



Haplotypic Associations and Differentiation of MHC Class II Polymorphic Alu Insertions at Five Loci With HLA-DRB1 Alleles in 12 Minority Ethnic Populations in China

Yina Cun^{1†}, Lei Shi^{1†}, Jerzy K. Kulski², Shuyuan Liu¹, Jia Yang³, Yufen Tao¹, Xinwen Zhang³, Li Shi^{3*} and Yufeng Yao^{1*}

¹ Department of Immunogenetics, Institute of Medical Biology, Chinese Academy of Medical Sciences and Peking Union Medical College, Kunming, China, ² Faculty of Health and Medical Sciences, University of Western Australia Medical School, Crawley, WA, Australia, ³ Yunnan Key Laboratory of Vaccine Research and Development on Severe Infectious Disease, Institute of Medical Biology, Chinese Academy of Medical Sciences and Peking Union Medical College, Kunming, China

OPEN ACCESS

Edited by:

Pierre Pontarotti,
Centre National de la Recherche
Scientifique (CNRS), France

Reviewed by:

Pierre-Antoine Gourraud,
Université de Nantes, France
Antonio Amaiz-Villena,
Universidad Complutense de Madrid,
Spain

*Correspondence:

Li Shi
shili.imb@gmail.com
Yufeng Yao
leoyyf@gmail.com;
yufeng_yao@imbcams.com.cn

† These authors have contributed
equally to this work

Specialty section:

This article was submitted to
Evolutionary and Population Genetics,
a section of the journal
Frontiers in Genetics

Received: 09 December 2020

Accepted: 08 June 2021

Published: 07 July 2021

Citation:

Cun Y, Shi L, Kulski JK, Liu S,
Yang J, Tao Y, Zhang X, Shi L and
Yao Y (2021) Haplotypic Associations
and Differentiation of MHC Class II
Polymorphic Alu Insertions at Five
Loci With HLA-DRB1 Alleles in 12
Minority Ethnic Populations in China.
Front. Genet. 12:636236.
doi: 10.3389/fgene.2021.636236

The analysis of polymorphic variations in the human major histocompatibility complex (MHC) class II genomic region on the short-arm of chromosome 6 is a scientific enquiry to better understand the diversity in population structure and the effects of evolutionary processes such as recombination, mutation, genetic drift, demographic history, and natural selection. In order to investigate associations between the polymorphisms of HLA-DRB1 gene and recent Alu insertions (POALINs) in the HLA class II region, we genotyped HLA-DRB1 and five Alu loci (AluDPB2, AluDQA2, AluDQA1, AluDRB1, AluORF10), and determined their allele frequencies and haplotypic associations in 12 minority ethnic populations in China. There were 42 different HLA-DRB1 alleles for ethnic Chinese ranging from 12 alleles in the Jinuo to 28 in the Yugur with only DRB1*08:03, DRB1*09:01, DRB1*12:02, DRB1*14:01, DRB1*15:01, and DRB1*15:02 present in all ethnic groups. The POALINs varied in frequency between 0.279 and 0.514 for AluDPB2, 0 and 0.127 for AluDQA2, 0.777 and 0.995 for AluDQA1, 0.1 and 0.455 for AluDRB1 and 0.084 and 0.368 for AluORF10. By comparing the data of the five-loci POALIN in 13 Chinese ethnic populations (including Han-Yunnan published data) against Japanese and Caucasian published data, marked differences were observed between the populations at the allelic or haplotypic levels. Five POALIN loci were in significant linkage disequilibrium with HLA-DRB1 in different populations and AluDQA1 had the highest percentage association with most of the HLA-DRB1 alleles, whereas the nearby AluDRB1 indel was strongly haplotypic for only DRB1*01, DRB1*10, DRB1*15 and DRB1*16. There were 30 five-locus POALIN haplotypes inferred in all populations with H5 (no Alu insertions except for AluDQA1) and H21 (only AluDPB2 and AluDQA1 insertions) as the two predominant haplotypes. Neighbor joining trees and principal component analyses of the Alu and HLA-DRB1 polymorphisms showed that genetic diversity of these genomic markers is associated strongly with the population characteristics of language family, migration and sociality. This comparative study of HLA-DRB1 alleles and multilocus, lineage POALIN frequencies of Chinese

ethnic populations confirmed that POALINs whether investigated alone or together with the HLA class II alleles are informative genetic and evolutionary markers for the identification of allele and haplotype lineages and genetic variations within the same and/or different populations.

Keywords: HLA class II regions, POALIN, HLA-DRB1, polymorphism, haplotypes, Chinese ethnic populations

INTRODUCTION

The human major histocompatibility complex (MHC) class II genomic region on the short-arm of chromosome 6 contains highly polymorphic classical and non-classical human leukocyte antigen (HLA) class II genes (HLA-DRB1, -DRA, -DQA1, -DQB1, -DQA2, -DQB2, -DPA1, and -DPB1) involved in the regulation of the innate and adaptive immune system, autoimmunity, and transplantation (Shiina et al., 2004, 2009; Vandiedonck and Knight, 2009; Trowsdale, 2011). The extensive polymorphism of the HLA class II genes is studied widely and used to provide a better understanding of the diversity in population structure and the effects of evolutionary processes such as recombination, mutation, genetic drift, demographic history, and natural selection (Meyer et al., 2006; Traherne, 2008; Pierini and Lenz, 2018; Manczinger et al., 2019). For example, there are at least 2,909 HLA-DRB1 alleles distributed worldwide with the official sequences and designations provided by the IMGT/HLA database (Robinson et al., 2020). Consequently, the HLA-DRB1 alleles are genetic markers that are utilized often for the assessment of population structure and differentiation as well as providing information on interpopulation genetic exchange (gene flow) and other demographic events (Di and Sanchez-Mazas, 2011; Sanchez-Mazas et al., 2013, 2017; Sanchez-Mazas and Meyer, 2014; Gonzalez-Galarza et al., 2020). Moreover, the HLA-DRB1 alleles present intracellular or exogenous antigen peptides to CD4⁺ T cells that trigger and regulate the downstream immune responses to defend against pathogen invasion (Chaplin, 2010). Therefore, this highly polymorphic genomic marker might reveal changes associated with pathogen-mediated pressure on highly heterogeneous and diverse populations (Sun et al., 2015; Weiskopf et al., 2016).

In addition to polymorphic HLA class II genes, the MHC class II region has a number of polymorphic Alu insertions (POALINs) that are informative population ancestral lineage markers. They are insertion/deletions (either present or absent) at integration sites, which carry characteristic alleles or haplotypes inherited from different ancestral populations (Bennett et al., 2004; Kulski and Dunn, 2005; Ray et al., 2007). Alu retroelements (short interspersed nuclear elements) are among the class of genomic repetitive DNA elements that first appeared in primates about 65 million years ago and then amplified by retrotransposition to the present estimated one million copies per human genome (Lander et al., 2001; Batzer and Deininger, 2002). POALINs are useful lineage and evolutionary genetic markers for studying the origin and genomic diversity of human populations because (1) their allelic frequency distributions vary significantly among geographically different human populations (Deininger and Batzer, 1999;

Jorde et al., 2000; Watkins et al., 2001), and (2) they have an inherited identity by descent arising from a known initial ancestral state (no Alu insertion), whereby their presence and/or absence define the ancestral lineages within a population (Antunez-de-Mayolo et al., 2002).

Some MHC Alu family members were used previously as evolutionary molecular markers to infer the ancestral duplication history of HLA class I and class II gene copies (Mnukova-Fajdelova et al., 1994; Svensson et al., 1996; Kulski et al., 1999, 2000). Also, several studies reported on the frequencies and distribution of human-specific POALIN loci within the HLA class I region and on their inferred haplotypic associations with HLA-A, -B and -C loci in different populations (Dunn et al., 2002, 2003, 2005a,b, 2007; Yao et al., 2009, 2010; Kulski et al., 2011, 2019; Mastana et al., 2017; Singh et al., 2019). These associations reflect in part the different haplotypic structures of the MHC class I and class II regions and the linkage of multiple polymorphic loci, especially when extended over long stretches (1–3 Mb) of conserved genomic sequences in human populations known as ancestral haplotypes (Dawkins et al., 1999) or conserved extended haplotypes (Alper et al., 2006; Larsen et al., 2014). Although comparative DNA sequence analysis of the entire MHC genome region between two homozygous HLA haplotypes has indicated the presence of POALIN within the MHC class II region (Stewart et al., 2004), five human-specific POALIN (AluDPB2, AluDQA2, AluDQA1, AluDRB1, and AluORF10) frequencies at five loci in the MHC class II genomic region were determined previously only for Japanese, Australian Caucasians (Kulski et al., 2010) and Chinese Han in Yunnan province (Shi et al., 2014) populations. By comparing the data of the MHC class II five-loci POALINs in Chinese Han with Japanese and Caucasian data, marked differences were observed between the three ethnic groups at the allelic or haplotypic levels. In addition, each POALIN was in significant linkage disequilibrium (LD) and/or haplotypically associated (Kulski et al., 2020, 2021) with a variety of HLA-DRB1 alleles in Chinese Han in Yunnan province (Shi et al., 2014). These results showed that POALINs whether investigated alone or together with the HLA class II alleles are informative genetic markers for the identification of allele and haplotype lineages and variations within the same and/or different populations.

Beside the Chinese speaking Han majority, there are 55 officially recognized minority ethnic populations of China, which contribute to about 8% of the overall Chinese population and provide abundant genetic resources for POALIN–HLA inferred haplotype studies (Yao et al., 2010). The minority ethnic groups living in the south and southwest of China can be traced back to three major ancient groups: Di-Qiang, Bai-Pu, and Bai-Yue that speak the Tibeto-Burman, Mon-Khmer and Daic language

subfamilies, respectively (**Table 1**); whereas in the northwest of China, most ethnic groups speak the language of the Mongolian and Tujue Manchu-Tungusic subfamily, which is the Altaic language family (Guo, 2000). Although the anthropological, cultural and linguistic characteristics of some of these ethnic populations have been studied in detail (You, 1994; Guo, 2000; Chu et al., 2006), there are few published comparative investigations on the genetic diversity of these populations by genome-wide sequencing or genotyping methods (Di and Sanchez-Mazas, 2011). Therefore, the analyses of robust and reproducible genetic markers such as the POALINs and HLA-DRB1 alleles in small and isolated ethnic minority remains an important task to better understand the human genome and its genetic variability throughout the world.

The aim of present study was to elucidate the inferred haplotypic association between the MHC POALINs and classical HLA class II alleles by determining (1) genetic structures of the five MHC class II POALIN dimorphisms and HLA-DRB1 allele and haplotype frequencies in 12 minority ethnic populations in China, and (2) correlations between the genetic diversity and the four language families of these populations (**Table 1**). Among these 12 minority ethnic populations, 8 of them settled in Yunnan province together with the Han people (Han-Yunnan). The Han-Yunnan, speaking Chinese of the Sino-Tibetan language family, migrated from the northern region by various routes and at different times to Yunnan province and exhibited genetic characteristics of both northern and southern Chinese groups (Shi et al., 2006). Thus, we included the published data of Han-Yunnan, Japanese and Caucasians as reference populations in order to compare and correlate the genetic differentiation of the HLA-DRB1 alleles and the five POALINs between the populations and the language families by using

the DA genetic distance measure in phylogeny and principal component analysis (PCA).

MATERIALS AND METHODS

Ethics Statement

This study was approved by the Committee on the Ethics of Institute of Medical Biology, Chinese Academy of Medical Sciences, the batch number is YIKESHENGLUNZI [2012]12. Moreover, the protocol employed by this investigation was in accordance with the principles expressed in the Helsinki Declaration of 1975, which was revised in 2008. Written informed consents were obtained from each participant.

Subjects and Samples

A total of 1,201 unrelated individuals were recruited from 12 Chinese minority ethnic populations in China (**Figure 1**). The geographic location, sample size of each population, the language family to which they belong, and the ancient groups from which they originated are listed in **Table 1**. These populations are descended from four ancient Chinese groups and belong to four different language subfamilies (Guo, 2000; Yao et al., 2010; Di and Sanchez-Mazas, 2011) as outlined in the introduction and **Table 1**. The geographic origin, nationalities, and pedigree (unrelated through at least three generations) of each individual were ascertained before sampling.

Genomic DNA and HLA-DRB1 Typing

Genomic DNA was extracted from peripheral lymphocytes using a QIAamp Blood Kit (Qiagen, Hilden, Germany). DNA samples

TABLE 1 | Geographic and language information for the 12 minority ethnic populations sampled in the current study.

Ethnic group	Sample size	Sample location	Ancient tribe	Language family	Language subfamily	Population size
Hani	149	Jingha village, Jinghong County, Yunnan Province, Southwest China	Di-Qiang (dq)	Sino-Tibetan	Tibeto-Burman (TB)	1,661,000
Jinuo	75	Jinuo Mountains Area, Jinghong City, Xishuangbanna Prefecture, Yunnan Province, Southwest China	Di-Qiang (dq)	Sino-Tibetan	Tibeto-Burman (TB)	23,143
Lisu	79	Fugong and Gongshan Counties, Nujiang Municipality, Yunnan Province, Southwest China	Di-Qiang (dq)	Sino-Tibetan	Tibeto-Burman (TB)	703,000
Nu	82	Fugong and Gongshan Counties, Nujiang Municipality, Yunnan Province, Southwest China	Di-Qiang (dq)	Sino-Tibetan	Tibeto-Burman (TB)	37,523
Jingpo	95	Dehong Dai-Jingpo Autonomous Prefectures, Yunnan Province, Southwest China	Di-Qiang (dq)	Sino-Tibetan	Tibeto-Burman (TB)	147,828
Bulang	109	Bulang Mountain Area, Menghai County, Yunnan Province, Southwest China	Baipu (bp)	Austo-Asiatic	Mon-Khmer (MK)	120,000
Wa	109	Canyuan and Ximeng Counties, Yunnan Province, Southwest China	Baipu (bp)	Austo-Asiatic	Mon-Khmer (MK)	430,000
Dai	121	Xishuangbanna Dai Autonomous Prefecture, Yunnan Province, Southwest China	Baiyue (by)	Sino-Tibetan	Daic (D)	1,261,000
Maonan	78	Xianan village of Huangjiang County, Guangxi Province, Southern China	Baiyue (by)	Sino-Tibetan	Daic (D)	101,000
Zhuang	101	Tiandeng County, Nanning City, Guangxi Zhuang Autonomous Province, Southern China	Baiyue (by)	Sino-Tibetan	Daic (D)	16,926,000
Tu	110	Huzu County, Qinghai Province, Northwest China	Mongolian (m)	Altaic	Mongolian (M)	290,000
Yugur	93	Sunan County, Gansu Province, Northwest China	Mongolian (m)	Altaic	Tujue (T)	14,378

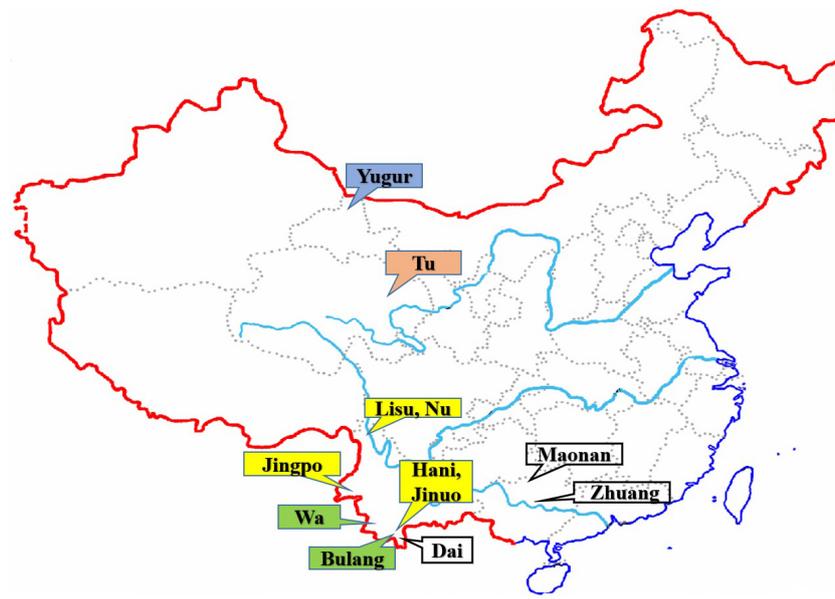


FIGURE 1 | The geographic locations of the 12 Chinese ethnic populations in China. The colored labeled boxes represent the ancient tribe, language family and subfamily for each population listed in **Table 1**. Yellow represent Di-Qiang, Sino-Tibetan, Tibeto-Burman. Green represent Baipu, Austro-Asiatic, Mon-Khmer. White represent Baiyue, Sino-Tibetan, Daic. Orange represent Mongolian, Altaic, Mongolian. Blue represent Mongolian, Altaic, Tujue.

were quantified with a NanoDrop ND-1000 spectrophotometer (NanoDrop Technologies, Wilmington, WI, United States) and adjusted to a concentration of 20 ng/L. The HLA-DRB1 genes were genotyped using a WAKFlow HLA typing kit (Wakunaga, Hiroshima, Japan) as in previous studies (Ogata et al., 2007; Shi et al., 2008, 2010a,b, 2011; Yao et al., 2012; Tao et al., 2020), which is based on polymerase chain reaction-sequence specific oligonucleotide probes (PCR-SSOP) coupled with multiple analyte profiling (xMAP) technology (Luminex System).

Alu and PCR Assay

The sense and antisense primers used for the PCR of the POALINs located in MHC II regions were previously reported (Kulski et al., 2010; Shi et al., 2014). As some of the previously published primers used for the PCR of the POALINs located in MHC II region have mutations in the Chinese Han in Yunnan province, new sense and antisense primer pairs were designed and used for the PCR of five POALINs located in the MHC II region (**Supplementary Table 1**). **Supplementary Figure 1** shows a map of the locations of the five POALINs with the HLA class II regions of the MHC on chromosome 6p21.3.

The PCR products were analyzed according to the fragments of different sizes by the presence or absence of an electrophoretic specific band in 2% agarose gel stained with ethidium bromide and visualized by ultraviolet light. The Alu-PCR methods clearly differentiate between an insertion and absence of insertion in heterozygous individuals based on distinctly different sized PCR products as shown in **Supplementary Figure 2**. The POALIN alleles are dimorphic structures whereby the absence of the Alu insertion at the Alu locus is the Alu*1 allele and the presence of the Alu insertion is the Alu*2 allele. The overall frequencies of

the Alu*2 (insertion) allele at each of the five loci were estimated from the genotypes as described below in the statistical section.

Allele Linkage Controls for Assessment of HLA-DRB1 Allele and POALIN Associations

To better assess the haplotypic associations between the POALINs and the HLA-DRB1 alleles, we examined their sequence linkages in 95 different MHC haplotype sequences (Kulski et al., 2021) that were sequenced, partially annotated and assembled from HLA-homozygous cell lines by Norman et al. (2017).

The FASTA files of the 95 MHC class I, II and III genomic sequences were downloaded from the archives at NCBI BioProject with the accession number PRJEB6763¹ and submitted to the RepeatMasker webserver² for output files of annotated members of the interspersed repetitive DNA families, their locations in the sequence and their relative similarity or identity in comparison to reference sequences of SINES, LINES, LTRs, ERVs, DNA elements, small RNA, and simple repeats (Kulski et al., 2021). The five MHC class II POALINs were easily identified within the RepeatMasker outputs on the basis of their location and flanking sequences and/or other repeats as previously described (Kulski et al., 2010). The HLA-DRB1 alleles for all of the 95 cell line sequences were determined and reported by Norman et al. (2017). **Supplementary Table 2** is a summary of the sequence linkages between the 5 POALIN and the HLA-DRB1 alleles that were determined in 90 of the sequenced

¹<https://www.ncbi.nlm.nih.gov/bioproject/>

²<http://www.repeatmasker.org/cgi-bin/WEBRepeatMasker>

haplotypes (Kulski et al., 2021). These were used as a comparative reference control to assist with a better interpretation of our results obtained for our haplotypic association analyses in 15 different populations.

Statistical Analysis

The frequencies of five POALINs were calculated from the genotyping data by the direct-counting method. For each locus, Hardy-Weinberg's equilibrium was assessed using the Guo and Thompson method (Guo and Thompson, 1992). The haplotypes were estimated by the maximum-likelihood method using the Pypop software (Lancaster et al., 2003, 2007). Pairwise LD of POALINs and HLA allele were calculated using the SHEsis software³ (Shi and He, 2005). The percentage association between a POALIN insertion and an HLA allele was calculated as the percentage of the total HLA allele frequency that was associated with the presence of the POALIN insertion at an inferred HLA class II gene/POALIN haplotype using the haplotype frequency data generated by the Pypop software (Lancaster et al., 2003, 2007). Percentage associations between HLA allele and POALIN insertion frequencies were considered to be very strong if between 80 and 100%, strong if over 50% and less than 79%, moderate if between 20 and 50%, and low or absent if less than 20% (Kulski et al., 2010; Shi et al., 2014). The differences in significance between the POALIN and its haplotype frequencies were determined by a contingency test (Fisher's exact test). Bonferroni correction was used for multiple testing. Statistical significance was defined at the 5% level.

Phylogenetic Analysis

Based on the POALIN allele, HLA-DRB1 allele frequencies and DRB1/AluDRB1 haplotypes of the different population, the DA was calculated using the Dispan software (Nei, 1973, 1978). The Mega 7.0 software was used to reconstruct the neighbor-joining (NJ) trees according to the DA (Tamura et al., 2007). Principal component analysis (PCA) was also performed based either on POALIN allele, or HLA-DRB1 frequencies using SPSS 16.0 software. POALIN allele and HLA-DRB1 allele frequencies were obtained from additional Japanese, Caucasian and Han-Yunnan populations (Kulski et al., 2010; Shi et al., 2014) for comparative phylogenetic analysis with the frequencies obtained for the 12 Chinese ethnic populations in this study.

RESULTS

HLA-DRB1 Allele Frequencies

We summarized the HLA-DRB1 allele frequencies in 15 populations in **Table 2** according to previous studies (Shi et al., 2006, 2008, 2010a,b, 2011; Ogata et al., 2007; Kulski et al., 2010; Yao et al., 2012; Tao et al., 2020). There were 57 different HLA-DRB1 alleles for the 15 populations ranging from 12 alleles in the Jinuo to 39 alleles in the Han-Yunnan. Only six HLA-DRB1 alleles were present in all 15 populations and these were DRB1*08:03, DRB1*09:01, DRB1*12:02,

DRB1*14:01, DRB1*15:01, and DRB1*15:02. There were sixteen low frequency, unique, solitary alleles for six populations; DRB1*03:05 (0.003), DRB1*08:27 (0.003), DRB1*09:09 (0.003), DRB1*11:31 (0.003), DRB1*11:52 (0.003), DRB1*12:19 (0.003), DRB1*13:28 (0.003), DRB1*14:32 (0.005) and DRB1*14:35 (0.003) in Han-Yunnan, DRB1*01:03 (0.011), DRB1*08:10 (0.003) and DRB1*11:03 (0.017) in Caucasians, DRB1*14:06 (0.01) in Japanese, DRB1*14:18 (0.015) in Zhuang, DRB1*14:25 (0.014) in Bulang, and DRB1*15:11 (0.011) in Jingo. The successive highest frequencies of DRB1 alleles were DRB1*16:02 in Dai, and DRB1*15:01 in Zhuang, and DRB1*14:01 and DRB1*16:02 in the Maonan. In addition, DRB1*12:02 was the most frequent in the Hani, Jinuo, Lisu, Nu, Jingo, Bulang, Wa, Maonan and Han-Yunnan ranging from 16% in Maonan to 55% in Bulang. The highest allelic frequency in the two Mongolian groups, Tu and Yugur, was DRB1*09:01 (12.7% and 13.4%, respectively) as same as Japanese (20%).

POALIN Allele Frequencies and Hardy-Weinberg's Equilibrium (HWE)

The five POALIN allele frequencies and the genotype counts in 12 Chinese minority populations, shown in **Table 3**, were compared statistically to those reported previously for the Japanese, Australian Caucasians (Kulski et al., 2010) and Chinese Han in Yunnan (Shi et al., 2014). The frequencies of five POALINs in 12 Chinese minority populations ranged from 0.359 to 0.514 (AluDPB2), 0 to 0.127 (AluDQA2), 0.777 to 0.995 (AluDQA1), 0.1 to 0.455 (AluDRB1) and 0.084 to 0.368 (AluORF10). The differences in significance between two populations for each POALIN frequency are shown in **Supplementary Table 3**.

Of all the five POALIN loci, the AluDQA1 locus showed a significant departure ($P < 0.01$ after Bonferroni's correction) from HWE in 10 minority populations, which were the Hani, Jinuo, Lisu, Nu, Jingo, Wa, Dai, Zhuang, Tu and Yugur (**Supplementary Table 4**). The data were similar to the results of Han in Yunnan (Shi et al., 2014) and Japanese (Kulski et al., 2010) that also showed that the AluDQA1 locus was not consistent with the HWE. The AluDPB2 locus showed a significant departure ($P < 0.01$ after Bonferroni's correction) from HWE in the Bulang; whereas the AluDQA2 locus showed a significant departure ($P < 0.01$ after Bonferroni's correction) from HWE in the Jinuo and Yugur.

POALIN Haplotype Frequencies

Table 4 shows the POALIN haplotypes for 12 Chinese minority populations, the Chinese Han in Yunnan (Shi et al., 2014), Japanese and Caucasians (Kulski et al., 2010). There were 30 five-locus POALIN haplotypes inferred in all 15 populations, with 11 in Hani, 11 in Jinuo, 15 in Lisu, 12 in Nu, 14 in Jingo, 10 in Bulang, 12 in Wa, 16 in Dai, 11 in Maonan, 16 in Zhuang, 19 in Tu, 18 in Yugur, 14 in Han-Yunnan, 14 in Japanese and 23 in Caucasians. All haplotypes were named H1-H30 and only five haplotypes were found in all 15 populations. These were the ancestral null H1 with no Alu insertions (AluDPB2*1: AluDQA2*1: AluDQA1*1: AluDRB1*1: AluORF10*1), and various haplotypes

³<http://analysis.bio-x.cn/myAnalysis.php>

TABLE 2 | HLA-DRB1 allele frequencies in 15 populations.

No.	DRB1 allele	Hani (n = 149)	Jinuo (n = 75)	Lisu (n = 79)	Nu (n = 82)	Jingpo (n = 95)	Bulang (n = 109)	Wa (n = 109)	Dai (n = 121)	Maonan (n = 78)	Zhuang (n = 101)	Tu (n = 110)	Yugur (n = 93)	Han-Yunnan (n = 186)	Japanese (n = 100)	Caucasians (n = 174)
1	DRB1*01:01	0.013		0.013	0.006	0.026						0.064	0.048	0.024	0.08	0.095
2	DRB1*01:02			0.006												0.014
3	DRB1*01:03															0.011
4	DRB1*03:01	0.007					0.018	0.058	0.045	0.129	0.023	0.065	0.035			0.126
5	DRB1*03:05													0.003		
6	DRB1*04:01			0.006			0.005					0.027	0.043	0.003	0.02	0.109
7	DRB1*04:02												0.005	0.003		0.006
8	DRB1*04:03	0.017	0.04		0.018	0.005	0.018	0.014	0.012	0.006		0.005	0.038	0.016	0.035	0.006
9	DRB1*04:04									0.013		0.005	0.016	0.003		0.043
10	DRB1*04:05			0.038	0.024	0.026	0.023	0.028	0.054	0.09	0.015	0.068	0.054	0.059	0.135	
11	DRB1*04:06			0.006	0.03	0.011	0.005	0.014	0.021	0.006	0.015	0.014	0.022	0.027	0.025	
12	DRB1*04:07			0.006											0.01	0.02
13	DRB1*04:08												0.011		0.01	0.003
14	DRB1*04:10			0.019	0.006			0.009				0.009		0.008	0.01	
15	DRB1*07:01	0.003	0.013	0.006	0.012	0.021	0.005	0.055	0.008			0.073	0.065	0.062		0.144
16	DRB1*08:01	0.003											0.016			0.014
17	DRB1*08:02			0.006	0.012							0.005	0.022	0.008	0.035	
18	DRB1*08:03	0.03	0.087	0.076	0.11	0.047	0.014	0.037	0.004	0.071	0.015	0.059	0.027	0.089	0.07	
19	DRB1*08:04												0.011			0.003
20	DRB1*08:09								0.004				0.005			
21	DRB1*08:10															0.003
22	DRB1*08:27													0.003		
23	DRB1*09:01	0.027	0.007	0.051	0.024	0.042	0.014	0.005	0.112	0.103	0.05	0.127	0.134	0.159	0.2	0.023
24	DRB1*09:09													0.003		
25	DRB1*10:01	0.003				0.016			0.008		0.02	0.045	0.054	0.003	0.005	0.006
26	DRB1*11:01	0.013	0.047	0.044	0.031	0.021		0.005	0.021	0.026	0.05	0.105	0.048	0.054	0.01	0.06
27	DRB1*11:03															0.017
28	DRB1*11:04											0.005	0.005	0.008		0.006
29	DRB1*11:06			0.006	0.037	0.016	0.05	0.018	0.004							
30	DRB1*11:31													0.003		
31	DRB1*11:52													0.003		
32	DRB1*12:01	0.003		0.006	0.018							0.077	0.054	0.011	0.02	0.026
33	DRB1*12:02	0.322	0.407	0.209	0.238	0.453	0.55	0.326	0.099	0.16	0.045	0.086	0.054	0.164	0.025	
34	DRB1*12:19													0.003		
35	DRB1*13:01			0.019	0.012			0.073			0.005	0.014	0.043	0.011		0.063
36	DRB1*13:02			0.013	0.006	0.011	0.005		0.025		0.015	0.018	0.038	0.038	0.03	0.049
37	DRB1*13:03		0.007		0.012		0.005		0.037	0.064	0.025					0.006
38	DRB1*13:28													0.003		
39	DRB1*14:01	0.208	0.153	0.108	0.14	0.026	0.032	0.014	0.112	0.122	0.139	0.045	0.043	0.027	0.04	0.017
40	DRB1*14:02											0.009		0.003		
41	DRB1*14:03				0.018							0.014	0.005	0.008	0.03	
42	DRB1*14:04	0.118		0.12	0.012	0.021	0.005	0.069	0.033	0.006	0.005			0.005		
43	DRB1*14:05	0.02	0.027						0.012	0.013	0.005	0.027	0.016	0.008	0.015	
44	DRB1*14:06														0.01	
45	DRB1*14:07			0.019										0.003		
46	DRB1*14:10	0.017		0.013												
47	DRB1*14:18										0.015					
48	DRB1*14:25						0.014									
49	DRB1*14:32													0.005		
50	DRB1*14:35													0.003		
51	DRB1*15:01	0.044	0.04	0.057	0.037	0.058	0.073	0.092	0.087	0.096	0.168	0.018	0.065	0.073	0.05	0.103
52	DRB1*15:02	0.07	0.167	0.108	0.085	0.1	0.124	0.087	0.136	0.058	0.139	0.041	0.022	0.035	0.125	0.011
53	DRB1*15:04	0.067	0.007	0.038	0.098	0.063	0.037	0.133						0.008		
54	DRB1*15:11					0.011										
55	DRB1*15:15				0.006	0.005										
56	DRB1*16:01			0.006								0.009				0.014
57	DRB1*16:02	0.013			0.006	0.021	0.023	0.005	0.153	0.122	0.149	0.009	0.011	0.022	0.01	

References: (Shi et al., 2006; Han-Yunnan); (Ogata et al., 2007; Maonan); (Shi et al., 2008; Jinuo, Wa); (Kulski et al., 2010; Japanese, Caucasians); (Shi et al., 2010a; Hani, Bulang); (Shi et al., 2010b; Dai); (Shi et al., 2011; Zhuang); (Yao et al., 2012; Lisu, Nu); (Tao et al., 2020; Jingpo).

TABLE 3 | The MHC POALIN allelic frequencies and genotype counts at five loci in 15 populations.

Alu allele or genotype	Alu allele frequency or genotype counts in 15 populations (n)														
	Hani (n = 149)	Jinuo (n = 75)	Lisu (n = 79)	Nu (n = 82)	Jingpo (n = 95)	Bulang (n = 109)	Wa (n = 109)	Dai (n = 121)	Maonan (n = 78)	Zhuang (n = 101)	Tu (n = 110)	Yugur (n = 93)	Han-Yunnan (n = 186) ^a	Japanese (n = 100) ^b	Caucasians (n = 174) ^b
Alu allele	Frequency														
AluDPB2*1	0.507	0.620	0.544	0.585	0.721	0.583	0.560	0.545	0.641	0.574	0.486	0.527	0.540	0.500	0.546
AluDPB2*2	0.493	0.380	0.456	0.415	0.279	0.417	0.440	0.455	0.359	0.426	0.514	0.473	0.460	0.500	0.454
AluDQA2*1	0.997	0.987	0.981	0.994	0.895	0.977	0.995	0.959	1.000	0.975	0.873	0.909	0.997	0.990	0.790
AluDQA2*2	0.003	0.013	0.019	0.006	0.105	0.023	0.005	0.041	0.000	0.025	0.127	0.091	0.003	0.010	0.210
AluDQA1*1	0.124	0.067	0.184	0.195	0.032	0.005	0.106	0.050	0.026	0.134	0.223	0.097	0.312	0.432	0.503
AluDQA1*2	0.876	0.933	0.816	0.805	0.968	0.995	0.894	0.950	0.974	0.866	0.777	0.903	0.688	0.568	0.497
AluDRB1*1	0.789	0.780	0.804	0.768	0.753	0.739	0.683	0.632	0.705	0.545	0.900	0.855	0.820	0.775	0.744
AluDRB1*2	0.211	0.220	0.196	0.232	0.247	0.261	0.317	0.368	0.295	0.455	0.100	0.145	0.180	0.225	0.256
AluORF10*1	0.916	0.900	0.873	0.896	0.826	0.830	0.862	0.632	0.641	0.634	0.859	0.882	0.825	0.855	0.764
AluORF10*2	0.084	0.100	0.127	0.104	0.174	0.170	0.138	0.368	0.359	0.366	0.141	0.118	0.175	0.145	0.236
Alu genotype	Counts														
AluDPB2 1, 1	38	29	24	29	50	44	35	37	34	30	28	23	57		
AluDPB2 1, 2	75	35	38	38	37	39	52	58	32	56	51	52	87		
AluDPB2 2, 2	36	11	17	15	8	26	22	26	12	15	31	18	42		
AluDQA2 1, 1	148	74	77	81	76	104	108	113	78	96	84	81	185		
AluDQA2 1, 2	1	0	1	1	18	5	1	6	0	5	24	7	1		
AluDQA2 2, 2	0	1	1	0	1	0	0	2	0	0	2	5	0		
AluDQA1 1, 1	14	3	13	15	3	0	5	4	1	13	21	5	28		
AluDQA1 1, 2	9	4	3	2	0	1	13	4	2	1	7	8	60		
AluDQA1 2, 2	126	68	63	65	92	108	91	113	75	87	82	80	98		
AluDRB1 1, 1	95	46	51	49	52	59	50	45	37	27	89	68	124		
AluDRB1 1, 2	45	25	25	28	39	43	49	63	36	56	20	23	57		
AluDRB1 2, 2	9	4	3	5	4	7	10	13	5	18	1	2	5		
AluORF10 1, 1	128	61	59	67	65	76	80	46	34	39	82	75	128		
AluORF10 1, 2	17	13	20	13	27	29	28	61	32	50	25	14	51		
AluORF10 2, 2	4	1	0	2	3	4	1	14	12	12	3	4	7		

^aThe data of five POALINs in Han-Yunnan were published previously (Shi et al., 2014).

^bThe data of five POALINs in Japanese and Caucasians were published previously (Kulski et al., 2010, 2011).

TABLE 4 | Haplotype frequencies of POALINs at five loci in 15 populations.

No.	5-loci Alu haplotypes					Haplotype frequencies in 15 populations														
	Alu DPB2	Alu DQA2	Alu DQA1	Alu DRB1	Alu ORF10	Hani	Jinuo	Lisu	Nu	Jingpo	Bulang	Wa	Dai	Maonan	Zhuang	Tu	Yugur	Han- Yunnan ^a	Japanese ^b	Caucasians ^b
H1	1	1	1	1	1	0.016	0.040	0.059	0.063	0.005	0.005	0.022	0.009	0.007	0.043	0.075	0.038	0.102	0.176	0.140
H2	1	1	1	1	2			0.006	0.010	0.011		0.009	0.015		0.056	0.012	0.005	0.031		
H3	1	1	1	2	1	0.007		0.006	0.006							0.016	0.006	0.014	0.021	0.040
H4	1	1	1	2	2														0.006	
H5	1	1	2	1	1	0.332	0.386	0.335	0.283	0.433	0.311	0.343	0.229	0.293	0.144	0.221	0.305	0.264	0.126	0.153
H6	1	1	2	1	2			0.010	0.050	0.037	0.063	0.020	0.062	0.155	0.031	0.040	0.052	0.039	0.051	0.023
H7	1	1	2	2	1	0.073	0.107	0.071	0.167	0.081	0.156	0.117	0.083	0.090	0.141	0.043	0.072	0.047	0.106	0.021
H8	1	1	2	2	2	0.075	0.074	0.043		0.092	0.025	0.049	0.125	0.096	0.140	0.010	0.004	0.040		0.033
H9	1	2	1	1	1			0.006					0.011			0.016	0.018		0.010	0.080
H10	1	2	1	1	2								0.000							0.004
H11	1	2	1	2	1	0.003											0.002			
H12	1	2	1	2	2															0.005
H13	1	2	2	1	1			0.006	0.006	0.052			0.006		0.010	0.050	0.020	0.003		0.013
H14	1	2	2	1	2										0.004	0.003				0.034
H15	1	2	2	2	1		0.013						0.004			0.001				
H16	1	2	2	2	2					0.010	0.023				0.006		0.005			
H17	2	1	1	1	1	0.094	0.021	0.099	0.108	0.011		0.070		0.010	0.021	0.087	0.028	0.159	0.091	0.091
H18	2	1	1	1	2		0.006		0.008	0.005			0.015	0.009	0.008	0.008			0.061	0.015
H19	2	1	1	2	1	0.004												0.005	0.052	0.053
H20	2	1	1	2	2															0.004
H21	2	1	2	1	1	0.347	0.319	0.238	0.205	0.156	0.352	0.206	0.204	0.195	0.206	0.297	0.320	0.179	0.198	0.095
H22	2	1	2	1	2		0.008	0.036	0.035		0.008	0.014	0.061	0.036	0.017	0.051	0.034	0.042	0.028	0.036
H23	2	1	2	2	1	0.040	0.014	0.045	0.058	0.045	0.007	0.104	0.066	0.046	0.064	0.013	0.027	0.051	0.036	0.011
H24	2	1	2	2	2	0.009	0.012	0.030		0.019	0.051	0.042	0.089	0.063	0.104		0.018	0.023		0.075
H25	2	2	1	1	1											0.009			0.010	0.060
H26	2	2	1	2	1															0.005
H27	2	2	1	2	2							0.005								0.007
H28	2	2	2	1	1			0.006		0.043			0.020		0.005	0.032	0.035			
H29	2	2	2	2	1												0.011			0.003
H30	2	2	2	2	2											0.017				

^aThe data of five POALINs in Han-Yunnan were published previously (Shi et al., 2014).

^bThe data of five POALINs in Japanese and Caucasians were published previously (Kulski et al., 2010, 2011).

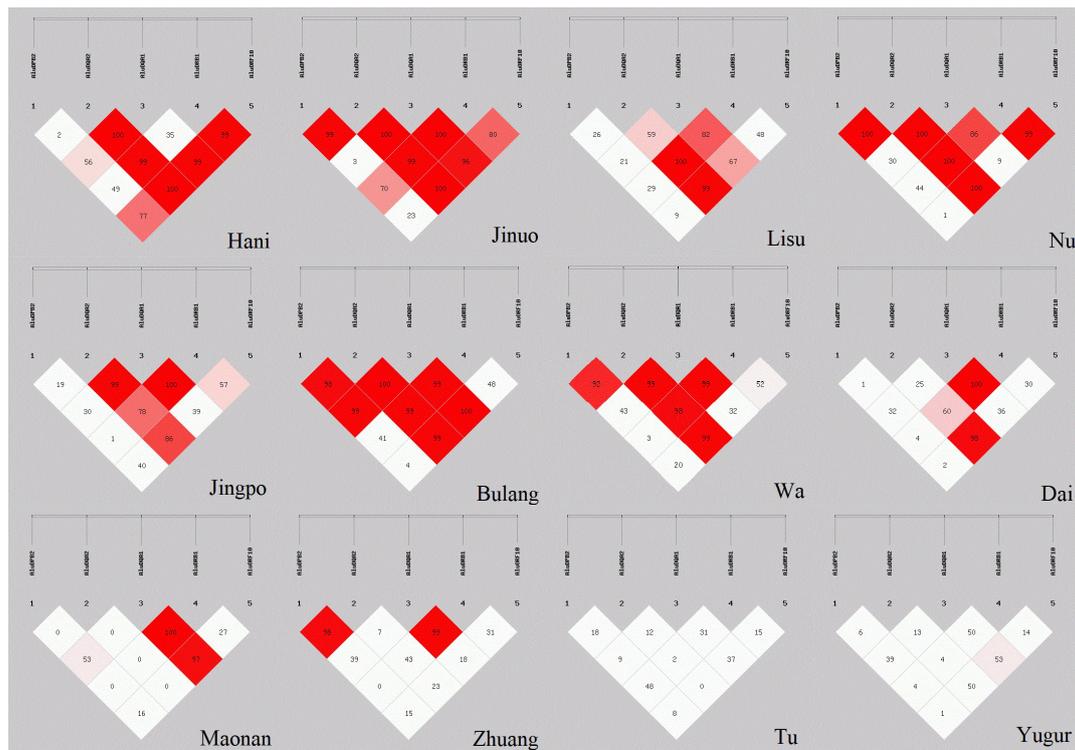


FIGURE 2 | LD estimations (D') among five POALINs within MHC II region for 12 Chinese ethnic populations.

with one to three Alu insertions; H5 (AluDQA1*2), H7 (AluDQA1*2: AluDRB1*2), H21 (AluDPB2*2: AluDQA1*2) and H23 (AluDPB2*2: AluDQA1*2: AluDRB1*2). The H5 (AluDQA1*2) and H21 (AluDPB2*2: AluDQA1*2) haplotypes were predominant in all 12 minority populations at frequency ranges of 0.144–0.433 and 0.158–0.352, respectively, which was the same as that for the Han-Yunnan, Japanese and Caucasians. There were seven haplotypes that were specific to only one particular population. These were three two-insertion haplotypes, three three-insertion haplotypes, and one five-insertion haplotype; H4 (AluDRB1*2: AluORF10*2) in the Japanese, H10 (AluDQA2*2: AluORF10*2), H12 (AluDQA2*2: AluDRB1*2: AluORF10*2), H20 (AluDPB2*2: AluDRB1*2: AluORF10*2), and H26 (AluDPB2*2: AluDQA2*2: AluDRB1*2) in Caucasians, H11 (AluDQA2*2: AluDRB1*2) in Hani, and H30 (AluDPB2*2: AluDQA2*2: AluDQA1*2: AluDRB1*2: AluORF10*2) in the Tu. The differences in significance between two populations for each haplotype frequency are shown in **Supplementary Table 5**.

The two most predominant haplotypes in all 15 populations were H5 (AluDPB2*1: AluDQA2*1: AluDQA1*2: AluDRB1*1: AluORF10*1) and H21 (AluDPB2*2: AluDQA2*1: AluDQA1*2: AluDRB1*1: AluORF10*1), both with the AluDQA1 insertion. Haplotype H6 (AluDPB2*1: AluDQA2*1: AluDQA1*2: AluDRB1*1: AluORF10*2) differentiated the Maonan from the other populations ($P < 0.01$ after Bonferroni's correction), whereas haplotype H7 (AluDPB2*1: AluDQA2*1: AluDQA1*2: AluDRB1*2: AluORF10*1) differentiated the

Caucasians from the other populations except for the Tu and Han-Yunnan ($P < 0.01$ after Bonferroni's correction). Also, haplotype H8 (AluDPB2*1: AluDQA2*1: AluDQA1*2: AluDRB1*2: AluORF10*2) differentiated the Caucasians from the other populations except from the Tu ($P < 0.01$ after Bonferroni's correction). The haplotype H18 (AluDPB2*2: AluDQA2*1: AluDQA1*1: AluDRB1*2: AluORF10*2) frequency was different between the Japanese and other populations but not from the Jinuo, Dai and Maonan ($P < 0.01$ after Bonferroni's correction). On the other hand, haplotype H19 (AluDPB2*2: AluDQA2*1: AluDQA1*1: AluDRB1*2: AluORF10*1) was observed only in four populations, with a significant difference obtained between Hani/Han-Yunnan and Japanese/Caucasians ($P < 0.01$ after Bonferroni's correction).

LD Analysis and Percentage Haplotypic Association Between POALINs and HLA Alleles

D' values for global LD between the five POALINs were calculated in twelve ethnic populations and are shown in **Figure 2**. LD values between the Alu loci were variable between the ethnic populations ranging from the absence of strong LD ($D' < 54\%$) between any of the Alu in the Yugur and Tu Mongolian populations to a strong LD ($D' > 0.8$) between four or five Alu in the Jinuo, Nu, Bulang and Wa. The Hani, Lisu and Jingpo had strong LD ($D' > 0.8$) between two or three

Alu insertions, whereas the Dai, Maonan and Zhuang of the ancient Baiyue tribe and the Daic subfamily language had only two Alu in strong LD.

Supplementary Table 6 shows the frequency of HLA-DRB1 alleles and class II POALINS and that the percentage associations between these POALIN and particular HLA-DRB1 alleles were at very high (80–100%), high (>50–79%), moderate (20–50%) and low (<20%) percentages. For example, all of the 19 HLA-DRB1 alleles in the Hani were associated with four of five of the Alu insertions at high to very high percentages: 16 alleles (except for HLA-DRB1*01:01, -DRB1*08:03 and -DRB1*10:01) associated at 67.3–100% with AluDQA1, 8 alleles (HLA-DRB1*01:01, -DRB1*04:03, -DRB1*08:03, -DRB1*09:01, -DRB1*12:01, -DRB1*14:01, -DRB1*14:04 and -DRB1*16:02) associated at 54.9–100% with AluDPB2, 4 alleles (HLA-DRB1*01:01, -DRB1*08:01, -DRB1*15:02 and -DRB1*15:04) associated at 57.1–100% with AluDRB1, and there was 100% association between AluDQA2 and HLA-DRB1*10:01, but at very low frequency (0.00336). The AluORF10 was associated with the Hani HLA-DRB1 alleles only at low to moderate levels.

Supplementary Table 7 shows a summary of the comparative percentage association between HLA-DRB1 alleles and the Alu class II POALINS in 12 ethnic populations (this study), Chinese Han in Yunnan (Shi et al., 2014), Japanese and Caucasians (Kulski et al., 2010) from previous studies. Overall, there was a strong similarity of haplotypic associations between AluDQA1 and HLA-DRB1 alleles in all fifteen populations.

Table 5 shows a summary of the percentage association between HLA-DRB1 alleles and AluDRB1. Overall, all the populations except for the Hani and the Dai have 83 to 100% association between the AluDRB1 insertion and HLA-DRB1*15 and HLA-DRB1*16. In comparison, the AluDRB1 insertion was linked to six of six homozygous cell lines with HLA-DRB1*01, seven of seven cell lines with -DRB1*16, 10 of 11 cell lines with -DRB1*15 and to none of the other 66 cell lines with nine other DRB1 lineage alleles (**Supplementary Table 2**). For the other Alu

insertions, HLA-DRB1*09 was not found in the Wa, but it had a moderate to very strong association (51–100%) with AluDPB2 in thirteen populations and a low association (31.7%) in the Lisu. For a comparison of the haplotypic associations with actual genomic sequence linkages, **Supplementary Table 2** shows the percentage linkage between these five POALIN with HLA-DRB1 alleles detected in the MHC class II haplotype sequences of 90 homozygous cell-lines (Kulski et al., 2021). Because of ancestral recombination at sites between various Alu loci and the DRB1 allelic loci, the linkages detected in the cell lines were not present in all the different Chinese ethnic populations, although the general trends are maintained between and within populations.

Phylogenetic Trees and PCA Plots

To compare the diversity of these ethnic populations, we constructed phylogenetic trees (**Figure 3**) and PCA plots (**Figure 4**) based on POALIN alleles, HLA-DRB1 alleles and DRB1-AluDRB1 haplotype frequencies. The topology for the NJ tree constructed using the DA of POALIN alleles (**Figure 3A**), revealed two distinct clusters: (1) the Dai, Zhuang and Maonan of the Daic subfamily in the Sino-Tibetan language family, and (2) the Jingpo of the Tibeto-Burman subfamily in the Sino-Tibetan language family with the Bulang stemming from the Wa, which are both part of the Mon-Khmer subfamily in the Austro-Asiatic language family. A third cluster was the stepwise grouping of Lisu, Nu, Hani and Jinuo of the Tibeto-Burman with the Mongolian Yugur of the Tujue subfamily in the Altaic language family inserted between the Hani and the Jinuo. The Han from Yunnan province grouped at the lower extremity of the 12 Chinese minority ethnic groups and away from the Japanese and the Caucasians that had grouped at the opposite end of the tree to that of the Daic cluster.

The topology of the NJ trees based on HLA-DRB1 allele frequencies and DRB1/AluDRB1 haplotypes were similar to each other (**Figures 3B,C**) and both revealed two distinct clusters: (1)

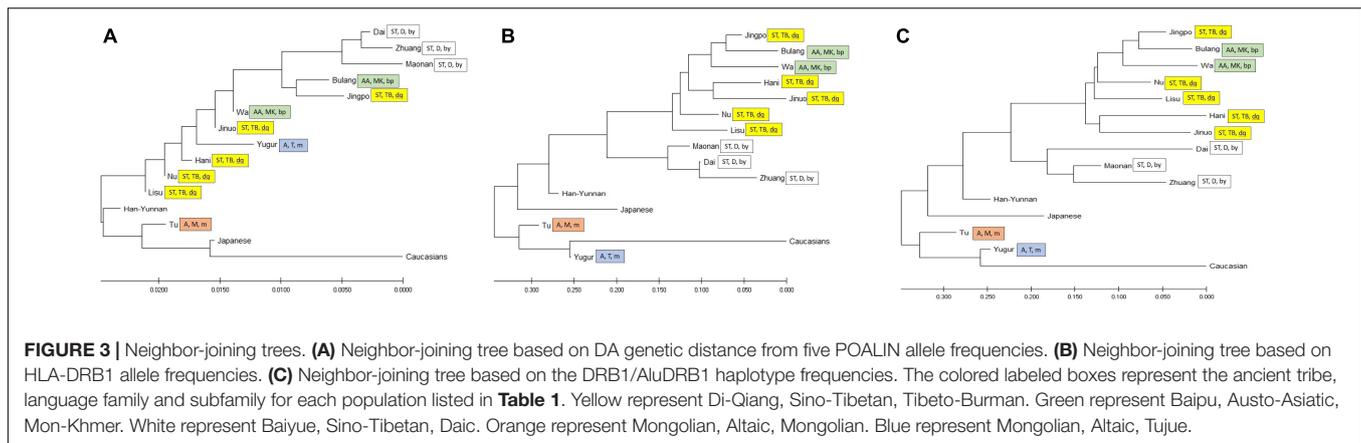
TABLE 5 | Percentage association between HLA-DRB1 alleles and AluDRB1.

DRB1 allele	Average of 10 ethnic groups ^a	Hani	Dai	Han-Yunnan ^b	Japanese ^c	Caucasian ^c
DRB1*01	64	68		100	75	100
DRB1*03	25		59	3		
DRB1*04	8		20	9	2	5
DRB1*07	11	50				
DRB1*08	10		50	6	5	
DRB1*09			13	3		
DRB1*10			100			50
DRB1*11			17			
DRB1*12	2		29	4		
DRB1*13	20		41	5		
DRB1*14	4	13	53	9		
DRB1*15	95	54	49	86	83	100
DRB1*16	94	50	26	87	100	100

^a10 ethnic groups include Jinuo, Lisu, Nu, Jingpo, Bulang, Wa, Maonan, Zhuang, Tu, and Yugur.

^bThe data in Han-Yunnan were published previously (Shi et al., 2014).

^cThe data in Japanese and Caucasian were published previously (Kulski et al., 2010, 2011).



the Dai, Zhuang and Maonan of the Daic subfamily in the Sino-Tibetan language family, and (2) the Bulang and Wa of the Mon-Khmer subfamily in the Austo-Asiatic language family separated from the Jingpo, Hani, Jinuo, Nu and Lisu group of the Tibeto-Burman subfamily in the Sino-Tibetan language family. The Han-Yunnan population grouped between the Chinese minority populations and the Japanese and at a genetic distance away from the Mongolian Tu and Yugur and the Caucasians. In this regard, the POALIN and HLA-DRB1 allele frequencies both grouped the 13 Chinese ethnic populations into their respective subfamilies and language families. The main exception was that the POALIN frequencies separated the Tu and Yugur at a greater distance from each other (**Figure 3A**), whereas the HLA-DRB1 allele frequencies placed them more closely together between the Japanese and the Caucasians (**Figures 3B,C**).

The PCA plots for the POALIN alleles (**Figure 4A**), HLA-DRB1 alleles (**Figure 4B**) and DRB1-AluDRB1 haplotypes (**Figure 4C**) showed that the distinct linguistic clusters of the 15 populations in each of four quadrants are similar to those revealed by the NJ trees (**Figure 3**). These plots have placed the Jingpo closer to the Mon-Khmer subfamily than to the Tibeto-Burman subfamily from which the Jingpo are believed to have originated, and the genetic distance between the Mongolian Tu and Yugur is greater for the POALIN alleles than the HLA-DRB1 alleles and DRB1-AluDRB1 haplotypes. Also, the Caucasians are the genetic outgroup in relation to the 13 Chinese ethnic populations and the Japanese.

DISCUSSION

In this study, we examined the genetic variations of the five POALIN and HLA-DRB1 allele and haplotype frequencies to further elucidate the association between the MHC class II POALIN and the classical HLA-DRB1 allele frequencies in 12 Chinese minority populations. The HLA-DRB1 alleles are used widely and commonly for assessing the genetic structure and differences within and between different populations (Di and Sanchez-Mazas, 2011; Sun et al., 2015; Weiskopf et al., 2016; Gonzalez-Galarza et al., 2020). The frequency of the HLA-DRB1 alleles within the 12 Chinese minority populations were similar

to previous reports (Ogata et al., 2007; Shi et al., 2008, 2010b, 2011; Sun et al., 2015; Tao et al., 2020). On the other hand, the previous studies on the distribution and frequency of the MHC class II POALIN dimorphisms were limited to only three populations, the Caucasian, Japanese (Kulski et al., 2010), and Chinese Han in Yunnan (Shi et al., 2014), and this published data provided the three outlying comparative populations for the present study. Therefore, we have provided new data on the POALIN frequencies for 12 Chinese minority populations that were selected for genetic analysis because of their culture, known ancient history and connection to five distinct language subfamilies, the Tibeto-Burman, Mon-Khmer, Daic, Mongolian and Tujue (**Table 1**).

Phylogenetic trees and PCA (**Figures 3, 4**) show that the Alu insertion dimorphism, HLA-DRB1 alleles and the DRB1-AluDRB1 haplotype diversity are associated strongly with the population characteristics of language family, migration and sociality. The Daic family, including the Dai, Zhuang and Maonan, always clustered closely together based on the POALIN dimorphisms, HLA-DRB1 alleles and HLA DRB1-AluDRB1 haplotypes. The Tibeto-Burman subfamily of the Jinuo, Hani, Lisu and Nu have certain shared population characteristics due to their migration from the north, and therefore are genetically closer to the Yugur and Tu northern populations, which belong to Mongolian tribal family. Surprisingly, the Jingpo from Tibeto-Burman subfamily are genetically closer to the Mon-Khmer family (Bulang and Wa) than to other populations from Tibeto-Burman subfamily probably because these three populations have long lived closely together in the mountains of the western part of Yunnan and have been infected by similar pathogens from the infectious environment. For example, malaria is a serious infectious disease prevalent in China since 2700 BC, and Yunnan Province is a high incidence area of malaria, especially in the border area between China and Myanmar (Cox, 2010; Bi et al., 2013; Diouf et al., 2014). Similarly, the Jinuo and Bulang who live closely together within this same area, also may have undergone high selective pressure from malaria.

The five different POALIN dimorphic frequencies provide unique evolutionary and genetic information on the relationships between the 12 Chinese minority populations. The frequencies of AluDPB2, AluDQA2 and AluDQA1 in the Jingpo had significant

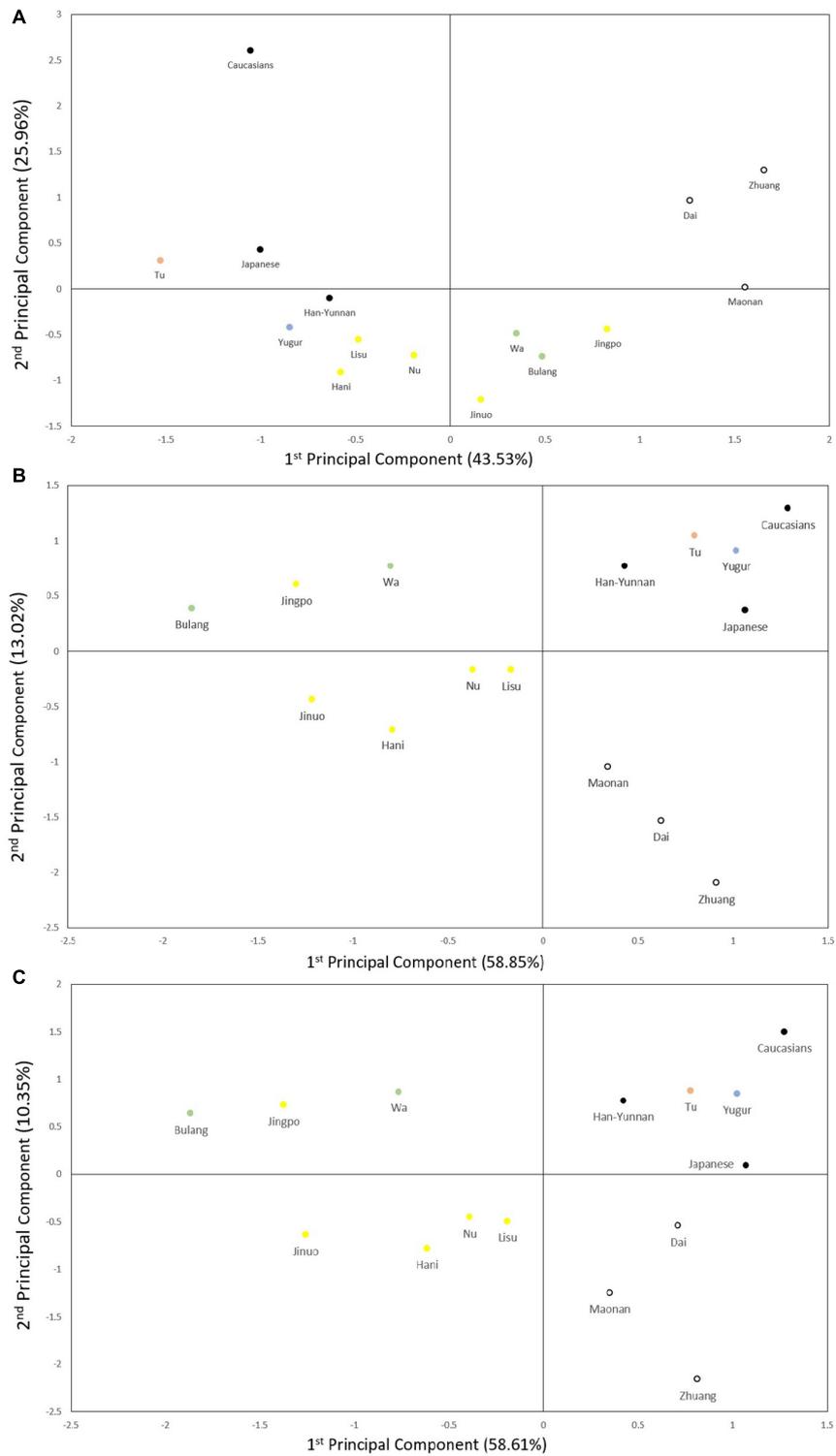


FIGURE 4 | Principal component analysis (PCA). **(A)** PCA based on five POALIN allele frequencies. Contributions of the first and second components were 43.53% and 25.96%, respectively. **(B)** PCA based on HLA-DRB1 allele frequencies. Contributions of first and second components were 58.85% and 13.02%. **(C)** PCA based on the DRB1/AluDRB1 haplotype frequencies. Contributions of first and second components were 58.61% and 10.35%. The colored dots represent the ancient tribe, language family and subfamily for each population listed in **Table 1**. Yellow represent Di-Qiang, Sino-Tibetan, Tibeto-Burman. Green represent Baiyu, Austro-Asiatic, Mon-Khmer. White represent Baiyue, Sino-Tibetan, Daic. Orange represent Mongolian, Altaic, Mongolian. Blue represent Mongolian, Altaic, Tujue.

differences with the other four populations ($P < 0.01$ after Bonferroni's correction) of the Tibeto-Burman subfamily. This suggests an expansion of these Alu insertions in the Jingpo people as a consequence of their different population histories or environmental effects. In comparison, the Bulang, a member of the Mon-Khmer family, had the highest POALIN frequency (0.995) for AluDQA1 in all 15 populations. This is the highest and closest to subpopulation genetic fixation for any of the MHC POALIN frequencies in world populations suggesting substantial long term population isolation. The frequencies of AluORF10 were higher in Dai, Maonan, and Zhuang (Daic subfamily in the Sino-Tibetan language family) than in the other nine Chinese minority populations. AluDQA1 was the highest POALIN frequency (0.777 and 0.903, respectively) in the Tu and the Yugur with a significant difference between these two populations ($P < 0.01$ after Bonferroni's correction). HLA-DRB1*09:01 had the strongest association (100%) with AluDQA1, and was the highest frequency (0.127 and 0.134) in the Tu and Yugur, respectively. According to historical records, all the Altaic language speaking groups such as the Tu and the Yugur who speak the Mongolian, Tujue, or Manchu-Tungusic sub-languages originated from the people and places overrun by the Mongol Empire and from the border adjacent to Northeastern China in the 13th century (Guo, 2000; Chu et al., 2006). HLA-DRB1*12:02 also had strong associations (88.7–100%) with AluDQA1 with a high frequency (0.160–0.550) in eight populations (Hani, Jinuo, Lisu, Nu, Jingpo, Bulang, Wa, and Maonan).

It is reported that the distribution of DRB1 allele frequencies for a Mongolian subpopulation in Yunnan was different to a Mongolian population of inner Mongolia and much closer to the Hani population of Yunnan (Sun et al., 2015). They hypothesized that the difference between the two Mongolian populations was due partly to gene flow and pathogen driven selection. We found a large differentiation between two Mongolian populations for the Alu alleles, but not for the HLA-DRB1 alleles. The Alu analysis placed the Mongolian Yugur within a cluster of the Di-Qiang subfamilies and at a substantial distance away from the Mongolian Tu, whereas the DRB1 allele frequencies for the two Mongolian populations placed them closer together at a genetic distance between the Japanese and Caucasians (Figures 3, 4). We attribute this difference between the two Mongolian populations for the Alu analysis mainly to a twofold difference in the AluDQA1*1 frequencies (Table 3). However, it is possible that the frequencies of particular DRB1 alleles of the two distinct Mongolian populations may have placed them closer together because of pathogen driven selection at that particular individual gene in contrast to the more independent and possibly less effective Alu loci. In this regard, the inheritance of identical by descent or identical by state genomic loci and/or haplotypes may in part be driven by selection, gene flow and various social and geographic factors, but has yet to be defined and investigated using a greater variety of different genomic markers for comparative analyses.

Overall, the branching patterns of the interrelationships between the populations and population clusters were similar for the Alu and DRB1 allelic frequencies, although the genetic

distances between particular populations were substantially different. Most of these similarities are likely due to the haplotypic characteristics between the Alu dimorphism and the DRB1 alleles (Kulski et al., 2010, 2021), as exemplified in this study with a comparison between the NJ trees of the HLA-DRB1 alleles and HLA-DRB1-AluDRB1 haplotypes (Figure 3). It is clear from this and previous studies that the closer the dimorphic Alu is to the HLA-DRB1 locus the stronger the haplotypic linkage/association and recombination resistance (Kulski et al., 2010, 2011, 2021). This seems to be the case for AluDRB1 that is most strongly associated with HLA-DRB1*15 and -DRB1*16 (Table 5) and is located within 14 kb of the HLA-DRB1 locus. In contrast, AluORF6 and AluDP2, which are 233 kb and 536 kb, respectively, from the HLA-DRB1 locus (Supplementary Figure 1), are associated with many different DRB1 alleles possibly because many more recombination events had occurred between their loci. The five genotyped and haplotyped 'lineage by descent' dimorphic Alu described in this study provide clues to the diversity of the MHC class II region of the 12 Chinese minority populations. However, further studies using fully phased genomic sequences of the MHC class II region within these historically small ethnic communities that are still strongly linked together by ancestry, culture and language might provide a better understanding of these POALIN haplotypic associations within the context of human MHC class II diversity, identity by descent (and/or by chance or state), haplotype shuffling and ancestral recombinations (Dawkins et al., 1999; Alper et al., 2006; Larsen et al., 2014).

The POALINs in the current study are all members of the young Alu subfamily, with AluDQA1 and AluDRB1 belonging to the AluY subgroup and AluDQA2, AluDPB2 and AluORF10 belonging to the youngest AluYa5 or AluYb8 subgroup (Kulski et al., 2010). AluDQA1 appears to be the oldest of the five POALINs on the basis of having the highest POALIN frequency in the 15 populations (Table 3) and its association with most of the HLA-DRB1 supertypes (Supplementary Table 7). Thus, the AluDQA1 insertion was distributed widely in the Chinese ethnic populations and associated strongly as a haplotype with all or most of the HLA-DRB1 alleles. The frequency of AluDQA2 was higher in the Caucasians than in the Chinese populations or Japanese. The hypothesis that AluDQA2 may have originated in Caucasians (Kulski et al., 2010; Shi et al., 2014) is confirmed by the present study.

The frequencies of AluDRB1 were the highest in the Dai, Maonan, Zhuang, which belong to Tibeto-Burman language subfamily. The AluDRB1 with the frequency range from 0.10–0.455 had a strong association with HLA-DRB1*15 and -DRB1*16 in most populations. However, there was a significantly lower % association of <55% between the AluDRB1 insertion and HLA-DRB1*15 or HLA-DRB1*16 in Hani and Dai compared to the other 11 Chinese ethnic groups including the Han-Yunnan, and the Japanese and Caucasians (Table 5). This could be due to primer mutation with allelic dropout, an AluDRB1 deletion, recombination events, or a high level of interbreeding among members of the population with the HLA-DRB1*15 or HLA-DRB1*16 haplotype that was missing the AluDRB1 insertion in the founding group. By comparison, the AluDRB1 insertion

is very much limited by linkage (or association) to the HLA-DRB1*01, -DRB1*10 (DR1 supertypes), -DRB1*15, and the -DRB1*16 (DR51 supertypes) allelic lineages, which occurred after their separation from the DR8, DR52 and DR53 supertypes (Kulski et al., 2010). On this basis, the AluDQA1 insertion must have happened much earlier than the AluDRB1 insertion during human evolution and population expansions. These results confirm that the AluDRB1 insertion probably originated in an ancestral HLA-DRB1 allele as a progenitor of the DR51 supertypes (Kulski et al., 2010), which contained HLA-DRB1*15 and -DRB1*16 (Andersson, 1998; Gibbons et al., 2004).

AluDPB2 has a frequency range from 0.278 to 0.574 in fifteen populations, with low- to high-level percentage associations with many different HLA-DRB1 alleles (**Supplementary Table 7**). This greater number of associations between AluDPB2 and HLA-DRB1 than between AluDRB1 and HLA-DRB1 is probably because the AluDPB2 locus is 536 kb from the HLA-DRB1 locus with the likelihood of numerous ancient recombination events occurring in between the two loci (**Supplementary Figure 1**; Kulski et al., 2021). The AluORF10 had a strong association with HLA-DRB1*15 only in Caucasians (89.1%). In contrast, the AluORF10 was associated strongly with HLA-DRB1*16 in eight East Asian populations (Jingpo, Wa, Maonan, Zhuang, Tu, Yugur, Han-Yunnan, and Japanese); whereas HLA-DRB1*16 was absent in the Caucasian population. This suggests at least one or more recombination events at an unidentified junction between the AluORF10 and HLA-DRB1 locations in the ancestral progenitors of the DR51 supertypes.

Although this study focused on Alu and HLA-DRB1 evolutionary genetic markers and population structure and was not related directly to medical or health issues, it is noteworthy that the Alu indels could have enhancer and other regulatory roles that affect the expression of HLA class II genes and/or other genes in the MHC and elsewhere in the human genome (Hasler and Strub, 2006; Moolhuijzen et al., 2010; Spirito et al., 2019; Goubert et al., 2020; Kulski et al., 2021). Many Alu elements of the AluJ, AluS, and AluY subfamilies are transcriptionally active with highly expressed self-cleaving ribozyme activity during T-cell activation and thermal and endoplasmic reticulum stress (Hernandez et al., 2020). Furthermore, Wang et al. (2017) identified two Alu indels Alu-5072 and Alu-5075 in the class II region as potential enhancers for HLA-DRB5, and HLA-DQB1-AS1 associated with phenotypes of lymphoma, Hodgkin lymphoma and chronic hepatitis B infection, respectively (Wang et al., 2017). In this regard, Alu-5057 is probably the AluDRB1 indel at the 5' end of HLA-DRB1 (**Supplementary Figure 1**). Thus, the question remains whether the other four Alu indels described in this study also have enhancer functions as Wang et al. (2017) reported for Alu-5072 and Alu-5075 (Wang et al., 2017). On the basis of these published findings, the transcriptional activity and role of Alu in the human MHC during epigenetic regulation needs to be investigated and better defined. Also, the Alu indels both as genotypes and haplotypes within the MHC could have important functions in cancer, autoimmunity and immunity to infections that have yet to be addressed and investigated.

We used a set of five-locus POALINs from the MHC class I region as lineage markers in a previous study to determine the haplotypic association and differentiation of MHC class I polymorphic Alu insertions and HLA-B/Cw alleles in seven Chinese ethnic populations (Yao et al., 2010). The POALIN markers that we used in this study were limited to five loci in the MHC class II region, but were a sufficient number to effectively micro-differentiate between 15 populations. The advantages of these POALIN lineage markers within the MHC class I and class II regions are their applicability – they are well defined, cheap to prepare and administer in the laboratory, and they produce results that are reasonably easy to interpret. In future work, the more widely studied MHC class I 5-loci POALINs (Yao et al., 2010; Abeid et al., 2019; Kulski et al., 2019), other autosomal Alu loci (Antunez-de-Mayolo et al., 2002), and STR loci (Garcia-Obregon et al., 2011) could be included in the haplotype analyses to broaden the genetic distances and diversity between and within the various populations.

In conclusion, the unique finding in this study, not previously reported, is that the MHC class II POALIN and HLA-DRB1 allele frequencies both grouped the 12 Chinese minority ethnic populations into their respective subfamilies and language families. When compared with the previously reported data of the Chinese Han in Yunnan, Japanese and Caucasians, it is evident that the POALINs in MHC class II, like the polymorphic class I and class II HLA genes, are informative genetic and haplotype markers, which can be used cheaply and simply in studies of population diversity, forensic medicine and disease research.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

AUTHOR CONTRIBUTIONS

LiS and YY conceived and designed the research. YC, SL, JY, YT, and XZ performed the experiments. YC and JK analyzed the data. SL and YT collected the samples. YC, YY, LiS, and JK wrote and revised the manuscript. All authors read and approved the final version of the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by the grant from the Yunnan Provincial Science and Technology Department (2008CC021) and the Special Funds for high-level health talents of Yunnan Province (D-201669, L-201615, and H-2018014). The funders had no role in the design of the study, data collection and analysis, decision to publish, or preparation of the manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2021.636236/full#supplementary-material>

Supplementary Figure 1 | Map of the location of the five POALIN within the MHC class II region. **(a)** Positions of the Alu indels AluORF10, AluDRB1, AluDQB1, AluDQA2, and AluDPB2 on the top horizontal line relative to the positions of the HLA class II genes. The horizontal arrows indicate 5' to 3' coding direction. **(b)** Magnification of the genomic location between the HLA-DRB1 and HLA-DQA1 genes to indicate the relative positions of the AluDRB1 and AluDQA1 insertions that are located ~13.7 kb and 36.1 kb from the 5' end of HLA-DRB1, respectively. It is a computer image taken from the online UCSC browser at <https://genome.ucsc.edu/cgi-bin/hgGateway> for chr6:32,565,856–32,632,194 covering 66,339 bp and including

representations of Curated RefSeq, gene expression profile from GTEx RNA-seq, ENCODE cCREs, and Repeat Elements by RepeatMasker and labeled to show the positions of the AluDRB1 and AluDQA1 insertions along the SINE row.

Supplementary Figure 2 | The Electrophoresis results for different sized PCR products of Five MHC POALINs. The POALIN alleles are dimorphic structures whereby the absence of the Alu insertion at the Alu locus is the Alu*1 allele and the presence of the Alu insertion is the Alu*2 allele. 1: the PCR products of AluORF10 1, 1; 2: the PCR products of AluORF10 1, 2; 3: the PCR products of AluORF10 2, 2; 4: the PCR products of AluDQA1 2, 2; 5: the PCR products of AluDQA1 1, 2; 6: the PCR products of AluDQA1 1, 1; 7: the PCR products of AluDPB2 2, 2; 8: the PCR products of AluDPB2 1, 2; 9: the PCR products of AluDPB2 1, 1; 10: the PCR products of AluDRB1 1, 1; 11: the PCR products of AluDRB1 1, 2; 12: the PCR products of AluDRB1 2, 2; 13: the PCR products of AluDQA2 1, 1; 14: the PCR products of AluDQA2 1, 2; 15: the PCR products of AluDQA2 2, 2; M1: 1000bp Marker; M2: 2000bp Marker.

REFERENCES

- Abeid, S. N., Motrane, M., Farhane, H., and Harich, N. (2019). Alu elements within the human major histocompatibility class I region in the Comoros Islands: genetic variation and population relationships. *Ann. Hum. Biol.* 46, 169–174. doi: 10.1080/03014460.2019.1620854
- Alper, C. A., Larsen, C. E., Dubey, D. P., Awdeh, Z. L., Fici, D. A., and Yunis, E. J. (2006). The haplotype structure of the human major histocompatibility complex. *Hum. Immunol.* 67, 73–84. doi: 10.1016/j.humimm.2005.11.006
- Andersson, G. (1998). Evolution of the human HLA-DR region. *Front. Biosci.* 3:d739–d745. doi: 10.2741/a317
- Antunez-de-Mayolo, G., Antunez-de-Mayolo, A., Antunez-de-Mayolo, P., Papiha, S. S., Hammer, M., Yunis, J. J., et al. (2002). Phylogenetics of worldwide human populations as determined by polymorphic Alu insertions. *Electrophoresis* 23, 3346–3356. doi: 10.1002/1522-2683(200210)23:19<3346::AID-ELPS3346<3.0.CO;2-J
- Batzler, M. A., and Deininger, P. L. (2002). Alu repeats and human genomic diversity. *Nat. Rev. Genet.* 3, 370–379. doi: 10.1038/nrg798
- Bennett, E. A., Coleman, L. E., Tsui, C., Pittard, W. S., and Devine, S. E. (2004). Natural genetic variation caused by transposable elements in humans. *Genetics* 168, 933–951. doi: 10.1534/genetics.104.031757
- Bi, Y., Yu, W., Hu, W., Lin, H., Guo, Y., Zhou, X. N., et al. (2013). Impact of climate variability on Plasmodium vivax and Plasmodium falciparum malaria in Yunnan Province, China. *Parasit. Vectors* 6:357. doi: 10.1186/1756-3305-6-357
- Chaplin, D. D. (2010). Overview of the immune response. *J. Allergy Clin. Immunol.* 125(2 Suppl. 2), S3–S23. doi: 10.1016/j.jaci.2009.12.980
- Chu, J. Y. J., Huang, X., and Sun, H. (2006). “China Nationalities,” in *Genetic Diversity in Chinese Populations*, eds L. Jin and J. Chu (Shanghai: Shanghai Science and Technology Press).
- Cox, F. E. (2010). History of the discovery of the malaria parasites and their vectors. *Parasit Vectors* 3, 5. doi: 10.1186/1756-3305-3-5
- Dawkins, R., Leelayuwat, C., Gaudieri, S., Tay, G., Hui, J., Cattle, S., et al. (1999). Genomics of the major histocompatibility complex: haplotypes, duplication, retroviruses and disease. *Immunol. Rev.* 167, 275–304. doi: 10.1111/j.1600-065x.1999.tb01399.x
- Deininger, P. L., and Batzler, M. A. (1999). Alu repeats and human disease. *Mol. Genet. Metab.* 67, 183–193. doi: 10.1006/mgme.1999.2864
- Di, D., and Sanchez-Mazas, A. (2011). Challenging views on the peopling history of East Asia: the story according to HLA markers. *Am. J. Phys. Anthropol.* 145, 81–96. doi: 10.1002/ajpa.21470
- Diouf, G., Kpanyen, P. N., Tokpa, A. F., and Nie, S. (2014). Changing landscape of malaria in China: progress and feasibility of malaria elimination. *Asia Pac. J. Public Health* 26, 93–100. doi: 10.1177/1010539511424594
- Dunn, D. S., Choy, M. K., Phipps, M. E., and Kulski, J. K. (2007). The distribution of major histocompatibility complex class I polymorphic Alu insertions and their associations with HLA alleles in a Chinese population from Malaysia. *Tissue Antigens* 70, 136–143. doi: 10.1111/j.1399-0039.2007.00868.x
- Dunn, D. S., Naruse, T., Inoko, H., and Kulski, J. K. (2002). The association between HLA-A alleles and young Alu dimorphisms near the HLA-J, -H, and -F genes in workshop cell lines and Japanese and Australian populations. *J. Mol. Evol.* 55, 718–726. doi: 10.1007/s00239-002-2367-4
- Dunn, D. S., Ota, M., Inoko, H., and Kulski, J. K. (2003). Association of MHC dimorphic Alu insertions with HLA class I and MIC genes in Japanese HLA-B48 haplotypes. *Tissue Antigens* 62, 259–262. doi: 10.1034/j.1399-0039.2003.00092.x
- Dunn, D. S., Romphruk, A. V., Leelayuwat, C., Bellgard, M., and Kulski, J. K. (2005a). Polymorphic Alu insertions and their associations with MHC class I alleles and haplotypes in the northeastern Thais. *Ann. Hum. Genet.* 69(Pt 4), 364–372. doi: 10.1046/j.1529-8817.2005.00183.x
- Dunn, D. S., Tait, B. D., and Kulski, J. K. (2005b). The distribution of polymorphic Alu insertions within the MHC class I HLA-B7 and HLA-B57 haplotypes. *Immunogenetics* 56, 765–768. doi: 10.1007/s00251-004-0745-3
- Garcia-Obregon, S., Alfonso-Sanchez, M. A., Gomez-Perez, L., Perez-Miranda, A. M., Arroyo, D., de Pancorbo, M. M., et al. (2011). Microsatellites and Alu elements from the human MHC in Valencia (Spain): analysis of genetic relationships and linkage disequilibrium. *Int. J. Immunogenet.* 38, 483–491. doi: 10.1111/j.1744-313X.2011.01037.x
- Gibbons, R., Dugaiczky, L. J., Girke, T., Duistermars, B., Zielinski, R., and Dugaiczky, A. (2004). Distinguishing humans from great apes with AluYb8 repeats. *J. Mol. Biol.* 339, 721–729. doi: 10.1016/j.jmb.2004.04.033
- Gonzalez-Galarza, F. F., McCabe, A., Santos, E., Jones, J., Takeshita, L., Ortega-Rivera, N. D., et al. (2020). Allele frequency net database (AFND) 2020 update: gold-standard data classification, open access genotype data and new query tools. *Nucleic Acids Res.* 48, D783–D788. doi: 10.1093/nar/gkz1029
- Goubert, C., Zevallos, N. A., and Feschotte, C. (2020). Contribution of unfixed transposable element insertions to human regulatory variation. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 375:20190331. doi: 10.1098/rstb.2019.0331
- Guo, D. D. J. (2000). *Summarization of Chinese Nationalities in Zhonghua mingzu zhishi tonglan*. Kunming: Yunnan Education Press.
- Guo, S. W., and Thompson, E. A. (1992). Performing the exact test of Hardy-Weinberg proportion for multiple alleles. *Biometrics* 48, 361–372.
- Hasler, J., and Strub, K. (2006). Alu elements as regulators of gene expression. *Nucleic Acids Res.* 34, 5491–5497. doi: 10.1093/nar/gkl706
- Hernandez, A. J., Zovoilis, A., Cifuentes-Rojas, C., Han, L., Bujisic, B., and Lee, J. T. (2020). B2 and ALU retrotransposons are self-cleaving ribozymes whose activity is enhanced by EZH2. *Proc. Natl. Acad. Sci. U S A* 117, 415–425. doi: 10.1073/pnas.1917190117
- Jorde, L. B., Watkins, W. S., Bamshad, M. J., Dixon, M. E., Ricker, C. E., Seielstad, M. T., et al. (2000). The distribution of human genetic diversity: a comparison of mitochondrial, autosomal, and Y-chromosome data. *Am. J. Hum. Genet.* 66, 979–988. doi: 10.1086/302825
- Kulski, J. K., and Dunn, D. S. (2005). Polymorphic Alu insertions within the Major Histocompatibility Complex class I genomic region: a brief review. *Cytogenet. Genome Res.* 110, 193–202. doi: 10.1159/000084952
- Kulski, J. K., Gaudieri, S., and Dawkins, R. L. (2000). Using alu J elements as molecular clocks to trace the evolutionary relationships between duplicated HLA class I genomic segments. *J. Mol. Evol.* 50, 510–519. doi: 10.1007/s002390010054

- Kulski, J. K., Gaudieri, S., Martin, A., and Dawkins, R. L. (1999). Coevolution of PERB11 (MIC) and HLA class I genes with HERV-16 and retroelements by extended genomic duplication. *J. Mol. Evol.* 49, 84–97. doi: 10.1007/pl00006537
- Kulski, J. K., Mawart, A., Marie, K., Tay, G. K., and AlSafar, H. S. (2019). MHC class I polymorphic Alu insertion (POALIN) allele and haplotype frequencies in the Arabs of the United Arab Emirates and other world populations. *Int. J. Immunogenet.* 46, 247–262. doi: 10.1111/iji.12426
- Kulski, J. K., Shigenari, A., and Inoko, H. (2011). Genetic variation and hitchhiking between structurally polymorphic Alu insertions and HLA-A, -B, and -C alleles and other retroelements within the MHC class I region. *Tissue Antigens* 78, 359–377. doi: 10.1111/j.1399-0039.2011.01776.x
- Kulski, J. K., Shigenari, A., Shiina, T., and Inoko, H. (2010). Polymorphic major histocompatibility complex class II Alu insertions at five loci and their association with HLA-DRB1 and -DQB1 in Japanese and Caucasians. *Tissue Antigens* 76, 35–47. doi: 10.1111/j.1399-0039.2010.01465.x
- Kulski, J. K., Suzuki, S., and Shiina, T. (2020). SNP-Density Crossover Maps of Polymorphic Transposable Elements and HLA Genes Within MHC Class I Haplotype Blocks and Junction. *Front. Genet.* 11:594318. doi: 10.3389/fgene.2020.594318
- Kulski, J. K., Suzuki, S., and Shiina, T. (2021). Haplotype Shuffling and Dimorphic Transposable Elements in the Human Extended MHC Class II Region. *Front. Genet.* 2021:665899. doi: 10.3389/fgene.2021.665899
- Lancaster, A., Nelson, M. P., Meyer, D., Single, R. M., and Thomson, G. (2003). PyPop: a software framework for population genomics: analyzing large-scale multi-locus genotype data. *Pac. Symp. Biocomput.* 2003, 514–525.
- Lancaster, A. K., Single, R. M., Solberg, O. D., Nelson, M. P., and Thomson, G. (2007). PyPop update—a software pipeline for large-scale multilocus population genomics. *Tissue Antigens* 69(Suppl. 1), 192–197. doi: 10.1111/j.1399-0039.2006.00769.x
- Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., et al. (2001). Initial sequencing and analysis of the human genome. *Nature* 409, 860–921. doi: 10.1038/35057062
- Larsen, C. E., Alford, D. R., Trautwein, M. R., Jalloh, Y. K., Tarnacki, J. L., Kunnenkeri, S. K., et al. (2014). Dominant sequences of human major histocompatibility complex conserved extended haplotypes from HLA-DQA2 to DAXX. *PLoS Genet.* 10:e1004637. doi: 10.1371/journal.pgen.1004637
- Manczinger, M., Boross, G., Kemeny, L., Muller, V., Lenz, T. L., Papp, B., et al. (2019). Pathogen diversity drives the evolution of generalist MHC-II alleles in human populations. *PLoS Biol.* 17:e3000131. doi: 10.1371/journal.pbio.3000131
- Mastana, S. S., Bhatti, J. S., Singh, P., Wiles, A., and Holland, J. (2017). Genetic variation of MHC Class I polymorphic Alu insertions (POALINs) in three sub-populations of the East Midlands, UK. *Ann. Hum. Biol.* 44, 562–567. doi: 10.1080/03014460.2017.1302507
- Meyer, D., Single, R. M., Mack, S. J., Erlich, H. A., and Thomson, G. (2006). Signatures of demographic history and natural selection in the human major histocompatibility complex Loci. *Genetics* 173, 2121–2142. doi: 10.1534/genetics.105.052837
- Mnukova-Fajdelova, M., Satta, Y., O'Huigin, C., Mayer, W. E., Figueroa, F., and Klein, J. (1994). Alu elements of the primate major histocompatibility complex. *Mamm. Genome* 5, 405–415. doi: 10.1007/BF00357000
- Moolhuijzen, P., Kulski, J. K., Dunn, D. S., Schibeci, D., Barrero, R., Gojobori, T., et al. (2010). The transcript repeat element: the human Alu sequence as a component of gene networks influencing cancer. *Funct. Integr. Genomics* 10, 307–319. doi: 10.1007/s10142-010-0168-1
- Nei, M. (1973). Analysis of gene diversity in subdivided populations. *Proc. Natl. Acad. Sci. U S A* 70, 3321–3323. doi: 10.1073/pnas.70.12.3321
- Nei, M. (1978). Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics* 89, 583–590.
- Norman, P. J., Norberg, S. J., Guethlein, L. A., Nemat-Gorgani, N., Royce, T., Wroblewski, E. E., et al. (2017). Sequences of 95 human MHC haplotypes reveal extreme coding variation in genes other than highly polymorphic HLA class I and II. *Genome Res.* 27, 813–823. doi: 10.1101/gr.213538.116
- Ogata, S., Shi, L., Matsushita, M., Yu, L., Huang, X. Q., Shi, L., et al. (2007). Polymorphisms of human leucocyte antigen genes in Maonan people in China. *Tissue Antigens* 69, 154–160. doi: 10.1111/j.1399-0039.2006.00698.x
- Pierini, F., and Lenz, T. L. (2018). Divergent Allele Advantage at Human MHC Genes: Signatures of Past and Ongoing Selection. *Mol. Biol. Evol.* 35, 2145–2158. doi: 10.1093/molbev/msy116
- Ray, D. A., Walker, J. A., and Batzer, M. A. (2007). Mobile element-based forensic genomics. *Mutat. Res.* 616, 24–33. doi: 10.1016/j.mrfmmm.2006.11.019
- Robinson, J., Barker, D. J., Georgiou, X., Cooper, M. A., Flicek, P., and Marsh, S. G. E. (2020). IPD-IMGT/HLA Database. *Nucleic Acids Res.* 48, D948–D955. doi: 10.1093/nar/gkz950
- Sanchez-Mazas, A., Buhler, S., and Nunes, J. M. (2013). A new HLA map of Europe: Regional genetic variation and its implication for peopling history, disease-association studies and tissue transplantation. *Hum. Hered.* 76, 162–177. doi: 10.1159/000360855
- Sanchez-Mazas, A., and Meyer, D. (2014). The relevance of HLA sequencing in population genetics studies. *J. Immunol. Res.* 2014:971818. doi: 10.1155/2014/971818
- Sanchez-Mazas, A., Nunes, J. M., Middleton, D., Sauter, J., Buhler, S., McCabe, A., et al. (2017). Common and well-documented HLA alleles over all of Europe and within European sub-regions: A catalogue from the European Federation for Immunogenetics. *HLA* 89, 104–113. doi: 10.1111/tan.12956
- Shi, L., Huang, X. Q., Shi, L., Tao, Y. F., Yao, Y. F., Yu, L., et al. (2011). HLA polymorphism of the Zhuang population reflects the common HLA characteristics among Zhuang-Dong language-speaking populations. *J. Zhejiang Univ. Sci. B* 12, 428–435. doi: 10.1631/jzus.B1000285
- Shi, L., Kulski, J. K., Zhang, H., Dong, Z., Cao, D., Zhou, J., et al. (2014). Association and differentiation of MHC class I and II polymorphic Alu insertions and HLA-A, -B, -C and -DRB1 alleles in the Chinese Han population. *Mol. Genet. Genomics* 289, 93–101. doi: 10.1007/s00438-013-0792-2
- Shi, L., Ogata, S., Yu, J. K., Ohashi, J., Yu, L., Shi, L., et al. (2008). Distribution of HLA alleles and haplotypes in Jinuo and Wa populations in Southwest China. *Hum. Immunol.* 69, 58–65. doi: 10.1016/j.humimm.2007.11.007
- Shi, L., Shi, L., Yao, Y. F., Matsushita, M., Yu, L., Huang, X. Q., et al. (2010a). Genetic link among Hani, Bulang and other Southeast Asian populations: evidence from HLA -A, -B, -C, -DRB1 genes and haplotypes distribution. *Int. J. Immunogenet.* 37, 467–475. doi: 10.1111/j.1744-313X.2010.00949.x
- Shi, L., Xu, S. B., Ohashi, J., Sun, H., Yu, J. K., Huang, X. Q., et al. (2006). HLA-A, HLA-B, and HLA-DRB1 alleles and haplotypes in Naxi and Han populations in southwestern China (Yunnan province). *Tissue Antigens* 67, 38–44. doi: 10.1111/j.1399-0039.2005.00526.x
- Shi, L., Yao, Y. F., Shi, L., Matsushita, M., Yu, L., Lin, Q. K., et al. (2010b). HLA alleles and haplotypes distribution in Dai population in Yunnan province, Southwest China. *Tissue Antigens* 75, 159–165. doi: 10.1111/j.1399-0039.2009.01407.x
- Shi, Y. Y., and He, L. (2005). SHEsis, a powerful software platform for analyses of linkage disequilibrium, haplotype construction, and genetic association at polymorphism loci. *Cell Res.* 15, 97–98. doi: 10.1038/sj.cr.7290272
- Shiina, T., Hosomichi, K., Inoko, H., and Kulski, J. K. (2009). The HLA genomic loci map: expression, interaction, diversity and disease. *J. Hum. Genet.* 54, 15–39. doi: 10.1038/jhg.2008.5
- Shiina, T., Inoko, H., and Kulski, J. K. (2004). An update of the HLA genomic region, locus information and disease associations: 2004. *Tissue Antigens* 64, 631–649. doi: 10.1111/j.1399-0039.2004.00327.x
- Singh, G., Sandhu, H. S., Sharma, R., Srinivas, Y., Matharoo, K., Singh, M., et al. (2019). Genetic variation and population structure of five ethnic groups from Punjab, North-West India: Analysis of MHC class I polymorphic Alu insertions (POALINs). *Gene* 701, 173–178. doi: 10.1016/j.gene.2019.03.057
- Spirito, G., Mangoni, D., Sanges, R., and Gustincich, S. (2019). Impact of polymorphic transposable elements on transcription in lymphoblastoid cell lines from public data. *BMC Bioinformatics* 20(Suppl. 9):495. doi: 10.1186/s12859-019-3113-x
- Stewart, C. A., Horton, R., Allcock, R. J., Ashurst, J. L., Atrazhev, A. M., Coggill, P., et al. (2004). Complete MHC haplotype sequencing for common disease gene mapping. *Genome Res.* 14, 1176–1187. doi: 10.1101/gr.2188104
- Sun, H., Yang, Z., Lin, K., Liu, S., Huang, K., Wang, X., et al. (2015). The Adaptive Change of HLA-DRB1 Allele Frequencies Caused by Natural Selection in a Mongolian Population That Migrated to the South of China. *PLoS One* 10:e0134334. doi: 10.1371/journal.pone.0134334
- Svensson, A. C., Setterblad, N., Pihlgren, U., Rask, L., and Andersson, G. (1996). Evolutionary relationship between human major histocompatibility complex HLA-DR haplotypes. *Immunogenetics* 43, 304–314. doi: 10.1007/BF02440998

- Tamura, K., Dudley, J., Nei, M., and Kumar, S. (2007). MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol. Biol. Evol.* 24, 1596–1599. doi: 10.1093/molbev/msm092
- Tao, Y., Shi, L., Liu, S., Yao, Y., and Shi, L. (2020). Distribution of HLA-A, HLA-B, HLA-C, and HLA-DRB1 alleles and haplotypes in Jingpo minority in Yunnan province of China. *Hum. Immunol.* 81, 267–268. doi: 10.1016/j.humimm.2020.04.008
- Traherne, J. A. (2008). Human MHC architecture and evolution: implications for disease association studies. *Int. J. Immunogenet.* 35, 179–192. doi: 10.1111/j.1744-313X.2008.00765.x
- Trowsdale, J. (2011). The MHC, disease and selection. *Immunol. Lett.* 137, 1–8. doi: 10.1016/j.imlet.2011.01.002
- Vandiedonck, C., and Knight, J. C. (2009). The human Major Histocompatibility Complex as a paradigm in genomics research. *Brief Funct. Genomic Proteomic* 8, 379–394. doi: 10.1093/bfpg/elp010
- Wang, L., Norris, E. T., and Jordan, I. K. (2017). Human Retrotransposon Insertion Polymorphisms Are Associated with Health and Disease via Gene Regulatory Phenotypes. *Front. Microbiol.* 8:1418. doi: 10.3389/fmicb.2017.01418
- Watkins, W. S., Ricker, C. E., Bamshad, M. J., Carroll, M. L., Nguyen, S. V., Batzer, M. A., et al. (2001). Patterns of ancestral human diversity: an analysis of Alu-insertion and restriction-site polymorphisms. *Am. J. Hum. Genet.* 68, 738–752. doi: 10.1086/318793
- Weiskopf, D., Angelo, M. A., Grifoni, A., O'Rourke, P. H., Sidney, J., Paul, S., et al. (2016). HLA-DRB1 Alleles Are Associated With Different Magnitudes of Dengue Virus-Specific CD4+ T-Cell Responses. *J. Infect. Dis.* 214, 1117–1124. doi: 10.1093/infdis/jiw309
- Yao, Y., Shi, L., Shi, L., Kulski, J. K., Chen, J., Liu, S., et al. (2010). The association and differentiation of MHC class I polymorphic Alu insertions and HLA-B/Cw alleles in seven Chinese populations. *Tissue Antigens* 76, 194–207. doi: 10.1111/j.1399-0039.2010.01499.x
- Yao, Y., Shi, L., Shi, L., Lin, K., Tao, Y., Yu, L., et al. (2009). Polymorphic Alu insertions and their associations with MHC class I alleles and haplotypes in Han and Jinuo populations in Yunnan Province, southwest of China. *J. Genet. Genomics* 36, 51–58. doi: 10.1016/S1673-8527(09)60006-0
- Yao, Y., Shi, L., Tao, Y., Kulski, J. K., Lin, K., Huang, X., et al. (2012). Distinct HLA allele and haplotype distributions in four ethnic groups of China. *Tissue Antigens* 80, 452–461. doi: 10.1111/tan.12007
- You, Z. (1994). *History of Yunnan Nationalities in Yunnan Mingzu Shi*. Kunming: Yunnan University Press.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Cun, Shi, Kulski, Liu, Yang, Tao, Zhang, Shi and Yao. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.