International Congress Series 1288 (2006) 171–173

www.ics-elsevier.com

The AMOVA analyses and phylogenetic relationships of Pakistani population using Y STRs

S. Hadi *, B. Bekaert, W. Goodwin

University of Central Lancashire, Preston, UK

Abstract. Pakistan is a large country with a diverse population consisting of four significant population groups besides many smaller isolated populations. Male population samples were collected from indigenous populations and the extracted DNA was used to amplify seven Y STRs. Allele frequencies within each population were similar as were the haplotype frequencies. The data showed that individual populations were diverse in spite of a high rate of consanguineous marriages. This work indicates that, even with smaller number of loci, significant amount of genetic and phylogenetic information can be gained from the data. © 2005 Elsevier B.V. All rights reserved.

Keywords: Pakistani; Y STR; AMOVA

1. Introduction

Pakistan is a large country inhabited by a number of ethnically diverse populations. The present day major populations consist of Punjabi, Sindhi, Pushtoon and Baluchi groups. Many smaller populations also exist in Pakistan including Makranis in the coastal regions and the Kalash and Brosho in the northern mountainous regions.

The aim of this work was to compare the Y STR haplotype genetic diversity of the different Pakistani populations. This would make it possible to assess the impact of Y STR haplotype diversity on the application of these markers for human identification. Y chromosome STRs have been previously used to infer population relationships. The second aim was to study the relationship of different Pakistani populations.

---

* Corresponding author. Dept. of Forensic and Investigative Sciences, University of Central Lancashire, Preston PR1 2HE, UK. Tel.: +44 1772 894395; fax: +44 1772 894981.

E-mail address: shadi@uclan.ac.uk (S. Hadi).

doi:10.1016/j.ics.2005.11.059

2. Materials and methods

Blood and Buccal samples were collected in Pakistan. DNA extraction was done using Purgene® DNA extraction kit in accordance with manufacturer's instructions. Seven Y STRs (DYS 19, 389I, 389II, 390, 391, 392 and 393) comprising the Yh1 haplotype were amplified using published primers [1] and in house multiplex reactions. All PCR reactions were analysed using ABI 310 automated DNA analyser. In house sequenced allelic ladders and the ROX 500 standard were used. The Y STR data was analysed using Arlequin software [2].

3. Statistical analyses

The analysis of molecular variance (AMOVA) developed for genetic analysis was conducted in order to look into mutational differences between the loci in different populations and in different arbitrary groups of populations. The pairwise F_{ST} was generated using Arlequin and this was used as the genetic distance for estimating the phylogenetic relationships of Pakistani populations using PHYLIP software [3]. UPGMA and Neighbourhood Joining trees were generated.

4. Results and discussion

Haplotypes were constructed for each individual and the populations were then compared using AMOVA. The haplotype diversity for most populations was ~0.98. Distance and phylogenetic statistics indicated a much closer relationship between Punjabi

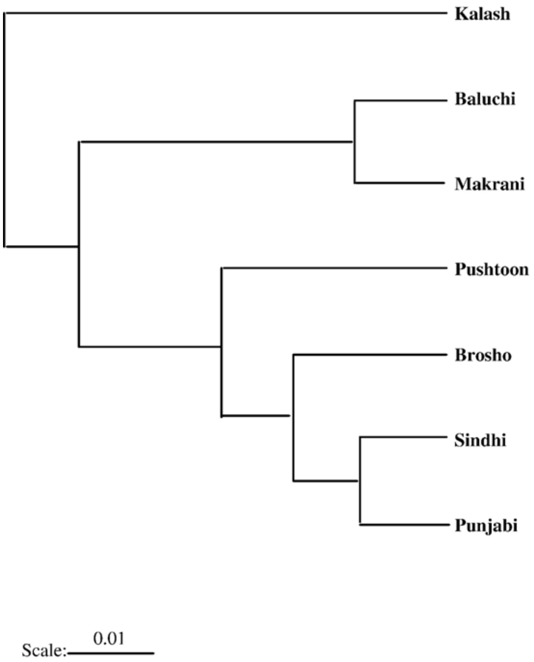

Fig. 1. Phylogenetic tree showing relationship of different Pakistani populations using F_{ST} as distance.

Table 1
AMOVA results for Pakistani populations

Group I: Kalash and group 2: all other Pakistani populations

Source of variation	*df*	Sum of squares	Variance	% variation
Among groups	1	69.171	0.10701	4.51
Among populations within groups	5	126.806	0.14212	5.99
among individuals within populations	555	2356.724	2.12317	89.5

and Sindhi populations (Fig. 1). The analysis indicated closer relationship of the Kalash and the Pakistani populations than had been thought due to cultural, historical and linguistic differences. AMOVA of Pakistani populations revealed that only 6% of the genetic variation was between the Kalash and all other Pakistani populations when they were treated as two groups (Table 1). The variation decreased between the Kalash and the Pakistani populations to 3.84% when Brosho were excluded from the Pakistani group.

This was also supported through F_{ST} analysis after correction as proposed earlier [4]. The corrected F_{ST} for Pak-Brosho pair was 0.024, whereas that for Pak-Kalash pair was 0.028. This may be due in part to high levels of genetic drift, particularly in Kalash which number ~5000. This was in sharp contrast to the values higher than 0.06 reported in the literature for other geographically closely located populations, e.g. in the Central Asia [5]. The AMOVA and distance analyses showed that Pakistani populations were more closely related to each other than to the Kalash as a group (Fig. 1). The phylogenetic analysis could not differentiate Brosho from other Pakistani populations.

5. Conclusions

These results are similar to the results of a larger study on Pakistani population where SNPs and 15 STRs were typed [6]. This study reported that Pakistani populations were closely related and that the linguistic differences must have arose after the establishment of Y haplogroup patterns. Same haplotypes and similar allele frequencies in different populations point towards consistent gene flow among these populations and though in case of Kalash the effect of genetic drift cannot be excluded.

References

[1] M. Kayser, et al., Int. J. Leg. Med. 110 (1997) 125–133.
[2] S. Schneider, D. Roessli, L. Excoffier, Dept. of Anthropology, University of Geneva, 2000.
[3] J. Felsenstein, Cladistics 5 (1989) 164–166.
[4] A. Perez-Lezaun, et al., J. Mol. Evol. 45 (1997) 265–270.
[5] A. Perez-Lezaun, et al., Am. J. Hum. Genet. 65 (1999) 208–219.
[6] R. Qamar, et al., Am. J. Hum. Genet. 70 (2002) 1107–1124.