



# Y-STR haplotypes and the genetic structure of Pathan populations in FATA and NWFP of Pakistan

Hwan Young Lee<sup>1</sup> · Jeong Eun Sim<sup>1</sup> · Ajin Choi<sup>1</sup> · Allah Rakha<sup>2</sup> · Woo Ick Yang<sup>1</sup> · Kyoung-Jin Shin<sup>1</sup>

<sup>1</sup>Department of Forensic Medicine, Yonsei University College of Medicine, Seoul, Korea

<sup>2</sup>Department of Forensic Sciences, University of Health Sciences, Lahore, Pakistan

## Introduction

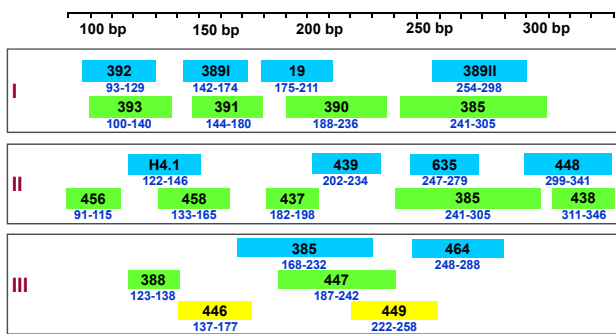
Cultural and linguistic affiliations divide the people of Pakistan into 16 ethnic groups with diverse origins. The evolutionary antiquity and endogamy of Pakistani populations generate a high degree of genetic differentiation and structuring. Hence the establishment of regional or ethnic databases is required to obtain the most reliable and conservative frequency estimates for forensic purposes and to estimate population history. 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, 22 Y-STRs were analyzed in 270 unrelated Pathans from the N.W.F.P. and Federally Administered Tribal Areas (FATA) of Pakistan, and the haplotype data were compared with other Pathan groups living in adjacent regions.

## Materials and Methods

### Samples

Blood samples were collected from 270 unrelated male Pathan volunteers in the N.W.F.P and FATA of Pakistan; 234 are from N.W.F.P. and 36 are from FATA. 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 AmpFISTR<sup>®</sup> Yfiler<sup>™</sup> 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>).

### Analysis of population data and interpopulation comparisons

Haplotype diversity indices were calculated according to Nei and discrimination capacity was determined by dividing the number of observed haplotypes by the number of total investigated individuals. Arlequin, v. 3.5.1.2 and YHRD website (<http://www.yhrd.org/Analysis>) were used to generate  $\Phi_{st}$  values. An analysis of molecular variance (AMOVA) using the Arlequin software was conducted using 1,000 permutation replicates and the RST genetic distance, which takes into account the probability of recurrent mutation for calculating distance. The Pathan samples from N.W.F.P. and FATA in Pakistan were compared with Afghanistan Pathans (YHRD accession YA003701), North and South Afghanistan Pathans (YA003702 and YA003703), Pathans from Pakistan (YA003218), Afridi Pathans from Uttar Pradesh in India (YA003686), and Yousafzai Pathans from Khyber Pakhtunkhwa in Pakistan (YA003748). In addition, full population sets were used for AMOVA, i.e. 1885 haplotypes from India, 444 from Iran and 789 from Pakistan. Since all haplotype overlaps in only 7 markers (DYS19, DYS389 I-II, DYS390, DYS391, DYS392 and DYS393), the 7 markers were used for the analysis. Results were visualized in an MDS plot as available from the YHRD website with slight modification.

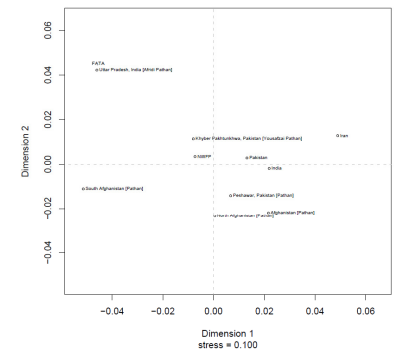
## Results

Table 1. Pairwise comparison between the Pathan populations

Population	Afghanistan (Pathan)	Khyber Pakhtunkhwa, Pakistan (Yousafzai Pathan)	North Afghanistan (Pathan)	Peshawar, Pakistan (Pathan)	South Afghanistan (Pathan)	Uttar Pradesh, India (Afridi Pathan)	India	Iran	Pakistan	NWFP, Pakistan (Pathan)	FATA, Pakistan (Pathan)
Afghanistan (Pathan)	-	0.0002	0.0895	0.104	0	0	0.0049	0	0.0001	0	0.0004
Khyber Pakhtunkhwa, Pakistan (Yousafzai Pathan)	0.0816	-	0.0023	0.0119	0.0001	0.0013	0.0004	0	0.0106	0.1741	0.0009
North Afghanistan (Pathan)	0.0185	0.0672	-	0.193	0.0036	0.0012	0.0081	0	0.0069	0.0422	0.0025
Peshawar, Pakistan (Pathan)	0.0093	0.0317	0.0076	-	0.0001	0.0017	0.0047	0	0.0058	0.036	0.0028
South Afghanistan (Pathan)	0.1235	0.051	0.0468	0.0656	-	0.0001	0	0	0	0.0004	0.0033
Uttar Pradesh, India (Afridi Pathan)	0.116	0.0634	0.0878	0.065	0.0653	-	0	0	0.0002	0.0074	0.7952
India	0.0121	0.0345	0.0295	0.0162	0.0772	0.0723	-	0	0.0001	0	0
Iran	0.0656	0.0582	0.1051	0.0685	0.1529	0.1493	0.0156	-	0	0	0
Pakistan	0.0278	0.0181	0.0306	0.0154	0.0616	0.0639	0.0062	0.0212	-	0.0009	0.0002
NWFP, Pakistan (Pathan)	0.0525	0.0048	0.0206	0.0129	0.0264	0.0364	0.0263	0.0661	0.0115	-	0.0137
FATA, Pakistan (Pathan)	0.1229	0.0735	0.0811	0.0684	0.05	-0.0129	0.085	0.1686	0.0716	0.0344	-

0st values are below the diagonal and the p-values above the diagonal

Fig. 1. MDS plot showing various Pathan groups



## Discussion

- In haplotype analysis for the 22 Y-STRs in 234 Pathan males from N.W.F.P., 200 different haplotypes were observed with discrimination capacity of 0.8547 and overall haplotype diversity of 0.9957. In haplotype analysis for the 22 Y-STRs in 36 Pathan males from FATA, 32 different haplotypes were observed with discrimination capacity of 0.8889 and overall haplotype diversity of 0.9889.
- In analysis of molecular variance between the neighboring Pathan populations inhabiting Afghanistan, Pakistan and India, the Pathan population from N.W.F.P. did not show significant difference from Yousafzai Pathans of Khyber Pakhtunkhwa, Pakistan, and the Pathan population from FATA did not show significant difference from Afridi Pathans in India.
- Considering that the Afridi Pathans of Pakistan mainly inhabit rough hilly area covering most of the Khyber Agency, FR Peshawar and FR Kohat in FATA, the close genetic distance between Afridi Pathans in India and the Pathan population from FATA of the present study with their large genetic distance from other Pathan populations should reveal a considerable regional stratification between different Pathan population groups from Afghanistan, Pakistan and India but high homogeneity between Pathan populations sharing the history of inhabiting the same region.