Forensic Science International: Genetics 6 (2012) 437–446

Contents lists available at SciVerse ScienceDirect

Forensic Science International: Genetics

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

Y-chromosomal microsatellite diversity in three culturally defined regions of historical Tibet

Tenzin Gayden [a,b,1], Areej Bukhari [a,b,1], Shilpa Chennakrishnaiah [a,c], Oliver Stojkovic [d], Rene J. Herrera [a,*]

[a] Department of Molecular and Human Genetics, College of Medicine, Florida International University, Miami, FL 33199, USA
[b] Department of Biological Sciences, Florida International University, Miami, FL 33199, USA
[c] Department of Chemistry and Biochemistry, Florida International University, Miami, FL 33199, USA
[d] Institute of Forensic Medicine, Faculty of Medicine, University of Belgrade, Deligradska 31, 11000 Belgrade, Serbia

ARTICLE INFO

Article history:
Received 25 April 2011
Received in revised form 8 August 2011
Accepted 13 September 2011

Keywords:
Y-STR
Yfiler
Tibet
U-Tsang
Amdo
Kham

ABSTRACT

In the present study, we analyzed 17 Y-STR loci in 350 Tibetan males from three culturally defined regions of historical Tibet: Amdo (88), Kham (109) and U-Tsang (153). A total of 299 haplotypes were observed, 272 (90.9%) of which were unique. Only one Y-STR profile is shared across the three Tibetan groups and, incidentally, is also the most frequent haplotype (4.0%), represented by two, five and seven individuals from U-Tsang, Kham and Amdo, respectively. The overall haplotype diversity for the three Tibetan populations at 17 Y-STR loci was 0.9978 and the corresponding values for the extended (11-loci) and minimal (9-loci) haplotypes were 0.9935 and 0.9909, respectively. Both neighbor-joining and *Rst* pairwise analyses suggest a close genetic relationship between the Amdo and Kham populations, while U-Tsang is genetically distinct from the aforementioned groups. The results demonstrate that the 17 Y-STR loci analyzed are highly polymorphic in all three Tibetan populations examined and hence useful for forensic cases, paternity testing and population genetic studies.

1. Introduction

The Tibetan Plateau in Central Asia has remained relatively isolated throughout history mostly due to its encapsulation on three sides by the highest mountain ranges in the world, including the KunLun and Tang La ranges to the north, the Karakoram and Ladakh Mountains to the west, and the Himalayas in the south. A break in the mountainous terrain to the east serves as a narrow migratory route in and out of the region. These unique geographical features have played an important role in shaping the genetic landscapes of the Tibetan populations. Previous studies, for example, have revealed that the Himalayan Mountain range acts as a biased bidirectional barrier to gene flow, limiting genetic influence from the Indian subcontinent [1,2].

Prior to the Chinese invasion in 1959, Tibet covered an area of about 2.5 million km^2, encompassing three main provinces: Amdo in the northeast, Kham in the southeast and U-Tsang in west and central Tibet. In 1965, China created the Tibet Autonomous Region (TAR), reducing Tibet's area to 1.2 million km^2, and assimilating parts of Amdo and Kham into the adjacent Chinese provinces of Qinghai, Gansu, Sichuan and Yunnan [3]. Although the official language of Tibet is Chinese, the majority of the population speaks the native Tibetan language, which belongs to the Tibeto-Burman subgroup of the Sino-Tibetan family. The people of Amdo, Kham, and U-Tsang each speak a different dialect of Tibetan. They are united, however, by their devout practice of Buddhism, introduced to the country in the seventh century C.E. during the rule of Songtsen Gampo [4].

Archaeological records indicate late Paleolithic inhabitation of the Tibetan plateau [3], while Y-chromosomal data [1,5,6] suggest that the peopling of the highland occurred during the Neolithic period. Recent articles on mtDNA genome diversity in Tibetan populations [7,8] revealed evidence of successful late Paleolithic settlement on the plateau, thereby bridging the gap between the findings from genetic and archaeological studies.

The human Y-chromosome is a powerful molecular tool for forensic and population genetic studies [9]. In addition to Single Nucleotide Polymorphism (SNP) and insertion/deletion (indel) sites, the non-recombining region of the Y-chromosome (NRY) contains a number of short tandem repeat (STR) loci, which, besides their forensic applications, are now used to investigate the evolution, migration and genetic diversity of modern human populations [10–16]. In the present study, we report the allelic frequencies for 17 Y-STR loci (DYS19, DYS389I, DYS389II, DYS390, DYS391, DYS392, DYS393, DYS385a/b, DYS437, DYS438, DYS439, DYS448, DYS456, DYS458, DYS635 and GATA H4) of three Tibetan populations. The

* Corresponding author at: Department of Molecular and Human Genetics, College of Medicine, Florida International University, University Park, OE 304, Miami, FL 33199, USA. Tel.: +1 305 348 1258; fax: +1 305 348 1259.
E-mail address: herrerar@fiu.edu (R.J. Herrera).
[1] These two authors contributed equally to the study.

1872-4973/$ – see front matter © 2011 Elsevier Ireland Ltd. All rights reserved.
doi:10.1016/j.fsigen.2011.09.002

aim of the current investigation is to evaluate the forensic and population genetic applications of the aforementioned 17 Y-chromosomal microsatellite loci in three different populations from historical Tibet, namely Amdo ($N = 88$), Kham ($N = 109$) and U-Tsang ($N = 153$). In addition, the data from this study were compared with previously published, geographically targeted reference populations from the Himalayas, South Central Asia, Southeast Asia, Central Asia and Northeast Asia using the 11-loci extended haplotypes (DYS19, DYS389I, DYS389II, DYS390, DYS391, DYS392, DYS393, DYS385a/b, DYS438, and DYS439) to assess genetic relationships among them.

2. Materials and methods

2.1. Sample collection and DNA isolation

Blood samples were collected with informed consent from 350 unrelated Tibetan males from three culturally defined regions, namely Amdo ($N = 88$), Kham ($N = 109$) and U-Tsang ($N = 153$). Genealogical history for at least two generations was recorded for each donor. Samples were collected in accordance with the ethical guidelines specified by the institutions involved in this study. DNA was extracted using the standard phenol-chloroform method, ethanol-precipitated as described previously [17] and stored at −80 °C.

2.2. Reference populations and previously reported Y-STR data

Allelic frequencies from a total of 23 previously published populations (Table 1) [10,18–35] were chosen for comparison across the 11 Y-STR loci (DYS19, DYS389I, DYS389II, DYS390, DYS391, DYS392, DYS393, DYS385a/b, DYS438, and DYS439), given that the alleles for the remaining loci typed in this study (DYS437, DYS448, DYS456, DYS458, DYS635 and GATA H4) were not reported for all of the geographically-targeted reference populations. In addition, three individuals from Amdo, 20 from Kham and 133 from U-Tsang were incorporated in this study from previously reported data [10].

Table 1
Reference populations.

Population	Abbreviation	*n*	References
Himalayas			
Amdo (Tibet)	AMD	88	Present study
Kham (Tibet)	KHM	109	Present study
U-Tsang (Tibet)	UTS	153	Present study
Bhutan	BHU	856	Parkin et al. [18]
China Qinghai	CQI	167	Zhu et al. [19]
Kathmandu (Nepal)	KAT	77	Gayden et al. [10]
Lhasa	LHA	112	Zhang et al. [20]
Newar (Nepal)	NEW	66	Gayden et al. [10]
Nepal	NEP	769	Parkin et al. [21]
Central Asia			
Buryat	BUR	215	Wozniak et al. [22]
Lipezkaja (Russia)	LIP	47	Fechner et al. [23]
Mongolia	MON	96	Zhu et al. [24]
North Afghanistan	AFN	43	Lacau et al. [25]
Pensenskaja (Russia)	PEN	81	Fechner et al. [23]
South Afghanistan	AFS	146	Lacau et al. [25]
Northeast Asia			
China Shangdong	CSH	131	Yan et al. [26]
China Ningxia	CNI	143	Guo et al. [27]
Japan	JAP	381	Hashiyada et al. [28]
Korea	KOR	301	Park et al. [29]
Southeast Asia			
China Guangdong	CGU	120	Hu [30]
China Sichuan	CSI	237	Zhang et al. [31]
Malaysia	MAL	334	Chang et al. [32]
Taiwan	TAI	200	Huang et al. [33]
South Central Asia			
Bangladesh	BAN	216	Alam et al. [34]
Haryana	HAR	84	Nagy et al. [35]
Punjab	PUN	80	Nagy et al. [35]

2.3. DNA amplification and STR genotyping

DNA samples were amplified in a multiplex reaction at 17 Y-STR loci in an Eppendorf Master gradient cycler (Eppendorf AG, Hamburg, Germany) using the AmpF*l*STR® Yfiler kit (Applied Biosystems, Foster City, CA) [36] according to the manufacturer's specifications. Amplicons were subsequently separated by multi-capillary electrophoresis on an ABI Prism 3130xl Genetic Analyzer using the ABI GeneScan 500 LIZ internal size standard as a basis for comparison. Fragment sizes were obtained using the software GeneMapper® v3.1 (Applied Biosystems, Foster City, CA) [36] and alleles were determined through comparison with an allelic ladder provided by the manufacturer (Applied Biosystems, Foster City, CA).

2.4. Quality control

Our laboratory has participated in the Y-STR haplotype reference database (YHRD) [37] quality assurance exercise by typing the YHRD core loci as well as additional loci DYS437, DYS448, DYS456, DYS458, DYS635 and Y-GATA-H4 (Certificate dated: July 09, 2010). The accession numbers generated by the YHRD for the three Tibetan populations studied are YA003694 for Amdo, YA003695 for Kham and YA003696 for U-Tsang.

2.5. Statistical and phylogenetic analyses

Allelic frequencies were calculated for the three Tibetan populations (Amdo, Kham and U-Tsang) using the gene counting method [38]. Gene and haplotype diversities for each population were assessed using the software package Arlequin v3.1 [39]. Haplotype diversities were calculated at the minimal 9-, extended 11- and Yfiler 17-loci levels, excluding chromosomes carrying null alleles. Discrimination capacity (DC) and fraction of unique haplotypes (FUH) were estimated as the percent proportions of different and unique haplotypes, respectively, within a given population. Diversity parameters including dw_{min} (minimum diversity within the population), mw_{max} (maximum matching probability within the population), mw_{min} (minimum matching probability within the population], mb_{min} [minimum matching probability between two populations) and db_{max} (maximum diversity between two populations) were calculated for the Tibetan populations as described previously [40,41]. The ratio mw_{max}/mb_{min} gives an estimate for the upper limit of how many times more probable it is to find a match within a population rather than between two populations. The ratio mw_{min}/mb_{min} gives an estimate for the lower limit of the same parameter. Arlequin v3.1 [39] was utilized to determine the pairwise genetic distance (*Rst*) between a given pair of populations based on 11 Y-STR loci (DYS19, DYS389I, DYS389II, DYS390, DYS391, DYS392, DYS393, DYS385a/b, DYS438, and DYS439). *Rst* values were ascertained at a significance level of 0.01 using 10,000 repetitions [42]. A Bonferroni adjustment ($\alpha = 0.01/325 = 0.000031$) was employed to compensate for potential type I errors. Samples carrying null alleles and microvariants were excluded from *Rst* calculations.

A correspondence analysis (CA) plot including all 23 reference populations was generated using the program NTSYSpc 2.02i [43]. A neighbor joining (NJ) dendrogram was constructed utilizing the PHYLIP v3.6 program [44], with the statistical robustness of the phylogenetic relationships within the tree assessed using bootstrap analysis with 1000 iterations. For all population comparisons, the repeat length of DYS389II was obtained by subtracting the size of the corresponding DYS389I allele.