Heliyon 9 (2023) e21229

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

Unveiling the genetic landscape of high-altitude adaptive ethnic groups with polymorphic markers: Implications of comprehensive forensic appraisals and population genetic investigations

Ming Zhao, Wei Cui, Hui Xu, Yunying Zhang, Bofeng Zhu *

Guangzhou Key Laboratory of Forensic Multi-Omics for Precision Identification, School of Forensic Medicine, Southern Medical University, 510515, Guangzhou, Guangdong, China

A R T I C L E I N F O

Keywords:
Deletion/Insertion polymorphism
Qinghai Tibetan
Tibet Tibetan
Genetic structure analysis
Population genetics

A B S T R A C T

Background: Deletion/insertion polymorphisms (DIPs), a novel class of biomarker, have been widely utilized in forensic areas for individual identification, paternity tests, and ancestral origin inference due to its applicability to degraded samples and low mutation rates. Despite the availability of a well-established commercial kit, the Investigator® DIPplex kit (Qiagen), certain loci exhibit limited levels of polymorphisms in East Asian populations, particularly in Chinese populations.

Objective: This dissertation seeks to undertake a comprehensive evaluation about the forensic efficiency of a self-developed multiplex amplification system in high-altitude adaptive ethnic groups of China. Healthy unrelated Tibetan individuals residing in Tibet Autonomous Region and Qinghai Province were genotyped using previously reported 43 deletion/insertion polymorphism loci. Forensic statistical analyses including allele frequencies and forensic parameters were conducted in the two Tibetan groups, and the genetic relatedness of the studied groups with reference populations from the 1000 Genomes Project Phase 3 were investigated.

Results: Forensic statistical results showed that the polymorphism information content values of the 43 deletion/insertion polymorphism loci in the two Tibetan groups exceeded 0.35. Moreover, the combined power of discrimination using the 43 deletion/insertion polymorphism loci was calculated to be 0.9999999999999999984 in the Qinghai Tibetan group and 0.9999999999999999921 in the Tibet Tibetan group. The cumulative power of exclusion using the 43 deletion/insertion polymorphism loci was calculated to be 0.999782512 in the Qinghai Tibetan group and 0.999886205 in the Tibet Tibetan group. Analysis of population genetics demonstrated that the two studied Tibetan groups shared close genetic relationships with East Asia populations.

Conclusion: The set of 43 deletion/insertion polymorphism loci exhibited remarkable forensic efficacy, rendering it a promising tool for forensic practice. Population genetic analyses indicated that the two Tibetan groups had closer genetic affinities to East Asian populations.

* Corresponding author.
E-mail address: zhubofeng7372@126.com (B. Zhu).

https://doi.org/10.1016/j.heliyon.2023.e21229
Received 6 January 2023; Received in revised form 11 September 2023; Accepted 18 October 2023
Available online 25 October 2023

1. Introduction

Qinghai-Tibetan Plateau is located in the southwest of China and is commonly referred to as the 'Roof of the World' or the 'Third Pole of the World'. It has the highest altitude and is also one of the broadest plateaus on earth [1,2]. Tibetans, the high-altitude adaptive group, predominantly live in the Qinghai, Tibet, and Sichuan provinces of China [3]. The distinctive geographical characteristic of the plateau is probably one of the reasons for its unique population features. The Tibetan groups have received considerable attention in recent years. Studies have been widely carried out on Tibetan groups, encompassing genetic variation, population genetics, evolutionary history, phenotypic analysis, and archaeology. Multiple researchers have revealed the genetic variations and genetic relationships among Tibetan groups from diverse regions using various molecular genetic markers, such as autosomal short tandem repeats (STRs) [4], single nucleotide polymorphisms (SNPs) [5], sex chromosome STR loci [6], and autosomal deletion/insertion polymorphism (DIP) loci et al [7,8]. Moreover, these molecular genetic markers have also been gradually applied in forensic practice for individual identification and paternity tests [9]. The STRs and SNPs have been widely utilized in human identification, family investigation, and ancestral origin inference [10,11]. In order to acquire higher accuracy and efficiency in individual identification, the Combined DNA Index System (CODIS) Core Loci Working Group has augmented the number of core STR loci [12, 13]. Nevertheless, the limitations of STRs, such as the issues of detecting complete profiles from the highly degraded samples and the high mutation rates, remain challenging [14,15]. Compared with STRs, SNPs with relatively conservative mutation rates in the human genome, can be amplified with shorter DNA segments. Thus, SNPs to a certain extent facilitate the resolution of the aforementioned problems. Nevertheless, the time-consuming, expensive, and cumbersome experimental steps of the SNP typing analysis methods (SNaPshot, TaqMan, pyrosequencing technology, next-generation sequencing, etc.) may not be conducive to the promotion and application of routine forensic cases [16,17]. Besides, the above-mentioned molecular markers (STRs, SNPs) may not comprehensively elucidate the genetic features of the Tibetan groups. And further exploration is necessary to gain a more comprehensive understanding of the history and population structure and evolution of Tibetan groups in China.

DIPs, a novel class of biomolecular markers, combine the advantages of STRs and SNPs. DIPs also have lower mutation rates, and their shorter amplification products can be genotyped on capillary electrophoresis platform. Thus, DIPs yield tremendous superiority in forensic routine applications [18,19]. Furthermore, DIP loci can also be a valuable tool for ancestry origin inference [20]. The Investigation® DIPplex kit (Qiagen), which encompasses 30 DIP loci, is currently one of the most commonly used commercial DIP multiplex amplification kit designed specifically for forensic degrated biological material [21]. However, this kit may not be well-suited for the East Asia populations, and some loci demonstrated the limited levels of polymorphisms [22–25]. To address this issue, we previously reported a self-developed system that can simultaneously amplify 43 DIP loci. In addition, our system has been validated in the Northwest district Hui group in China (NHW) [26]. Besides, all the amplicons in this panel are less than 200 bp, which is suitable for profiling the degraded DNA sample. Therefore, this study aims to conduct a comprehensive evaluation of the forensic utilities of our self-developed system in two Tibetan groups in order to further ascertain its robustness in forensic DNA analysis. Moreover, we also investigated the population structure of the Tibetan groups and performed relevant population genetic analyses using 43 DIP loci genotype data of the reference populations from the 1000 Genomes Project Phase 3.

2. Material and methods

2.1. The information of the two studied Tibetan groups and reference populations in this study

Blood samples of healthy unrelated individuals in Tibetan groups (QHT, n = 155; TT, n = 86) from China were collected after written informed contents, and their self-declared ancestry information was acquired. In this study, we obtained the approval from the Ethics Committee of Xi'an Jiaotong University Health Science Center (approved number: 2019-1039) and Southern Medical University (2023-KY-097-02).

The 26 reference populations were selected from the 1000 Genomes Project Phase 3 and their genotype data of 43 DIP loci were downloaded on the Ensembl online website (https://www.ensembl.org/index.html) [27]. In addition, the genotype data of 43 DIP loci from the Hui group were also collected [26]. And all data above-mentioned were collated for subsequent population genetic analyses. The reference dataset comprises 3037 individuals from five distinct continental regions, namely Africa, East Asia, America, Europe, and South Asia. There are 661 individuals from seven African populations including LWK, ESN, ACB, GWD, YRI, MSL and ASW; 1123 individuals from six East Asian populations including NWH, CDX, CHB, KHV, JPT and CHS; 503 individuals from five European populations including TSI, IBS, GBR, CEU and FIN; 347 individuals from four American populations including PEL, CLM, MXL and PUR; and 489 individuals from five South Asian populations including STU, ITU, PJL, BEB and GIH. Detailed information (population full names, population sizes, and data sources) for the studied groups and reference populations were collated and listed in Table S1.

2.2. DNA extraction, PCR amplification and DIP genotyping

The amplification system contains 43 loci based on the criteria described by Jin et al. [26]. The Chelex-100 method was utilized for genomic DNA extraction. The PCR reaction was conducted in accordance with the guideline from the previously reported study using a GeneAmp PCR system 9700 thermal cycler (Thermo Fisher Scientific, South San Francisco, CA, US) [26]. The amplification products were recognized and analyzed by ABI 3500xL Genetic Analyzer (Thermo Fisher Scientific, South San Francisco, US). The DIP allele typing was conducted by GeneMapper ID-X software v1.5 (Thermo Fisher Scientific, South San Francisco, US).

2.3. Quality control

Throughout the entire experiment process, we utilized a positive (DNA 9947A and 9948) and a negative control (deionized water) to guarantee the standardization of the experimental procedure and the reliability of the results.

2.4. Statistics analysis

The STR Analysis for Forensics (STRAF) online web app was designed for statistical analyses in forensic genetics (http://cmpg.unibe.ch/shiny/STRAF/). The STRAF was used to conduct the Hardy-Weinberg equilibrium tests (*p*-HWE), and calculate the frequencies of alleles, and relevant forensic statistical parameters of the 43 loci in Tibetan groups [28,29]. The linkage disequilibrium (LD) tests and the pairwise fixation index (F_{ST}) values by the 43 DIP loci were performed on the Genepop version 4.7 [30]. The DISPAN program was utilized to calculate the pairwise *Nei's* genetic distance (D_A) values based on the 43 DIP loci [31]. The phylogenetic relationships between the Tibetan groups and 27 reference populations were investigated, and two phylogenetic trees were constructed. One tree was established based on the pairwise D_A values using the neighbor-joining method by the MEGA 7.0, and another was built based on the pairwise D_A values by FigTree v1.4.3 with the unweighted pair group method with arithmetic mean (UPGMA) method [32]. The heatmap of the insertion allele frequencies for 29 populations was conducted by *R* version 4.0.4. The heatmaps of pairwise F_{ST} values and pairwise D_A values for the two Tibetan groups and 27 worldwide reference populations were shown by TBtools [33]. The multidimensional scaling (MDS) was visualized using the pairwise F_{ST} values by *R* version 4.0.4 [34]. We conducted principal component analysis (PCA) by collating the insertion allele frequencies of 43 DIP loci in 29 populations worldwide (*R* version 4.0.4). Furthermore, the PCA for individuals was conducted using the genotypes of the 43 DIP loci in Tibetan groups and three intercontinental populations (African, East Asian, and European) by *R* version 4.0.4. STRUCTURE version 2.3.4 was used to conduct the population genetic component analyses. The online website STRUCTURE HARVESTER (http://taylor0.biology.ucla.edu/structureHarvester/) was applied to acquire the optimal *K* value for the STRUCTURE results. The STRUCTURE results were visualized using CLUMPP version 1.1.222 [35] and Distruct version 1.1.23 [36].

3. Results

3.1. Evaluations of HWE and LD tests of 43 DIP loci in QHT and TT groups

The *p*-HWE and LD tests were conducted on 43 DIP loci in the two Tibetan groups to evaluate the applicability of each locus in forensic practice. QHT and TT groups exhibited no significant deviations from HWE (Table S2) for 43 DIP loci after the Holm-Bonferroni sequential correction ($p > 0.05/43 = 0.0012$). No significant differences were observed in LD tests (Tables S3 and S4) for pairwise loci in the QHT, TT groups after the Holm-Bonferroni sequential correction ($p > 0.05/903 = 0.000055$). All loci were suitable for the subsequent forensic statistical analyses.

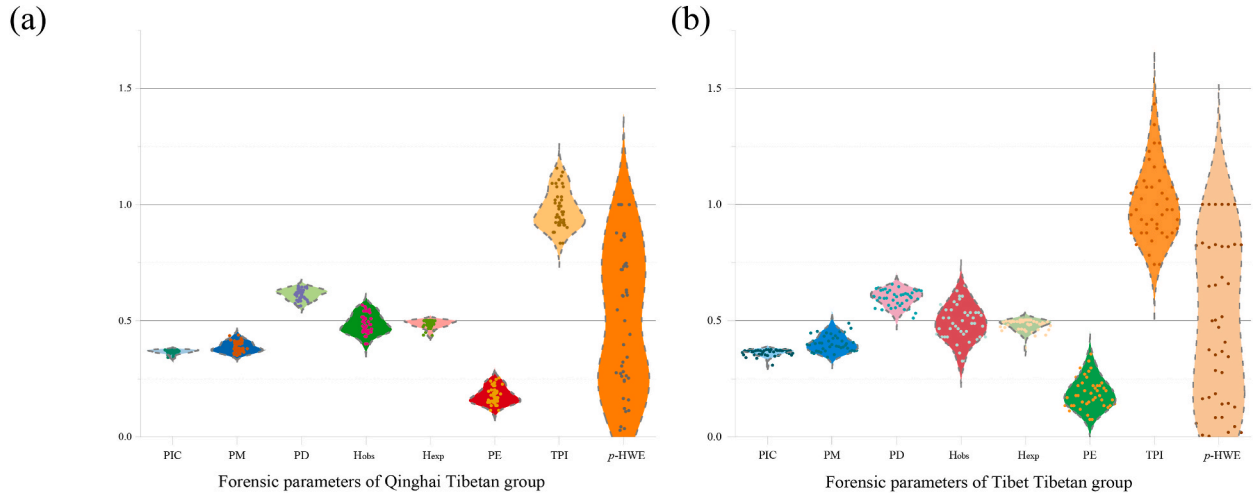

Fig. 1. Violin plots of forensic parameters for the 43 DIP loci in Qinghai Tibetan and Tibet Tibetan groups from China. (a) The distributions of PIC, PM, PD, H_{obs}, H_{exp}, PE, TPI and *p*-HWE values of the 43 DIP loci in Qinghai Tibetan group were shown in the violin plots. (b) The distributions of PIC, PM, PD, H_{obs}, H_{exp}, PE, TPI and *p*-HWE values of the 43 DIP loci in Tibet Tibetan group were shown in the violin plots. PIC, polymorphism information content; PM, probability of match; PD, power of discrimination; H_{obs}, observed heterozygosity; H_{exp}, expected heterozygosity; PE, power of exclusion; TPI, typical paternity index; *p*-HWE, *p* value of the Hardy-Weinberg Equilibrium test.

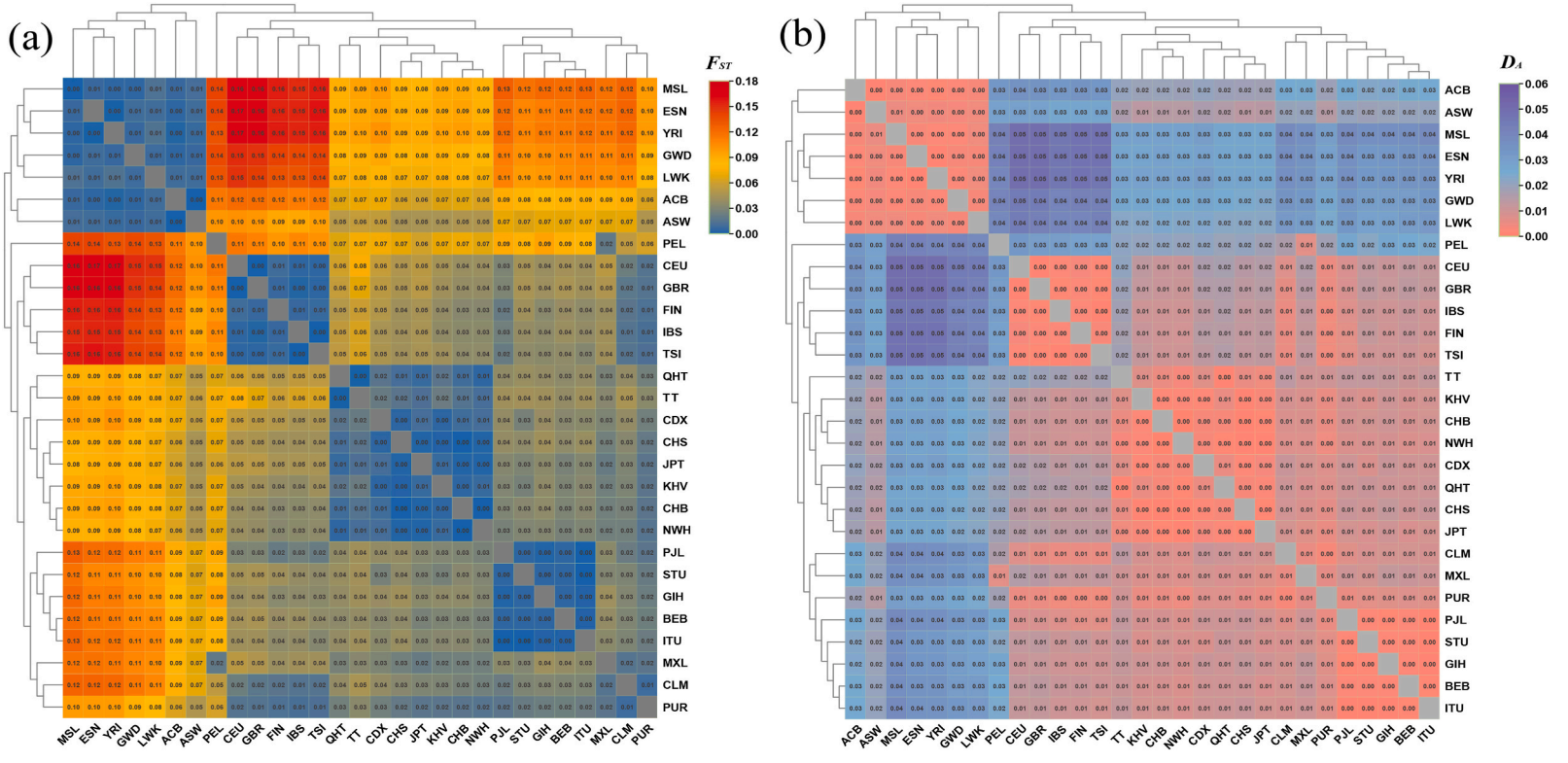

Fig. 2. Heatmaps of the pairwise F_{ST} and D_A values of the studied groups and 27 reference populations. (a) Heatmap of the pairwise F_{ST} values. (b) Heatmap of the pairwise D_A values. QHT, Qinghai Tibetan, China; TT, Tibet Tibetan, China; NWH, Northwest district Hui, China; CDX, Chinese Dai in Xishuangbanna, China; CHB, Han Chinese in Bejing, China; CHS, Southern Han Chinese, China; KHV, Kinh in Ho Chi Minh City, Vietnam; JPT, Japanese in Tokyo, Japan; CEU, Utah residents with Northern and Western European ancestry; FIN, Finnish in Finland; GBR, British in England and Scotland; IBS, Iberian populations in Spain; TSI, Toscani in Italy; CLM, Colombian in Medellin, Colombia; MXL, Mexican Ancestry in Los Angeles, California; PEL, Peruvian in Lima, Peru; PUR, Puerto Rican in Puerto Rico; PJL, Punjabi in Lahore, Pakistan; GIH, Gujarati Indian in Houston, TX; ITU, Indian Telugu in the UK; STU, Sri Lankan Tamil in the UK; BEB, Bengali in Bangladesh; ACB, African Caribbean in Barbados; ASW, African Ancestry in Southwest US; ESN, Esan in Nigeria; GWD, Gambian in Western Division, The Gambia; LWK, Luhya in Webuye, Kenya; MSL, Mende in Sierra Leone; YRI, Yoruba in Ibadan, Nigeria.

3.2. Results of the forensic statistical analyses of two Tibetan groups

The results of the forensic statistical analyses in QHT and TT groups from China were listed in Table S2. In the QHT group (Fig. 1a), the range of the insertion allele frequency values was observed between 0.3194 (rs146880183) and 0.6774 (rs10533337). The observed heterozygosity (H_{obs}) values of the 43 DIP loci were all over 0.40, and the expected heterozygosity (H_{exp}) values of the 43 DIP loci were greater than 0.4850. In the TT group (Fig. 1b), the frequencies of insertion alleles varied from 0.3245 (rs146880183) to 0.7742 (rs10533337). H_{obs} values were all greater than 0.3256, and the average H_{obs} and H_{exp} values of 43 DIP loci were 0.4908 and 0.4801, respectively.

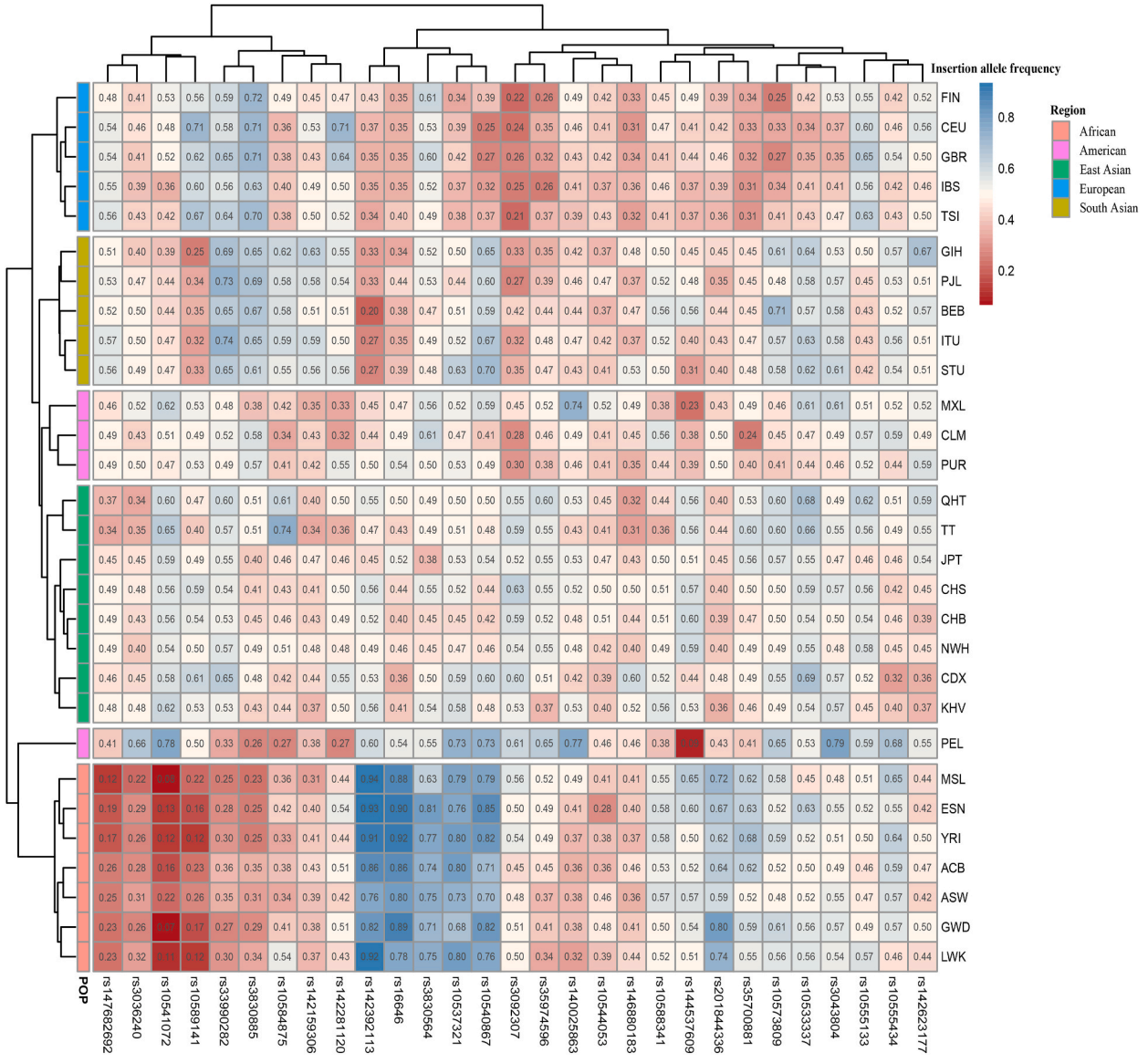

Fig. 3. Heatmap of the insertion allele frequencies of 43 DIP loci in the studied groups and 27 reference populations. The color depth of each box in the heatmap represents the insertion allele frequency value of a single locus in a population. The clustering analyses were shown at the top and left sides of heatmap. QHT, Qinghai Tibetan, China; TT, Tibet Tibetan, China; NWH, Northwest district Hui, China; CDX, Chinese Dai in Xishuangbanna, China; CHB, Han Chinese in Bejing, China; CHS, Southern Han Chinese, China; KHV, Kinh in Ho Chi Minh City, Vietnam; JPT, Japanese in Tokyo, Japan; CEU, Utah residents with Northern and Western European ancestry; FIN, Finnish in Finland; GBR, British in England and Scotland; IBS, Iberian populations in Spain; TSI, Toscani in Italy; CLM, Colombian in Medellin, Colombia; MXL, Mexican Ancestry in Los Angeles, California; PEL, Peruvian in Lima, Peru; PUR, Puerto Rican in Puerto Rico; PJL, Punjabi in Lahore, Pakistan; GIH, Gujarati Indian in Houston, TX; ITU, Indian Telugu in the UK; STU, Sri Lankan Tamil in the UK; BEB, Bengali in Bangladesh; ACB, African Caribbean in Barbados; ASW, African Ancestry in Southwest US; ESN, Esan in Nigeria; GWD, Gambian in Western Division, The Gambia; LWK, Luhya in Webuye, Kenya; MSL, Mende in Sierra Leone; YRI, Yoruba in Ibadan, Nigeria.

The polymorphism information content (PIC) values in the two studied Tibetan groups for the 43 DIP loci exceeded 0.35. Moreover, the combined power of discrimination (CPD) using 43 DIP loci was calculated to be 0.9999999999999999984 in QHT group and 0.9999999999999999921 in TT group. The cumulative power of exclusion (CPE) using 43 DIP loci was calculated to be 0.999782512 in the QHT group and 0.999886205 in the TT group. (Table S2).

3.3. Interpopulation differentiation analyses based on the 43 DIP loci for Tibetan groups and reference populations

The genetic differentiations among the two Tibetan groups and 27 reference populations were assessed by the pairwise F_{ST} and D_A values (Tables S5 and S6). The results were visualized in Fig. 2. For the QHT group (Fig. 2a), the minimum F_{ST} value (0.0012) was recorded between the QHT and TT groups, followed by the NWH group (0.0051), JPT population (0.0076), and CHB population (0.0092), while the largest F_{ST} value (0.0909) was found between QHT group and YRI population. For the TT group (Fig. 2a), we found the smallest F_{ST} value (0.0076) was with the QHT group, followed by the NWH group (0.0105), JPT population (0.0127), and CHB population (0.0143), while the largest F_{ST} value (0.0905) was also found between TT group and YRI population.

For the QHT group (Fig. 2b), the highest D_A value (0.0274) was between the QHT group and YRI population, while the lowest D_A value (0.0014) was found between QHT and TT groups, followed by the NWH group (0.0036), JPT population (0.0046), and CHB population (0.0050). For the TT group (Fig. 2b), the highest D_A value (0.0228) was between the TT group and MSL population, while the lowest D_A value (0.0014) was found between TT and QHT groups, followed by NWH group (0.0018), JPT population (0.0029), and CHB population (0.0033). We could observe that the Tibetan groups exhibited smaller genetic differentiations with East Asian populations, compared with the reference populations from African, European American, and South Asian (Fig. 2).

We further investigated the diversities of allele frequencies for the 43 DIP loci for all 29 populations in this study, exhibited in Fig. 3 and Table S7. The depth of color in the heatmap corresponded to the range of insertion allele frequencies. We could see that the clustering analysis on the left side of the heatmap (Fig. 3) revealed that the 29 populations were basically assigned into five subbranches based on their allele frequency distributions. South Asian, East Asian and European populations were divided into three relatively independent parts. PEL population gathered with the African populations, and the remaining American populations clustered into one part. The QHT and TT groups gathered with East Asian populations. It was evident that compared with populations in other continents, East Asian populations exhibited modest allele frequencies (0.40–0.60) of 43 DIP loci, as indicated by the clustering analysis on the top of the heatmap.

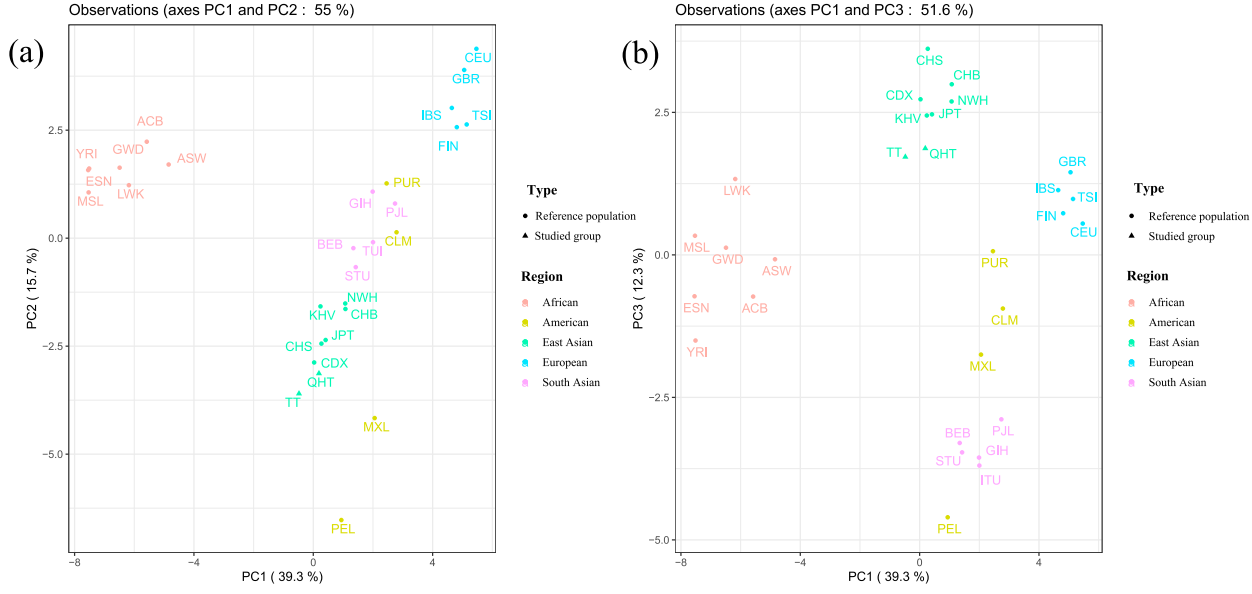

Fig. 4. PCA plots at population level for the studied groups and 27 reference populations worldwide based on insertion allele frequencies. The PCA results of population level for QHT, TT groups and 27 reference populations were visualized based on (a) PC1 and PC2, (b) PC1 and PC3. PC1, PC2 and PC3, the top three principal components; QHT, Qinghai Tibetan, China; TT, Tibet Tibetan, China; NWH, Northwest district Hui, China; CDX, Chinese Dai in Xishuangbanna, China; CHB, Han Chinese in Bejing, China; CHS, Southern Han Chinese, China; KHV, Kinh in Ho Chi Minh City, Vietnam; JPT, Japanese in Tokyo, Japan; CEU, Utah residents with Northern and Western European ancestry; FIN, Finnish in Finland; GBR, British in England and Scotland; IBS, Iberian populations in Spain; TSI, Toscani in Italy; CLM, Colombian in Medellin, Colombia; MXL, Mexican Ancestry in Los Angeles, California; PEL, Peruvian in Lima, Peru; PUR, Puerto Rican in Puerto Rico; PJL, Punjabi in Lahore, Pakistan; GIH, Gujarati Indian in Houston, TX; ITU, Indian Telugu in the UK; STU, Sri Lankan Tamil in the UK; BEB, Bengali in Bangladesh; ACB, African Caribbean in Barbados; ASW, African Ancestry in Southwest US; ESN, Esan in Nigeria; GWD, Gambian in Western Division, The Gambia; LWK, Luhya in Webuye, Kenya; MSL, Mende in Sierra Leone; YRI, Yoruba in Ibadan, Nigeria.

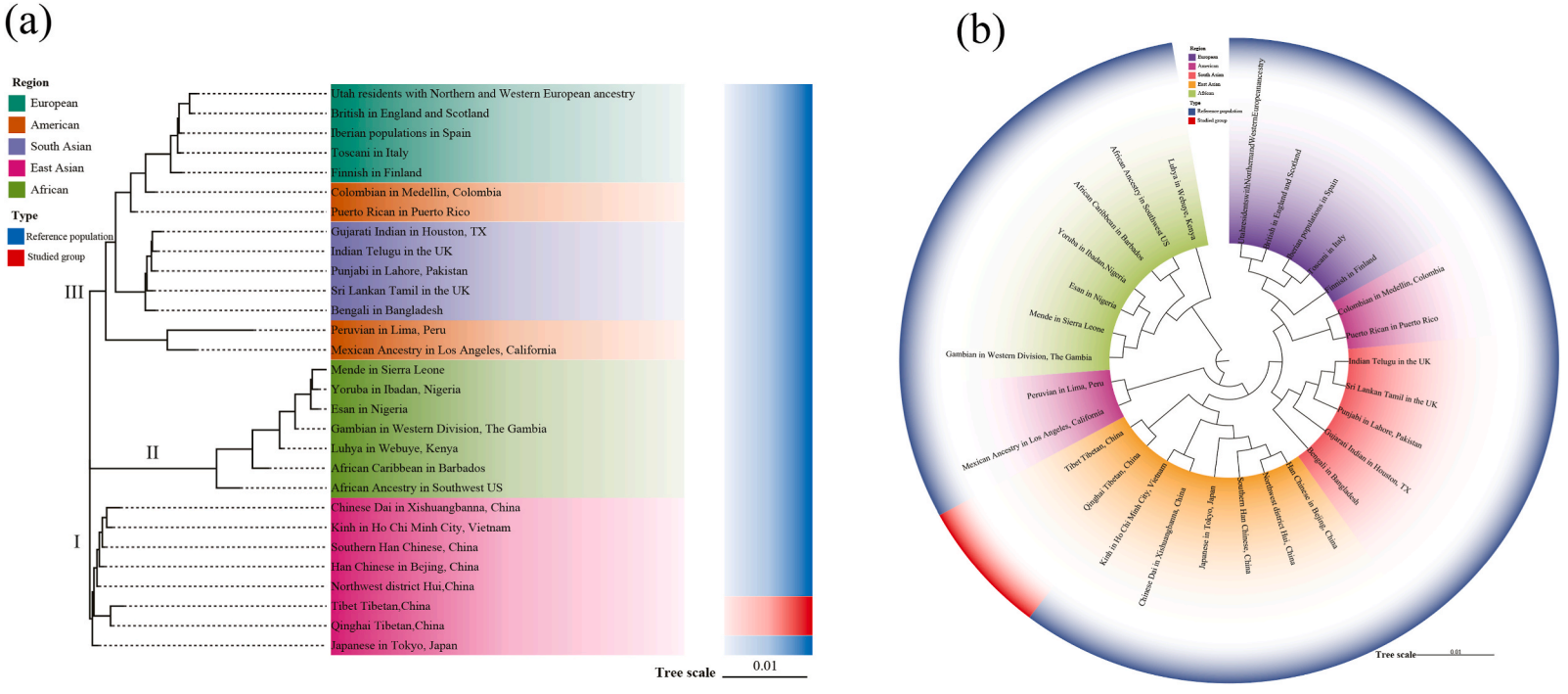

Fig. 5. Reconstructions of two phylogenetic trees of two Tibetan groups and 27 reference worldwide populations. (a) A phylogenetic tree of two Tibetan groups and 27 reference worldwide populations using the neighbor-joining method. (b) A phylogenetic tree of two Tibetan groups and 27 reference worldwide populations using the UPGMA method.

3.4. Principal component analysis of the 43 DIP loci in two Tibetan groups and 27 reference populations

PCA plays a crucial role in population genetics analysis by facilitating dimensionality reduction, visualizing genetic relationships, detecting sample outliers, and conducting genetic association studies. The population genetic relationships among 29 populations worldwide could be exhibited by PCA (Fig. 4; Fig. S1). The top three principal components accounted for 67.3 % of the total variation (PC1:39.3 %, PC2:15.7 %, and PC3:12.3 %). From the PCA plot at the population level (Fig. 4a), it was evident that the African populations formed a relatively distinct cluster that could be easily distinguished from populations in other continents; European populations gathered on the right PCA plot. South Asian and PEL populations gathered closely, and the rest of the American populations clustered together (Fig. 4a). The PC1 and PC2 could distinguish African and European populations (Fig. 4a), while PC1 and PC3 could basically distinguish European, African, East Asian, and South Asian populations (Fig. 4b). The QHT and TT groups clustered with East Asian populations (CDX, CHB, and JPT populations) and had the distant distances from African and European populations (Fig. 4). The PCA plot (Fig. S1) was generated based on the individual genotypes of the QHT, TT groups and three-continent populations (African, European, and East Asian populations). The first three principal components explained 17.1 % of genetic variation at the individual level, and the individuals from East Asian populations to large extent overlapped with those from Tibetan groups (Fig. S1).

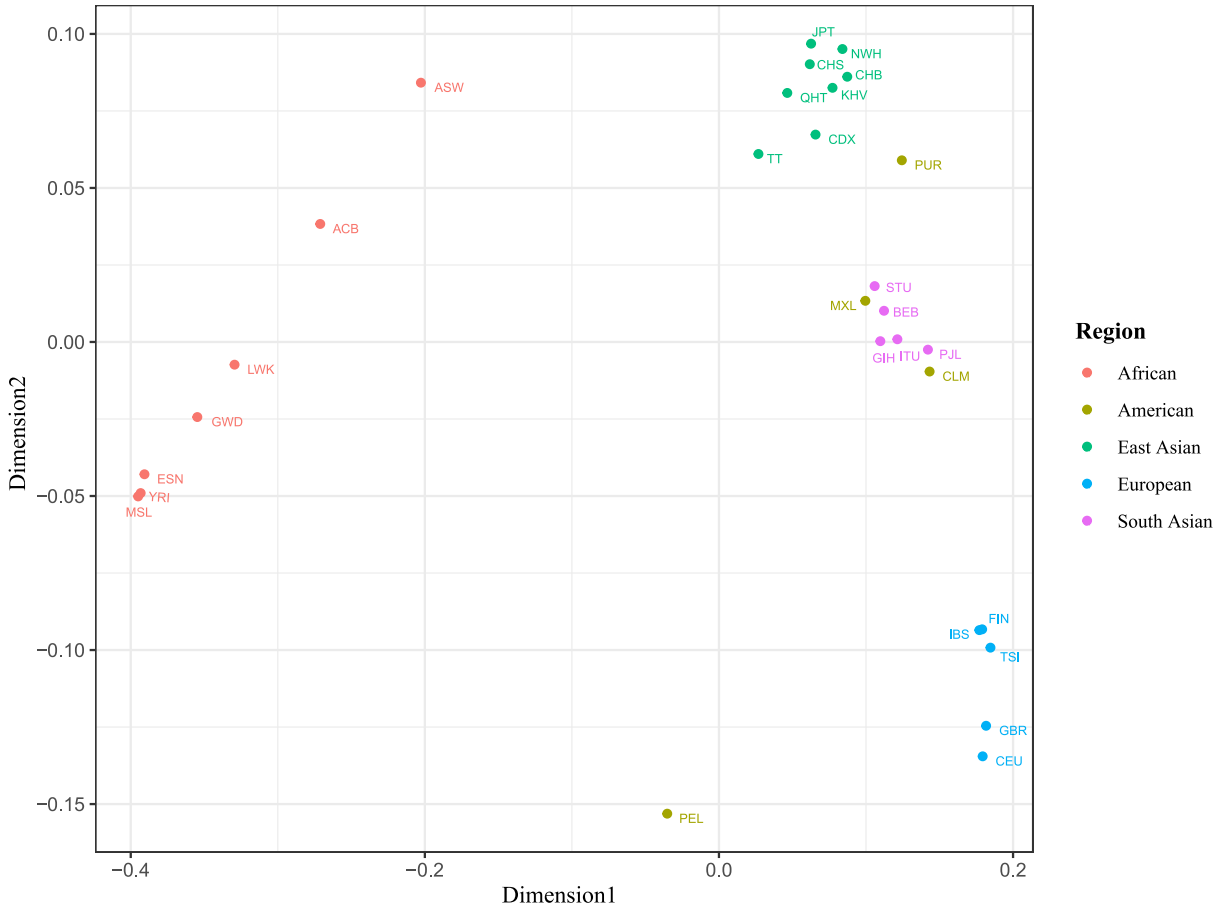

Fig. 6. MDS analysis of the two studied Tibetan groups and 27 reference populations based on the pairwise F_{ST} values. QHT, Qinghai Tibetan, China; TT, Tibet Tibetan, China; NWH, Northwest district Hui, China; CDX, Chinese Dai in Xishuangbanna, China; CHB, Han Chinese in Bejing, China; CHS, Southern Han Chinese, China; KHV, Kinh in Ho Chi Minh City, Vietnam; JPT, Japanese in Tokyo, Japan; CEU, Utah residents with Northern and Western European ancestry; FIN, Finnish in Finland; GBR, British in England and Scotland; IBS, Iberian populations in Spain; TSI, Toscani in Italy; CLM, Colombian in Medellin, Colombia; MXL, Mexican Ancestry in Los Angeles, California; PEL, Peruvian in Lima, Peru; PUR, Puerto Rican in Puerto Rico; PJL, Punjabi in Lahore, Pakistan; GIH, Gujarati Indian in Houston, TX; ITU, Indian Telugu in the UK; STU, Sri Lankan Tamil in the UK; BEB, Bengali in Bangladesh; ACB, African Caribbean in Barbados; ASW, African Ancestry in Southwest US; ESN, Esan in Nigeria; GWD, Gambian in Western Division, The Gambia; LWK, Luhya in Webuye, Kenya; MSL, Mende in Sierra Leone; YRI, Yoruba in Ibadan, Nigeria.

3.5. Phylogenetic tree reconstructions and multidimensional scaling among Tibetan groups and 27 reference populations

Two phylogenetic trees were built using the pairwise D_A values to gain insight into the genetic relationships of the QHT, TT groups and 27 worldwide reference populations. In Fig. 5a, the tree was basically separated into three main branches. Branch I consisted of all the above-mentioned East Asian populations, including QTH and TT groups; branch II included all African populations; branch III was in the higher part of the tree and contained European, American, and South Asian populations. The European populations were at the top of the plot, while the branch under the European populations basically included South Asian and American populations. The

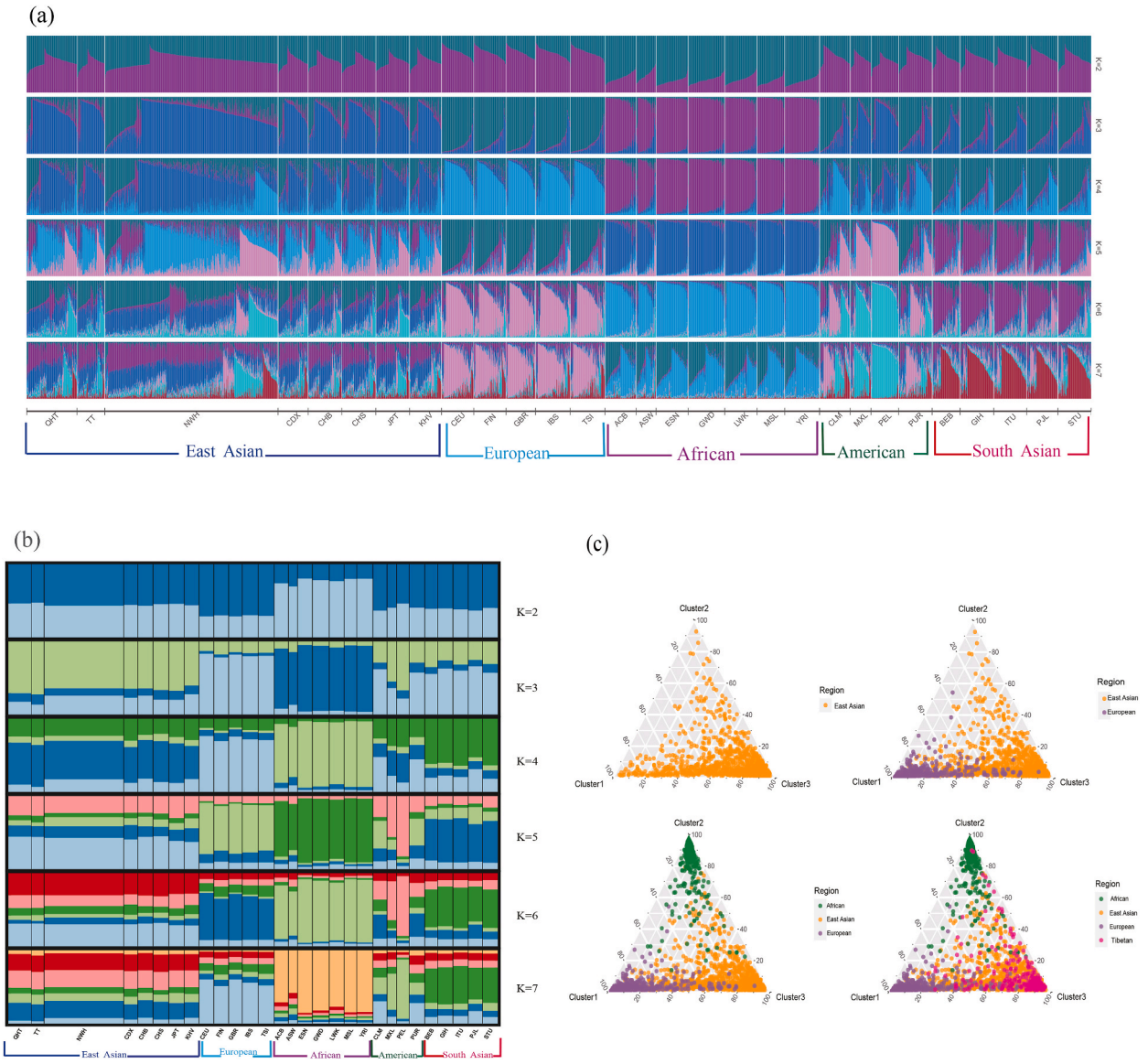

Fig. 7. STRUCTURE analyses of the two studied groups and 27 reference populations. (a) STRUCTURE analyses of the QHT, TT groups and 27 reference populations based on the 43 DIP loci at the individual level (K = 2–7). (b) STRUCTURE analyses of the QHT, TT groups and 27 reference populations based on the 43 DIP loci at the population level (K = 2–7). (c) The ternary plots of cluster analyses for ancestral components of Tibetan groups when K = 3. QHT, Qinghai Tibetan, China; TT, Tibet Tibetan, China; NWH, Northwest district Hui, China; CDX, Chinese Dai in Xishuangbanna, China; CHB, Han Chinese in Bejing, China; CHS, Southern Han Chinese, China; KHV, Kinh in Ho Chi Minh City, Vietnam; JPT, Japanese in Tokyo, Japan; CEU, Utah residents with Northern and Western European ancestry; FIN, Finnish in Finland; GBR, British in England and Scotland; IBS, Iberian populations in Spain; TSI, Toscani in Italy; CLM, Colombian in Medellin, Colombia; MXL, Mexican Ancestry in Los Angeles, California; PEL, Peruvian in Lima, Peru; PUR, Puerto Rican in Puerto Rico; PJL, Punjabi in Lahore, Pakistan; GIH, Gujarati Indian in Houston, TX; ITU, Indian Telugu in the UK; STU, Sri Lankan Tamil in the UK; BEB, Bengali in Bangladesh; ACB, African Caribbean in Barbados; ASW, African Ancestry in Southwest US; ESN, Esan in Nigeria; GWD, Gambian in Western Division, The Gambia; LWK, Luhya in Webuye, Kenya; MSL, Mende in Sierra Leone; YRI, Yoruba in Ibadan, Nigeria.

clustering result of the tree shown in Fig. 5b remained substantially consistent with those of Fig. 5a, demonstrating each population was basically gathered following their respective continent origins.

As depicted in Fig. 6, a MDS plot was also generated to further excavate the detailed genetic relatedness among the 29 populations. African populations scattered on the left part of the MDS plot, while European populations clustered in the lower right corner of the MDS plot. American and South Asian populations gathered on the right part of the MDS plot. QHT and TT groups congregated with East Asian populations in the upper right part of the MDS plot, which was basically consistent with the PCA results.

3.6. Interpopulation structure analysis on basis of 43 DIP loci in the two Tibetan groups and 27 reference populations

As shown in Fig. 7, the population ancestral component analyses for Tibetan groups and 27 reference populations were conducted by STRUCTURE software. At the individual level (Fig. 7a), African and non-African populations could be identified at $K = 2$. For the population level at $K = 3$ (Fig. 7b), African populations presented mainly deep blue ancestral component; European populations displayed mainly light blue ancestral component; and East Asian populations presented mainly lightweight green ancestral component. It was relatively straightforward to distinguish among African, East Asian, and European populations when K was three (Fig. 7b), which was consistent with the presentation of results at the individual level (Fig. 7a). Besides, the ancestral components of the QHT and TT groups was similar to those of East Asian populations. Furthermore, the STRUCTURE result was uploaded on the online website STRUCTURE HARVESTER (http://taylor0.biology.ucla.edu/structureHarvester/) to acquire the optimal K value, and it turned out that the most appropriate K value was three according to the result of Delta K plot (Fig. S2). The ternary plot was displayed in Fig. 7c when K was three at the individual level for East Asian, African, European populations, and two Tibetan groups. The green, orange, and purple colors represent the ancestral components of the African, East Asian, and European populations, respectively. The pink color represented the ancestral components of QHT and TT groups, which were basically clustered with East Asian populations.

4. Discussion

Tibetan groups, located on the Qinghai-Tibet Plateau with the highest average altitude in the world, represent one of the most ancient Chinese groups and hold an irreplaceable status in the history of China. Shedding light on the population genetic features of the Tibetan groups and exploring their ancestral origins would be a vital step toward an enhanced comprehension of the historical origins of the Chinese populations.

The genetic traits of the Tibetan groups in Qinghai and Tibet regions were investigated using an innovative amplification system containing 43 DIP loci previously published by our team. When compared with the population data generated by the Investigator® DIPplex kit (Qiagen) [24], the studied 43 DIP system yielded higher CPD (QHT: 0.9999999999999999984; TT: 0.9999999999999999921) and CPE values (QHT: 0.999782512, TT: 0.999886205) in the Tibetan groups. The H_{obs} and H_{exp} values of 43 DIP loci exceeded 0.45 in QHT and TT groups, and high PIC values (>0.35) of 43 DIP loci were observed in both groups. It is demonstrated that this multiplex amplification system exhibited considerably greater levels of polymorphisms in two Tibetan groups. The diversity distributions of allele frequencies of 43 DIP loci among 29 populations exhibited that most loci could be suitable for individual identification due to their relatively modest allele frequencies in East Asian populations.

We conducted an analysis of the genetic features and relationships among 29 populations on the basis of pairwise F_{ST} and D_A values. These results indicated the QHT and TT groups exhibited the smallest genetic distances with East Asian populations, implying that they possessed high levels of genetic similarities according to the population genetic analyses. We employed PCA, a method for classical feature extraction and data representation [37], to analyze genetic differentiations and associations among different populations. The results indicated that the top three principal components contributed 67.3 % of the variation among all populations, which enabled us to distinguish among different populations from three continents. The phylogenetic trees were reconstructed to probe the phylogenetic relationships among the 29 populations. Furthermore, in the results of PCA plots and the phylogenetic trees, QHT and TT groups clustered with East Asian populations, implying they probably possessed a certain genetic similarity among them. The MDS plot was substantially consistent with the PCA results. The STRUCTURE analysis results demonstrated that the proportions of ancestral components in the Tibetan groups were similar to those of the East Asian populations. In a word, our present study, which utilized a multiplex amplification system containing 43 DIP loci, confirmed that the Tibetan groups had close genetic affinities with each other, and they shared close genetic relationships with East Asian populations.

We could also draw support for our findings from historical records and published reports. A considerable amount of studies have investigated the genetic relationships between Tibetan groups and reference populations worldwide based on STRs on the autosomes [6,38–40]. Tibetan groups in different regions, to a large extent, were genetically similar to each other and shared close genetic relationships with East Asian populations. Similar results have been obtained from the population genetic studies of Tibetan groups based on other autosomal DIP loci [7,41,42]. Xie et al. discovered that based on 30 DIP loci, QHT and TT groups exhibited the close genetic affinities with Chinese various populations, including Han population and Hui group [41]. Similarly, the current research results we obtained were basically consistent with those obtained by Liu et al. [7], demonstrating that QHT and TT groups were more genetically similar to each other than other reference populations based on 35 DIP loci. Jin et al. investigated the Tibetan groups using 39 ancestry informative DIP markers and discovered the genetic similarities between Tibetan groups and East Asian populations [42].

However, in order to substantiate the system's robustness in forensic applications across diverse populations, it is imperative that a larger sample size of East Asian population data is utilized. Additionally, the population genetic characteristics of the Tibetan groups, as revealed by this system, lacks comprehensiveness and requires further exploration and discussion in tandem with diverse molecular genetic markers.

The Late Paleolithic period on the Tibetan Plateau is believed to be one of the origins of modern Tibetans, as evidenced by the signs of human activities during that time [43–45]. This perspective is further supported by several studies on mitochondrial polymorphisms [46,47]. The EPAS1 gene is believed to play a significant role in reducing hypoxia, and it is considered one of the key genes involved in the adaptation of Tibetan groups to the high-altitude plateau environment [48]. Linguistic studies have shown that Tibetan groups and Chinese-speaking people share a common ancestral root, which likely diverged approximately 6000 years ago [49]. Research on genetics using exon sequencing has estimated that the divergence time between the Tibetan and Chinese Han populations was about 2750 years [50]. Additionally, to a certain extent, social advancements and economic developments have increased cultural exchanges and migrations among residents in the Qinghai-Tibetan Plateau and its neighboring regions [51]. This intermingling of cultures and populations may have contributed to the close genetic relationships observed among the populations in this region.

5. Conclusion

We assessed the forensic performance of this innovative self-developed amplification system containing 43 DIP loci by caculating their forensic parameters in the Qinghai Tibetan and Tibet Tibetan groups and exploring their population genetic relationships with reference populations. Our results showed these DIP loci were the remarkably high-level genetic polymorphisms, indicating that the system could be well-suited for the individual identifications in these two studied groups. The population genetic analyses indicated that the two Tibetan groups from Qinghai and Tibet regions exhibited closer relationships with East Asian populations.

Data availability

All the data are available from the corresponding author.

Funding statement

The study was funded by grants from National Key R&D Program of China (2022YFC3302004), 2022YFC3302004-1.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

We sincerely like to express the greatest gratitude and thanks to the volunteers who had contributed samples for our current study.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.heliyon.2023.e21229.

References

[1] L.H. Royden, B.C. Burchfiel, R.D. van der Hilst, B>The geological evolution of the Tibetan Plateau</B&gt, Science 321 (2008) 1054–1058.
[2] J. Wen, J. Zhang, Z.-L. Nie, Y. Zhong, H. Sun, Evolutionary diversifications of plants on the Qinghai-Tibetan plateau, Front. Genet. 5 (2014) 4.
[3] F.R. Lorenzo, C. Huff, M. Myllymäki, et al., A genetic mechanism for Tibetan high-altitude adaptation, Nat. Genet. 46 (2014) 951–956.
[4] G. He, Z. Wang, X. Zou, et al., Genetic diversity and phylogenetic characteristics of Chinese Tibetan and Yi minority ethnic groups revealed by non-CODIS STR markers, Sci. Rep. 8 (2018) 5895, https://doi.org/10.1038/s41598-018-24291-5.
[5] J. Liu, Z. Wang, G. He, et al., Massively parallel sequencing of 124 SNPs included in the precision ID identity panel in three East Asian minority ethnicities, Forensic Sci Int Genet 35 (2018) 141–148.
[6] G. He, Z. Wang, Y. Su, et al., Genetic structure and forensic characteristics of Tibeto-Burman-speaking Ü-Tsang and Kham Tibetan Highlanders revealed by 27 Y-chromosomal STRs, Sci. Rep. 9 (2019) 7739, https://doi.org/10.1038/s41598-019-44230-2.
[7] Y. Liu, X. Jin, S. Mei, et al., Insights into the genetic characteristics and population structures of Chinese two Tibetan groups using 35 insertion/deletion polymorphic loci, Mol. Genet. Genom. 295 (2020) 957–968.
[8] L. Li, Y. Ye, F. Song, Z. Wang, Y. Hou, Genetic structure and forensic parameters of 30 InDels for human identification purposes in 10 Tibetan populations of China, Forensic Sci Int Genet 40 (2019) e219–e227.
[9] B.R. McCord, Q. Gauthier, S. Cho, et al., Forensic DNA analysis, Anal. Chem. 91 (2019) 673–688, https://doi.org/10.1021/acs.analchem.8b05318.
[10] B. Budowle, A. van Daal, Forensically relevant SNP classes, Biotechniques 44 (10) (2008) 603–608, https://doi.org/10.2144/000112806.
[11] R. Grover, H. Jiang, R.S. Turingan, J.L. French, E. Tan, R.F. Selden, FlexPlex27—highly multiplexed rapid DNA identification for law enforcement, kinship, and military applications, Int. J. Leg. Med. 131 (2017) 1489–1501.
[12] Y. Fang, Y. Guo, T. Xie, et al., Forensic molecular genetic diversity analysis of Chinese Hui ethnic group based on a novel STR panel, Int. J. Leg. Med. 132 (2018) 1297–1299.
[13] L.K. Hennessy, N. Mehendale, K. Chear, et al., Developmental validation of the GlobalFiler® express kit, a 24-marker STR assay, on the RapidHIT® System, Forensic Sci. Int.: Genetics 13 (2014) 247–258.
[14] C.D. Campbell, J.X. Chong, M. Malig, et al., Estimating the human mutation rate using autozygosity in a founder population, Nat. Genet. 44 (2012) 1277–1281.
[15] Y. Yang, B. Xie, J. Yan, Application of next-generation sequencing technology in forensic science, Dev. Reprod. Biol. 12 (2014) 190–197.

[16] J. Hampe, A. Wollstein, T. Lu, et al., An integrated system for high throughput TaqMan™ based SNP genotyping, Bioinformatics 17 (2001) 654–655.
[17] A.-C. Syvänen, Accessing genetic variation: genotyping single nucleotide polymorphisms, Nat. Rev. Genet. 2 (2001) 930–942.
[18] R.E. Mills, C.T. Luttig, C.E. Larkins, et al., An initial map of insertion and deletion (INDEL) variation in the human genome, Genome Res. 16 (2006) 1182–1190.
[19] R. Pereira, C. Phillips, C. Alves, A. Amorim, A. Carracedo, L. Gusmão, A new multiplex for human identification using insertion/deletion polymorphisms, Electrophoresis 30 (2009) 3682–3690, https://doi.org/10.1002/elps.200900274.
[20] Q. Lan, C. Shen, X. Jin, et al., Distinguishing three distinct biogeographic regions with an in-house developed 39-AIM-InDel panel and further admixture proportion estimation for Uyghurs, Electrophoresis 40 (2019) 1525–1534.
[21] B.L. LaRue, J. Ge, J.L. King, B. Budowle, A validation study of the Qiagen Investigator DIPplex® kit; an INDEL-based assay for human identification, Int. J. Leg. Med. 126 (2012) 533–540, https://doi.org/10.1007/s00414-012-0667-9.
[22] 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® DIPplex kit, Forensic Sci Int Genet 11 (2014) e12–e14.
[23] 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–e25.
[24] Y. Guo, C. Shen, H. Meng, et al., Population differentiations and phylogenetic analysis of Tibet and Qinghai Tibetan groups based on 30 InDel loci, DNA Cell Biol. 35 (2016) 787–794, https://doi.org/10.1089/dna.2016.3395.
[25] 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® DIPplex kit, Forensic Sci. Int.: Genetics 11 (2014) e12–e14, https://doi.org/10.1016/j.fsigen.2014.03.017.
[26] R. Jin, W. Cui, Y. Fang, et al., A novel panel of 43 insertion/deletion loci for human identifications of forensic degraded DNA samples: development and validation, Front. Genet. 12 (2021), 610540, https://doi.org/10.3389/fgene.2021.610540.
[27] A. Auton, L.D. Brooks, R.M. Durbin, et al., A global reference for human genetic variation, Nature 526 (2015) 68–74, https://doi.org/10.1038/nature15393.
[28] A. Gouy, M. Zieger, STRAF-A convenient online tool for STR data evaluation in forensic genetics, Forensic Sci Int Genet 30 (2017) 148–151.
[29] M. Nei, Estimation of average heterozygosity and genetic distance from a small number of individuals, Genetics 89 (1978) 583–590.
[30] F. Rousset, genepop'007: a complete re-implementation of the genepop software for Windows and Linux, Mol Ecol Resour 8 (2008) 103–106.
[31] M. Nei, F. Tajima, Y. Tateno, Accuracy of estimated phylogenetic trees from molecular data, J. Mol. Evol. 19 (1983) 153–170.
[32] S. Kumar, G. Stecher, K. Tamura, MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets, Mol. Biol. Evol. 33 (2016) 1870–1874.
[33] C. Chen, H. Chen, Y. Zhang, et al., TBtools: an integrative toolkit developed for interactive analyses of big biological data, Mol. Plant 13 (2020) 1194–1202.
[34] M.A. Cox, T.F. Cox, Multidimensional Scaling, Springer, 2008, pp. 315–347.
[35] M. Jakobsson, N.A. Rosenberg, CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure, Bioinformatics 23 (2007) 1801–1806, https://doi.org/10.1093/bioinformatics/btm233.
[36] L. Porras-Hurtado, Y. Ruiz, C. Santos, C. Phillips, A. Carracedo, M.V. Lareu, An overview of STRUCTURE: applications, parameter settings, and supporting software, Front. Genet. 4 (2013) 98, https://doi.org/10.3389/fgene.2013.00098.
[37] J. Yang, D. Zhang, A.F. Frangi, J-y Yang, Two-dimensional PCA: a new approach to appearance-based face representation and recognition, IEEE Trans. Pattern Anal. Mach. Intell. 26 (2004) 131–137.
[38] H.-B. Yao, C.-C. Wang, J. Wang, et al., Genetic structure of Tibetan populations in Gansu revealed by forensic STR loci, Sci. Rep. 7 (2017) 1–6.
[39] Z. Li, J. Zhang, H. Zhang, Z. Lin, J. Ye, Genetic polymorphisms in 18 autosomal STR loci in the Tibetan population living in Tibet Chamdo, Southwest China, Int. J. Leg. Med. 132 (2018) 733–734.
[40] G. He, Y. Li, X. Zou, et al., X-chromosomal STR-based genetic structure of Sichuan Tibetan minority ethnicity group and its relationships to various groups, Int. J. Leg. Med. 132 (2018) 409–413.
[41] T. Xie, Y. Guo, L. Chen, et al., A set of autosomal multiple InDel markers for forensic application and population genetic analysis in the Chinese Xinjiang Hui group, Forensic Sci. Int.: Genetics 35 (2018) 1–8, https://doi.org/10.1016/j.fsigen.2018.03.007.
[42] X.-Y. Jin, C.-M. Shen, C. Chen, et al., Ancestry informative DIP loci for dissecting genetic structure and ancestry proportions of Qinghai Tibetan and Tibet Tibetan groups, Mol. Biol. Rep. 47 (2020) 1079–1087.
[43] X.L. Zhang, B.B. Ha, S.J. Wang, et al., The earliest human occupation of the high-altitude Tibetan Plateau 40 thousand to 30 thousand years ago, Science 362 (2018) 1049–1051, https://doi.org/10.1126/science.aat8824.
[44] D.B. Madsen, M. Haizhou, P.J. Brantingham, et al., The late Upper Paleolithic occupation of the northern Tibetan Plateau margin, J. Archaeol. Sci. 33 (2006) 1433–1444.
[45] X. Qi, C. Cui, Y. Peng, et al., Genetic evidence of paleolithic colonization and neolithic expansion of modern humans on the Tibetan plateau, Mol. Biol. Evol. 30 (2013) 1761–1778.
[46] B. Yuan, W. Huang, D. Zhang, New evidence for human occupation of the northern Tibetan Plateau, China during the Late Pleistocene, Chin. Sci. Bull. 52 (2007) 2675–2679.
[47] M. Zhao, Q.-P. Kong, H.-W. Wang, et al., Mitochondrial genome evidence reveals successful Late Paleolithic settlement on the Tibetan Plateau, Proc. Natl. Acad. Sci. USA 106 (2009) 21230–21235.
[48] E. Huerta-Sánchez, X. Jin, Asan, et al., Altitude adaptation in Tibetans caused by introgression of Denisovan-like DNA, Nature 512 (2014) 194–197, https://doi.org/10.1038/nature13408.
[49] W.S.-Y. Wang, Language and the evolution of modern humans, The origins and past of modern humans (1998) 247–262.
[50] X. Yi, Y. Liang, E. Huerta-Sanchez, et al., Sequencing of 50 human exomes reveals adaptation to high altitude, Science 329 (2010) 75–78.
[51] G.E. Clarke, The movement of population to the west of China: Tibet and Qinghai, in: J.M. Brown, R. Foot (Eds.), Migration: the Asian Experience, Palgrave Macmillan UK London, 1994, pp. 221–257.