Forensic Science International: Genetics 56 (2022) 102633

Contents lists available at ScienceDirect

Forensic Science International: Genetics

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

Research paper

Genomic history and forensic characteristics of Sherpa highlanders on the Tibetan Plateau inferred from high-resolution InDel panel and genome-wide SNPs

Mengge Wang [a,b,1], Weian Du [c,d,1], Renkuan Tang [e], Yan Liu [f], Xing Zou [g,h], Didi Yuan [e], Zheng Wang [h], Jing Liu [h], Jianxin Guo [i], Xiaomin Yang [i], Jing Chen [j], Meiqing Yang [j], Xianpeng Zhang [k], Lan-Hai Wei [i], Haibing Yuan [l,m,*], Hui-Yuan Yeh [n,**], Chuan-Chao Wang [i,**], Chao Liu [a,b,d,***], Guanglin He [i,n,****,1]

[a] Guangzhou Forensic Science Institute, Guangzhou 510030, PR China
[b] Faculty of Forensic Medicine, Zhongshan School of Medicine, Sun Yat-sen University, No. 74 Zhongshan Road II, Guangzhou 510080, PR China
[c] AGCU ScienTech Incorporation, Wuxi 214174, PR China
[d] School of Forensic Medicine, Southern Medical University, Guangzhou 510515, China
[e] Department of Forensic Medicine, College of Basic Medicine, Chongqing Medical University, Chongqing 400016, PR China
[f] School of Basic Medical Sciences, North Sichuan Medical College, Nanchong, Sichuan 637100, PR China
[g] College of Basic Medicine, Chongqing University, Chongqing 400016, PR China
[h] Institute of Forensic Medicine, West China School of Basic Science and Forensic Medicine, Sichuan University, Chengdu 610000, PR China
[i] Department of Anthropology and Ethnology, Institute of Anthropology, National Institute for Data Science in Health and Medicine, State Key Laboratory of Marine Environmental Science, State Key Laboratory of Cellular Stress Biology, School of Life Sciences, Xiamen University, Xiamen 361005, PR China
[j] Department of Forensic Medicine, Guizhou Medical University, Guiyang 550000, PR China
[k] Institute of Biological Anthropology, Jinzhou Medical University, Jinzhou 121000, PR China
[l] National Demonstration Center for Experimental Archaeology Education and Department of Archaeology, Sichuan University, Chengdu 610200, PR China
[m] School of Archaeology and Museology & National Demonstration Center for Experimental Archaeology Education, Sichuan University, Chengdu, Sichuan 610064, PR China
[n] School of Humanities, Nanyang Technological University, Nanyang 639798, Singapore

ARTICLE INFO

Keywords:
InDels
Genome-wide SNPs
Forensic genetics
Genetic admixture history
Ancestral composition
Natural selection

ABSTRACT

Sherpa people, one of the high-altitude hypoxic adaptive populations, mainly reside in Nepal and the southern Tibet Autonomous Region. The genetic origin and detailed evolutionary profiles of Sherpas remain to be further explored and comprehensively characterized. Here we analyzed the newly-generated InDel genotype data from 628 Dingjie Sherpas by merging with 4222 worldwide InDel profiles and collected genome-wide SNP data (approximately 600K SNPs) from 1612 individuals in 191 modern and ancient populations to explore and reconstruct the fine-scale genetic structure of Sherpas and their relationships with nearby modern and ancient East Asians based on the shared alleles and haplotypes. The forensic parameters of 57 autosomal InDels (A-InDels) included in our used new-generation InDel amplification system showed that this focused InDel panel is informative and polymorphic in Dingjie Sherpas, suggesting that it can be used as the supplementary tool for forensic personal identification and parentage testing in Dingjie Sherpas. Descriptive findings from the PCA, ADMIXTURE, and TreeMix-based phylogenies suggested that studied Nepal Sherpas showed excess allele sharing

* Corresponding author at: National Demonstration Center for Experimental Archaeology Education and Department of Archaeology, Sichuan University, Chengdu 610200, PR China.
** Corresponding authors.
*** Corresponding author at: Guangzhou Forensic Science Institute, Guangzhou 510030, PR China.
**** Corresponding author at: Department of Anthropology and Ethnology, Institute of Anthropology, National Institute for Data Science in Health and Medicine, State Key Laboratory of Marine Environmental Science, State Key Laboratory of Cellular Stress Biology, School of Life Sciences, Xiamen University, Xiamen 361005, PR China.
E-mail addresses: yuanbenhb@scu.edu.cn (H. Yuan), hyyeh@ntu.edu.sg (H.-Y. Yeh), wang@xmu.edu.cn (C.-C. Wang), liuchaogzf@163.com (C. Liu), Guanglinhescu@163.com (G. He).
[1] Mengge Wang, Weian Du and Guanglin He contributed equally to this work and should be considered co-first authors.

https://doi.org/10.1016/j.fsigen.2021.102633
Received 11 August 2021; Received in revised form 13 October 2021; Accepted 10 November 2021
Available online 18 November 2021

with neighboring Tibeto-Burman Tibetans. Furthermore, patterns of allele sharing in *f*-statistics demonstrated that Nepal Sherpas had a different evolutionary history compared with their neighbors from Nepal (Newar and Gurung) but showed genetic similarity with 2700-year-old Chokhopani and modern Tibet Tibetans. *QpAdm*/*qpGraph*-based admixture sources and models further showed that Sherpas, core Tibetans, and Chokhopani formed one clade, which could be fitted as having the main ancestry from late Neolithic Qijia millet farmers and other deep ancestries from early Asians. Chromosome painting profiles and shared IBD fragments inferred from fineSTRUCTURE and ChromoPainter not only confirmed the abovementioned genomic affinity patterns but also revealed the fine-scale genetic microstructures among Sino-Tibetan speakers. Finally, natural-selection signals revealed via iHS, nSL and iHH12 showed natural selection signatures associated with disease susceptibility in Sherpas. Generally, we provided the comprehensive landscape of admixture and evolutionary history of Sherpa people based on the shared alleles and haplotypes from the InDel-based genotype data and high-density genome-wide SNP data. The more detailed genetic landscape of Sherpa people should be further confirmed and characterized via ancient genomes or single-molecule real-time sequencing technology.

1. Introduction

Sherpa is one of the Tibeto-Burman-speaking groups native to the most mountainous regions of China, Nepal, India, and Bhutan. Sherpa highlanders are renowned as excellent mountaineers and guides for their hardiness, expertise, and experience at high altitudes, who are considered as a good candidate for illuminating the genomic mechanism of high-altitude adaptation and have played an invaluable role in the exploration of peopling processes of the Himalayas [1–3]. However, when and where the Sherpa people originated and settled in the high-altitude regions of the Himalayas remains unclear. Besides, their genetic relationships with adjoining Tibetans also keep contentious and more genomic evidence was needed to support or disprove the hypothesis of the common origin of Sherpas and Tibetans. Genetic observations from matrilineal lineages demonstrated that Tibetans or Tibetan-related ancestors contributed considerably to the gene pool of Sherpa people [2,4,5]. A previous study of mitochondrial DNA (mtDNA) diversity on Sherpas reported considerable South Asian genetic components existing in Zhangmu Sherpas [6], but maternal lineages of other Sherpas living in Nepal and Tibet Autonomous Region of China revealed that matrilineal haplogroups with South Asian origin occurred in these studied Sherpas with a minor frequency [4]. Genetic findings based on Y-chromosomal single nucleotide polymorphisms (SNPs) showed that Sherpas shared most of their paternal lineages with indigenous Tibetans, as haplogroups D and O were their dominant founding paternal lineages [4]. Among non-uniparental markers, the genetic structure of the Nepal Sherpas and neighboring Nepalese populations based on the genome-wide SNPs indicated that the Sherpas were a remarkably isolated population, with little gene flow from surrounding Nepalese populations [7]. Further well-characterized genomic landscape of Tibeto-Burman and Indo-Aryan communities from the remote Nepalese valleys also provided evidence for Sherpa isolation [8]. A recent study focused on Tibetan highlanders in China based on the whole-genome sequencing data indicated that a small portion of Tibetans' and Sherpas' ancestral components originated from separate ancient populations, which were estimated to have lived approximately 7000–11000 years before present (YBP) [9]. Lu et al. also demonstrated that Sherpas and Tibetans shared more recent ancestors than one of them with Han Chinese [9]. Conversely, Jeong et al. pointed out that Sherpas and Han Chinese served as dual ancestral populations of Tibetans and supported that modern Tibetans were the recent admixture result of ancestral sources related to Sherpas and Hans [10]. Generally, some genetic evidence supported the Nepal-Tibet migrations, while some others supported the Tibet-Nepal migrations; some genetic findings suggested that Tibetans were the ancestral populations of the Sherpas, while some others indicated that Tibetans were a mixture of ancestral populations related to the Sherpas and Han Chinese. The disparity in these perspectives reflects a fundamental division in the demographic history of Sherpas, suggesting that the genetic makeup of Sherpas remains to be elucidated.

Notably, genetic researches focused on the Sherpas were mainly performed in the fields of archaeology [11,12], molecular anthropology [7–9], medical genetics [3,5], and genetic genealogy [2,4], only a few forensic-related studies focused on the Nepal Sherpas were conducted based on the low-density short tandem repeats (STRs) [13,14]. Insertion and deletion polymorphisms (InDels), the second most abundant polymorphism across the human genome with low mutation rates and small amplicon lengths [15], combining the desirable features of both SNPs and STRs. The prominent properties enable this molecule to be genotyped via capillary electrophoresis (CE) platform. In addition, InDels cause similar levels of variation as SNPs and do not involve repetitive sequences, which can avoid stutter artifacts that may complicate the interpretation of STR profiles [16]. Reportedly, the 1000 Genomes Project (1KGP) has characterized ~3.6 million short InDels [17] and the analysis of high-coverage genome sequences of 54 diverse human populations included in the HGDP-CEPH panel identified ~8.8 million small InDels [18], which promoted the practical application of InDel loci in forensic investigations [19–21]. In the last few years, several commercialized InDel kits have been designed, such as the Investigator DIPplex kit developed by Qiagen [22–24], the AGCU InDel 50 kit [20,25,26], and the AGCU InDel 60 kit developed by AGCU ScienTech Incorporation. However, the practicability and efficiency of the updated version of the InDel amplification system (the AGCU InDel 60 kit) have not been validated in geographically/ethnolinguistically different populations.

Here, we first sought to explore the forensic characteristics and genetic relationships of Dingjie Sherpas and their neighbors on the Tibetan Plateau using the AGCU InDel 60 kit. The sampled Sherpas live in Dingjie County, Shigatse, Tibet Autonomous Region, which borders Sankhuwasabha and Taplejung Districts of Nepal to the south and the Sikkim State of India to the southeast and is one of the four counties that comprise the Qomolangma National Nature Preserve. Next, we comprehensively characterized the evolutionary genomic history of Nepal Sherpas based on one of the most representative modern and ancient genome-wide datasets from two analysis strategies: allele frequency-based shared alleles (PCA, ADMIXTURE, and *f*-statistics) and haplotype-based shared ancestry chunks (ChromoPainter and fineSTRUCTURE). We demonstrated the forensic effectiveness of our used InDel panel in highland adaptive Dingjie Sherpas. And then we explored the genetic structure, population admixture history of Dingjie and Nepal Sherpas based on the genetic variations of shared InDels and SNPs and found a strong genetic affinity between the studied Sherpas and Tibetans. Finally, we demonstrated that Nepal Sherpas derived their major ancestry from Neolithic Yellow River farmers and harbored a similar evolutionary history with Tibet Tibetans based on the shared drift in the *qpGraph*-based admixture graph and shared haplotypes in the fineSTRUCTURE-based chromosome painting.

2. Materials and methods

2.1. Sample preparation, DNA extraction, and genotyping

Bloodstain samples from 250 males and 378 females were collected

from Sherpas living in Dingjie County, Shigatse, Tibet Autonomous Region after receiving written informed consent (Fig. 1A). All participants were required to be the indigenous Sherpas whose ancestors have lived in Tibet Autonomous Region for at least three generations. This study was approved by the Ethical Committee of North Sichuan Medical College and Xiamen University, and all procedures were performed following the recommendations of the Declaration of Helsinki [27].

Human genomic DNA (gDNA) was extracted using the QIAamp DNA Mini Kit (QIAGEN, Germany) and quantified using the Qubit dsDNA HS Assay Kit (Thermo Fisher Scientific) on an Invitrogen Qubit 3.0 fluorometer following the manufacturer's recommendations. The gDNA was subsequently diluted to 2.0 ng/μL and stored at −20 °C until amplification. The control DNA 9947A and 9948 were adopted as reference samples.

Fifty-seven A-InDels and three sex-determination loci (rs76041101, rs199815934, and Amel) were amplified simultaneously using the AGCU InDel 60 kit on a ProFlex 96-Well PCR System (Thermo Fisher Scientific) following the manufacturer's protocol. The rough physical positions of all genotyped loci are provided in Fig. S1A. We used 25 μL multiplex PCR reaction volume, which included reaction mix (10 μL), primers (5 μL), U-Taq Enzyme (1 μL), sdH2O (8 μL) and template (1 μL) (AGCU ScienTech Incorporation, Wuxi, Jiangsu, China). ProFlex 3 × 32-Well PCR System (Thermo Fisher Scientific) was used to amplify the targeted InDels with the following amplification parameters: initial denaturation at 95 °C for 5 min for enzyme activation; then 28 cycles of 30 s at 94 °C, 1 min at 60 °C, 1 min at 62 °C; the final extension at 72 °C for 10 min. A mixture of 0.5 μL AGCU Marker SIZ-500, 12 μL HiDi™ formamide and 1 μL AGCU InDel 60 kit ladder were used to run capillary electrophoresis. The Applied Biosystems 3500xl Genetic Analyzer with the POP-4® Polymer (Thermo Fisher Scientific, Waltham, Massachusetts, USA) was utilized to separate the amplified products after the obtained mixture denaturation (95 °C for 3 min) and then immediately chilled on ice (3 min). The GeneMapper ID-X v.1.5 software (Thermo Fisher Scientific) was applied to allocate alleles.

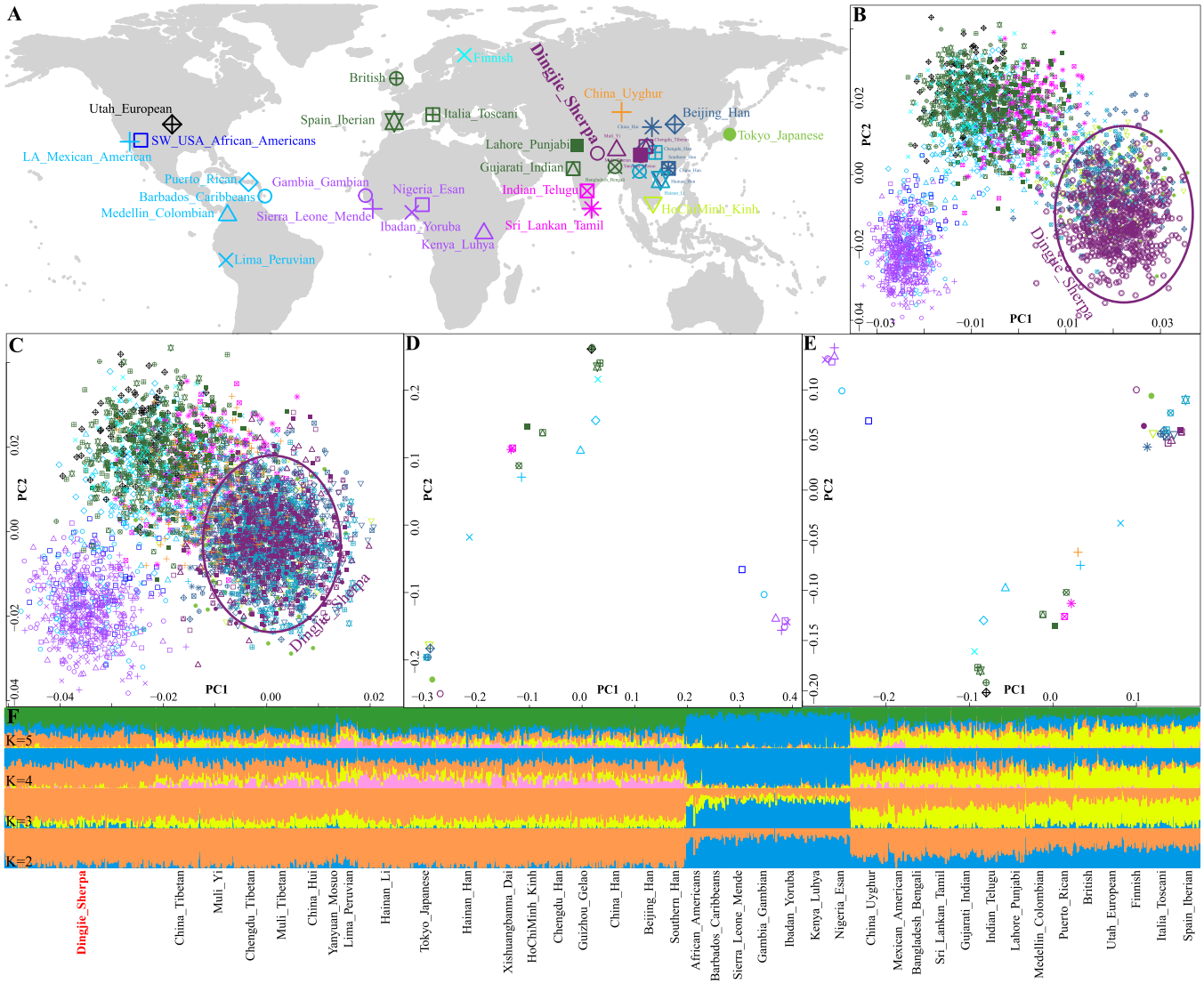

Fig. 1. Geographical information of all involved populations and genetic relationship between Dingjie Sherpas and reference populations based on InDel markers. (A) The geographical positions of all involved populations. (B) Principal component analysis (PCA) among 27 worldwide populations (Dingjie Sherpas and 26 populations from the 1KGP) based on the top two components extracted from the genetic variations of 57 InDels included in the merged 57 InDel dataset. (C) PCA results among 4850 individuals from 39 worldwide populations based on the genetic polymorphisms of 39 InDels included in the merged 39 InDel dataset. (D-E) PCA analysis results among worldwide populations based on the allele frequency distributions based on the two merged InDel datasets. (F) Model-based ADMIXTURE results showed the genetic compositions of 27 worldwide populations with the K values ranging from 2 to 5. The model with the predefined five ancestral sources had the lowest cross-validation error. (For interpretation of the references to colour in this figure, the reader is referred to the web version of this article.)

2.2. InDel and genome-wide SNP datasets composition

To reconstruct the genomic history of Sherpas and investigate the genetic relatedness between the studied Sherpas and reference populations, we generated three datasets for the population genetic analysis: The merged 57 InDel dataset, the merged 39 InDel dataset, and the publicly available Human Origins dataset (HO dataset). We first merged the newly generated InDel data of 57 A-InDels with the genotypes of 57 A-InDels in 2054 individuals from 26 worldwide populations retrieved from the 1000 Genomes Phase III release [17], which produced the ***merged 57 InDel dataset*** (3132 genotypes, Table S1). Furthermore, we combined the merged 57 InDel dataset with the genotype data of 47 A-InDels included in the AGCU InDel 50 kit from four Sinitic- (532 Hans and 129 Huis), five Tibeto-Burman- (432 Tibetans, 134 Yis, and 37 Mosuos), one Turkic- (154 Uyghurs), and two Tai-Kadai-speaking (84 Gelaos and 216 Lis) populations [26,28]. The overlapping loci of the AGCU InDel 50 kit and the AGCU InDel 60 kit were 39 A-InDels. Thus, we referred to the second dataset as the ***merged 39 InDel dataset***. The detailed information of reference populations is listed in Table S1 and the geographical locations of all involved populations are displayed in Fig. 1A. We then collected genome-wide SNP data from modern Eurasian populations genotyped using the Affymetrix Human Origins array [29–31] and ancient eastern Eurasian populations genotyped via whole-genome sequencing or targeted-sequencing [32–36] as the direct ancient ancestral sources to explore the genetic contribution and possible existing genetic continuity and admixture processes. We finally included 378 Tai-Kadai people from 29 populations, 91 Hmong-Mien people from 8 populations, 64 Sinitic people from 11 populations, 297 Tibeto-Burman people from 33 populations, 224 Austroasiatic people from 18 populations, 115 Austronesian populations from 13 populations, 30 Japanese, 6 Korean, 202 Altaic people from 24 populations and 205 ancient East Asians from 53 populations in the following population genetic history exploration (Table S1). The detailed information of the HO dataset is displayed in Allen Ancient DNA Resource (AADR) from Reich Lab (https://reich.hms.harvard.edu/allen-ancient-dna-resource-aadr-downloadable-genotypes-present-day-and-ancient-dna-data).

2.3. Data analysis

2.3.1. InDel-based analysis of genetic variations

The exact tests of linkage disequilibrium (LD) and Hardy-Weinberg equilibrium (HWE), as well as the estimations of observed heterozygosity (Ho) and expected heterozygosity (He) were conducted using the Arlequin v.3.5.2.2 [37]. The browser-based software of STR Analysis for Forensics (STRAF) was adopted to calculate the allele frequencies and forensic statistical parameters: discrimination power (DP), probability of exclusion (PE), polymorphism information content (PIC), and typical paternity index (TPI) [38]. We used the Gendist package built into the PHYLIP v.3.698 to estimate Nei's genetic distances [39]. We subsequently conducted genotype-based principal component analysis (PCA) using the genome-wide Complex Trait Analysis (GCTA) tool [40] and frequency-based PCA using the Multivariate Statistical Package (MVSP) v.3.22 [41]. The model-based clustering analysis was performed via the ADMIXTURE v.1.3.0 [42] with 10-fold cross-validation (cv) in 100 bootstraps to dissect the ancestral components of Dingjie Sherpas and reference East Asians. To infer the patterns of population splits and mixtures among multiple populations, we ran TreeMix v.1.13 [43] with migration events ranging from 0 to 10 to build a maximum likelihood tree. We also conducted the analysis of molecular variance (AMOVA) using the Arlequin v3.5.2.2 by grouping involved populations according to geography, language family, or ethnicity classifications.

2.3.2. Allele-based shared ancestry inferred from the high-density genome-wide data

Focused on the high-density SNP dataset, we used the smartpca program built into the EIGENSOFT v.6.1.4 [29] to carry out PCA analysis. East Asian modern populations from Altaic, Sino-Tibetan, Austronesian, Austroasiatic, Hmong-Mien, and Tai-Kadai language families were used to construct the basic genetic landscape of East Asians. And ancient East Asians were projected onto modern PCA space with two additional parameters (numoutlieriter: 0 and lsqproject: YES). PLINK v.1.9 [44] was applied to prune SNPs in strong linkage disequilibrium with the following parameters (–indep-pairwise 200 25 0.4). We then conducted model-based clustering analysis using the ADMIXTURE v.1.3.0 [42] with 10-fold cross-validation (–cv = 10) and predefined ancestry sources ranging from 2 to 20. We also built a bifurcating tree with the maximum likelihood algorithm based on the genome-wide allele frequency data using the TreeMix v.1.13 [45].

We used the *qp3Pop* program [29] of the ADMIXTOOLS to perform outgroup-f_3(Source, Sherpa; Mbuti) to explore the shared genetic drift and conduct admixture-f_3(Source1, Source2; Sherpa) to explore the status of genetic admixture and corresponding proxy ancestral sources. We further used the *qpDstat* package [29] of the ADMIXTOOLS to perform f_4-statistics of the forms f_4(Eurasian1, Eurasian2; Sherpa, Outgroup) and f_4(Eurasian1, Sherpa; Eurasian2, Mbuti) to investigate genetic affinity, admixture, and continuity with different comparative populations. The potential ancestral sources and corresponding admixture proportions were estimated using the *qpWave/qpAdm* packages [29] and graph-based admixture modeling was conducted using the *qpGraph* software [29].

2.3.3. Shared haplotype chunks inferred from the fineSTRUCTURE and ChromoPainter

Haplotype phasing was carried out employing the statistical-based method implemented in the SHAPEIT v2 [46] using default parameters. The phased haplotype data were used to paint the target individual's chromosome conditional on all included genetic material from other populations as the potential donor populations. The inference framework of chromosome painting of recipient chromosomes in the context of extensive donor chromosomes was used and conducted via a statistical algorithm instrumented in the ChromoPainter v2 and ChromoCombine v2 [47]. Individual-level model-based Bayesian clustering was carried out using the fineSTRUCTURE v4 [47] based on the output of ChromoPainter. The model-based likelihood and PCA based on our used coancestry matrix were estimated using the R packages of fineSTRUCTURELibrary, fineSTRUCTUREExample, ape, and XML. Finally, natural selection signals were explored via selscan based on the phased haplotype data and Identity by Descent (IBD) fragments were calculated using Refined IBD [46].

3. Results

3.1. Overview of forensic characteristics and population genetic structure inferred from the InDel panel

3.1.1. Allele frequency distribution and forensic statistical parameters

The present study has generated the first batch of InDel-based genotype data of 628 Dingjie Sherpas (Table S2). The allele frequency and corresponding forensic parameters are listed in Table S3. Two InDel loci (rs3064355 and rs5787309) showed significant departure from HWE after conducting Bonferroni correction (Table S3, $p < 0.0009$), five pairs of InDels (rs5787309 and rs34287950, rs67426579 and rs151335218, rs5897566 and rs76158822, rs35464887 and rs76158822, rs72085595 and rs34529638) located on the same chromosome showed LD after applying Bonferroni correction (Table S4, $p < 0.00003$). However, all these loci were in line with linkage equilibrium (LE) in lowland East Asians from the 1KGP and Tai-Kadai-speaking Li (unpublished data). Our results revealed a potential unique genetic architecture of Dingjie Sherpas, which would be further explored and discussed in the subsequent population admixture history reconstruction. Deviations from LE of autosomal STRs and SNPs in the highland East Asians were also observed in previous studies [48,49]. Thus, considering the patterns of

the observed LD in Dingjie Sherpas, we subsequently estimated the combined forensic parameters both on the pruned marker set (48 A-InDels) and all 57 A-InDels. Focused on the locus-based forensic parameters, the allele frequencies of insertions fluctuated within 0.1449–0.9076. The values of Ho and He ranged from 0.1656 (rs10607699) to 0.5526 (rs34529638) and 0.1678 (rs10607699) to 0.5004 (rs146875868). The values of DP and PE varied from 0.2921 (rs10607699) to 0.6319 (rs145941537) and 0.0211 (rs10607699) to 0.2378 (rs34529638). The measured value of PIC was in the range of 0.1536 (rs10607699) to 0.3750 (rs146875868), the minimum value of TPI was 0.5992 at rs10607699, whereas the maximum reached 1.1174 at rs34529638. For the combined forensic parameters calculated based on all 57 A-InDels, the total probability of discrimination power (TDP) value reached 1–1.5732E-22, and the combined probability of exclusion (CPE) value was approximately 0.9999. After pruning markers in LD, we reevaluated the combined forensic parameters in Dingjie Sherpas and we found this pruned InDel panel was also informative and polymorphic in Dingjie Sherpas (TDP: 1–5.1461E-19 and CPE: 0.9996).

3.1.2. *Genetic similarity and differences between studied Sherpas and reference populations*

Genetic affinities between Dingjie Sherpas and reference worldwide populations were explored via Nei's genetic distances, PCA, ADMIXTURE, and TreeMix analyses. The estimated 57-InDel-based genetic distances indicated that Dingjie Sherpas were genetically close to East Asian reference populations (Table S5), including northern and southern Han Chinese, Yunnan Dais, and Japanese. The 39-InDel-based genetic distances revealed that studied Sherpas were genetically close to previously published Tibeto-Burman-speaking populations residing in the Tibetan Plateau (Table S6, Yanyuan Mosuo: 0.0248, Muli Tibetan:

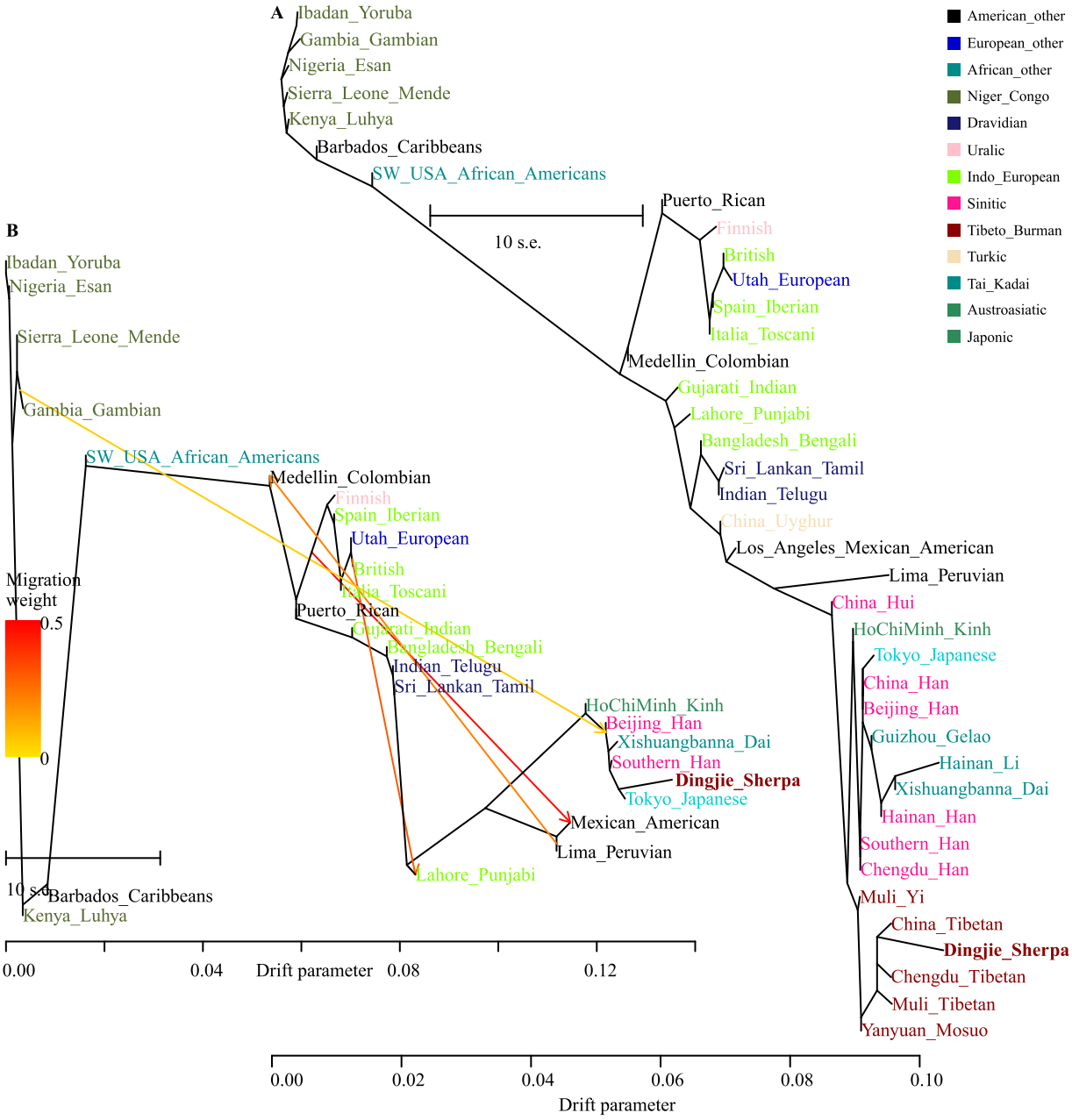

Fig. 2. Phylogenetic relationship among worldwide populations. (A) TreeMix-based phylogenetic relationship among 39 populations with zero migration event. (B) TreeMix-based phylogenetic relationship among 27 populations with four migration events. Populations were color-coded based on their geographical affinity and linguistic affinity.

0.0269, and Muli Yi: 0.0298). The 57-InDel-genotype-based and 39-InDel-genotype-based PCAs (Fig. 1B–C and S2A–B) only identified three rough clusters: African cluster, East Asian cluster, and a mixed cluster including European, American, and South Asian populations. Generally, the studied Sherpas were overlapped with Sino-Tibetan-speaking individuals. The frequency-based PCAs (Fig. 1D–E and S2C–D) identified four geography-related clusters: African cluster, European cluster, South Asian cluster, and East Asian cluster, the American populations were scattered along the PC2 or PC3. Specifically, the Dingjie Sherpas showed a close genetic relationship with East Asian reference populations, especially Sino-Tibetan-speaking populations and Tokyo Japanese.

We further performed model-based ADMIXTURE analysis to roughly model the ancestral components of the Dingjie Sherpas (Fig. 1F). At K = 2, we observed two distinct ancestral components deriving from non-African (orange component) and African populations (blue component). At K = 3, the yellow ancestral component occurred in Chinese Uyghur, South Asian, American, and European populations. As the K values increased, two additional ancestries, including East Asian-dominant ancestral component (showed in green) and non-Dingjie_Sherpa-related ancestral component (showed in pink), were observed. Obviously, ancestral compositions of the Dingjie Sherpas were similar to other Tibeto-Burman speakers. Furthermore, we built maximum likelihood topologies to model the gene flow between target populations and proxy sources. The 39-InDel-based phylogenetic tree demonstrated that the Dingjie Sherpas grouped with Tibeto-Burman-speaking populations and showed strong genetic affinities with Tibetans (Fig. 2A and Fig. S3A). The 57-InDel-based topology revealed that the Dingjie Sherpas clustered with East Asian reference populations and showed a relatively close genetic relationship with Tokyo Japanese (Fig. 2B and Fig. S3B). Moreover, we observed gene flow events from Gambia_Gambian-related ancestry into East Asians, from Lima_Peruvian-related ancestry into Medellin_Colombian, from European-related ancestry into Mexican_American, and from Utah_European-related ancestry into Lahore_Punjabi. To explore the factors that might have played roles in shaping the InDel diversity among geographically, linguistically and ethnically diverse populations, we conducted a series of AMOVA analyses and found that the among-group variations were slightly higher when dividing worldwide populations into geographically different groups, compared with that among linguistically different groups (Table S7). We also observed that the among-group variations were slightly higher when dividing East Asian populations into ethnically different groups, compared with that among language-related and altitude-related groups.

3.2. Genetic origins of Sherpa people from the Tibetan Plateau inferred from genome-wide SNP data

3.2.1. Genetic structure and affinity inferred from PCA, ADMIXTURE, and treemix-based phylogeny

Although the forensic genetic markers, such as STRs and InDels, possessed powerful statistical performance in the personal identification and parentage testing, the limited resolution (the number of the genetic markers is low) in deep ancestry dissection has hindered its application in molecular anthropology and population genomic history reconstruction. We also observed the possible differentiated genetic history of Dingjie Sherpas and lowland East Asians based on the InDel-based forensic parameter estimation (LD). Thus, to gain new insight into the genetic makeup of Sherpa highlanders, we retrieved genomic data of 597,569 genome-wide SNPs of Nepal Sherpas from one of the most comprehensively representative datasets, which includes all publicly available ancient and modern Eurasian genomes from the AADR [30,31,34,36,50]. We first explored the genetic admixture history based on the shared independent alleles and shared haplotype chunks consisting of the successive SNPs. We also provided the direct genetic relationship between Nepal Sherpas and their geographically close ancient ancestors from the Tibetan Plateau (Chokhopani, Samdzong and Mebrak [33]), Yellow River millet farmers (Qijia Jinchankou and Lajia people [36], Henan Yangshao and Longshan people [36], and Shandong Houli people [35]) and other ancient East Asians [34]. Raw data of 597,569 variants in 1612 individuals from 191 populations were used to conduct a formal test (*f*-statistics) and LD-pruned data of 249,785 variants were used to conduct descriptive analyses.

To explore the general patterns of genetic relatedness between the Nepal Sherpas and reference East Asian populations, we first performed PCA analyses based on all East Asian populations or all Sino-Tibetan subpopulations. All published Sino-Tibetan (Tibeto-Burman and Sinitic) people were our focus to characterize the genetic history and relationship of Sherpas and their neighbors (Fig. 3A). The HO-based East Asian PCA (Fig. 3B and S4) identified two genetic clines: the southern cline consisting of Austroasiatic-, Tai-Kadai-, Austronesian-, Tibeto-Burman-(mainly in Southeast Asia), and Hmong-Mien-speaking populations, and the northern one consisting of the Nepal Sherpas, Sinitic-, Japonic-, Koreanic-, Tibeto-Burman- (mainly in the Tibetan Plateau), Tungusic-, Mongolic-speaking populations, and ancient individuals from Russia and Mongolia. We found that Tibeto-Burman people from the Tibetan Plateau and mainland Southeast Asia were separated into two clusters (purple color-coded), and the Nepal Sherpas clustered with Tibetan populations. The Sino-Tibetan-related PCAs (only including 361 people from 44 Sino-Tibetan-speaking populations) based on the top three components revealed obvious population subclusters within Sino-Tibetan-speaking populations: Newar-related, Karen-related, low-altitude Sino-Tibetan-related, and high-altitude Tibeto-Burman-related genetic clines (Fig. 3C–E). The Nepal Sherpas showed close genetic relationships with Tibet Tibetans (especially Tibetan_Shigatse and Tibetan_Shannan).

Clustering patterns inferred from the ADMIXTURE results among Sino-Tibetan-speaking populations revealed multiple sources of their ancestral components (Fig. 3F). Under the model of four predefined ancestral sources with the lowest cross-validation error (K = 4), we found four specific ancestral components that were respectively maximized in Tibet Tibetans and Nepal Sherpas (yellow), Sila and LoLo (blue), Karen-related populations (pink), and Southeast Asian Lahu-related populations (orange). Model-based clustering of five predefined ancestral sources confirmed the Sila-specific ancestry, which also occurred in relatively low-altitude Tibetan populations (except Tibetan_Yunnan), Qiangs, Lisus, PhuLas, Vietnam Lahus, Congs, and HaNhis. Model-based admixture results showed a similar genetic composition between Sherpas and Tibetans, suggesting their close genetic relationship and shared common ancestral evolutionary history. The maximum likelihood trees with migration events indicated that the Nepal Sherpas was aggregated with high-altitude Tibetan populations to form the high-altitude-related clade, which then clustered with Qiangs and relatively low-altitude Tibetan populations (Fig. S5), which further confirmed the similar genetic composition of Sherpas and Tibetans. Here, we also identified evidence of the documented gene flow from Newar_Nepal to Southeast Asian Tibeto-Burman-speaking populations, from Vietnam_Lahu to the common ancestor of LoLo, PhuLa and Lahu people, from Tamang_Nepal to the common ancestor of Sherpas and central and southern Tibet Tibetans (Tibetan_Lhasa, Tibetan_Shannan, and Tibetan_Shigatse), from Han_Guangdong to Han_Fujian, which suggested that extensive genetic admixture occurred among Sino-Tibetan-speaking populations shaped these genetically attested diversities.

3.2.2. Shared alleles between Sherpas and other East Asians inferred from *f*-statistics

To explore the inter-population genetic relatedness between the studied Nepal Sherpas and other East Asians, we estimated pairwise Fst genetic distances among 44 Sino-Tibetan-speaking groups (Table S8). We found that the Nepal Sherpas had close genetic relationships with Tibetans on the Tibetan Plateau (Tibetan_Shannan: 0.0036,

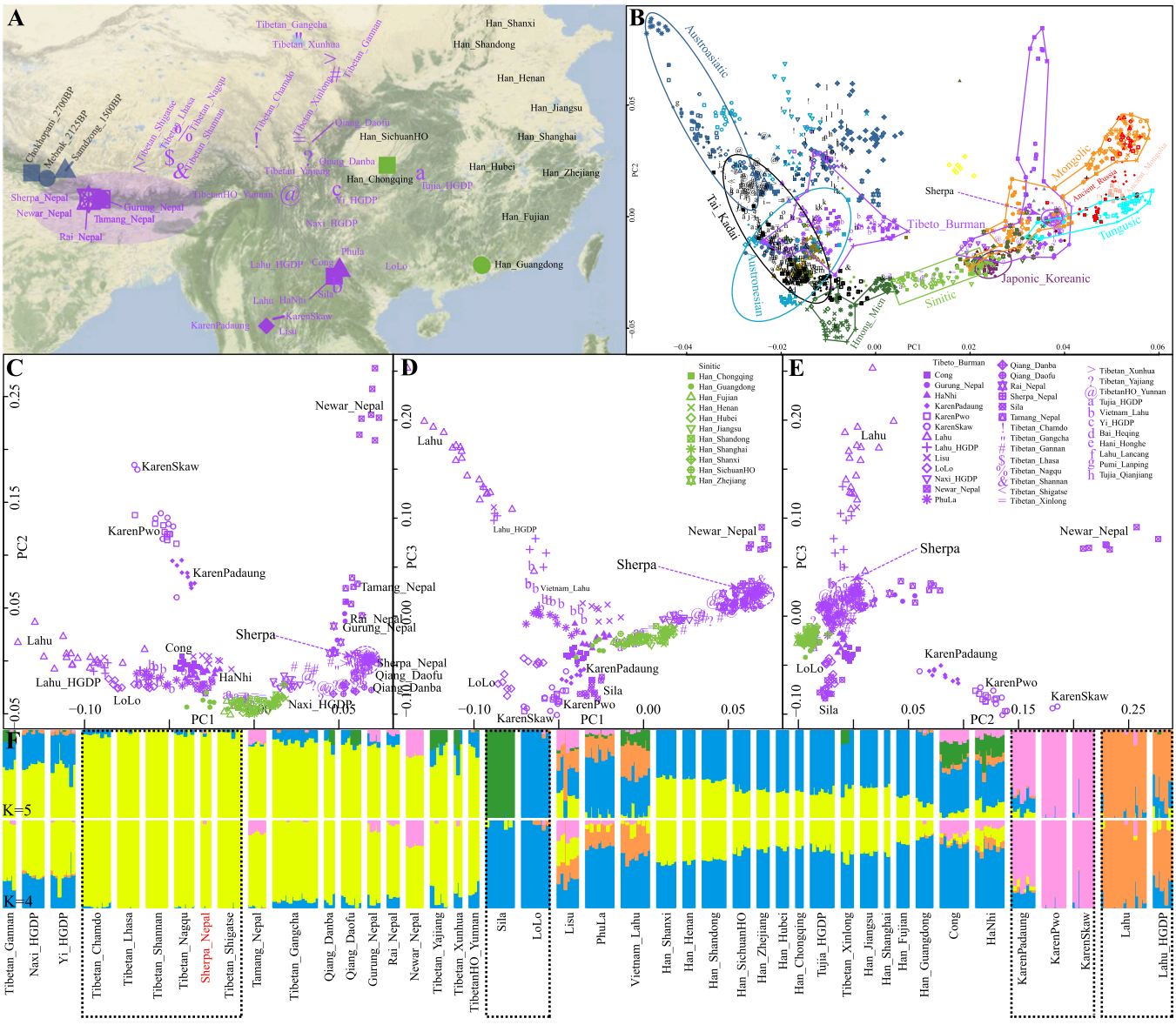

Fig. 3. Genetic structure and geographical positions of populations used in the population genomic analysis based on the genome-wide data. (A) Geographical positions of 44 Sino-Tibetan populations and three Nepal ancient populations, five Nepal populations (including one Sherpa population) were emphasized. (B) PCA analysis among 1612 modern and ancient East Asians based on the Human Origins dataset, ancient populations from Nepal, China, Japan, Mongolia, and southern Siberia were projected onto the top two components. The detailed population labels were presented in Fig. S4. (C–E) PCA results among 361 Sino-Tibetan-speaking individuals from 44 populations. (F) Ancestry composition of 361 Sino-Tibetan people based on unlinked SNPs. (For interpretation of the references to colour in this figure, the reader is referred to the web version of this article.)

Tibetan_Lhasa: 0.0040, Tibetan_Shigatse: 0.0043, Tibetan_Nagqu: 0.0047, and Tibetan_Chamdo: 0.0059). We subsequently measured the shared genetic drift (Table S9) via the outgroup f_3-statistics of the form f_3(Sherpa_Nepal, Reference modern and ancient populations; Mbuti), in which larger f_3 values denoted more shared ancestry and a closer genetic relationship between target and reference populations. We observed that the Nepal Sherpas possessed the most shared alleles with Qiang groups (Qiang_Danba: 0.2894 and Qiang_Daofu: 0.2886), high-altitude Tibetan populations (Tibetan_Lhasa: 0.2886, Tibetan_Nagqu: 0.2885, Tibetan_Shigatse: 0.2884, Tibetan_Shannan: 0.2883, and Tibetan_Chamdo: 0.2882), as well as the middle/late Neolithic ancient populations (Wuzhuangguoliang_LN: 0.2902, Jinchankou_LN: 0.2888, and Miaozigou_MN: 0.2883). The identified genetic affinity suggested the close genetic correlations between Sherpas and Tibetans, as well as between Sherpas and Yellow River miller farmers. We next performed admixture f_3-statistics of the form f_3(Source1, Source2; Target populations) to identify plausible ancestral sources of the Nepal Sherpas (Table S10). Surprisingly, we did not observe significant negative Z-scores (< -3), which denoted that there was no evidence to support that Nepal Sherpas were descended from a recent genetic admixture event of our used modern and ancient source1 and source2 or they underwent strong genetic drift recently. No-fitted ALDER-based curves and models have further confirmed the relatively isolated genetic structure of Nepal Sherpas. We found that only the combination of (Mebrak_2125BP, BanChiang_IA) could produce relatively negative Z-scores (< -1.5), indicating their close genetic relationship.

Four population-based analyses can provide strong power to validate our proposed topologies and potential existing gene flow events and gene flow directions [29]. Thus, we next conducted *qpWave* analyses to validate the genetic continuity between Nepal ancients and Sherpas, and Sherpas and Tibetans using one set of distinct outgroups (Fig. 4A) and more powerful outgroups (Fig. 4B). P values of pairwise qpWave larger

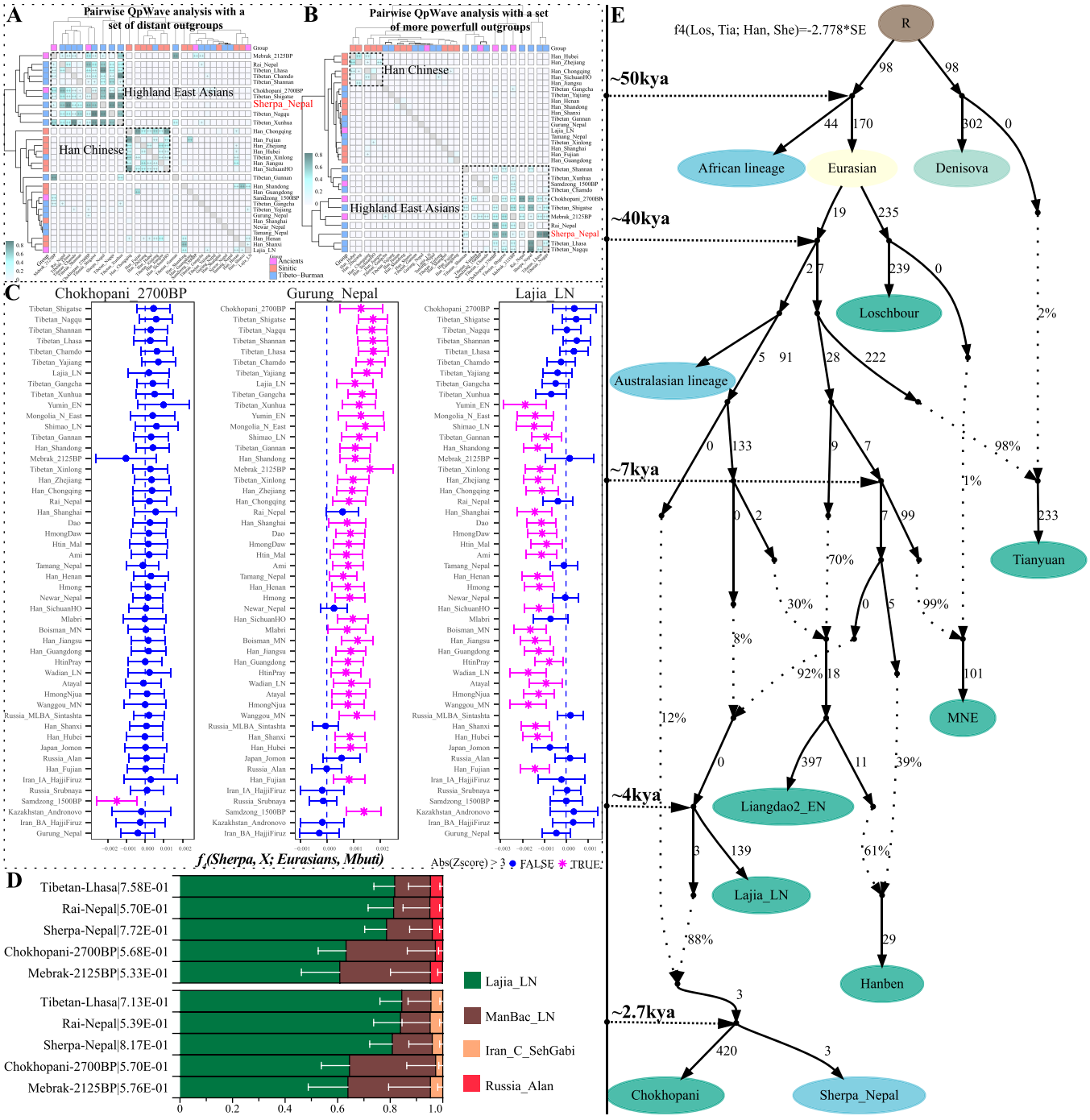

Fig. 4. Population genetic analysis results based on the shared alleles in the combination of *f*-statistics. (A-B) Pairwise qpWave analysis based on distant outgroup set (Mbuti, Russia_Ust_Ishim, Russia_Kostenki14, Papuan, Australian, Mixe, Russia_MA1_HG, Onge, Atayal, Mongolia_N_East, Bianbian_EN, Wuqi_EN, Qihe_EN, Liangdao2_EN, Yumin_EN, Russia_EBA_Yamnaya_Samara and Xiaowu_MN) and more powerful outgroup set (Mbuti, Russia_Ust_Ishim, Russia_Kostenki14, Papuan, Australian, Mixe, Russia_MA1_HG, Onge, Atayal, Mongolia_N_East, Bianbian_EN, Wuqi_EN, Qihe_EN, Liangdao2_EN, Yumin_EN, Yamnaya_Samara_EBA and Xiaowu_MN) displayed genetic heterogeneity and homogeneity. (**C**) Four-population-based *f*-statistics of the form f_4(Sherpa, Chokhopani_2700/Gurung Nepal/Lajia_LN; Eurasians, Mbuti) showed the genetic continuity or heterogeneity between Sherpas and their ancestors and modern neighbors. (**D**) QpAdm-based admixture results of modern and ancient highland East Asians used nine outgroups (Mbuti, Russia_Ust_Ishim_HG, Russia_Kostenki14, Papuan, Australian, Mixe, Russia_MA1_HG, Onge, and Atayal). (**E**) Deep evolutionary history of modern Sherpas. The dotted line denoted the reconstructed admixture events and the corresponding admixture proportions were marked on it. The shared genetic drift of branch length was marked using 1000 *f_2. The time-scale was marked based on the recalibrated years of ancient samples, which was the simplified scale.

than 0.05 denoted that two left populations could be explained via one common ancestral source without additional gene flows compared with our used outgroups (distant outgroups: Mbuti, Ust_Ishim, Kostenki14, Papuan, Australian, Mixe, MA1, Onge, Atayal and Samara Yamnaya; additional outgroups in more powerful outgroups: Mongolia_N_East, Bianbian_EN, Wuqi_EN, Qihe_EN, Liangdao2_EN, Yumin_EN, and Xiaowu_MN). Genetic continuity was identified between Nepal Sherpas and 2700-year-old Chokhopani and 2125-year-old Mebrak. Besides, genetic similarities between Sherpas and Tibetans from Shannan, Shigatse, Lhasa, and Naqu cities were also confirmed via pairwise qpWave

analysis (Fig. 4B). Additionally, we performed a series of symmetry-f_4-statistics of the form f_4(Eurasian1, Eurasian2; Sherpa_Nepal, Mbuti) and found that the Nepal Sherpas shared more alleles with Tibetan groups, northern Han Chinese (Shanxi, Henan, and Shandong), ancient populations (Samdzong_1500BP, Mebrak_2125BP, and Chokhopani_2700BP) on the Tibetan Plateau, and middle/late Neolithic ancient northern East Asians when compared with modern and ancient southern East Asians, Southeast Asians, South Asians, and ancient individuals not mentioned above, such as the most negative Z-score of f_4(Sintashta, Tibetan_Shigatse; Sherpa_Nepal, Mbuti) =−62.023*SE. To clarify if additional ancestral sources contributed to the gene pool of Nepal Sherpas, affinity-f_4-statistics of the form f_4(Ancestral source candidates, Sherpa_Nepal; Reference populations, Mbuti) were conducted. When ancient individuals on the Tibetan Plateau and modern Tibetan groups were adopted as the ancestral sources, no significant negative Z-scores were observed in f_4(High-altitude Tibetan populations/Mebrak_2125BP/Chokhopani_2700BP, Sherpa_Nepal; Reference population, Mbuti), which demonstrated that high-altitude Tibetans, Mebrak_2125BP, and Chokhopani_2700BP formed one clade with Nepal Sherpas and these ancient populations might be the direct ancestors of the Nepal Sherpas (Fig. 4C). Specifically, additional gene flow related to ancient Yellow River millet farmers and modern northern East Asians into the Nepal Sherpas was identified in f_4(Samdzong_1500BP, Sherpa_Nepal; Tibetan_Yajiang/HmongDaw/Han_Shandong/Mongolia_N_East/Dao/Mlabri/Tibetan_Gannan/Shimao_LN, Mbuti), f_4(Tibetan_Gangcha/Tibetan_Xunhua/Tibetan_Gannan, Sherpa_Nepal; Tamang_Nepal/Rai_Nepal/Gurung_Nepal, Mbuti), and f_4(Tibetan_Gannan, Sherpa_Nepal; Newar_Nepal/Mongolia_N_East/Yumin_EN, Mbuti). The Z-scores of f_4(Ancient East Asians, Sherpa_Nepal; Non-Tibetan-related reference population, Mbuti) revealed that the combinations of (Mongolia_N_East, Sherpa_Nepal; Tamang_Nepal/Rai_Nepal/Gurung_Nepal/Newar_Nepal/HtinPray, Mbuti) and (Yumin_EN, Sherpa_Nepal; Tamang_Nepal/Rai_Nepal/Atayal, Mbuti) could produce significant negative Z-scores.

3.2.3. Estimation of ancestral composition and admixture proportion

Considering the patterns of admixture events that were observed in the *f*-statistics results, we modeled the minimum number of proxy ancestral populations using the *qpWave* and estimated the corresponding ancestry proportions using the *qpAdm*. The two-way admixture models showed that the Nepal Sherpas could be modeled as an admixture of 0.246 ManBac_LN-like ancestry and 0.754 Lajia_LN-like ancestry (p_rank1: 0.0503), 0.543 ManBac_LN-like ancestry and 0.457 Mongolia_N_East-like ancestry (p_rank1: 0.0565), or 0.036 Sintashta-like ancestry and 0.964 Lajia_LN-like ancestry (p_rank1: 0.4077). Compared with the ancient populations on the Tibetan Plateau, the Nepal Sherpas possessed more Lajia_LN-like ancestral components. While no obvious differences in the proportion of Lajia_LN-like ancestry were observed between the Nepal Sherpas and high-altitude Tibetan groups. The results of three-way admixture models indicated that the Nepal Sherpas could be fitted as having 0.786 Lajia_LN-like ancestry, 0.174 ManBac_LN-like ancestry, and 0.040 Alan-like ancestry, or having 0.808 Lajia_LN-like ancestry, 0.152 ManBac_LN-like ancestry, and 0.040 Iran_C_SehGabi-like ancestry (Fig. 4D). Similarly, we found that the studied Nepal Sherpas, high-altitude Tibetans, and Nepalese populations harbored more Lajia_LN-like ancestry than ancient groups on the Tibetan Plateau. We finally used the *qpGraph* to further test if the phylogeny-based model with population splits and mixtures could provide a reasonable fit to the data using the combination of all f_4-statistics. We obtained the best-fitting model for the Nepal Sherpas with an absolute Z-score of 2.778 (Fig. 4E). The optimal model revealed that the Nepal Sherpas had ~88% ancestry from the ancestor of Lajia_LN and ~12% ancestry from the Australasian lineage. Ancestral admixture sources and corresponding admixture proportions inferred from *qpAdm* and *qpGraph* consistently supported Nepal Sherpas had a similar evolutionary history as Tibetan people, harboring the major ancestry of East Asians related to Neolithic millet farmers from the Yellow River Basin.

3.3. Finer-scale population demographic history and natural selection signals inferred from the shared haplotype

FineSTRUCTURE model-based analyses based on the linked coancestry matrix of two different datasets were conducted to explore the population structure of Nepal Sherpas and their neighbors: one large dataset consisted of 361 people from 44 Sino-Tibetan populations and the other small one included 121 people from 18 Sinitic and highland Tibeto-Burman-speaking populations. Two-dimensional plots based on the top two components of haplotype-based PCA showed a clear separated position between Karen and other Sino-Tibetan people (Fig. 5A). The third component separated Lahus from other included reference populations. The fourth and fifth components separated Sila, Lahu, and Lolo populations from other people (Fig. 5B–J). These separated populations were representative ancestral sources that harbored the maximized ancestry proportion revealed in the model-based ADMIXTURE. Overall, Nepal Sherpas were clustered together with Tibetans. Heatmap and the corresponding clustering patterns based on the coancestry matrix showed five major branches (Fig. 5K): Han Chinese population branch was clustered closely with lowland Tibetan-Burman branch and then grouped with Lahu branch; Highland Tibetan-Burman speakers from the Tibet Autonomous Region and Nepal were separated. We also investigated interactions between populations by analyzing the number and length of shared IBD segments. The large and long shared IBD segments showed wide interaction and/or recent common ancestor sharing of the Nepal Sherpas with Tibetan groups, Tamang and Rai from Nepal, and Qiang groups (Fig. 5L).

The patterns of population structure among the 18 populations revealed from the fineSTRUCTURE-based PCA showed a clear population substructure consistent with their ethnic or geographical divisions. PCA plots from the first two components showed five genetic homogeneous clusters, including Newar_Nepal, Taman_Nepal and Gurung_Nepal, Sherpa_Nepal and Rai_Nepal, Yis, and Han Chinese (Fig. S6A). The third component separated the Rai_Nepal and Sherpa_Nepal, and the fourth component separated northern and southern Han Chinese populations. Generally, Nepal Sherpas possessed closer relationships with Tamang, Rai, and Gurung groups compared with others (Fig. S6B–C). These genetic affinities between Nepal Sherpas and Rais and Tamangs were further confirmed via the heatmap and clustering patterns of pairwise coincidence (Fig. S6D). Nepal populations were clustered into genetically homogeneous populations consistent with their ethnic divisions, and Han Chinese were divided into northern, central, and southern groups. Lowland Yis were clustered separately (Fig. S6D). Generally, the combined fine-scale clustering patterns based on the coancestry matrix in ChromoPainter and fineSTRUCTURE analyses among Tibeto-Burman people (Fig. 5) and Sino-Tibetan people (Fig. S6) not only illuminated that finer-scale population substructures within Sino-Tibetan-speaking populations but also indicated that the Nepal Sherpas had the closest genetic relatedness with geographically close Tibetans, such as Tibetan_Shigatse in Fig. 5.

We also explored the potential natural-selection signals in the Nepal Sherpas based on the phased haplotypes using the integrated haplotype score (iHS), nSL, and integrated haplotype homozygosity pooled (iHH12). Positive selection signals were observed on chromosomes 1, 5, 6, 9, 10, 12, and 22 (Figs. S6E and S7), these observed SNPs were correlated with Cysteine and Glycine-Rich Protein 1 (CSRP1), UDP Glycosyltransferase Family 3 Member A1 (UGT3A1), Major Histocompatibility Complex, Class I, B (HLA-B), GLIS Family Zinc Finger 3 (GLIS3), Supervillin (SVIL), Carboxypeptidase M (CPM), and Cat Eye Syndrome Chromosome Region, Candidate 5 (CECR5), respectively. Recent genetic studies have suggested that the EPAS1 and EGLN genes underwent the adaptive admixture or adaptive archaic introgression in the high-altitude groups [51,52]. However, we did not identify obvious

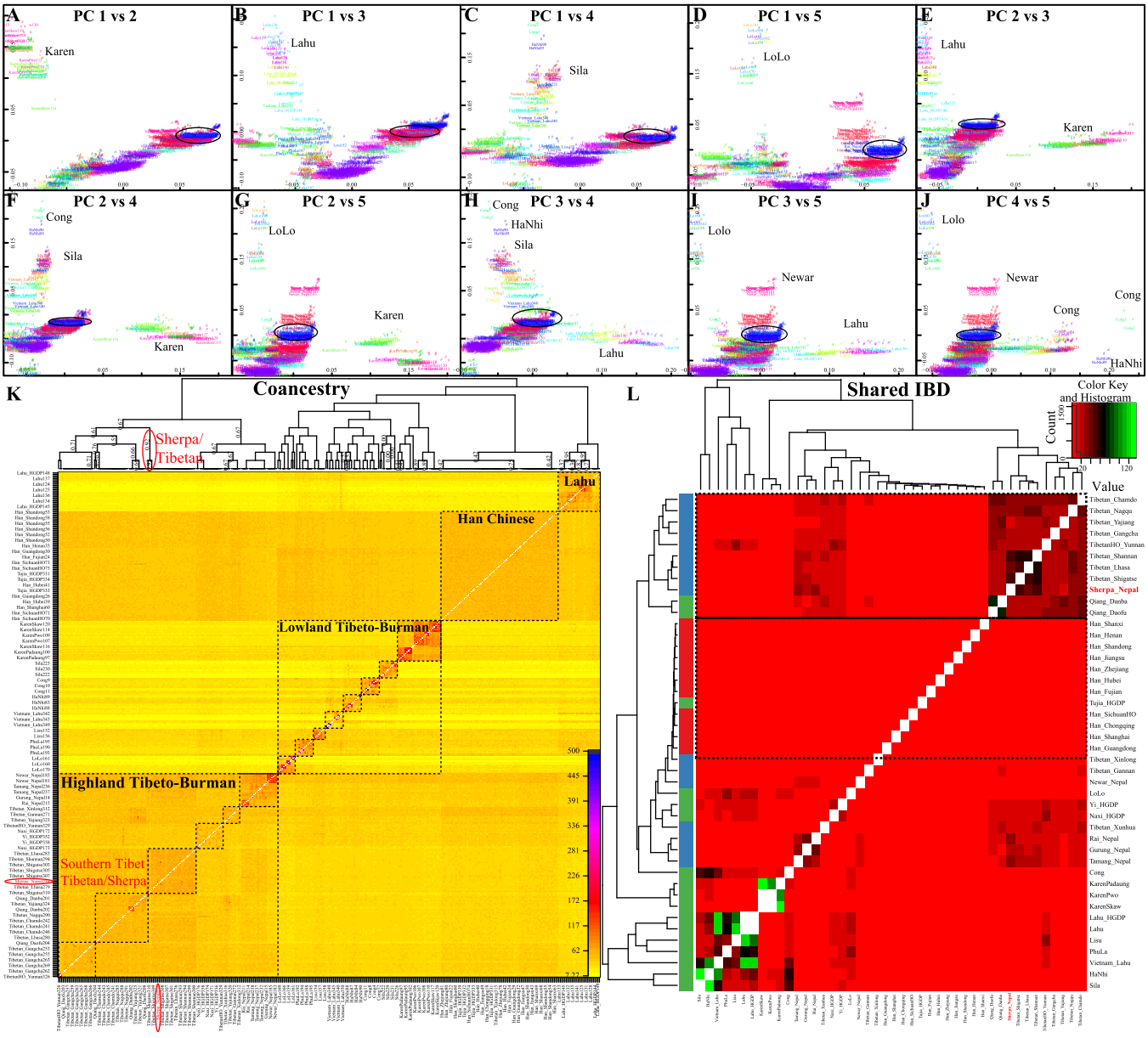

Fig. 5. Shared ancestry among 44 Sino-Tibetan-speaking populations based on shared ancestry chunks. (A–J) IBD-based genetic relationships were reconstructed based on the top five components. (K) Coancestry matrix inferred from the fineSTRUCTURE-reconstructed IBD numbers. (L) IBD fragments among 44 populations inferred from the Refined IBD analysis.

natural selection signals of the putative high-altitude adaptive variants localized in chromosomes 1 and 2 in Nepal Sherpas as lower iHS values observed in Fig. S6E. Thus, a further whole-genome sequencing project based on geographically different Sherpa people with larger sample size and single-molecule real-time sequencing technology should be conducted focused on Sherpa people in the next step, as the genetic findings in Tibetans [53].

4. Discussion

The Tibetan Plateau is widely known as the Third Pole of the world and is characterized by extreme climatic and environmental conditions [54]. As one of the most challenging environments ever occupied by anatomically modern humans (AMHs) [55], the timing and mechanisms of the colonization of Tibetan highlanders (Tibetans, Sherpas, Monpas, Lhobas, and Dengs) have attracted wide attention from various research fields, including genetics, Palaeo-genomics, archeology, anthropology, and geology [3,9,34,56–58]. Over the last few decades, with the deepening of researches in these fields, remarkable progress has been made in the exploration of the demographic history of Tibetans [34,56,58,59]. However, as one of the high-altitude adaptive populations living around the Himalayas, the demographic history and fine-scale genetic background of Sherpas need to be comprehensively explored. The previous genetic analyses focused on the peopling of Tibetan Plateau mainly based on allele-based shared ancestry or only modern genomes without the comprehensively comparative analysis based on both modern and ancient references [9,60,61]. Our recent population genetic analyses have started to directly explore the genetic contribution from ancient East Asians to modern Tibet Tibetans based on ancient genomes, but the genetic history of Sherpa people and their forensic characteristics still need to be further studied [48,62,63]. In the present study, we first performed an InDel-based forensic and population genetic study to explore the genetic relatedness between Dingjie Sherpas and worldwide reference populations. Furthermore, we conducted genome-wide

SNP-based genetic analyses among 191 modern and ancient populations to dissect the fine-scale genetic structure of Nepal Sherpas based on the shared patterns of independent SNPs or reconstructed haplotypes.

The results of tests of HWE revealed that two InDel loci showed a departure from HWE after applying Bonferroni correction, which may be attributed to the population substructure, purifying selection, inbreeding, or copy number variation [64], which was consistent with high IBD fragments within Sherpa people. The allele frequency distributions and corresponding forensic parameters of 57 A-InDels in the Dingjie Sherpas indicated that several InDels (such as rs10607699, rs145577149, and rs66477007, among others) were not polymorphic in the targeted Dingjie Sherpa, but the values of TDP and CPE obtained based on the pre-LD-pruned InDel panel and LD-pruned InDel panel demonstrated that the AGCU InDel 60 kit is suitable for forensic individual identification in the Dingjie Sherpas. We should note that geographically different population data and forensic parameter estimation in geographically different Sherpa people should be conducted to explore their homogeneity or heterogeneity of forensic features and population genetic structures, as some population substructures were identified via the mtDNA evidence [5,6]. Recent genetic analyses based on the whole-genome sequencing data or genome-wide array-based SNP data also showed the significant genetic differentiation between lowland East Asians and highland East Asians, and also identified population stratification within geographically different ethnic groups [9,48,61,65]. Considering the highland East Asian in the Tibetan Plateau is rich ethnolinguistic diversity, here, we emphasized the importance to develop the InDel-based forensic amplification systems with high polymorphic and informative for both highland and lowland East Asians.

The InDel-based close genetic affinity between the Dingjie Sherpas and high-altitude Tibeto-Burman-speaking populations (especially Tibetans) supported that they originated from the same ancestral populations or they obtained substantial gene flows from Tibetan highlanders [4,5,9]. The ancestral composition obtained from model-based ADMIXTURE analysis revealed that the European-dominant and Li-dominant ancestries might be absent in the gene pool of the Dingjie Sherpas, indicating that its ancestral components were mainly derived from northern East Asians. Additionally, we did not observe gene flow events from proxy sources into the Dingjie Sherpas, which suggested that more InDel-based genetic profiles from geographically/ethnically/linguistically different populations and genome-wide data from worldwide modern and ancient populations are indispensable to explore the fine-scale genetic makeup of the Dingjie Sherpas.

Additionally, The patterns of genetic clustering inferred from the genome-wide SNP-based PCA, ADMIXTURE, and TreeMix analyses revealed the population substructures among Sino-Tibetan-speaking groups in East Asia, Southeast Asia, and South Asia, and also indicated that the targeted Nepal Sherpas had close genetic affinities with high-altitude Tibetans and Nepal ancient individuals but had relatively distant genetic relationships with modern Nepalese populations, which supported the genetic findings that geographically different Tibeto-Burman-speaking groups have experienced complex and extensive admixtures [7,8,34] and the Nepal Sherpas possessed limited gene flow from surrounding Nepalese populations [7]. Furthermore, these genetic observations highlighted the long-term stability of the Tibetan highlanders' genetic makeup [33]. Genetic similarity and continuity between the Nepal Sherpas and ancient populations and Tibetan groups on the Tibetan Plateau were further confirmed by their excess shared alleles revealed in *f*-statistics, *qpWave/qpAdm*-based admixture models, and *qpGraph*-based evolutionary topology. Moreover, we found that Lajia_LN-related ancestry contributed the most to the gene pool of Nepal Sherpas and modern Tibetan populations, suggesting that Neolithic millet farmers played a pivotal role in the formation of modern Sherpas and Tibetans, which was consistent with previous genetic findings [34,36,48]. The haplotype-based genetic observations further confirmed the recent common ancestor sharing between the Sherpas and Tibetans. Additionally, we observed several positive selection signals not associated with high-altitude adaptation across the genomes of Nepal Sherpas, which may result from the limited sample sizes.

5. Conclusion

Sherpa people are one of the officially unrecognized ethnolinguistic groups, like the Gejia, Dongjia, and Xijia in Guizhou province. Here, we first generated the first batch of InDel-based population reference data of Dingjie Sherpas, which could promote the further potential for InDel-based forensic applications in highland Tibetan Plateau populations. The estimated allele frequencies and forensic parameters of different InDel combinations suggested their potential application in Dingjie Sherpas as the observed highly informative and polymorphic features. Besides, we provided genetic evidence from both InDels and genome-wide SNPs and suggested that Dingjie and Nepal Sherpa people shared a common admixture and evolutionary history with core Tibetans and other Tibeto-Burman-speaking populations from surrounding regions. Additionally, we identified the genetic affinity between Nepal Sherpas and modern Tibetans and ancient populations on the Tibetan Plateau, and this genetic homogeneity among them was further confirmed via the PCA, ADMIXTURE, Fst, and allele-sharing profiles in *f*-statistics and haplotype-based shared ancestry fragments. Finally, consistent with their modern and ancient neighbors (such as Chokhopani and Tibetans), Nepal Sherpas possessed strong lowland East Asian affinity and was formed via the admixture of major ancestry related to the Yellow River Basin Neolithic millet farmers and minor from the indigenous ancestry related to the early East Asians. In short, as the identified unique genetic admixture in Sherpas, a further project focused on the whole-genome sequencing data in geographically different Sherpas should be conducted to characterize the finer-scale landscapes of the evolutionary history, biological adaptation, and archaic adaptive introgression of Sherpas, as well as reconstruct the genome-wide atlas of forensics-related STRs, SNPs, InDels, microhaplotypes, and copy number variants.

Acknowledgments

This study was supported by grants from the China Postdoctoral Science Foundation (2021M691879), the Science and Technology Program of Guangzhou, China (2019030016), the "Double First Class University Plan" key construction project of Xiamen University (the origin and evolution of East Asian populations and the spread of Chinese civilization), National Natural Science Foundation of China (NSFC 31801040), Nanqiang Outstanding Young Talents Program of Xiamen University (X2123302), the Major project of National Social Science Foundation of China (20&ZD248), the European Research Council (ERC) grant to Dan Xu (ERC-2019-ADG-883700-TRAM). S. Fang and Z. Xu from Information and Network Center of Xiamen University are acknowledged for the help with the high-performance computing. The funders had no role in study design, data collection, and analysis, preparation of the manuscript, or decision to submit the manuscript for publication. We thank all volunteers who participated in this project and also thank Wibhu Kutanan (Department of Biology, Faculty of Science, Khon Kaen University), Dang Liu, and Mark Stoneking (Department of Evolutionary Genetics, Max Planck Institute for Evolutionary Anthropology) for sharing genotypes of modern populations from Vietnam, Thailand, and Laos.

Disclosures

The authors declare no competing interests.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.fsigen.2021.102633.

References

[1] G.A. Gnecchi-Ruscone, P. Abondio, S. De Fanti, S. Sarno, M.G. Sherpa, P.T. Sherpa, G. Marinelli, L. Natali, M. Di Marcello, D. Peluzzi, D. Luiselli, D. Pettener, M. Sazzini, Evidence of polygenic adaptation to high altitude from tibetan and sherpa genomes, Genome Biol. Evol. 10 (11) (2018) 2919–2930.

[2] L. Kang, H.X. Zheng, F. Chen, S. Yan, K. Liu, Z. Qin, L. Liu, Z. Zhao, L. Li, X. Wang, Y. He, L. Jin, mtDNA lineage expansions in Sherpa population suggest adaptive evolution in Tibetan highlands, Mol. Biol. Evol. 30 (12) (2013) 2579–2587.

[3] L. Deng, C. Zhang, K. Yuan, Y. Gao, Y. Pan, X. Ge, Y. He, Y. Yuan, Y. Lu, X. Zhang, H. Chen, H. Lou, X. Wang, D. Lu, J. Liu, L. Tian, Q. Feng, A. Khan, Y. Yang, Z.-B. Jin, J. Yang, F. Lu, J. Qu, L. Kang, B. Su, S. Xu, Prioritizing natural-selection signals from the deep-sequencing genomic data suggests multi-variant adaptation in Tibetan highlanders, Natl. Sci. Rev. 6 (6) (2019) 1201–1222.

[4] S. Bhandari, X. Zhang, C. Cui, Bianba, S. Liao, Y. Peng, H. Zhang, K. Xiang, H. Shi, Ouzhuluobu, Baimakongzhuo, Gonggalanzi, S. Liu, Gengdeng, T. Wu, X. Qi, B. Su, Genetic evidence of a recent Tibetan ancestry to Sherpas in the Himalayan region, Sci. Rep. 5 (2015) 16249.

[5] L. Kang, H.X. Zheng, M. Zhang, S. Yan, L. Li, L. Liu, K. Liu, K. Hu, F. Chen, L. Ma, Z. Qin, Y. Wang, X. Wang, L. Jin, MtDNA analysis reveals enriched pathogenic mutations in Tibetan highlanders, Sci. Rep. 6 (2016) 31083.

[6] L. Kang, C.-C. Wang, F. Chen, D. Yao, L. Jin, H. Li, Northward genetic penetration across the Himalayas viewed from Sherpa people, Mitochondrial DNA Part A 27 (1) (2016) 342–349.

[7] A.M. Cole, S. Cox, C. Jeong, N. Petousi, D.R. Aryal, Y. Droma, M. Hanaoka, M. Ota, N. Kobayashi, P. Gasparini, H. Montgomery, P. Robbins, A. Di Rienzo, G. L. Cavalleri, Genetic structure in the Sherpa and neighboring Nepalese populations, BMC Genom. 18 (1) (2017) 102.

[8] G.A. Gnecchi-Ruscone, C. Jeong, S. De Fanti, S. Sarno, M. Trancucci, D. Gentilini, A.M. Di Blasio, M.G. Sherpa, P.T. Sherpa, G. Marinelli, M. Di Marcello, L. Natali, D. Peluzzi, D. Pettener, A. Di Rienzo, D. Luiselli, M. Sazzini, The genomic landscape of Nepalese Tibeto-Burmans reveals new insights into the recent peopling of Southern Himalayas, Sci. Rep. 7 (1) (2017) 15512.

[9] D. Lu, H. Lou, K. Yuan, X. Wang, Y. Wang, C. Zhang, Y. Lu, X. Yang, L. Deng, Y. Zhou, Q. Feng, Y. Hu, Q. Ding, Y. Yang, S. Li, L. Jin, Y. Guan, B. Su, L. Kang, S. Xu, Ancestral Origins and Genetic History of Tibetan Highlanders, Am. J. Hum. Genet. 99 (3) (2016) 580–594.

[10] C. Jeong, G. Alkorta-Aranburu, B. Basnyat, M. Neupane, D.B. Witonsky, J. K. Pritchard, C.M. Beall, A. Di Rienzo, Admixture facilitates genetic adaptations to high altitude in Tibet, Nat. Commun. 5 (2014) 3281.

[11] M. Aldenderfer, Peopling the Tibetan plateau: insights from archaeology, High Alt. Med. Biol. 12 (2) (2011) 141–147.

[12] J. Qiu, Who are the Tibetans? Am. Assoc. Adv. Sci. (2015).

[13] M. Ota, Y. Droma, B. Basnyat, Y. Katsuyama, H. Asamura, H. Sakai, H. Fukuhsima, Allele frequencies for 15 STR loci in Tibetan populations from Nepal, Forensic Sci. Int 169 (2–3) (2007) 234–238.

[14] I. Amoruso, L. Caenazzo, S. Presciuttini, E. Gregianin, A. Bozzato, G. Caravello, Y-chromosome haplotypes and clan structure of the Sherpa of the Solukhumbu (Nepal): preliminary ethnogenetic considerations, Anthr. Sci. (2013), 121122.

[15] J.M. Mullaney, R.E. Mills, W.S. Pittard, S.E. Devine, Small insertions and deletions (INDELs) in human genomes, Hum. Mol. Genet. 19 (R2) (2010) R131–R136.

[16] M. Kayser, P. de Knijff, Improving human forensics through advances in genetics, genomics and molecular biology, Nat. Rev. Genet 12 (3) (2011) 179–192.

[17] C. Genomes Project, A. Auton, L.D. Brooks, R.M. Durbin, E.P. Garrison, H.M. Kang, J.O. Korbel, J.L. Marchini, S. McCarthy, G.A. McVean, G.R. Abecasis, A global reference for human genetic variation, Nature 526 (7571) (2015) 68–74.

[18] A. Bergstrom, S.A. McCarthy, R. Hui, M.A. Almarri, Q. Ayub, P. Danecek, Y. Chen, S. Felkel, P. Hallast, J. Kamm, H. Blanche, J.F. Deleuze, H. Cann, S. Mallick, D. Reich, M.S. Sandhu, P. Skoglund, A. Scally, Y. Xue, R. Durbin, C. Tyler-Smith, Insights into human genetic variation and population history from 929 diverse genomes, Science 367 (6484) (2020).

[19] Q. Lan, C. Shen, X. Jin, Y. Guo, T. Xie, C. Chen, W. Cui, Y. Fang, G. Yang, B. Zhu, Distinguishing three distinct biogeographic regions with an in-house developed 39-AIM-InDel panel and further admixture proportion estimation for Uyghurs, Electrophoresis 40 (11) (2019) 1525–1534.

[20] 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. 40 (2019) 64–73.

[21] K. Sun, Y. Ye, T. Luo, Y. Hou, Multi-indel analysis for ancestry inference of sub-populations in China, Sci. Rep. 6 (2016) 39797.

[22] R. Pereira, C. Phillips, C. Alves, A. Amorim, A. Carracedo, L. Gusmao, A new multiplex for human identification using insertion/deletion polymorphisms, Electrophoresis 30 (21) (2009) 3682–3690.

[23] 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–e14.

[24] 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. Leg. Med. 126 (4) (2012) 533–540.

[25] M. Wang, G. He, S. Gao, F. Jia, X. Zou, J. Liu, S. Wang, Z. Ye, Y. Hou, Z. Wang, Molecular genetic survey and forensic characterization of Chinese Mongolians via the 47 autosomal insertion/deletion marker, Genomics 113 (4) (2021) 2199–2210.

[26] M. Wang, W. Du, G. He, S. Wang, X. Zou, J. Liu, C. Liu, C. Liu, Z. Wang, Revisiting the genetic background and phylogenetic structure of five Sino-Tibetan-speaking populations: insights from autosomal InDels, Mol. Genet. Genom.: MGG 295 (4) (2020) 969–979.

[27] A. Nicogossian, O. Kloiber, B. Stabile, The revised world medical association's declaration of Helsinki 2013: enhancing the protection of human research subjects and empowering ethics review committees, World Med. Health Policy 6 (1) (2014) 1–3.

[28] J. Liu, W. Du, M. Wang, C. Liu, S. Wang, G. He, Z. Wang, Forensic features, genetic diversity and structure analysis of three Chinese populations using 47 autosomal InDels, Forensic Sci. Int. Genet. 45 (2020), 102227.

[29] 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.

[30] W. Kutanan, D. Liu, J. Kampuansai, M. Srikummool, S. Srithawong, R. Shoocongdej, S. Sangkhano, S. Ruangchai, P. Pittayaporn, L. Arias, M. Stoneking, Reconstructing the human genetic history of mainland Southeast asia: insights from genome-wide data from Thailand and Laos, Mol. Biol. Evol. 38 (8) (2021) 3459–3477.

[31] D. Liu, N.T. Duong, N.D. Ton, N. Van Phong, B. Pakendorf, N. Van Hai, M. Stoneking, Extensive ethnolinguistic diversity in Vietnam reflects multiple sources of genetic diversity, Mol. Biol. Evol. 37 (9) (2020) 2503–2519.

[32] C. Jeong, K. Wang, S. Wilkin, W.T.T. Taylor, B.K. Miller, J.H. Bemmann, R. Stahl, C. Chiovelli, F. Knolle, S. Ulziibayar, D. Khatanbaatar, D. Erdenebaatar, U. Erdenebat, A. Ochir, G. Ankhsanaa, C. Vanchigdash, B. Ochir, C. Munkhbayar, D. Tumen, A. Kovalev, N. Kradin, B.A. Bazarov, D.A. Miyagashev, P.B. Konovalov, E. Zhambaltarova, A.V. Miller, W. Haak, S. Schiffels, J. Krause, N. Boivin, M. Erdene, J. Hendy, C. Warinner, A dynamic 6,000-Year genetic history of Eurasia's eastern steppe, Cell 183 (4) (2020) 890–904, e29.

[33] C. Jeong, A.T. Ozga, D.B. Witonsky, H. Malmstrom, H. Edlund, C.A. Hofman, R. W. Hagan, M. Jakobsson, C.M. Lewis, M.S. Aldenderfer, A. Di Rienzo, C. Warinner, Long-term genetic stability and a high-altitude East Asian origin for the peoples of the high valleys of the Himalayan arc, Proc. Natl. Acad. Sci. U.S.A. 113 (27) (2016) 7485–7490.

[34] C.C. Wang, H.Y. Yeh, A.N. Popov, H.Q. Zhang, H. Matsumura, K. Sirak, O. Cheronet, A. Kovalev, N. Rohland, A.M. Kim, S. Mallick, R. Bernardos, D. Tumen, J. Zhao, Y.C. Liu, J.Y. Liu, M. Mah, K. Wang, Z. Zhang, N. Adamski, N. Broomandkhoshbacht, K. Callan, F. Candilio, K.S.D. Carlson, B.J. Culleton, L. Eccles, S. Freilich, D. Keating, A.M. Lawson, K. Mandl, M. Michel, J. Oppenheimer, K.T. Ozdogan, K. Stewardson, S. Wen, S. Yan, F. Zalzala, R. Chuang, C.J. Huang, H. Looh, C.C. Shiung, Y.G. Nikitin, A.V. Tabarev, A. A. Tishkin, S. Lin, Z.Y. Sun, X.M. Wu, T.L. Yang, X. Hu, L. Chen, H. Du, J. Bayarsaikhan, E. Mijiddorj, D. Erdenebaatar, T.O. Iderkhangai, E. Myagmar, H. Kanzawa-Kiriyama, M. Nishino, K.I. Shinoda, O.A. Shubina, J. Guo, W. Cai, Q. Deng, L. Kang, D. Li, D. Li, R. Lin, Nini, R. Shrestha, L.X. Wang, L. Wei, G. Xie, H. Yao, M. Zhang, G. He, X. Yang, R. Hu, M. Robbeets, S. Schiffels, D.J. Kennett, L. Jin, H. Li, J. Krause, R. Pinhasi, D. Reich, Genomic insights into the formation of human populations in East Asia, Nature 591 (7850) (2021) 413–419.

[35] M.A. Yang, X. Fan, B. Sun, C. Chen, J. Lang, Y.C. Ko, C.H. Tsang, H. Chiu, T. Wang, Q. Bao, X. Wu, M. Hajdinjak, A.M. Ko, M. Ding, P. Cao, R. Yang, F. Liu, B. Nickel, Q. Dai, X. Feng, L. Zhang, C. Sun, C. Ning, W. Zeng, Y. Zhao, M. Zhang, X. Gao, Y. Cui, D. Reich, M. Stoneking, Q. Fu, Ancient DNA indicates human population shifts and admixture in northern and southern China, Science 369 (6501) (2020) 282–288.

[36] C. Ning, T. Li, K. Wang, F. Zhang, T. Li, X. Wu, S. Gao, Q. Zhang, H. Zhang, M. J. Hudson, G. Dong, S. Wu, Y. Fang, C. Liu, C. Feng, W. Li, T. Han, R. Li, J. Wei, Y. Zhu, Y. Zhou, C.C. Wang, S. Fan, Z. Xiong, Z. Sun, M. Ye, L. Sun, X. Wu, F. Liang, Y. Cao, X. Wei, H. Zhu, H. Zhou, J. Krause, M. Robbeets, C. Jeong, Y. Cui, Ancient genomes from northern China suggest links between subsistence changes and human migration, Nat. Commun. 11 (1) (2020) 2700.

[37] 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.

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

[39] J.D. Retief, Phylogenetic Analysis using PHYLIP, Bioinformatics Methods and Protocols, Springer, 2000, pp. 243–258.

[40] J. Yang, S.H. Lee, M.E. Goddard, P.M. Visscher, GCTA: a tool for genome-wide complex trait analysis, Am. J. Hum. Genet. 88 (1) (2011) 76–82.

[41] W.L. Kovach, MVSP: A Multivariate Statistical Package for Windows. Version 3.22, Kovach Computing Services, Pentraeth, 2013.

[42] D.H. Alexander, J. Novembre, K. Lange, Fast model-based estimation of ancestry in unrelated individuals, Genome Res. 19 (9) (2009) 1655–1664.

[43] J. Pickrell, J. Pritchard, Inference of population splits and mixtures from genome-wide allele frequency data, Nat. Preced. (2012), 1-1.

[44] C.C. Chang, C.C. Chow, L.C. Tellier, S. Vattikuti, S.M. Purcell, J.J. Lee, Second-generation PLINK: rising to the challenge of larger and richer datasets, Gigascience 4 (2015) 7.

[45] J.K. Pickrell, J.K. Pritchard, Inference of population splits and mixtures from genome-wide allele frequency data, PLoS Genet. 8 (11) (2012), e1002967.

[46] B.L. Browning, S.R. Browning, Improving the accuracy and efficiency of identity-by-descent detection in population data, Genetics 194 (2) (2013) 459–471.

[47] D.J. Lawson, G. Hellenthal, S. Myers, D. Falush, Inference of population structure using dense haplotype data, PLoS Genet. 8 (1) (2012), e1002453.

[48] Y. Liu, M. Wang, P. Chen, Z. Wang, J. Liu, L. Yao, F. Wang, R. Tang, X. Zou, G. He, Combined low-/high-density modern and ancient genome-wide data document genomic admixture history of high-altitude East Asians, Front. Genet. 12 (2021), 582357.

[49] G. He, Z. Wang, Y. Su, X. Zou, M. Wang, J. Liu, Y. Hou, Genetic variation and forensic characterization of highland Tibetan ethnicity reveled by autosomal STR markers, Int. J. Leg. Med. 132 (4) (2018) 1097–1102.
[50] X. Mao, H. Zhang, S. Qiao, Y. Liu, F. Chang, P. Xie, M. Zhang, T. Wang, M. Li, P. Cao, R. Yang, F. Liu, Q. Dai, X. Feng, W. Ping, C. Lei, J.W. Olsen, E.A. Bennett, Q. Fu, The deep population history of northern East Asia from the Late Pleistocene to the Holocene, Cell 184 (12) (2021) 3256–3266, e13.
[51] X. Zhang, K.E. Witt, M.M. Banuelos, A. Ko, K. Yuan, S. Xu, R. Nielsen, E. Huerta-Sanchez, The history and evolution of the Denisovan-EPAS1 haplotype in Tibetans, Proc. Natl. Acad. Sci. U.S.A. 118 (22) (2021), e2020803118.
[52] E. Huerta-Sanchez, X. Jin, Asan, Z. Bianba, B.M. Peter, N. Vinckenbosch, Y. Liang, X. Yi, M. He, M. Somel, P. Ni, B. Wang, X. Ou, Huasang, J. Luosang, Z.X. Cuo, K. Li, G. Gao, Y. Yin, W. Wang, X. Zhang, X. Xu, H. Yang, Y. Li, J. Wang, J. Wang, R. Nielsen, Altitude adaptation in Tibetans caused by introgression of Denisovan-like DNA, Nature 512 (7513) (2014) 194–197.
[53] C. Quan, Y. Li, X. Liu, Y. Wang, J. Ping, Y. Lu, G. Zhou, Characterization of structural variation in Tibetans reveals new evidence of high-altitude adaptation and introgression, Genome Biol. 22 (1) (2021) 159.
[54] T. Yao, L.G. Thompson, V. Mosbrugger, F. Zhang, Y. Ma, T. Luo, B. Xu, X. Yang, D. R. Joswiak, W. Wang, Third pole environment (TPE), Environ. Dev. 3 (2012) 52–64.
[55] P.J. Brantingham, G. Xing, Peopling of the northern Tibetan Plateau, World Archaeol. 38 (3) (2006) 387–414.
[56] M. Wang, Z. Wang, G. He, S. Wang, X. Zou, J. Liu, F. Wang, Z. Ye, Y. Hou, Whole mitochondrial genome analysis of highland Tibetan ethnicity using massively parallel sequencing, Forensic Sci. Int.: Genet. 44 (2020), 102197.
[57] L. Ren, G. Dong, F. Liu, J. d'Alpoim-Guedes, R.K. Flad, M. Ma, H. Li, Y. Yang, Y. Liu, D. Zhang, G. Li, J. Li, F. Chen, Foraging and farming: archaeobotanical and zooarchaeological evidence for Neolithic exchange on the Tibetan Plateau, Antiquity 94 (375) (2020) 637–652.
[58] M.C. Meyer, M.S. Aldenderfer, Z. Wang, D.L. Hoffmann, J.A. Dahl, D. Degering, W. R. Haas, F. Schlütz, Permanent human occupation of the central Tibetan Plateau in the early Holocene, Science 355 (6320) (2017) 64–67.
[59] Y.-C. Li, J.-Y. Tian, F.-W. Liu, B.-Y. Yang, K.-S.-Y. Gu, Z.U. Rahman, L.-Q. Yang, F.-H. Chen, G.-H. Dong, Q.-P. Kong, Neolithic millet farmers contributed to the permanent settlement of the Tibetan Plateau by adopting barley agriculture, Natl. Sci. Rev. 6 (5) (2019) 1005–1013.
[60] E. Arciero, T. Kraaijenbrink, Asan, M. Haber, M. Mezzavilla, Q. Ayub, W. Wang, Z. Pingcuo, H. Yang, J. Wang, M.A. Jobling, G. van Driem, Y. Xue, P. de Knijff, C. Tyler-Smith, Demographic history and genetic adaptation in the himalayan region inferred from genome-wide SNP genotypes of 49 populations, Mol. Biol. Evol. 35 (8) (2018) 1916–1933.
[61] C. Zhang, Y. Lu, Q. Feng, X. Wang, H. Lou, J. Liu, Z. Ning, K. Yuan, Y. Wang, Y. Zhou, L. Deng, L. Liu, Y. Yang, S. Li, L. Ma, Z. Zhang, L. Jin, B. Su, L. Kang, S. Xu, Differentiated demographic histories and local adaptations between Sherpas and Tibetans, Genome Biol. 18 (1) (2017) 115.
[62] G.L. He, M.G. Wang, Y.X. Li, X. Zou, H.Y. Yeh, R.K. Tang, X.M. Yang, Z. Wang, J. X. Guo, T. Luo, J. Zhao, J. Sun, R. Hu, L.H. Wei, G. Chen, Y.P. Hou, C.C. Wang, Fine-scale north-to-south genetic admixture profile in Shaanxi Han Chinese revealed by genome-wide demographic history reconstruction, J. Syst. Evol. (2021) 0.
[63] M. Wang, X. Zou, H.-Y. Ye, Z. Wang, Y. Liu, J. Liu, F. Wang, H. Yao, P. Chen, R. Tao, S. Wang, L.-H. Wei, R. Tang, C.-C. Wang, G. He, Peopling of Tibet Plateau and multiple waves of admixture of Tibetans inferred from both modern and ancient genome-wide data, bioRxiv (2020), 2020.07.03.185884.
[64] B. Chen, J.W. Cole, C. Grond-Ginsbach, Departure from Hardy Weinberg equilibrium and genotyping error, Front. Genet. 8 (2017) 167.
[65] G.L. He, M.G. Wang, X. Zou, P.Y. Chen, Z. Wang, Y. Liu, H.B. Yao, L.H. Wei, R. K. Tang, C.C. Wang, H.Y. Yeh, Peopling history of the Tibetan Plateau and multiple waves of admixture of tibetans inferred from both ancient and modern genome-wide data, Front. Genet. 12 (1634) (2021).