DNA Sequence Variation in the Human Y Chromosome: Functions and Dysfunctions

SUDHIR KUMAR
Department of Zoology, University of Lucknow, Lucknow-226 007

(Received 20 September 2006; Accepted 12 October 2006)

Human Y chromosome is clonally inherited and show extensive variation owing to large-scale mutations. Studies have shown that distal-Yq heterochromatin changed its length 12 times, the TSPY gene array, 23 times, the 3.6-Mb IR3/IR3 region underwent sequence re-orientation 12 times and the AZFc region rearranged 20 times. Molecular lineages of the Y chromosome encompassing 1.3 million years or 52,000 generations have confirmed high rate of mutations responsible for extensive structural polymorphism. In addition, large-scale copy number variation of the Y linked gene(s) has further fueled this process. Some of these changes may affect fertility status of the males. Therefore, it is logical to look for a correlation between organizational variations of the human Y chromosome and occurrence of male infertility. Information on this line would facilitate DNA based diagnosis and augment genetic counseling to the affected patients.

Key Words: Human Y chromosome, Male infertility, Azoospermia factor(s), Paracrine control, copy number polymorphism, microdeletions.

Introduction

Human Y chromosome contains about 70 million nucleotides but harbors least number of genes compared to any other chromosome. The male specific region, MSY, comprising 95% of the Y chromosome represents a mosaic of heterochromatic and three classes of euchromatic (X-transposed, X-degenerate and ampliconic) sequences (Fig. 1). Thus far, a total of 156 transcriptional units, 78 protein-coding genes and 27 distinct proteins have been ascribed to the Y chromosome. The MSY euchromatic sequences show frequent gene conversion. Of the eight palindromes localized on the human Y chromosome (Fig. 1d), six harbor vital testis specific genes. Characterization of palindromic complexes on the long arm of Y chromosome encompassing AZFb and AZFc regions and

Fig. 1: Schematic representation of the Y chromosome showing pseudoautosomal and heterochromatic regions (a). Enlarged view of a 24-Mb MSY extending from the proximal boundary of the Yp pseudoautosomal region to the proximal boundary of the large heterochromatic region of Yq (b). Shown are classes of euchromatic and heterochromatic sequences. A 1-Mb bar indicates the scale of the diagram. Gene, pseudogene and interspersed repeat elements of three euchromatic classes are given in (c) and palindromes P1-P8 in (d). Taken from Skaletsky et al. Nature 423, 825-837 (19 June 2003).
identification of HERV15 class of endogenous retroviruses close to AZFa region have facilitated our understanding on the organization of azoospermia factors [3, 50]. Deletion of any of the three azoospermia (AZFa, AZFb or AZFc) factor(s) and some still unidentified regulatory elements located elsewhere in the genome have been suspected to be responsible for male infertility. Considerable overlap of the AZFb and AZFc regions encompassing a number of genes and transcripts has been detected. In a recent study, a total of 1,099 proteins, co-purified with spermatogenetic chromatin have been identified and these proteins vital for DNA compaction and chromosomse segregation show high degree of evolutionary conservation [8]. However, it is not clear as to how many of these are encoded by the genes present on the Y chromosome. Thus, information on the exact number of genes or the types of mutations prevalent in the infertile male is still not available. Similarly, effects of loss or gain of the Y chromosome linked loci or their copy number polymorphism in the context of infertility, repeated abortion or sex chromosome related anomalies are not known. In a clinical setting, a couple experiencing repeated abortion may be subjected to DNA analysis in addition to clinical diagnosis. If the Y chromosome mosaicism or deletion of the DYZ1 is encountered in the father’s DNA, that would be a cause of concern. Thus, DNA based diagnosis would augment genetic counselling to the affected couples. Present article provides a brief summary on this line focusing also on the fast emerging newer tools of genome analysis and their applications.

The Y chromosome may be delineated into several broad sections encompassing, (i) pseudoautosomal boundary regions (PABY), (ii) a pericentric region on the short arm harboring the sex determining (SRY) gene, (iii) an euchromatic region (DYS1) on the proximal long arm (iv) a heterochromatic region (DYZ1) on the distal long arm and (v) an important DYZ3 region (Fig. 2).

Fig. 2: Schematic representation showing five different loci of the human Y chromosome encompassing both the arms. The Yqh region is shown as a major repeat DYZ1 (For details, see Bashamboo et al. 2005)

DYZ3 region is critical for the survival and propagation of the Y chromosome since it harbors the centromeric sequences. Besides these delineations, large sets of primers encompassing palindromes [43, 45] have been used for STS based analysis to assess genetic variation. In the process, several DNA markers have been found to have multiple locations in the non-recombining regions of the human Y chromosome [57]. These STS based markers may be used to screen patients DNA samples encompassing critical region(s) representing gonadal sex reversal, Turner syndrome, graft rejection and spermatogenic failure. Similarly, screening of 100–500 DNA samples, encompassing cases of repeated abortion, prostate cancer or males exposed to natural background radiation may be used for establishing genotype–phenotype correlation. With the availability of complete sequence information (http://www.nature.com/nature/focus/ychromosome/), more comprehensive screening approach encompassing new genes or gene families would go a long way to understand the genomics of the human Y chromosome.

Repeat DNA on the Y Chromosome

A total of 44 loci linked to the human Y chromosome have been described (http://www.ncbi.nlm.nih.gov/Omim/mimstats.html). Besides functional genes, Y chromosome includes the alphoid repeats, the major human SINE (Alu repeats) and additional 15 families of satellite sequences [1, 2, 6]. The alphoid sequences are clustered tandemly near the centromere on the Y chromosome and can be distinguished from those on the other chromosomes. Majority of the Y chromosome Alu repeats has little similarity with genomic consensus Alu sequences. In contrast, the Y LINE repeats cannot be distinguished from the LINEs found on the other chromosomes. Thus, Y chromosome Alu repeat sequences seem to have evolved together with other male specific (heteromorphic) sequences. We identified two such heteromorphic (male specific) sequences, though not involved in sex determination but showed cross hybridization with a few mammalian species [2, 6]. This indicates that some of the repeat sequences present on the human Y chromosome may have their homologues across the species. In addition, Y chromosome seems to have acted as a repository of the heterochromatic sequences. In earlier studies, of the 758 STS markers used for the development of high-resolution Y chromosome map, 136 from the NRY region were based on repetitive DNA [57]. These set of markers may be used as potential candidates to uncover sequence polymorphism and their possible biological roles.

Studies focused on this line have enabled identification of palindromic complexes P1 to P8 (Fig. 1d) on the long arm of human Y chromosome
encompassing azoospermia (AZF) factors. Deletions of these palindromic sequences have been found to correlate with spermatogenic failure [42]. P5/proximal-P1 deletions encompass up to 6.2 Mb sequences and 32 genes and transcripts whereas, P5/distal-P1 deletion encompass up to 7.7 Mb and 42 genes and transcripts. Extensive STS based analysis of these palindromic complexes has demonstrated that AZFb and AZFc regions are not independent, as reported earlier, but show overlap [42]. However, AZFa region, which spans about 0.8 Mb sequence is independent of the AZFb and AZFc regions [54]. Analysis of small but still undetected deletion in the patients is envisaged to uncover the roles of individual gene or gene families involved in regulation of spermatogenesis. Studies have shown de novo point mutation in the Y linked USP9Y gene in an azoospermic man [54]. Though infertility may also be caused by mutation(s) in the autosomal genes, the ones controlling paracrine systems or involved in signal transduction [35].

**Organisation of DYZ1 Sequences**

Besides alphoid repeats (Miklos and John, 1979), Y chromosome harbors DYZ1 (Fig. 1 and 2), another major repeat, detected as 3.4 kb fragments in HaeIII digest of male genomic DNA [9]. This largely contains a pentameric motif “5TTCCA3” and in normal males, approximately, 3000–4000 copies are estimated to be present [32]. Whether DYZ1 copy number variation has any role to play in the processes of regulation of spermatogenesis, male fertility, sustenance of full-term pregnancy or overall reproductive potential of the human males is not clear. However, DYZ1 seems to be affected showing massive copy number variation (CNV) in cases of repeated abortion, males exposed to natural background radiation and prostate cancer [36, 37]. Further, DYZ1 was found to undergo programmed sequence modulation resulting in alteration of restriction sites for RsaI enzyme, which is restored back sometimes after fertilization [39].

The 5’TCCA3’ sequences are the core element of the DYZ1 fraction, organized structurally in sex specific manner in humans but absent in the non-human systems[2].

In our study, FISH conducted with metaphase chromosomes of male having prostate cancer showed DYZ1 arrays on the chromosome 10 and varying signals of the same in the interphase nuclei (Fig. 3). In subsequent studies on DNA samples from the cases of repeated abortion, DYZ1 showed conspicuous loss of the sequences at three regions (Fig. 4a-d). Similarly, DNA samples from semen and blood of the males

![Fig. 3: Fluorescent in situ Hybridization (FISH) of the DYZ1 with metaphase chromosomes of human males representing prostate cancer (a-d). Note the DYZ1 signal on autosomes 10 and Y chromosome in a-b, absence of the same in the interphase nuclei (c) and on the Y chromosome in panel d (arrows).](image)
Sudhir Kumar

exposed to natural background radiation (NBR) also showed deletion of the DYZ1 in some cases (Fig. 4e). Thus, DYZ1 is affected in the abnormal genomes and often show copy number variation. In cases of repeated abortion, DYZ1 copies were found to be far below the normal range showing as few as 95 (Fig. 5a).

Surprisingly, males exposed to NBR showed more copies of the DYZ1 in many samples compared to that in the normal ones (Fig. 5b). Though the mechanism is not clear, this study suggests that DYZ1 repeat is affected in case of repeated abortion and males exposed to NBR[34]. NBR males showing higher copies of DYZ1 were fertile though it is not clear for how long they would be able to maintain their fertility status. Surprisingly, all the changes took place in somatic DNA and not in that of germline. We construed that germline DNA is protected by some innate mechanism. Irrespective of the mechanism involved, copy number variation (CNV) may be used as an important parameter to assess the status of DYZ1 in a given cell population to establish genotype phenotype correlation.

Y Chromosome Related Anomalies

Involvement of Y in sex chromosome related anomalies like XY gonadal dysgenesis (Swyer syndrome), XYY males, recurrent spontaneous abortions (RSA) and Turner Syndrome (TS) is well established. At birth, the XY female patients with gonadal dysgenesis (Swyer syndrome) appear to be normal but at puberty, they develop streak gonads, do not menstruate and lack secondary sexual characters. Patients are of normal stature and have no somatic stigmata of Turner syndrome. The height of patients with XY gonadal dysgenesis (unusually higher for females) is probably explained by androgen production in the streak gonad whereas clitoromegaly is present in some cases. Three forms of Swyer syndrome have been suggested [30]. These are (i) sporadic testicular agenesis syndrome (STAS) detected in the H-Y negative individuals, (ii) familial testicular agenesis syndrome (FTAS), also in the H-Y negative individuals but showing an X-linked recessive pedigree and (iii) familial testicular dysgenesis syndrome (FTDS) where the patients are H-Y positive but have female phenotype and streak gonads which may contain testis-like tumoral structures. The phenotype of STAS and FTAS is identical even though the mutation is probably on the Y chromosome in STAS and on the X chromosome in FTAS [30].

The XYY syndrome with an extra Y chromosome is an abnormality associated with tall stature, low intelligence, delayed speech and some learning problems. The characteristics of XYY syndrome are often very subtle and do not indicate any serious disorder. Therefore, males with XYY chromosomes are either undiagnosed or misdiagnosed. Thus far, the XYY syndrome has been diagnosed on the basis of chromosomal analysis. It would be informative if such cases are subjected to FISH and

Fig. 4: Schematic representation showing specific deletion points within the DYZ1 array (a) and corresponding individual deletions in the aborted fetus (b-d) and germline and blood samples of the males exposed to background radiations (e). As control, genomic DNA amplification with β actin primers is shown in panel B. Sample ID is given on top of each panel and molecular marker (M) is shown in base pairs (For details, See Pathak et al. 2006)
Real Time PCR analyses to assess if both the Ys are identical with respect to Y-linked loci and their major repeat elements. Since all the cells may not have XYY chromosome constitution, it is likely that some patients show mosaicism. These mosaicisms may be detected by FISH as reported earlier [34]. DNA analysis to assess loss/gain or sequence modulation of the critical Y-linked loci from the couple experiencing repeated abortions and that from the aborted fetus might generate useful information. Studies have shown that long Y chromosome may affect the viability of the zygote but not the fertility of its carrier [15]. Studies conducted on DNA samples from repeat abortion showed distinct deletion in the DYZ1 [34]. Thus, it is reasonable to infer that intact human Y chromosome is needed for the sustenance of full term pregnancies of the male fetus.

**Candidate Turner Syndrome Genes**

Turner syndrome (TS) with 45, XO karyotype is the only monosomy of the X chromosome known to occur 1 in 2000–5000 female live births [17]. Over 90% of the pure O concept uses are eliminated during prenatal development and postnatally, chromosomal mosaicism of both X and Y chromosomes are observed. The most common karyotypes are, 45,X, isochromosome X, 45,X/45,XX, 45,X/46,XY, 45,X,i(Xq), 46,Xt(X;X), 46,Xi(X;Y), 46,Xi(Xp), 46,Xi,Xq, r(X), 46,Xr(X) (including small ring X chromosome), 45,X/46,XX, 45,X/47,XXX, 46,X/46,XX/47,XXX, 45,X/46,XY, and 45,X/51446,XXq + 46,X + mar(X) 47,XXq+, + mar(X) 46,X + mar(Y) 47,XXq+, + mar(Y). In a patient with stigmata of Turner syndrome, a Y-derived aberrant marker chromosome showed rare indels in the DYS1 region [7].

Several molecular mechanisms have been proposed to explain the Turner syndrome (TS) phenotype including the chromosome imbalance, genomic imprinting and haploinsufficiency. In normal women with two X’s in each cell, genes on one X are permanently switched off to give women the same dose of X-based genes as that in men. The exceptions to this rule are the X-linked housekeeping genes having their counterparts on the Y chromosome [38]. The housekeeping genes remain switched on in both the X’s so as to equal the male cell’s dose from its X and Y chromosomes. Thus, the abnormalities in Turner’s syndrome may arise due to haploinsufficiency because the product of single X chromosome is only half the dose needed for various housekeeping functions [60]. In accordance with this hypothesis, Turner syndrome loci were originally assigned to be present in the pseudoautosomal region on the short arm of X and Y chromosomes [33, 21, 49]. The only TS feature consistently present in patients with small distal Xp deletions is short stature, suggesting that loci responsible for other features lie outside of the pseudoautosomal region [53].

One reason for poor genotype–phenotype correlation may be the high prevalence of mosaicism, essential for the viability of the X chromosome monosomy (Held et al. 1992). There are evidences favoring Yq and distal Yp as the site for TS lymphedema gene (Fisher et al. 1990; Barbaux et al. 1995). Apart from ZFY, RPS4Y seems to be another candidate gene in this region for TS. There are evidences against RPS4Y haploinsufficiency as the sole cause of TS [14]. However, these findings do not rule out a role for RPS4X/RPS4Y haploinsufficiency in certain phenotypes associated with

---

**Fig. 5:** A bar diagram showing copy number status of the DYZ1 arrays in samples representing cases of repeated abortion, (RA), prostate cancer (PC) (a) and males exposed to natural background radiation (NBR) (b). Note the copy number variations of the DYZ1 arrays amongst different RA and NBR samples (For details, See Pathak et al. 2006)
monosomy of the X chromosome including lethality or lymphatic abnormalities [60]. Owing to a broad spectrum of physical and physiological abnormalities observed in TS and inconsistent phenotypes, it is reasonable to infer that a large number of autosomal genes but relatively fewer Y-linked genes are involved in these anomalies. However, much effort would be required to uncover the exact number of genes implicated in TS. In our study focusing on five loci encompassing both the arms of the Y chromosome (see Fig. 2), we screened 19 TS samples. Of these, 8 were positive for PABY, 7 SRY, 11 DY2Z3, 9 DYS1 and 7 DY2Z1. No two TS samples showed identical pattern but maximum number of samples were positive for DY2Z3 [39]. Screening of a large number of samples is envisaged to uncover better picture with respect to most frequently affected loci in case of TS.

**Y Chromosome and Gonadoblastoma**

Despite all efforts, mapping of the GBY gene has proven to be unexpectedly complex. It is likely that several genes present in multiple copies distributed across the Y chromosome are responsible for gonadoblastoma [58]. Isolation of many Y-linked genes and high resolution physical map of the Y chromosome are envisaged to be helpful for studying multiple candidate genes for GBY locus [26, 57]. The small region defined by Tsuichiya et al.1995 contains seven known genes, amelogenin Y (AMELY), RNA binding motif (RBM), protein kinase Y (PRKY), protein tyrosine phosphatases (PTP)-BL related Y (PRY), testis transcripts Y1 and Y2 (TTY1 and TTY2) also referred to as TTYY1 and TTYY2, and testis specific protein Y-encoded (TSPY). Both TTY1 and TTY2 consist of repetitive DNA lacking any apparent protein coding sequences [26]. AMELY, like its X-linked homologue, AMELX, encodes an enamel protein [28] expressed exclusively in the developing tooth bud [48]. Thus, these gene products are unlikely to affect cell proliferation in the gonad. RBM is a repeated gene with functional sequences that are mostly located in the interval 6 on the long arm [41]. Only a few copies are located within the postulated GBY critical region. The PRKY and its X homologue, PRKX, encode members of the c-AMP dependent serine-threonine protein kinase family [22, 47]. The PRY is a repeated gene whose product displays some similarity to PTP-BL gene [26]. Other copies of PRY are found in the interval 6 outside of the GBY critical region. Protein kinases are known to act as signal transducers for growth factors and cytokine receptors, whereas phosphatases counter balance their effects [18, 31]. Thus, an imbalance in the expression between PRKY and PRY may hamper normal growth regulatory functions. However, biological significance of these proteins in the context of gonadoblastoma is yet to be understood. The remaining candidate locus for gonadoblastoma is a multicopy TSPY (gene for testis specific protein on Y) located primarily in the GBY critical region at the interval 3 [4, 59]. Several homologous copies have also been mapped on proximal intervals 4 and 5 of the long arm. TSPY expresses in gonadoblastoma tissues and has several features fitting the profile of the GBY locus [58]. Most of its repeat units have been mapped to the critical region of GBY including intervals 3E–3G and 4/5 [46, 58]. Further, its expression in spermatogonial cells in normal testis suggests a function during early spermatogenesis. TSPY is unregulated in gonadoblastoma tissues as well as in certain testicular and prostate cancers [58]. Interestingly, TSPY expression is drastically reduced in the dysgenetic gonads of the patients with 46, XY Tfm, who rarely develops gonadoblastoma. Despite evidences pointing towards a number of genes as possible candidates for the postulated GBY locus, there is no conclusive evidence in support of any such gene(s). Thus, it would be relevant to monitor the presence/absence and if possible, expression of several recently uncovered Y-linked genes using STS based marker systems [48]. Coupled with this, assessment of copy number status of candidate genes using Real Time PCR and FISH may be yet another complementary approach to nail the GBY locus on the Y chromosome.

**Y Chromosome Haplotyping and SNP**

A correlation between the occurrences of Y chromosome related SNPs with their resultant diseased phenotypes is useful to develop understanding on its biological roles. SNPs in the repetitive sequences have been described in males with deletion of azoospermia complex, highlighting the significance of these repeat elements in male infertility [44, 56]. Based on Y chromosome SNP haplotypes, gene flow among caste, tribe, and the migrant Siddi populations of Southern part of India has been reported [40]. This approach was adopted to uncover Y-chromosome variation in a Norwegian population [11]. In addition, micro- and minisatellites have been used to unravel the evolutionary history of the Y chromosome since changes in microsatellite length occur more frequently than the emergence of new unique event polymorphisms (UEPs) [20].

**Conclusions**

Perusal of literature provides an understanding on the genes/factors associated with human male infertility albeit without throwing light on the exact number and type of such genes and their overall mutational status. Although infertility is seen in both the sexes, but in majority of the cases, it remains confined to males contributing to about 10–15% in a given population. Characterization of additional Y- linked and autosomal candidate genes with respect to mutations prevalent therein would provide a better picture on the actual
genesis of infertility. A comprehensive physiological and genetic testing of the patients would establish a more meaningful genotype phenotype correlation. Finally, copy number polymorphism of Y-linked loci and major repeat elements using real time PCR in normal males and those suffering from sex chromosome related anomalies would prove to be informative. The male infertility is caused due mutations in the genes not only on the Y chromosome but also on the autosomes in addition to failures of several physical and physiological attributes including paracrine controls. Thus, search for candidate genes on the autosomes would equally be useful.

Acknowledgements

The Author is grateful to Vice-Chancellor and Head, Department of Zoology, University of Lucknow, Lucknow for encouragement and Dr. Sher Ali, National Institute of Immunology, New Delhi for his critical comments on the manuscript. Thanks are also due to DST, New Delhi for the award of SERC Visiting Fellowship.

References

2. S Ali, S Gauri and S Bala A synthetic oligonucleotide probe (5VTTTC-CA3V) n uncovers a male specific hybridization pattern in the human genome Mol Cell Probes 6 (1992) 521-525
17. E B Hook and D Warburton The distribution of chromosomal genotypes associated with Turner’s syndrome, live birth prevalence rates and evidence for diminished fetal mortality and severity in genotypes associated with structural X abnormalities or mosaicism Hum Genet 64 (1983) 24-27, 866
18. T Hunter Anti-phasatases take the stage Nat Genet 18 (1999) 303-305
27. E S Lander, L M Linton, B Birren, C Nusbaum, M C Zody, J Baldwin, K Devon, M Doyle, W Fitz Hugh et al.


31. M P Myers and NK Tonks PTEN, sometimes taking it off can be better than putting it on Am J Hum Genet 61 (1997) 1234-1238


34. D Pathak, S Premi, J Srivastava, P J Chandy and S Ali Genomic Instability of the DYZ1 Repeat in Patients with Y chromosome Anomalies and Males Exposed to Natural Background Radiation. DNA Res (2006) 00:000-000 (in press)


45. E C Salido, P H Yen, K Koprivnikar, L C Yu and L J Shapiro The human enamel protein gene amelogenin is expressed from both the X and the Y chromosomes Am J Hum Genet 50(2) (1992) 303-316


49. E Schwinger, F Schaaff and H Wedemann Analysis of sex and delta F508 in single amniocytes using primer extension preamplification Hum Genet 98(2) (1996) 158-161


52. H Skakelstyk et al. The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes Nature 423 (2003) 825-837


