

YAP, signature of an African–Middle Eastern migration into northern India

Suraksha Agrawal^{1,*}, Faisal Khan¹, Atul Pandey¹, Manorama Tripathi¹ and Rene J. Herrera²

¹Department of Medical Genetics, Sanjay Gandhi Postgraduate Institute of Medical Sciences, Raebareilly Road, Lucknow 226 014, India

²Department of Biological Sciences, Florida International University, Miami, Florida 33199, USA

YAP, an *Alu* insertion polymorphism found on human Y-chromosome is present in two lineages worldwide, corresponding to M145/M203/SRY4064 (haplogroup E) and M145/M203/M174 (haplogroup D) polymorphisms respectively. First lineage belonging to haplogroup D is specific to Japan and other Southeast Asian populations, while haplogroup E is confined to Sub-Saharan African, Middle Eastern and Southern European populations. In the present study, 1021 Y-chromosomes belonging to nine different populations of North India were analysed for YAP insertion and four other single nucleotide polymorphisms (SNPs) to delineate the two lineages. Out of nine populations only one, i.e. Shiya Muslims revealed presence of YAP element at a frequency of 11%. Further analysis based on four additional SNPs revealed that all the YAP+ve samples could be categorized under African/Middle East-specific haplogroup E lineage. Interestingly, Sunni Muslims who historically have the same origin, i.e. from the Middle east showed a complete lack of YAP+ve lineage similar to other castes. We hypothesize that unlike Sunnis, Shiya Muslims due to their lesser number and less admixture with other caste groups of India, still carry the ancestral YAP+ve lineage, which in all probabilities is one of the founder haplogroups. All Middle Eastern populations show the presence of this lineage in almost similar frequency. Our study shows the presence of YAP+ve lineage in North Indian populations, reflecting an African/Middle Eastern migration into North India.

THE Y-chromosome – a single haploid entity passed from father to son, is a highly suitable marker to trace the patrilineal migration due to its uniparental transmission, lack of recombination and its high sensitivity despite small effective population size. At least three different types of polymorphism have been reported on the Y-chromosome: contraction–expansion mutation at tandem repeats markers¹, single nucleotide polymorphism (SNP) mutation^{2–5} and insertions of reiterated elements (e.g. YAP polymorphism)⁶. The YAP marker originated when an *Alu* repetitive element retro transposed at the DYS 287 locus at location Yq11. The use of this polymorphic *Alu* insertion (PAI) in human population studies has been bolstered by DNA sequencing studies that show that the YAP element is inserted be-

tween the same two base pairs on eight Y-chromosomes from different geographical origins⁷. This observation indicated that YAP is the product of a single insertion event and therefore, useful to draw the relationships by descent.

The frequency of Y-chromosomes carrying YAP element (YAP+ve) is highest among sub-Saharan African populations (82–95%), followed by North African populations (50–70%), and is rare among Europeans (1–10%)^{2–5,7}. Most Asian populations examined so far lack this element. An exception to this pattern is the presence of YAP element in the Middle East and Central Asian populations (12–27%)^{8,9}. YAP element is present among Japanese populations in relatively higher frequency (27–85.7%)¹⁰. Interestingly, YAP+ve individuals in these two diverse geographical areas of Africa and Japan have different lineages.

In the present study, an attempt has been made to investigate the presence of YAP+ve lineages among different North Indian population groups. People of the Indian subcontinent are one of the most heterogeneous populations in the world, having enormous genetic, cultural and linguistic diversity¹¹. Present-day Indian populations are differentiated by 20 major languages, hundreds of dialects, various religions, and are subdivided into numerous (approximately 2000) castes and tribal groups¹². Among the non-tribal groups, a vast majority of the ethnic groups (about 82%) follow Hindu religion and are socially organized into castes and sub-castes¹³. Tribal groups are outside the caste hierarchy and constitute about 8.08% of the total Indian population¹¹. Muslims are the second largest non-tribal religious group. Muslims of the Indian subcontinent are of two types, Sunni and Shiya, and constitute approximately 12% of total Indian population¹⁴. The other important fact about Indian culture and the society is that India has experienced various migrational events, invasions and massive gene flow from Central Asia, South China and Southeast Asia¹⁵, in particular, migration of the Indo-Aryan speakers (Indo-Europeans) from the steppes of Central Asia, who settled in the Indus Valley (northwestern India) ~4000 yrs BP¹⁶. Migration of the Indo-Aryan speakers led to the southward retreat of Dravidians in India^{16–18}.

The present study has been envisaged to ascertain the presence of the YAP insertion in various North Indian groups of Brahmins, Bhargavas, Chaturvedis, Kayastha, Rastogis, Vaish, Mathurs, Sunni and Shiya Muslims. YAP+ve lineage has been further analysed for M-145 and M-203 markers, which are equivalent to YAP insertion and M-174 and SRY-4064 markers, to delineate the two different lineages specific to African/Middle East Asian and East Asian/Japanese populations respectively.

A total of nine populations belonging to different caste and religious groups were selected from Uttar Pradesh (UP), a state in North India. Seven of these belong to Hindu caste groups, while rest are from the Muslim sects (Sunni and Shiya). Among the Hindu groups three populations, namely Bhargavas, Chaturvedis and Brahmins are upper caste groups, who follow strict surname endogamy.

*For correspondence. (e-mail: suraksha@sippi.ac.in)

The other four populations, namely Kayastha, Rastogies, Mathurs and Vaish belong to middle caste groups. These four populations claim to practice group endogamy. Both sects of Muslims practice consanguinity.

A total of 1800 unrelated individuals were randomly selected, 200 from each of the nine population groups, viz. Bhargavas, Chaturvedis, Brahmins, Kayastha, Rastogies, Mathurs, Vaish, Sunni and Shiya Muslims. Before sample collection, regional addresses and detailed computerized lists of the populations were prepared. Random numbers were generated with the help of a computer and samples were collected from different collection sites of UP–Lucknow, Kanpur, Faizabad, Basti, Gonda and Agra. Whole blood was collected in EDTA vacutainer tubes. Three-generation pedigree charts were prepared to assure un-relatedness in all the groups. Blood samples were collected after obtaining informed consent from the subjects and ethical clearance from the institute.

DNA was extracted by phenol chloroform method, as described by Comey *et al.*¹⁹ and purified by ethanol precipitation. A total of 1021 male samples – Bhargavas ($n = 86$), Chaturvedis ($n = 104$), Brahmins ($n = 114$), Kayastha ($n = 132$), Rastogies ($n = 136$), Mathurs ($n = 108$), Vaish ($n = 65$), Sunni ($n = 124$) and Shiya ($n = 154$) were typed for four Y-chromosome markers.

Presence of an *Alu* element at the YAP locus was detected by PCR amplification using flanking primers 5'CAGGGGAAGATAAAGAAATA3' and 5'ACTGCTAAAAGGGGATGGAT3' against unique sequences, as described by Hammer *et al.*⁶. The amplicons were then size-fractionated on 2% agarose 1X TBE to detect the presence of the 340 bp YAP insertion.

All the samples carrying YAP insertion were analysed for M-145 G → A transition and M-203 G → C transversion. All YAP+ve samples were amplified for M-145 and M-203 polymorphisms using flanking primers, as previously described⁴. Sequencing was carried out using forward primer by means of standard cycle sequencing using florescent dideoxy nucleotides. The fluorescently labelled extension product was run on ABI 310 DNA fragment size analyser. The data obtained were then analysed using sequencing 3.4.1 software (Applied Biosystems). Mutation in each sample was detected by aligning the sequences with control sample using CLUSTAL W software.

All samples having YAP insertion, 'A' allele of M-145 and 'C' allele of M-203 were then analysed for two different lineages. The first lineage or haplogroup D corresponds to T → C transition at M-174 and other lineage or haplogroup E corresponds to G → A transition³ at SRY-4064. Initially, all samples were analysed for the presence of the T → C transition at M-174 using standard sequencing. YAP+ve samples with ancestral allele (T) at M-174 were then analysed for the presence of G → A transition at SRY-4064. Subsequent to PCR amplification, the amplicons were digested with *Bsr*B1. The intact site corresponds to the presence of 'A' allele.

Out of the 1021 male samples analysed for YAP insertion, only 17 were YAP+ve. Interestingly, all the 17 samples were from Shiya Muslims. None of the other population groups showed YAP insertion (Table 1). Moreover, none of the earlier reports has documented the presence of YAP insertion among North Indians. All the YAP+ve samples exhibited the G → A transition at M-145 and the T → C transition at M-203.

All 17 YAP+ve Shiya samples possessed 'A' allele at SRY-4064 locus, indicating the presence of the YAP/

Table 1. Comparison of frequency of YAP insertion and two YAP+ve lineages among various Indian and world populations

Population	No. of samples	Haplogroup DE (%)	Haplogroup D (%)	Haplogroup E (%)
Studied populations				
Bhargavas	86	–	–	–
Chaturvedis	112	–	–	–
Brahmins	104	–	–	–
Kayastha	132	–	–	–
Rastogies	136	–	–	–
Mathurs	108	–	–	–
Vaish	65	–	–	–
Shia Muslims	154	11.30	–	11.30
Sunni Muslims	124	–	–	–
Other North Indian populations¹³				
Brahmins	17	–	–	–
Chamars	18	–	–	–
Muslims	19	–	–	–
Rajputs	35	–	–	–
South Indian populations²⁰				
Vizag Brahmins	41	–	–	–
Peruru Brahmins	44	–	–	–
Kammas	40	–	–	–
Siddis	13	38.40	Not analysed	Not analysed
Tribal populations				
Bagota ¹⁶	23	–	–	–
Poroja ¹⁶	20	–	–	–
Chenchu ²²	41	–	–	–
Koya ²²	41	–	–	–
Middle East Asian populations				
Iraqis ⁸	139	12.2	–	12.2
Lebanese ⁸	31	25.8	–	25.8
Turks ⁸	30	16.6	–	16.6
Syrians ⁹	20	20.0	–	20.0
Palestinians ⁹	73	19.0	–	19.0
Saudi Arabians ⁹	21	5.0	–	5.0
African populations⁶				
Zulu	47	81	–	81
Pygmies	24	79	–	79
Khosians	68	46	–	46
East Asian populations				
Japanese Ainu	16	87.5	87.5	–
Japanese Honshu	82	36.6	36.6	–
Japanese Kiyushu	104	27.9	27.9	–
Japanese Okinawa	45	55.6	55.6	–
Thais	34	2.9	2.9	–

M145/M203/SRY4064 lineage (haplogroup E), which is found at a high frequency among African populations and among Middle Eastern groups. Presence of the 'T' allele at M-174 locus confirms absence of haplogroup D and thus indicates that the source of YAP+ve lineage among Shiya Muslims is of African origin, rather than of East Asian origin.

The present report documents the presence of YAP+ve lineage among any contemporary North Indian population. Further, the presence of YAP/M145/M03/SRY4064 (haplogroup E) lineage clearly indicates that the source of this lineage among Shiya Muslims is either from Africa or Middle East. Other groups analysed in the present study showed complete lack of the YAP+ve lineage (Table 1).

Different studies based on Y-chromosome and mt-DNA have reported that both North and South Indian populations carry both Neolithic and Palaeolithic haplotypes, although depicting a relative homogeneity at mt-DNA, with most of the lineages belonging to Indian specific super-haplogroup 'M'^{14,20-24}. However, recent studies based on Y-chromosome markers have suggested that present-day North Indians are mostly Indo-Europeans, who have migrated from West Eurasia via the fertile crescent of Middle East Asia and entered India through the western corridor of the Indus Valley^{4,14,16,18,20}. These individuals populated most of the northern, central and western India, pushing the proto-Asian Dravidian inhabitants towards the south ~4000–3000 yrs BP^{16,18,20,21}. The proto-Asian Dravidians were ascendants of Elamites, inhabitants of Elam (southwestern Iran) and speakers of proto-Elamo Dravidian languages^{16,18}. They brought with them the skill of cattle farming, the concept of iron work, and also European and western Eurasian haplogroups^{16,18,25}. Significantly, the YAP+ve lineage, although present in Middle Eastern populations in a frequency of 11–25%, has never been reported in any other North Indian or South Indian populations^{8,9}. The only exception is Siddis, a migrant group of East African ancestry found in the southern state of Andhra Pradesh. Siddis show a high frequency of YAP+ve lineage (40%)^{26,27}.

Underhill *et al.*⁴ have explained, on the basis of a global survey of 134 Y-bi-allelic markers, that all Y-chromosomes that are not exclusively African contain the M-168 mutation, which might have originated within an East African population. M-168 lineages evolved into three distinct sub-clusters: one which acquired an *Alu* insertion (YAP at DYS287) as well as the M-145 and M-203 nucleotide substitutions, and two other lineages, defined by the distinct mutations RPS4Y/M216 and M89/M213. These three sub-clusters represent deep structuring of Y-chromosome diversity outside Africa²⁻⁵. Out of the three lineages, the branch that acquired the YAP/M145/M203 polymorphisms in Africa has been divided into two sub-clusters. One is now found in Africa and the Mediterranean regions and defined by the SRY4064/M96 mutations. The other is found in East Asia and is defined by the M-174 mutation.

The YAP/M145/M203/SRY4064 lineages are found with high frequencies in Africa, and relatively high frequencies in the Middle East and southern Europe (characterized by the M35/M215 mutations), with occasional occurrences in Central Asia, Pakistan and America²⁻⁵. However, YAP/M145/M203/M174 is exclusively Asian and carries the ancestral alleles for SRY-4064 and M96. This lineage is now mostly confined to Japan (27.5–87.5%) and Tibet (1.5–2.7%)¹⁰, where it occurs at a high frequency; however, it is found scattered throughout Southeast Asia with a low frequency²⁸. On the basis of YAP and SRY-4064 nested cladistic analysis, Underhill *et al.*⁴ have concluded that the YAP/M145/M203 polymorphisms have an African origin, as the M-174 ancestral allele among YAP+ve lineages is found exclusively in Africa.

The presence of YAP+ve haplogroup E among Shiya Muslims with a frequency of 11% is strikingly similar to various Middle Eastern populations, including Iraqis (10.8%), Labanese (25.8%), Syrians (10%), Palestinians (17%), Saudi Arabians (5%) and Israelis (20.3%)^{8,9}. All these populations carry most of the haplogroups common to the one found in West Eurasia along with YAP+ve lineage of sub-Saharan African populations. Underhill *et al.*⁴ have suggested that an East African population with a sub-clad of the African YAP/M145/M203/SRY4064 cluster expanded into the Middle East and the Arab communities within the last few thousand years. Further, farmers may have introduced these lineages from the Middle East into southern Europe and to a lesser extent to northern India and Pakistan during the Neolithic expansion⁴. The presence of YAP+ve lineage and the striking similarity in frequency between Indian Shiya Muslims and other Middle East populations have their basis in the history of the origin and expansion of Muslim populations in India.

Muslims represent the second largest non-tribal population group in India (12% of total population of India). There are two sects of Muslims found worldwide – Sunni and Shiya. The Sunni Muslims predominate and constitute more than 90% of the Muslims found in India. Muslim rulers from the borders of Iran, Turkey and Afghanistan invaded India between AD 1000 and 1527 on several occasions^{11-13,29}. They spread and flourished in different parts of the country, converting upper and middle caste Hindu populations to Islam¹⁴. Muslims, unlike the Hindus of India, do not follow caste hierarchy, but they practice a high level of consanguinity and the two sects rarely marry among each other.

Our results demonstrated that YAP+ve lineage is present only in Shiya and not in Sunni Muslims and also in other caste groups, despite the fact that both groups have a common origin historically²⁹. Sunni Muslims have ruled different parts of India for around 3000 years and during their reign, have spread within the subcontinent. Despite practising consanguinity, they married outside their religion. This is supported by various studies^{13,14} that Indian Muslims, specifically North Indian Muslims, have high ge-

netic similarity with other Hindu caste populations. High rate of admixture with local caste populations and high extinction rate of Y-haplotype² might account for the apparent absence of YAP+ve lineages among Sunni Muslims. Shiya Muslims on the other hand, due to their less numbers, may have remained more culturally and genetically isolated within their communities. Under these conditions, ancestral Y haplogroups may have survived and persist till date. However, analysis of RPS4Y-'T' chromosome that occupies a high hierarchical position of haplogroup 'C' in the Y-binary tree³, will provide a clear lineage, as Zerjal *et al.*³⁰ have reported that about 8% of males belonging to 16 different populations from Pacific to Caspian Sea show haplogroup 'C' lineage, often considered as genetic legacy of the Mongols.

The present study explores the presence of YAP+ve lineage in Shiya. Our study suggests that this could have been one of the founder lineages as the Shiites migrated into India, as revealed by the presence of YAP/M145/M203/SRY4064 lineage and its nearly similar frequency to those found among Middle Eastern populations, which are in all probability the source of Indian Muslim populations. Historians have demonstrated in the past that Shiya Muslims were less effective in converting other groups of people to their religion compared to Sunni Muslims.

Examination of additional diagnostic Y-specific haplogroups in the YAP-ve lineage as well as mt-DNA lineages may help further ascertaining the male-driven gene flow observed in the present study.

23. Kivisild, T. *et al.*, *Am. J. Hum. Genet.*, 2003, **72**, 313–332.
24. Palanichamy, M. G. *et al.*, *Am. J. Hum. Genet.*, 2004, **75**, 966–978.
25. Metspalu, M. *et al.*, *BMC Genet.*, 2004, **5**, 26.
26. Ramana, G. V., Singh, L. and Chakraborty, R., *Hum. Biol.*, 2001, **73**, 71–80.
27. Ramana, G. V., Su, B., Jin, L., Singh, L., Wang, N., Underhill, P. and Chakraborty, R., *Eur. J. Hum. Genet.*, 2001, **9**, 695–700.
28. Su, B. *et al.*, *Genet.*, 1999, **65**, 1718–1724.
29. Hutton, J. H., *Caste in India*, Oxford University Press, London, 1961.
30. Zerjal, T. *et al.*, *Am. J. Hum. Genet.*, 2003, **72**, 717–724.

ACKNOWLEDGEMENTS. This project was supported by the Indian Council of Medical Research, New Delhi. We thank Sanjay Gandhi Postgraduate Institute of Medical Sciences, Lucknow for providing various laboratory facilities and assistance.

Received 8 November 2004; revised accepted 24 February 2005

Co-seismic and post-seismic displacements in Andaman and Nicobar Islands from GPS measurements

Sridevi Jade*, M. B. Ananda, P. Dileep Kumar and Souvik Banerjee

CSIR Centre for Mathematical Modelling and Computer Simulation, Belur Campus, Bangalore 560 037, India

We calculate the displacements of four sites in Andaman and Nicobar islands from Diglipur (13.16°N) to Car Nicobar (9.22°N), using GPS measurements made at these sites in September 2003 and repeated in February 2005, assuming that the inter-seismic displacements at all these sites can be represented by the 14 ± 2 mm/yr convergence between CARI near Port Blair and Bangalore, reported earlier on the basis of GPS measurements made by us between 1996 and 1999. Since the latest measurements were made after about a month of the great Sumatra event of 26 December 2004, and several moderate earthquake ruptures had since occurred adding to the co-seismic surface displacements, the values reported here also represent the contributions of after-shocks. Rigorous analysis of the two epoch GPS datasets from these sites yields precise displacement vectors, among which the one at Car Nicobar has the largest horizontal magnitude (6.49 ± 0.009 m to the WSW) with a significant 1.1 m subsidence. The horizontal displacement at Chatham Island near Port Blair is also similarly oriented, but smaller (3.53 ± 0.010 m) with reduced subsidence. Further northeast of Port Blair, the Havelock Island site shows an even smaller hori-

*For correspondence. (e-mail: sridevi@cmmacs.ernet.in)

1. Scozzari, R. *et al.*, *Am. J. Hum. Genet.*, 1999, **65**, 829–846.
2. Jobling, M. M. and Tyler-Smith, C., *Nature Rev. Genet.*, 2003, **4**, 598–612.
3. Y Chromosome consortium, *Genome Res.*, 2002, **12**, 339–348.
4. Underhill, P. A. *et al.*, *Ann. Hum. Genet.*, 2001, **65**, 43–62.
5. Underhill, P. A. *et al.*, *Nature Genet.*, 2000, **26**, 358–361.
6. Hammer, M. F. *et al.*, *Genetics*, 1997, **145**, 787–805.
7. Hammer, M. F. *et al.*, *Mol. Biol. Evol.*, 1998, **15**, 427–441.
8. Al-Zahery, N., Semino, O., Benuzzi, G., Magri, C., Passarino, G., Torrioni, A. and Santachiara-Benerecetti, A. S., *Mol. Phylogenet. Evol.*, 2003, **28**, 458–472.
9. Deka, R. *et al.*, *Genome Res.*, 1996, **6**, 1177–1184.
10. Tajima, A. *et al.*, *Hum. Genet.*, 2002, **110**, 80–88.
11. Roychoudhury, S. *et al.*, *Hum. Genet.*, 2001, **109**, 339–350.
12. Puppala, S. and Crawford, M. H., *HOMO*, 1996, **47**, 73–84.
13. Karve, I., *Kinship Organization in India*, Asia Publishing House, Bombay, 1968.
14. Majumder, P., *J. Biosci.*, 2001, **4**, 533–545.
15. Balakrishna, V., *J. Hum. Evol.*, 1978, **7**, 67–75.
16. Quintana Mercı, L. *et al.*, *Am. J. Hum. Genet.*, 2004, **74**, 827–845.
17. Cavalli-Sforza, *The Great Human Diaspora: A History of Human Diversity and Evolution*, Smithsonian Institution Press, Washington DC, 1996, pp. 51–69.
18. Quintana Mercı, L. *et al.*, *Am. J. Hum. Genet.*, 2001, **68**, 537–542.
19. Comey, C. T., Koons, B. W., Presley, K. W., Smerick, J. B., Sobieralski, C. A. and Stanely, D. M., *J. Forensic. Sci.*, 1994, **39**, 1254–1269.
20. Bamshad, M. *et al.*, *Genome Res.*, 2001, **11**, 994–1004.
21. Bamshad, M. J. *et al.*, *Nature*, 1998, **395**, 651–652.
22. Cordaux, R., Aunger, R., Bentley, G., Nasidze, I., Sirajuddin, S. M. and Stoneking, M., *Curr. Biol.*, 2004, **14**, 231–235.