27–9.

[42] W. Du, Z. Peng, C. Feng, B. Zhu, B. Wang, Y. Wang, C. Liu, L. Chen, Forensic efficiency and genetic variation of 30 InDels in Vietnamese and Nigerian populations, Oncotarget 8 (51) (2017) 88934–88940.

[43] T. Mei, C.M. Shen, Y.S. Liu, H.T. Meng, Y.D. Zhang, Y.X. Guo, Q. Dong, X.X. Wang, J.W. Yan, B.F. Zhu, L.P. Zhang, Population genetic structure analysis and forensic evaluation of Xinjiang Uigur ethnic group on genomic deletion and insertion polymorphisms, Springerplus 5 (1) (2016) 1087.

[44] T. Kong, Y. Chen, Y. Guo, Y. Wei, X. Jin, T. Xie, Y. Mu, Q. Dong, S. Wen, B. Zhou, L. Zhang, C. Shen, B. Zhu, Autosomal InDel polymorphisms for population genetic structure and differentiation analysis of Chinese Kazak ethnic group, Oncotarget 8 (34) (2017) 56651–56658.

[45] Y. Guo, C. Chen, X. Jin, W. Cui, Y. Wei, H. Wang, T. Kong, Y. Mu, B. Zhu, Autosomal DIPs for population genetic structure and differentiation analyses of Chinese Xinjiang Kyrgyz ethnic group, Sci. Rep. 8 (1) (2018) 11054.

[46] H.T. Meng, Y.D. Zhang, C.M. Shen, G.L. Yuan, C.H. Yang, R. Jin, J.W. Yan, H.D. Wang, W.J. Liu, H. Jing, B.F. Zhu, Genetic polymorphism analyses of 30 InDels in Chinese Xibe ethnic group and its population genetic differentiations with other groups, Sci. Rep. 5 (2015) 8260.

[47] Y.D. Zhang, C.M. Shen, R. Jin, Y.N. Li, B. Wang, L.X. Ma, H.T. Meng, J.W. Yan, H. Dan Wang, Z.L. Yang, B.F. Zhu, Forensic evaluation and population genetic study of 30 insertion/deletion polymorphisms in a Chinese Yi group, Electrophoresis 36 (9-10) (2015) 1196–1201.

[48] C.H. Yang, C.Y. Yin, C.M. Shen, Y.X. Guo, Q. Dong, J.W. Yan, H.D. Wang, Y.D. Zhang, H.T. Meng, R. Jin, F. Chen, B.F. Zhu, Genetic variation and forensic efficiency of autosomal insertion/deletion polymorphisms in Chinese Bai ethnic group: phylogenetic analysis to other populations, Oncotarget 8 (24) (2017) 39582–39591.

[49] S. Ramachandran, O. Deshpande, C.C. Roseman, N.A. Rosenberg, M.W. Feldman, L.L. Cavalli-Sforza, Support from the relationship of genetic and geographic distance in human populations for a serial founder effect originating in Africa, Proc. Natl. Acad. Sci. U. S. A. 102 (44) (2005) 15942–15947.

[50] A. Gouy, M. Zieger, STRAF-A convenient online tool for STR data evaluation in forensic genetics, Forensic Sci. Int. Genet. 30 (2017) 148–151.

[51] L. Excoffier, H.E. Lischer, Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows, Mol. Ecol. Resour. 10 (3) (2010) 564–567.

[52] M. Nei, The theory of genetic distance and evolution of human races, Jinrui Idengaku Zasshi 23 (4) (1978) 341–369.

[53] B.S. Weir, C.C. Cockerham, Estimating F-Statistics for the analysis of population structure, Evolution 38 (6) (1984) 1358–1370.

[54] W.L. Kovach, MVSP-A MultiVariate Statistical Package for Windows, ver. 3.1, Kovach Computing Services, Pentraeth, Wales, U.K, 2007.

[55] J. Hansen, Using SPSS for windows and macintosh: analyzing and understanding data, Am. Stat. 59 (1) (2005), 113-113.

[56] S. Kumar, G. Stecher, K. Tamura, MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets, Mol. Biol. Evol. 33 (7) (2016) 1870–1874.

[57] G. Evanno, S. Regnaut, J. Goudet, Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study, Mol. Ecol. 14 (8) (2005) 2611–2620.

[58] N. Patterson, P. Moorjani, Y. Luo, S. Mallick, N. Rohland, Y. Zhan, T. Genschoreck, T. Webster, D. Reich, Ancient admixture in human history, Genetics 192 (3) (2012) 1065–1093.

[59] I. Ullah, J.K. Olofsson, A. Margaryan, M. Ilardo, H. Ahmad, M. Sikora, A.J. Hansen, M. Shahid Nadeem, N. Fazal, M. Ali, A. Buchard, B.E. Hemphill, E. Willerslev, M.E. Allentoft, High Y-chromosomal differentiation among ethnic groups of Dir and Swat districts, Pakistan, Ann. Hum. Genet. 81 (6) (2017) 234–248.

[60] A.K. Pathak, A. Kadian, A. Kushniarevich, F. Montinaro, M. Mondal, L. Ongaro, M. Singh, P. Kumar, N. Rai, J. Parik, E. Metspalu, S. Rootsi, L. Pagani, T. Kivisild, M. Metspalu, G. Chaubey, R. Villems, The genetic ancestry of modern Indus valley populations from Northwest India, Am. J. Hum. Genet. 103 (6) (2018) 918–929.

[61] A. Adnan, A. Rakha, K. Kasim, A. Noor, S. Nazir, S. Hadi, H. Pang, Genetic characterization of Y-chromosomal STRs in Hazara ethnic group of Pakistan and confirmation of DYS448 null allele, Int. J. Legal Med. (2018).

[62] S. Xu, L. Jin, A genome-wide analysis of admixture in Uyghurs and a high-density admixture map for disease-gene discovery, Am. J. Hum. Genet. 83 (3) (2008) 322–336.

[63] S. Xu, W. Huang, J. Qian, L. Jin, Analysis of genomic admixture in Uyghur and its implication in mapping strategy, Am. J. Hum. Genet. 82 (4) (2008) 883–894.

[64] Q. Feng, Y. Lu, X. Ni, K. Yuan, Y. Yang, X. Yang, C. Liu, H. Lou, Z. Ning, Y. Wang, D. Lu, C. Zhang, Y. Zhou, M. Shi, L. Tian, X. Wang, X. Zhang, J. Li, A. Khan, Y. Guan, K. Tang, S. Wang, S. Xu, Genetic history of Xinjiang's Uyghurs suggests bronze age multiple-way contacts in Eurasia, Mol. Biol. Evol. 34 (10) (2017) 2572–2582.