

## DHPLC Applications: Finding DNA variation on the Y chromosome

Qasim Ayub<sup>1</sup>, Aisha Mohyuddin<sup>1</sup>, Shagufta Khaliq<sup>1</sup>, Abdul Hameed<sup>1</sup>, Chris Tyler-Smith<sup>2</sup> and S. Qasim Mehdi<sup>1</sup>

<sup>1</sup>Biomedical and Genetic Engineering Division, Dr. A. Q. Khan Research Laboratories, P.O. Box 2891, Islamabad 44000, Pakistan.

<sup>2</sup>CRC Chromosome Molecular Biology Group, Department of Biochemistry, University of Oxford, Oxford, UK.

### Abstract

Denaturing High-Performance Liquid Chromatography (DHPLC) is a recently developed technique for the detection of single nucleotide polymorphisms (SNPs) and mutations. It involves the comparison between two or more DNAs as a mixture of denatured and reannealed PCR products. The methodology is based on the principle of reversed phase liquid chromatography and uses a unique DNA separation matrix. The exquisite sensitivity of the technique is determined by adjusting the oven temperature. Elution of DNA fragments is dependant on the chain length and sequence and can be predicted by computation. Under partially denaturing conditions heteroduplexes formed upon mixing, denaturing, and reannealing of two, or more, chromosomes that differ in sequence, are retained less than their corresponding homoduplexes and sequence variation is recognized by the appearance of two, or more, peaks in the chromatographs. Numerous SNPs have been identified on the non-recombinant portion of the Y chromosome by using this technique. To investigate the DNA variation within Pakistan 15 Y-SNPs, an Alu insertion, a LINE1 insertion and the 12f2 deletion, mapping on the non-recombining portion of the human Y chromosome, were typed in 834 Pakistani males. The combination of these biallelic markers identified 11 stable Y chromosomal lineages or Y 'haplogroups' in the Pakistani population. Haplogroup frequencies were generally similar to

those in neighboring geographical areas and indicate that there was a common pool of Y lineages within Pakistan that are predominantly from West and Central Asia.

**Keywords:** DHPLC, Y-polymorphisms, Y SNPs, Population genetics, Pakistan.

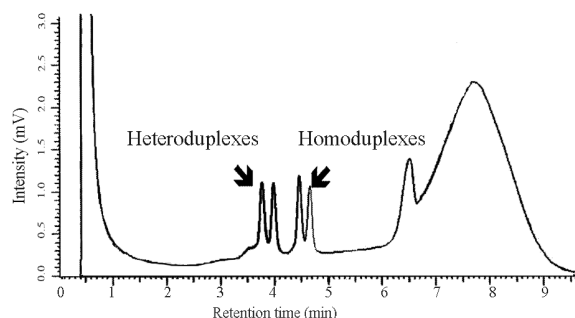
### INTRODUCTION

Denaturing High-Performance Liquid Chromatography (DHPLC) is a recently developed technique for the detection of single nucleotide polymorphisms (SNPs) and mutations. It involves the comparison between two or more DNAs as a mixture of denatured and reannealed PCR products. The PCR products do not require purification and single nucleotide substitutions, deletions and insertions have been detected successfully by on line monitoring within 2-6 minutes (For review see Xiao and Oefner 2001). The instrumentation is based on the principle of ion-paired reversed phase high-performance liquid chromatography and uses a unique DNA separation matrix. The proprietary matrix (DNASep column) consists of an alkylated non-porous poly-styrene divinyl benzene C-18 particles, 2-3 µm in diameter. The double stranded DNA molecules are retained on the column by binding with amphiphilic ions (0.1 M tri-ethylammonium acetate (TEAA) buffer, pH 7.0). The PCR products are eluted in 0.1 M TEAA buffer (pH 7.0) with a linear acetonitrile gradient at a flow rate of 0.9 ml/min. The exquisite sensitivity of the

Correspondence to: Qasim Ayub, Ph.D  
Tel: +92-51-9261138; Fax: +92-51-9261144;  
E-mail: sqmehdi@isb.comsats.net.pk

technique is determined by adjusting the oven temperature. The optimum temperature and elution profiles are dependant upon the chain length and sequence and can be predicted by computation (<http://insertion.stanford.edu/melt1.html/>). Under partially denaturing conditions heteroduplexes formed upon mixing, denaturing, and reannealing of two, or more, DNAs that differ in sequence, are retained less than their corresponding homoduplexes and sequence variation is recognized by the appearance of two, or more, peaks in the elution profiles (Figure 1). This methodology has been successfully applied for high throughput screening and identification of numerous Y chromosomal SNPs (Underhill *et al.* 1997; 2001). Briefly sequences of the non-recombinant portion of the human Y chromosome available in public databases (such as <http://www.ncbi.nlm.nih.gov/>) were screened *In silico* for unique sequences. Human repeat DNA sequences were identified by masking repeat sequences using the RepeatMasker2 software (<http://ftp.genome.washington.edu/cgi-bin/RM2>).

Primers were designed ([http://www-genome.wi.mit.edu/genome\\_software/other/primer3.html](http://www-genome.wi.mit.edu/genome_software/other/primer3.html)) to amplify 150-700 base pairs of unique sequences by PCR. Heteroduplex analyses were carried out by mixing, denaturing and reannealing of unpurified PCR product from a single reference male individual with an amplicon from another individual (Figure 1). Y-SNPs were identified by the appearance of two or more peaks in the elution profiles and confirmed by DNA sequencing.



**Figure 1.** The chromatographic elution profile of the sY81 A to G Y-SNP following DHPLC.

The Y chromosome carries the largest non-recombining segment in the human genome and this allows

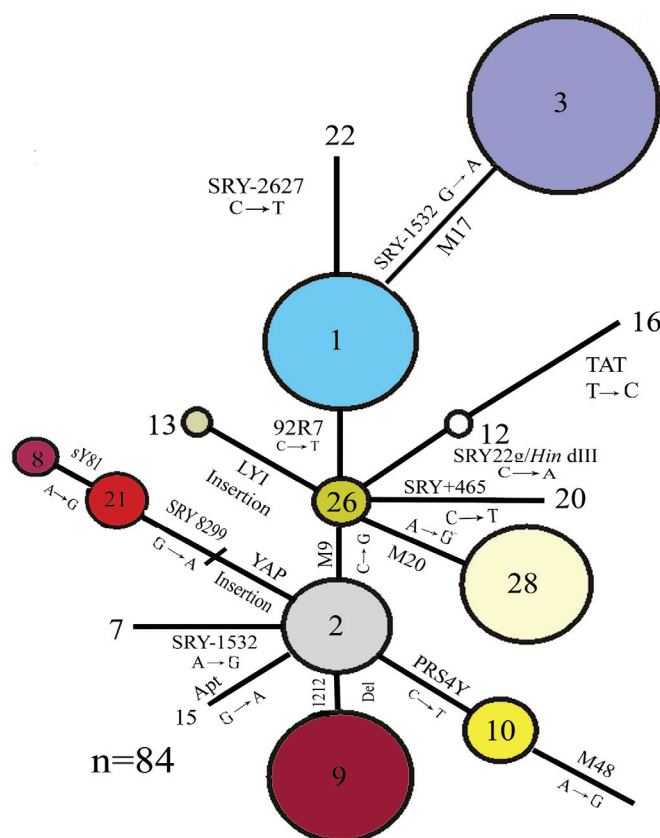
investigation of the male genetic history. To investigate the DNA variation within Pakistan 18 biallelic markers mapping on the non-recombining portion of the human Y chromosome were examined. These included 5 Y-SNPs that had been identified previously by DHPLC (Underhill *et al.* 1997).

## MATERIALS AND METHODS

**Samples:** The Y chromosomes of 834 unrelated Pakistani males belonging to fifteen ethnic groups of Pakistan were analyzed. The details about the various ethnic groups are given elsewhere (Mehdi *et al.* 1999). Informed consent was obtained from all participants in this study. DNA was extracted from EBV-transformed lymphoblastoid cell lines that were established for each individual.

**Y-Chromosome Binary Polymorphisms:** Eighteen biallelic markers, including 15 Y-SNPs, an *Alu* insertion (Hammer 1994; Hammer and Horai 1995), a LINE1 insertion (Santos *et al.* 2000) and the 12f2 deletion (Casanova *et al.* 1985), were examined. The SNPs studied were 92R7 C to T (Mathias *et al.* 1994); SRY-2627 C to T (Bianchi *et al.* 1997); SRY-1532 A to G to A (Kwok *et al.* 1996; Santos *et al.* 1999b; Whitfield *et al.* 1995); sY81 A to G (Seielstad *et al.* 1994); SRY-8299 G to A (Santos *et al.* 1999a); Apt G to A (Pandya *et al.* 1998); SRY +465 C to T (Shinka *et al.* 1999); LLY22g C to A, Tat T to C transition (Zerjal *et al.* 1997) and the RPS4Y C to T mutation (Bergen *et al.* 1999). Markers that were identified by DHPLC included M9, M11, M17, M20 and M48 (Underhill *et al.* 1997). The set of Y binary marker alleles carried by a single individual was referred to as the Y 'haplogroup' (Tyler-Smith 1999).

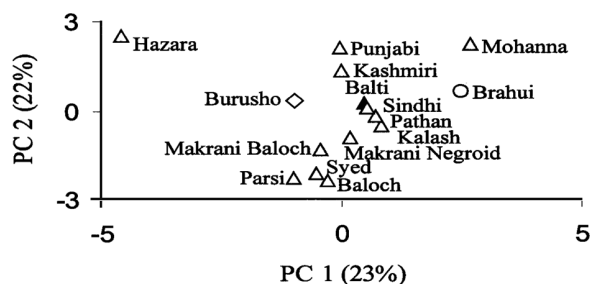
**Data Analysis:** Principal components analysis was carried out on haplogroup frequencies using the ViSta (Visual Statistics) system software version 6.0 (Young and Bann 1996). For graphic representation the first and second principal components were plotted using the Microsoft Office Suite Excel Package on Windows 2000.



**Figure 2. Phylogeny of Y haplogroups in Pakistan.** An unrooted maximum parsimony tree showing Y haplogroups (numbers in circles) and mutations (on lines). Circle area is proportional to haplogroup frequency in Pakistan.

## RESULTS

The combination of the biallelic markers identified 11 stable Y chromosomal lineages or Y ‘haplogroups’ in the Pakistani population (Figure 2). Of these five haplogroups (1, 2, 3, 9 and 28) accounted for 92.1% of the sample. Haplogroups 1, 3 and 9 were present in all Pakistani populations examined except for the Mohanna, Hazara and Punjabi respectively. Haplogroup 28 was found in most Pakistani ethnic groups except for the Hazara, Kashmiri and Punjabi. The south-western Pakistani populations (Baloch, Brahui, Makrani Baloch, Makrani Negroid, Parsi, and Sindhi) show higher frequencies of haplogroup 9 and the YAP+ haplogroups 21 and 8 compared to the northern populations (Balti, Burusho, Hazara, Kalash, Kashmiri, Pathan, Punjabi). However, no strong overall geographical clustering of haplogroup frequencies is apparent within the



**Figure 3. Principal components analysis of Y haplogroup frequencies within Pakistan.** Indo-European speakers are indicated by triangles, Sino-Tibetan by a filled triangle, Dravidian by a circle and the language isolate Burusho by a diamond.

country.

The principal components analysis (Figure 3) shows that the Pakistani populations cluster with each other except for the Hazara. It is striking that the Pakistani populations speaking a language isolate

(the Burushos), a Dravidian language (the Brahui), or a Sino-Tibetan language (the Balti) resembled the Indo-European-speaking majority and do not stand out in these analyses.

## DISCUSSION

To investigate the DNA variation within Pakistan 15 Y-SNPs, an *Alu* insertion (Hammer 1994), a LINE1 insertion (Santos *et al.* 2000) and the 12f2 deletion (Casanova *et al.* 1985) were examined in 834 male samples. Of these 5 SNPs had been identified by DHPLC. This technique has been successfully used to identify numerous such polymorphisms on the non-recombinant part of the human Y chromosome (Underhill *et al.* 1997; 2001).

The combination of these markers identified 11 stable Y chromosomal lineages or Y 'haplogroups' within Pakistan. The haplogroup frequencies within Pakistan were generally similar to those in neighboring geographical areas to the west and the Pakistani populations speaking a language isolate (the Burushos), a Dravidian language (the Brahui), or a Sino-Tibetan language (the Balti) resembled the Indo-European-speaking majority. The Hazaras are the most distinct among the Pakistani populations.

They are Mongoloid and claim descent from Genghis Khan's army. Their Y chromosomal DNA variation supports this claim. The greater genetic similarity of Pakistani populations to those in the west than to eastern populations is illustrated by the fact that four of the five frequent haplogroups in Pakistan (1, 2, 3 and 9) are also frequent in Western Asia and Europe but not in the Chinese or Japanese. These patterns can be accounted for by a common pool of Y lineages within Pakistan that arrived predominantly from West Asia and the Levant (Qamar *et al.* 1999, Qamar *et al.* 2002; Quintana-Murci *et al.* 2001). This is supported by historical records that describe the arrival of the Indo-European language speaking pastoral nomads from Western and Central Asia into present day Pakistan (Wolpert 2000). The fifth common haplogroup in Pakistan is haplogroup 28. It is delineated by the M20 A to G transition and was originally identified by DHPLC in a Pakistani

sample (Underhill *et al.* 1997). This haplogroup makes up to 13% of the Pakistani sample. It has rarely been found outside Pakistan and is probably associated with the spread of agriculture in the Indus Valley (Qamar *et al.* 2002). The use of DHPLC technology has greatly facilitated studies on human migration and evolution from the male perspective and shed light on the origins of modern human populations.

## Acknowledgments

This work was supported by a Wellcome Trust Collaborative Research Initiative Grant to SQM. We express our gratitude to the original DNA donors who made this study possible.

## References

- Bergen A W, Wang C-Y, Tsai J, Jefferson K, Dey C, Smith K D, Park S-C, Tsai S-J and Goldman D. (1999) An Asian-Native American paternal lineage identified by *RPS4Y* resequencing and by microsatellite haplotyping, *Ann. Hum. Genet.*, 63, 63-80.
- Bianchi N O, Bailliet G, Bravi C M, Carnese R F, Rothhammer F, Martinez-Marignac V L and Pena S D. (1997) Origin of Amerindian Y-chromosomes as inferred by the analysis of six polymorphic markers, *Am. J. Phys. Anthropol.*, 102, 79-89.
- Casanova M, Leroy P, Boucekkine C, Weissenbach J, Bishop C, Fellous M, Purrello M, Fiori G and Siniscalco M. (1985) A human Y-linked DNA polymorphism and its potential for estimating genetic and evolutionary distance, *Science*, 230, 1403-1406.
- Hammer M F. (1994) A recent insertion of an *Alu* element on the Y chromosome is a useful marker for human population studies, *Mol. Biol. Evol.*, 11, 749-761.
- Hammer M F and Horai S. (1995) Y chromosomal DNA variation and the peopling of Japan, *Am. J. Hum. Genet.*, 56, 951-962.
- Kwok C, Tyler-Smith C, Mendonca B B, Hughes I., Berkovitz G D, Goodfellow P N and Hawkins J R. (1996) Mutation analysis of the 2 kb 5' to SRY in XY females and XY intersex subjects, *J. Med. Genet.*, 33, 465-468.
- Mathias N, Bayes M and Tyler-Smith C. (1994) Highly informative compound haplotypes for the human Y chromosome, *Hum. Mol. Genet.*, 3, 115-123.
- Mehdi S Q, Qamar R, Ayub Q, Khaliq S, Mansoor A, Ismail M, Hammer M F, Underhill P A and Cavalli-Sforza L L. (1999) The origins of Pakistani populations: evidence from Y chromosome markers. In:



- Genomic Diversity: Applications in Human Population Genetics*, Papiha S. S., Dekka, R. and Chakraborty, R. ed., Kluwer Academic/Plenum Publishers, New York, PP. 83-90.
- Pandya A, King T E, Santos F R, Taylor P G, Thangaraj K, Singh L, Jobling M A and Tyler-Smith C. (1998) A polymorphic human Y-chromosomal G to A transition found in India, *Ind. J. Hum. Genet.*, 4, 52-61.
- Qamar R, Ayub Q, Khaliq S, Mansoor A, Karafet T, Mehdi S Q, Hammer M F. (1999) African and Levantine origins of Pakistani YAP+ Y chromosomes, *Hum. Biol.*, 71, 745-755.
- Qamar R, Ayub Q, Mohyuddin A, Helgason A, Mazhar K, Mansoor A, Zerjal T, Tyler-Smith C and Mehdi S Q. (2002) Y-chromosomal DNA variation in Pakistan, *Am. J. Hum. Genet.*, 70, 1002-1024.
- Quintana-Murci L, Krausz C, Zerjal T, Sayar S H, Hammer M F, Mehdi S Q, Ayub Q, Qamar R, Mohyuddin A, Radhakrishna U, Jobling M A, Tyler-Smith C and McElreavey K. (2001) Y-chromosome lineages trace diffusion of people and languages in southwestern Asia, *Am. J. Hum. Genet.*, 68, 537-542.
- Santos F R, Carvalho-Silva D R and Pena S D J. (1999a) PCR-based DNA profiling of human Y chromosomes. In *Methods and Tools in Biosciences and Medicine*, Epplen, J. T. and Lubjuhn, T. ed., Birkhauser Verlag, Basel, Switzerland, PP. 133-152.
- Santos F R, Pandya A, Tyler-Smith C, Pena S D, Schanfield M, Leonard W R, Osipova L, Crawford M H and Mitchell R J. (1999b) The central Siberian origin for native American Y chromosomes. *Am. J. Hum. Genet.*, 64, 619-628.
- Santos F R, Pandya A, Kayser M, Mitchell, R J, Liu A, Singh L, Destro-Bisol G, Novelletto A, Qamar R, Mehdi S Q, Adhikari R, Knijff P and Tyler-Smith C. (2000) A polymorphic L1 retroposon insertion in the centromere of the human Y chromosome, *Hum. Mol. Genet.*, 9, 421-430.
- Seielstad M T, Hebert J M, Lin A A, Underhill P A., Ibrahim M, Vollrath D and Cavalli-Sforza L L. (1994) Construction of human Y-chromosomal haplotypes using a new polymorphic A to G transition, *Hum. Mol. Genet.*, 3, 2159-1261.
- Shinka T, Tomita K, Toda T, Kotliarova S E, Lee J, Kuroki Y, Jin D K, Tokunaga K, Nakamura H and Nakahori Y. (1999) Genetic variations on the Y chromosome in the Japanese population and implications for modern human Y chromosome lineage, *J. Hum. Genet.*, 44, 240-245.
- Tyler-Smith C. (1999) Y-chromosomal DNA markers. In: *Genomic Diversity: Applications in Human Population Genetics*, Papiha S S, Dekka R and Chakraborty R. ed., Kluwer Academic/Plenum Publishers, New York, PP. 65-73.
- Underhill P A, Jin L, Lin A A, Mehdi S Q, Jenkins T, Vollrath D, Davis R W, Cavalli-Sforza L L and Oefner P J. (1997) Detection of numerous Y chromosome biallelic polymorphisms by denaturing high-performance liquid chromatography, *Genome Res.*, 7, 996-1005.
- Underhill P A, Passarino G, Lin A A, Shen P, Lahr M M, Foley R, Oefner P J and Cavalli-Sforza L L. (2001) The phylogeography of Y chromosome binary haplotypes and the origins of modern human populations, *Ann. Hum. Genet.*, 65, 43-62.
- Whitfield L S, Sulston J E and Goodfellow P N (1995) Sequence variation of the human Y chromosome, *Nature*, 378, 379-380.
- Wolpert S (2000) A New History of India (New York Oxford University Press).
- Xiao W and Oefner P J. (2001) Denaturing high-performance liquid chromatography: a review. *Hum. Mut.*, 17, 439-474.
- Young F W and Bann. M. (1996) A visual statistics system. In *Statistical Computing Environments for Social Researches*, Stine, R. A. and Fox, J. ed., Sage Publications, New York, PP. 207-236.
- Zerjal T, Dashnyam B, Pandya A, Kayser M, Roewer L, Santos F R, Schiefenhover W, Fretwell N, Jobling M A, Harihara S, Shimizu K, Semjiddmaa D, Sajantila A, Salo P, Crawford M H, Ginter E K, Evgrafov O V and Tyler-Smith C. (1997) Genetic relationships of Asians and Northern Europeans, revealed by Y-chromosomal DNA analysis, *Am. J. Hum. Genet.*, 60, 1174-1183.