Aisha Mohyuddin · Qasim Ayub · Peter A. Underhill
Chris Tyler-Smith · S. Qasim Mehdi

Detection of novel Y SNPs provides further insights into Y chromosomal variation in Pakistan

Received: 15 November 2005 / Accepted: 6 December 2005 / Published online: 10 February 2006
© The Japan Society of Human Genetics and Springer-Verlag 2006

Abstract Biallelic polymorphisms on the Y chromosome have been extensively used to study the history, evolution, and migration patterns of world populations. In this study we screened 8.5 kb of Y chromosomal DNA for single nucleotide polymorphisms (SNPs) in a panel of 95 male individuals belonging to different haplogroups. Five novel Y-SNPs (PK1–5) were identified, four in the Pakistani sample and one in an African sample. The ancestral state of each SNP was determined in two chimpanzee samples and a variety of Pakistani ethnic groups. In addition to these novel Y-SNPs 77 additional markers on the Y chromosome were analyzed to place the SNPs on the phylogenetic tree of Y chromosomal lineages and to further investigate extant human Y chromosomal variation within Pakistan. BATWING analysis gave an estimate of between 2,500 and 7,300 YBP for population expansion in Pakistan which coincides with the period of the Indus Valley civilizations.

Keywords Y-chromosome · Y-SNPs · Pakistan · DHPLC · Population genetics

Aisha Mohyuddin and Qasim Ayub contributed equally to this publication.

A. Mohyuddin · Q. Ayub · S.Q. Mehdi (✉)
Biomedical and Genetic Engineering Laboratories,
G. P. O Box 2891, 44000 Islamabad, Pakistan
E-mail: sqmehdi@comsats.net.pk
Tel.: +92-51-9261138
Fax: +92-51-9261144

P.A. Underhill
Department of Genetics, Stanford University,
Stanford, CA, USA

C. Tyler-Smith
The Wellcome Trust Sanger Institute,
Hinxton, Cambridge, UK

C. Tyler-Smith
Department of Biochemistry,
University of Oxford, Oxford, UK

Introduction

Sequence variants on the male-specific region of the human Y chromosome have been useful in examining the evolution and migration patterns of world populations, including Pakistan (Qamar et al. 2002). To further delineate the population sub-structure in Pakistan, 8.5 kb of Y-chromosomal DNA were screened in 93 Pakistani and 2 African samples for novel single nucleotide polymorphisms (SNPs) using denaturing high-performance liquid chromatography (DHPLC).

Materials and methods

Using sequences available in public databases, the non-recombining portion of the human Y chromosome was screened in silico. RepeatMasker2 software (Smit 1996) was used to exclude human repeat DNA sequences and primers were designed to amplify 150–700 bp of unique male sequences using Primer3 software (Rozen and Skaletsky 2000). Heteroduplex analyses were carried using a WAVE DNA Fragment Analysis System (Transgenomic, Crewe, UK) as described elsewhere (Underhill et al. 2000). Novel Y-SNPs were identified by the appearance of two or more peaks in the elution profiles and confirmed by DNA sequencing. The ancestral state of each SNP was determined in two chimpanzee samples. ARMS or RFLP-PCR assays were designed for rapid screening of these novel SNPs in the Pakistani population (supplementary Table 1).

Results

Four novel Y SNPs (PK2–5) were identified in the Pakistani samples and one (PK1) in the Africans (Table 1, Fig. 1). The African individuals belonged to

Table 1 Description of the novel Y SNPs

Clone	Marker	Nucleotide change	Site (bp)	Allele in Chimpanzee	Primers (5'–3')	
					Forward	Reverse
RP11-329C15	PK1	C → A	60	C	tcaactttcttaaatgattgtacgtt	tctgttcaggagaacctctatgg
RP11-329C15	PK2	T → C	342	T	tgtgtcctggtgtcttttgg	tgacaacagcatctcttaactgg
RP11-329C15	PK3	T → C	365	T	tgtgtcctggtgtcttttgg	tgacaacagcatctcttaactgg
NT_003571	PK4	A → T	83	A	ccatcctccatggctagt	gcttccaaggtgcccttat
RP11-65G9	PK5	C → T	186	T	ttcaaacacatgcttctgc	taaaaggaggaggactgc

haplogroup A2, that is restricted to southern Africa (Underhill et al. 2000). The PK2 polymorphism was observed in clade C3 Y-chromosomes in the Hazara and Burusho populations at frequencies of 43% and 9%, respectively. PK3, PK4 and PK5 polymorphisms represent new branches (L/-4, O2a1a and R1a1/-d respectively) of the Y phylogeny (Fig. 1). PK3 was found exclusively in the Kalash (23%). PK4 was detected in 4% of the Pathan samples of the AusoKhel sub-tribe. The PK5 transition was found in two unrelated Burusho individuals. This seems to be a recurrent mutation, because it was detected in the chimpanzee samples but was absent from a gorilla sample.

Discussion

In this study, five novel Y-SNPs were identified. The PK2 polymorphism, found in the Burusho and Hazara, distinguished between the northern and southern clade C3 lineages within Pakistan. The derived allele for this marker was found in individuals that were part of a previous study in which the Hazara samples formed a star cluster with 16 different populations (Zerjal et al. 2003). The Burusho samples did not fall within this cluster. We extrapolate that all chromosomes within the star cluster should be derived for this mutation. The Mongolian origin of the Hazara is well documented historically and genetically (Zerjal et al. 2003; Qamar et al. 2002), whereas not much is known about the origins of the Burusho. According to some, the Burusho are descendants of Greek soldiers that came to this area with Alexander the Great. Others describe them as descendants of Dards from Central Asia (Biddulph 1977). Haplogroup C chromosomes are not found in Greece (Francalacci et al. 2003; Rootsi et al. 2004) and studies with autosomal genetic markers suggest the Burusho are genetically closer to their geographical neighbours (Mansoor et al. 2004; Ayub et al. 2003). In an earlier study (Wells et al. 2001) populations from Tajikistan were shown to cluster with the Hunza Burusho. The presence of the PK2 polymorphism at high frequencies in both the Hazara and Burusho (43% and 9%, respectively) suggests that this Y-SNP may be an ancient polymorphism that probably arose in Central Asia before

the separation of these two populations. This is corroborated by a BATWING analysis (Wilson and Balding 1998). Incorporating data from 19 Y-SNPs, including the novel Y-SNPs and 16 microsatellite loci, gave TMRCA estimates of between 9,400 (5,200–17,200) YBP for the PK2 polymorphism.

PK3 was found solely in the Kalash population of Pakistan. They inhabit remote valleys in the Hindu Kush Mountains in the Northern Areas of Pakistan. Previous studies (Mansoor et al. 2004; Rosenberg et al. 2002) reported a Eurasian influence in this isolated population. Principal-component analyses (Cavalli-Sforza et al. 1994), carried out using a larger number of markers, in this study clustered the Kalash population with the Yadhavas (data obtained from Wells et al. 2001), a Dravidian speaking group from south India (Fig. 2). This could be because of shared Eurasian ancestry, as demonstrated in the earlier study (Wells et al. 2001) in which the Yadhavas grouped together with other Central Asian populations. Y-STR variation across 16 loci enabled estimation of the median TMRCA as approximately 3,400 (1,400–8,100) YBP, which corresponds to the time of invasion of the Indo-Pak subcontinent by Indo European-speaking tribes from Central Asia (Wolpert 2000). Analysis of the PK3 polymorphism in the Indian population could shed further light on this relationship.

PK4 detected in four Pathan samples represents a new branch, O2a1a of clade O (Fig. 1). It is absent in the remaining haplogroup O samples from Pakistan. The four Pathan individuals carrying this SNP belong to a sub tribe (AusoKhel) of one of the major Pathan tribes (Yousafzai) from the Dir area (between 71°20' and 72°30'E; 34°22' and 35°50'N) in the North West Frontier Province (NWFP) in Pakistan (Bokhari 1993). The presence of several male lineages in Pathans reflects the diversity that exists in this population, contrary to oral traditions that claim that the Pathans have a single male ancestor (Dorn 1999).

The PK5 transition found in two Burusho individuals represents a new branch, R1a1/-d, on the M17 background. These Burusho individuals, although unrelated and belonging to different villages, shared the same haplotype across all 16 Y-STRs, indicating that this may be a recent population-specific polymorphism. A TMRCA of 350 (14–1790) YBP was obtained for this polymorphism.

Fig. 1 Y chromosome phylogeny indicating the positions (*black arrows*) of the five novel Y-SNPs. The name of each haplogroup is given at the tip of lineage along with its frequency (*parenthesis*) in 869 Pakistani individuals. The haplogroup nomenclature is

according to the Y Chromosome Consortium (2002) and Jobling and Tyler-Smith (2003). The name of each polymorphism is shown along the branches

In this study 874 Pakistani individuals were analyzed for a large number (82) of Y chromosomal markers that included five novel Y-SNPs. Three of these SNPs identify population-specific lineages within Pakistan. Typing of these novel Y-SNPs (PK2,

PK3, and PK4) in Eurasian populations will provide a comprehensive portrait of the complex genetic architecture of extant Pakistani ethnic groups and shed light on their origins and migration patterns.

Fig. 2 Principal-components analysis of Y haplogroup frequencies in fourteen Pakistani populations, three Indian populations, and a Greek population. Indo-European speakers are indicated by *black triangles*, Sino-Tibetan speakers by a *white triangle*, Dravidian speakers by a *circle*, and the language-isolate Burusho by a *diamond*. The principal-component plots were constructed using SPSS version 10.0 software and the first and second principal components were plotted using the Microsoft Excel for Windows

References

- Ayub Q, Mansoor A, Ismail M, Khaliq S, Mohyuddin A, Hameed A, Mazhar K, Rehman S, Siddiqi S, Papaioannou M, Piazza A, Cavalli-Sforza LL, Mehdi SQ (2003) Reconstruction of the human evolutionary tree using polymorphic autosomal microsatellites. *Am J Phys Anthropol* 122:259–268
- Biddulph J (1977) *Tribes of the Hindoo Koosh*. Indus Publications, Karachi
- Bokhari A (1993) *A glossary of the people of Pakistan, Part I* NWFP. New Nisa Printers, Quetta
- Cavalli-Sforza LL, Menozzi P, Piazza A (1994) *The history and geography of human genes*. Princeton University Press, Princeton
- Dorn B (1999) *History of the Afghans*. Vanguard Books Pvt. Ltd, Lahore
- Francalacci P, Morelli L, Underhill PA, Lillie AS, Passarino G, Useli A, Madeddu R, Paoli G, Tofanelli S, Calò CM, Ghiani ME, Varesi L, Memmi M, Vona G, Lin AA, Oefner P, Cavalli-Sforza LL (2003) Peopling of three Mediterranean islands (Corsica, Sardinia, and Sicily) inferred by Y-chromosome biallelic variability. *Am J Phys Anthropol* 121:270–279
- Jobling MA, Tyler-Smith C (2003) The human Y chromosome: an evolutionary marker comes of age. *Nat Rev Genet* 4:598–612
- Mansoor A, Mazhar K, Khaliq S, Hameed A, Rehman S, Siddiqi S, Papaioannou M, Cavalli-Sforza LL, Mehdi SQ, Ayub Q (2004) Investigation of the Greek ancestry of populations from northern Pakistan. *Hum Genet* 114:484–490
- Qamar R, Ayub Q, Mohyuddin A, Helgason A, Mazhar K, Mansoor A, Zerjal T, Tyler-Smith C, Mehdi SQ (2002) Y-chromosomal DNA variation in Pakistan. *Am J Hum Genet* 70:1007–1024
- Rootsi S, Magri C, Kivisild T, Benuzzi G, Help H, Bermisheva M, Kutuev I et al (2004) Phylogeography of Y-chromosome haplogroup I reveals distinct domains of prehistoric gene flow in Europe. *Am J Hum Genet* 75:128–137
- Rosenberg NA, Pritchard JK, Weber JL, Cann HM, Kidd KK, Zhivotovsky LA, Feldman MW (2002) Genetic structure of human populations. *Science* 298:2381–2385
- Rozen S, Skaletsky H (2000) Primer3 on the www for general users and for biologist programmers. In: Krawetz S, Misener S (eds) *Bioinformatics methods and protocols: methods in molecular biology*. Humana Press, Totowa, pp 365–386
- Smit AFA (1996) *Structure and evolution of mammalian interspersed repeats*. PhD dissertation, University of Southern California, CA
- Underhill PA, Shen P, Lin AA, Jin L, Passarino G, Yang WH, Kauffman E, Bonne-Tamir B, Bertranpetit J, Francalacci P, Ibrahim M, Jenkins T, Kidd JR, Mehdi SQ, Seielstad MT, Wells RS, Piazza A, Davis RW, Feldman MW, Cavalli-Sforza LL, Oefner PJ (2000) Y chromosome sequence variation and the history of human populations. *Nat Genet* 26:358–361
- Wells RS, Yuldasheva N, Ruzibakiev R, Underhill PA, Evseeva I, Blue-Smith J, Jin L et al (2001) The Eurasian heartland: a continental perspective on Y-chromosome diversity. *Proc Natl Acad Sci USA* 98:10244–10249
- Wilson IJ, Balding DJ (1998) Genealogical inference from microsatellite data. *Genetics* 150:499–510
- Wolpert S (2000) *A new history of India*. Oxford University Press, New York
- Y Chromosome Consortium (2002) A nomenclature system for the tree of human Y-chromosomal binary haplogroups. *Genome Res* 12:339–348
- Zerjal T, Xue Y, Bertorelle G, Wells SR, Bao W, Zhu S, Qamar R, Ayub Q, Mohyuddin A, Fu S, Li P, Yuldasheva N, Ruzibakiev R, Xu J, Shu Q, Du R, Yang H, Hurles ME, Robinson E, Gerel-saikhan T, Dashnyam B, Mehdi SQ, Tyler-Smith C (2003) The genetic legacy of the Mongols. *Am J Hum Genet* 72:717–721