Genetic characterization of male lineages of Pathans from Pakistan

Hwan Young Lee[1] · Na Young Kim[1] · Jeong Eun Sim[1] · Ajin Choi[1] · Allah Rakha[2] · Myung Jin Park[1] · Ja Hyun An[1] · Woo Ick Yang[1] · Kyoung-Jin Shin[1]

[1]*Department of Forensic Medicine, Yonsei University College of Medicine, Seoul, Korea*
[2]*National Forensic Science Agency, National Police Bureau, Islamabad, Pakistan*

Introduction

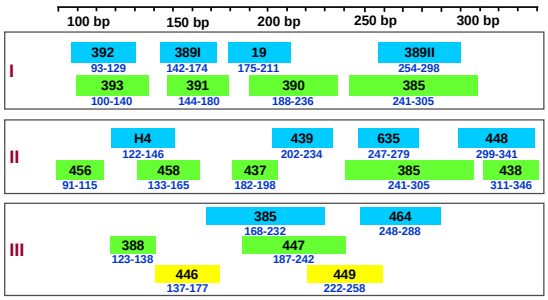
Cultural and linguistic affiliations divide the people of Pakistan into 16 ethnic groups with diverse origins. The evolutionary antiquity and endogamy of Pakistani populations generate a high degree of genetic differentiation and structuring. Hence to obtain the most reliable and conservative frequency estimates for forensic purposes and to estimate population history require that regional or ethnic databases be established. Major ethnic groups of Pakistan include the Punjabis, Pathans, Sindhis, Seraikis, Muhajirs, Balochis, Hindkowans, and Chitralis. The Pathans represent the tribes who speak Pashto (Eastern Iranian branch of the Indo-Iranian language family), inhabit mainly the North West Frontier Province (N.W.F.P.), adjoining tribal areas of Pakistan, and southern and eastern parts of Afghanistan, and are the second-largest ethnic group in Pakistan. In the present study, to better characterize and understand the male lineages of Pakistani populations, 22 Y-STRs and 18 Y-SNPs were analyzed in 270 unrelated Pathans from the N.W.F.P. and Federally Administered Tribal Areas (FATA) of Pakistan.

Materials and Methods

Samples

Blood samples were collected from 270 unrelated male Pathan volunteers in the N.W.F.P and FATA of Pakistan. All participants gave their informed consent orally or in writing after we explained the aims and procedures of the study. The Institutional Review Board of Severance Hospital, Yonsei University in Seoul, Korea approved this study. DNA was isolated from blood using QIAamp DNA Mini Kit (Qiagen, Hilden, Germany) according to the manufacturer's instructions.

Schematic of 3 in-house multiplex PCRs for analysis of 22 Y-STRs



The 22 Y-STRs were amplified using AmpFlSTR® Yfiler™ PCR amplification kit (Applied Biosystems, Foster City, CA, USA) and an in-house multiplex PCR system (Please refer to http://forensic.yonsei.ac.kr/protocols.html).
Haplotype diversity was calculated with the software package Arlequin 3.5.1.2

Y chromosomal haplogroup determination using multiplex SBE

Y chromosomal SNP selection and primer design for PCR and SBE

Fifteen (M40, M89, M201, M69, M304, M9, M20, M45, M242, M207, P231, M17, M479, M124 and M184) and 3 Y-SNPs (RPS4Y_{711}, JST002611 and M117) were selected to determine Y-chromosomal haplogroups E, F, G, H, J, K, L, P, Q, R, R1, R1a1a, R2, R2a and T, and C, O3a1c and O3a2c1a, that are present in South and East Asian populations, respectively. Primers for PCR amplification and subsequent SBE were designed using programs Primer 3 (http://frodo.wi.mit.edu/primer3/) and Batchprimer 3 (http://batchprimer3.bioinformatics.ucdavis.edu/index.html).

Multiplex PCR and multiplex SBE

Multiplex PCR: Several multiplex and monoplex PCRs were performed in a final volume of 25 μl that contained 1 ng of template DNA, 2.5 μl of Gold ST*R 10× buffer (Promega, Madison, WI, USA), 2.0 U (1.0 U for monoplex) of AmpliTaq Gold® DNA polymerase (Applied Biosystems) and primer mix. Thermal cycling was performed under the following conditions: 95°C for 11 min; 33 cycles of 94°C for 20 sec, 60°C for 1 min, 72°C for 30 sec; and a final extension of 72°C for 7 min. For the following SBE, 5.0 μl of PCR products was purified with ExoSAP-IT (USB, Cleveland, OH, USA).
Multiplex SBE: Several multiplex and monoplex SBE reactions were carried out with the SNaPshot™ Multiplex kit (Applied Biosystems) according to the manufacturer's instruction. After the SBE reactions, 1 U SAP (USB) was added to the extension products, and the mix was incubated at 37°C for 45 min followed at 80°C for 15 min to remove the unincorporated ddNTPs.
Detection system: ABI PRISM 310 Genetic Analyzer and GeneMapper ID 3.2 software (Applied Biosystems)

Results

Fig. 1. Phylogenetic tree of the 18 Y-chromosomal binary polymorphisms and with their corresponding haplogroup frequencies observed in Pathans from Pakistan

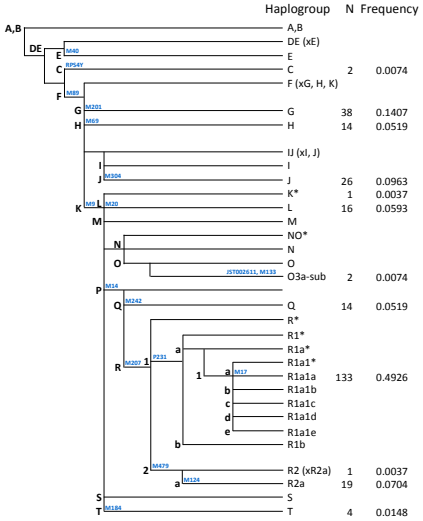


Fig. 2. Representative electropherograms of the multiplex SBE reactions for Y-haplogrouping of Pathans from Pakistan

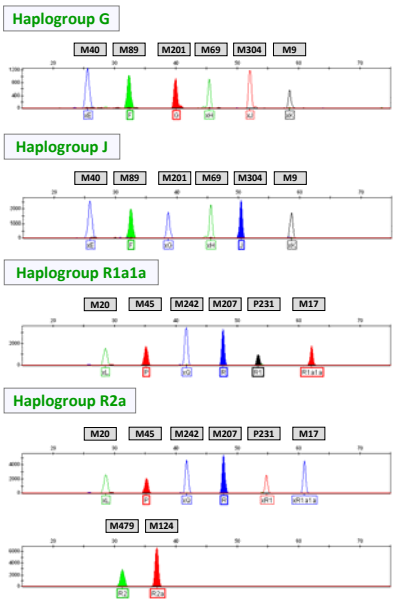


Table 1. Y-STR haplotype diversities and discrimination capacities in each haplogroup observed in Pathans from Pakistan

Haplogroup	No.	Haplotype diversity	Discrimination capacity*
C	2	1.0000	1.0000
G	38	0.9829	0.8421
H	14	1.0000	1.0000
J	26	0.9815	0.8077
K	1	-	-
L	16	0.9917	0.9375
O3a1c	1	-	-
O3a2c1a	1	-	-
Q	14	0.9890	0.9286
R1a1a	133	0.9892	0.7970
R2	1	-	-
R2a	19	1.0000	1.0000
T	4	1.0000	1.0000
Total	270	0.9968	0.8519

* Discrimination capacity was determined by dividing the number of observed haplotypes by the number of total investigated individuals.

Discussion

- **In haplotype analysis for the 22 Y-STRs in 270 Pathan males, 230 different haplotypes were observed with overall haplotype diversity of 0.9968. Among them, 211 (78.1%) were observed once; 13 (9.6%), twice; 2 (2.2%), three times; 3 (5.6%), five times; and one (4.4%) twelve times. The relatively low values for haplotype diversity and discrimination capacity in Pathans should be considered in the forensic interpretation of Y-STR data.**
- **With the developed multiplex SBE reactions and additional SNP analysis, 13 different Y-chromosomal haplogroups were defined within our samples, and the R1a1a haplogroup of South Asian origin was most frequently observed.**
- **Haplogroup distribution in our samples is mainly consistent with the previous study by Firasat et al. (EJHG 15:121), but to better characterize and compare the male lineages of Pakistani populations, further investigation of the other ethnic groups will be needed.**