Forensic Science International: Genetics 25 (2016) 125–131

Contents lists available at ScienceDirect

Forensic Science International: Genetics

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

Research paper

EMPOP-quality mtDNA control region sequences from Kashmiri of Azad Jammu & Kashmir, Pakistan

Allah Rakha[a,b,*], Min-Sheng Peng[c], Rui Bi[a], Jiao-Jiao Song[c], Zeenat Salahudin[b], Atif Adan[b], Muhammad Israr[d], Yong-Gang Yao[a]

[a] Key Laboratory of Animal Models and Human Disease Mechanisms of the Chinese Academy of Sciences & Yunnan Province, Kunming Institute of Zoology, Kunming, Yunnan 650223, China
[b] Department of Forensic Sciences, University of Health Sciences, Lahore, Pakistan
[c] State Key Laboratory of Genetic Resources and Evolution, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming 650223, China
[d] Department of Forensic Studies, University of Swat, Swat, Pakistan

ARTICLE INFO

Article history:
Received 20 May 2016
Received in revised form 24 August 2016
Accepted 26 August 2016
Available online 28 August 2016

Keywords:
Pakistan
Kashmiri
mtDNA
Control region
Forensic database

ABSTRACT

The mitochondrial DNA (mtDNA) control region (nucleotide position 16024-576) sequences were generated through Sanger sequencing method for 317 self-identified Kashmiris from all districts of Azad Jammu & Kashmir Pakistan. The population sample set showed a total of 251 haplotypes, with a relatively high haplotype diversity (0.9977) and a low random match probability (0.54%). The containing matrilineal lineages belonging to three different phylogeographic origins of Western Eurasian (48.9%), South Asian (47.0%) and East Asian (4.1%). The present study was compared to previous data from Pakistan and other worldwide populations (Central Asia, Western Asia, and East & Southeast Asia). The dataset is made available through EMPOP under accession number EMP00679 and will serve as an mtDNA reference database in forensic casework in Pakistan.

1. Introduction

Mitochondrial DNA analysis has become a very useful tool for human evolutionary studies and especially forensic casework in several circumstances when standard nuclear markers cannot be applied [1,2]. Forensic casework involving mtDNA depends on relevant but authentic databases for estimating the probability of random match. The EMPOP, at present, provides the best quality data representation from all over the world based on logical and phylogenetic measures admissible for forensic purposes [3].

Large-scale investigations of archaeological sites in Central Asia, Northern Pakistan and India revealed a typographical affinity between their cultures going as far back as major pre- and proto-historic periods. The legend relates to an early periodic movement of tribal people from Central Asia to the Kashmir Valley during the cold season when the valley was comparatively warm, which were later replaced by the influx of Aryans from the Punjab. There is historical evidence to the settlement of immigrants from Persia, Greece, Turkistan and Tibet, China. With the advent of Islam there was an influx of a large number of Sufis and Sayyids in the 14th century [4]. Colonies of Mughals, Pathans, Punjabis, and Paharis settled within comparatively recent times throughout the Jammu and Kashmir. Kashmiris through out the Jammu and Kashmir (India) and Azad Jammu and Kashmir (Pakistan) speak the Kashmiri language. By origin it is a Dravidian Burushaski language, but it has become predominately Indo-Aryan in character. Reflecting the history of area, the Kashmiri vocabulary is mixed, containing Dardic, Sanskrit, Punjabi, and Persian elements [5]. Recent waves of immigrants to Azad Jammu and Kashmir have also introduced Punjabi and Pashto to the main languages [6].

The available mtDNA sequence data from Pakistan is scarce, fragmentary and limited to a few samples from main ethnic groups [7–9]. However, limited number of sample size inhibits the accurate characterization of any population for forensic and genetic purposes. Moreover, there is no representation of Kashmiris from Azad Jammu and Kashmir in any of reference population databases. The present study is intended to characterize the diversity of the matrilineal lineages of current inhabitants of Azad Jammu and Kashmir by analyzing the entire mitochondrial DNA control region. With this analysis, we seek to contribute new mtDNA haplotype data, taking into account that the development

* Corresponding author at: Key Laboratory of Animal Models and Human Disease Mechanisms of the Chinese Academy of Sciences & Yunnan Province, Kunming Institute of Zoology, Kunming, Yunnan 650223, China.
E-mail addresses: dnaexpert@me.com, dnatypist@gmail.com (A. Rakha).

http://dx.doi.org/10.1016/j.fsigen.2016.08.009
1872-4973/

and improvement of databases constitute a major goal for consolidating the use of mtDNA for forensic purposes. We further analyzed the haplogroup distribution in Kashmiris from Azad Jammu and Kashmir (Pakistan) to corroborate, from the perspective of female genetic lineages, the ancestry composition of this highly mixed population.

The present work constitutes of 317 entire mtDNA control region sequences from randomly selected Kashmiris sampled at eight districts of Jammu & Kashmir, Pakistan. The generated mtDNA population data is deposited to the EMPOP database under the accession number EMP00679 for the worldwide use, and particularly as a reference database for mtDNA applications in forensic and missing person casework in Pakistan.

2. Materials and methods

2.1. Samples

Blood samples were collected from 317 unrelated Kashmiri males and females residing in different parts of Azad Jammu & Kashmir, Pakistan (Fig. 1).

Only individuals with self-reported Kashmiri origin of at least two generations back on the maternal side were included. Written informed consent was obtained from all the volunteer donors. In order to have full representation, samples were collected from different towns and cities of Azad Jammu & Kashmir. Personal information was treated anonymously. In addition, sample collection was conducted in accordance with the Institutional Review Board of University of Health Sciences, Lahore.

2.2. DNA extraction, amplification and sequencing

Genomic DNA extraction was carried out with the Axygen® AxyPrep™ Blood Genomic DNA Miniprep Kit following the manufacturer's protocol (Axygen Biosciences; CA, USA). The entire control region from nt16024 to nt576 was amplified by using primers (Table S4 in Supplementary material) as reported earlier [10]. PCR reactions were performed in 30 mL of reaction mixture containing 3 μL 10 × LA PCR Buffer II (Mg^{2+} Plus), 1.5 units of TaKaRa LA Taq (TaKaRa Bio Inc., Dalian, China), 400 μM of each dNTP, 0.2 μM of each primer, and 10 ng DNA. The amplification was run on the GeneAmp PCR System 9700 (Applied Biosystems, Foster City, CA, USA) under following conditions: one denaturation cycle of 94 °C for 5 min; 30 amplification cycles of 94 °C for 30 s, 60 °C for 30 s and 72 °C for 2 min; and one full extension cycle of 72 °C for 10 min. The PCR products were purified with Exonuclease I (TaKaRa) and Shrimp Alkaline Phosphatase (TaKaRa) in 10X Exonuclease I Buffer (TaKaRa) following reaction compositions recommended by manufacturer, incubating at 37 °C for 45 min, followed by enzyme deactivation at 95 °C for 15 min. Sequencing by capillary electrophoresis using the BigDye™ Terminator v3.0 Ready Reaction Cycle Sequencing Kit was performed on Applied Biosystems 3730xl DNA Analyzer (Thermo Fisher Scientific) according to manufacturer's manual.

2.3. Haplogroup assignment

Forward and reverse sequences were aligned and compared using SeqManNGen® version 12.0 (DNASTAR. Madison, WI) with the revised Cambridge Reference Sequences (rCRS) [11]. Quality of sequences was examined manually, and two analysts independently annotated deviations from the reference sequence. The recommended nomenclature for mtDNA typing was used for alignment of variants [12]. The haplogroup assignments were carried out using Mitotool (www.mitotool.org) [13], Haplogrep (www.haplogrep.uibk.ac.at) [14], and EMMA (www.empop.online) [15] based on PhyloTree builds 16 and 17 [16] with referring to the additional guidelines [1]. Haplogroup assignments were re-evaluated by manual inference and conservative MRCA status was assigned to each sequence to improve the predictions. All 317 mtDNA haplotypes were confirmed and validated by the EMPOP curators, being now available from EMPOP browser with the accession number (EMP00679) [3]. The sequences are also available on GenBank via accession numbers KX084069–KX084385.

2.4. Data analysis

For all calculations, insertions at nt16193, nt309, nt315, and nt573 were ignored unless otherwise mentioned. The number and

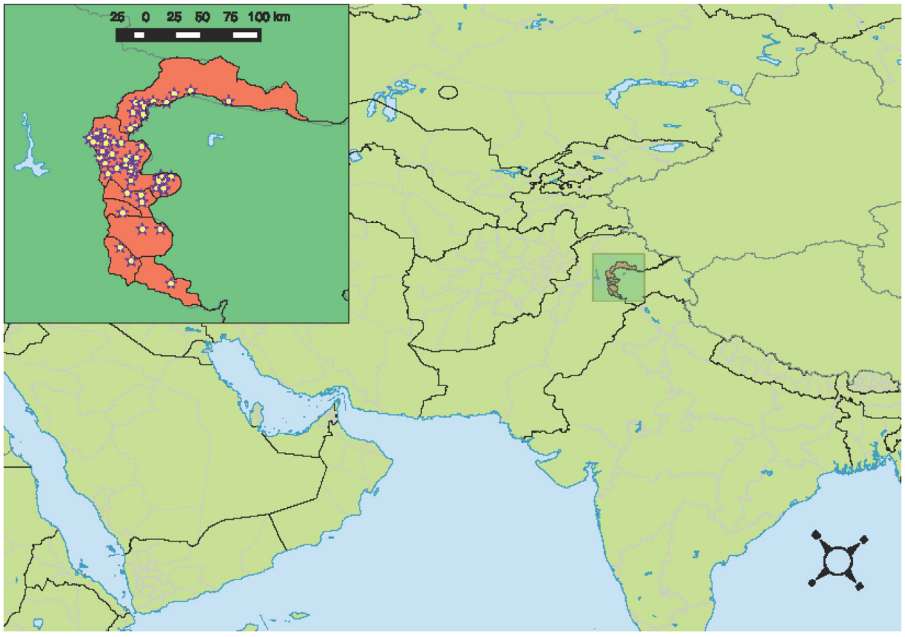

Fig. 1. Geographic location of Azad Jammu & Kashmir and inset map depicting the birthplaces of samples used for this study.